

Genomweite Charakterisierung der putativen Rekombinationshotspots zweier Makakenarten

-

***PRDM9* in der Primatenevolution**

Dissertation

zur Erlangung des Grades

Doktor der Naturwissenschaften

am Fachbereich Biologie

der Johannes Gutenberg-Universität in Mainz

Sacha Heerschop

geboren am 11.12.1987 in Sneek, Niederlande

Mainz, 2019

Dekan:

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung: 16.04.2020

Danksagung

Zusammenfassung

Die Speziation ist ein Prozess, der in der Keimbahn ansetzt. Veränderungen, die sich dort akkumulieren, werden an die nächsten Generationen weitergegeben und sind Ansatzpunkte der Evolution. PRDM9 stellt einen in trans wirksamen Regulator der Keimbahn-spezifischen Rekombination dar, der als putatives Speziationsgen gehandelt wird. Das Protein bestimmt während der Meiose I die Lokalisation und Markierung der Rekombinationshotspots über eine Interaktion seiner Zinkfingerdomäne mit den entsprechenden DNA-Zielstrukturen. Die DNA-bindenden Aminosäuren des PRDM9 weisen Signaturen positiver Selektion auf. Die dadurch schnelle Evolution und der kontinuierliche Abbau von Rekombinationshotspots können möglicherweise Grund für die Hybridsterilität bei Mäusen sein. Ein möglicher Zusammenhang von *PRDM9* mit dem Speziationsprozess stellt den theoretischen Überbau der vorgestellten Arbeiten dar.

Vor dem Hintergrund bekannter Genomsequenzen der rezent divergierenden Schwestergruppen, *Macaca mulatta* und *Macaca fascicularis*, wurden genomweit die Rekombinationshotspots auf Basis der Bindungseigenschaften des Proteins PRDM9 vergleichend untersucht. Die explorative Datenanalyse der Chromatinimmopräzipitation weist viele Gemeinsamkeiten der PRDM9-Zielstrukturen zwischen den zwei Makakenarten bezüglich der Eigenschaften, wie GC-Gehalt, Motiv und Lokalisation entlang der Chromosomen, auf. Dennoch überlappen die Loci selbst kaum. Einige Bindungsstellen sind in und nah bei Genen zu finden, die mit der akuten lymphatischen Leukämie (ALL) assoziiert werden, wobei ein Zusammenhang zwischen der ALL und PRDM9 bereits bekannt ist. Dies und eine Verknüpfung PRDM9s mit piRNA-Clustern (Piwi-interacting RNA-Cluster) könnten auf unbekannt Funktionen hinweisen.

Zu den weiteren Projekten gehören die Charakterisierung strepsirrhiner *PRDM9*-Zinkfingerdomänen von *Daubentonia madagascariensis* als Vertreter der tiefsten Verzweigung in der Divergenz madagassischer Lemuren und die Datierung des Duplikationsereignisses, das zu dem aus *PRDM9* entstandenen *PRDM7* führte, im Stammbaum der Primaten.

Die Ergebnisse zeigen, dass *PRDM7* auf der Linie zum letzten gemeinsamen Vorfahren der Catarrhini und nach Abspaltung der Platyrrhini entstanden ist. Somit ist die Duplikation des *PRDM9*-Gens deutlich jünger als bisher angenommen. Die phylogenetische Baumrekonstruktion auf Basis des mit strepsirrhinen Zinkfingerdomänen erweiterten Datensatzes spiegelt die Phylogenie der Primaten wider und spaltet sich in drei Cluster mit Vertretern der Strepsirrhini, Tarsier und Anthropeidea auf.

INHALT

1. EINLEITUNG UND ZIELSETZUNG.....	1
1.1 Einführung zu PRDM9	1
1.1.1 Entdeckung	1
1.1.2 Aufbau	1
1.1.3 Funktion	2
1.1.4 Evolution.....	4
1.1.5 PRDM9 und Speziation	5
1.1.6 PRDM9-assoziierte Krankheiten.....	7
1.2 <i>Macaca mulatta</i> und <i>Macaca fascicularis</i>	8
1.3 Zielsetzung.....	9
2. PRDM9 IN DER PRIMATENEVOLUTION	11
2.1 Die PRDM9-Zinkfinger der Strepsirrhini	11
2.1.1 Material und Methoden	11
2.1.1.1 Labor	11
2.1.1.2 Datenanalysen	12
2.1.2 Ergebnisse	13
2.1.3 Diskussion	15
2.2 Datierung des Duplikationsevents PRDM7/PRDM9	16
2.2.1 Material und Methoden	16
2.2.1.1 Labor	16
2.2.1.2 Datenanalysen	17
2.2.1.2.1 Sequenzen aus dem 1000-Genome-Projekt	17
2.2.1.2.2 Erstellung der Phylogenien	17
2.2.1.2.3 Sequenzvergleich mit needle	18
2.2.1.2.4 BLAST der PRDM7-Sequenzen in Cercopithecoidea	18
2.2.1.2.5 ensembl-Suche	19
2.2.2 Ergebnisse	19
2.2.2.1 PCR der internen PRDM7-Duplikation	19
2.2.2.2 PCR des Zinkfingerarrays	20
2.2.2.3 PRDM7/PRDM9-Phylogenien	20
2.2.2.4 Vergleich von Teilsequenzen mit needle	22
2.2.2.5 BLAST gegen Chromosom 20	23
2.2.2.6 ensembl-Suche.....	23
2.2.3 Diskussion	25
3. CHARAKTERISIERUNG DER PRDM9-BINDUNGSSTELLEN ALS PUTATIVE REKOMBINATIONSHOTSPOTS BEI ZWEI MAKAKENARTEN	26
3.1 Material und Methoden.....	26
3.1.1 Labor.....	26
3.1.1.1 RNA-Extraktion und cDNA-Synthese	26

3.1.1.2 Sequenzierung der Zinkfingerarrays	26
3.1.1.3 Western Blot	26
3.1.1.4 Chromatinimmunopräzipitation.....	27
3.1.1.5 Immunopräzipitation und Massenspektrometrie	28
3.1.2 Datenanalyse	28
3.1.2.1 Aufbereitung der Zinkfingersequenzen	28
3.1.2.2 Ermittlung der DNA-Bindungssequenzen anhand der Zinkfingerarrays	28
3.1.2.3 Analyse der ChIP-Sequenzen	28
3.1.2.3.1 Qualitätskontrolle der Rohdaten	28
3.1.2.3.2 Galaxy	28
3.1.2.3.3 Qualitätswerte.....	29
3.1.2.3.4 Peak Calling	29
3.1.2.3.5 Verteilung der Peaks im Genom	30
3.1.2.3.6 Annotation und Genfunktion	30
3.1.2.3.7 Motive	31
3.1.2.3.8 piRNA Cluster.....	31
3.1.2.3.9 Auswertung der Daten aus der Massenspektrometrie	31
3.2 Ergebnisse	32
3.2.1 Überprüfung PRDM9s auf RNA- und Proteinebene	32
3.2.2 Die Sequenzen der Zinkfingerarrays	33
3.2.3 Ergebnisse der Chromatinimmunopräzipitation	34
3.2.3.1 Übersicht über alle Reads	34
3.2.3.2 Anzahl der Peaks	35
3.2.3.3 Verteilung der Peaks	37
3.2.3.4 GC-Gehalt	39
3.2.3.5 Zusammenhang zwischen Verteilung und GC-Gehalt	40
3.2.3.6 Prozentuale Verteilung auf einzelne Chromosomen.....	42
3.2.3.7 Verteilung der H3K4me3- und DMC1-Peaks PRDM9-Peaks.....	45
3.2.3.8 Annotation PRDM9s.....	46
3.2.3.9 Gene Ontology	48
3.2.3.10 piRNA Cluster	49
3.2.3.11 Motiv-Konsensussequenz.....	51
3.2.3.11.1 Zusammenhang zwischen Motiv und Alu-Elementen.....	53
3.2.3.11.2 Vergleich der Motive: MEME vs. Binding Site Prediction.....	54
3.2.3.12 Interaktionspartner	54
3.2.3.13 H3K4me3.....	55
3.2.3.13.1 Testis, Haushalt	56
3.2.3.13.2 GO mit DAVID	57
3.2.3.13.3 Annotation.....	57
3.3 Diskussion	59
3.3.1 Reads, Peaks und Qualitätsmerkmale	59
3.3.2 Verteilung der PRDM9-Peaks und GC-Gehalt.....	59
3.3.3 Verteilung auf einzelne Chromosomen.....	60
3.3.3.1 Autosome und Telomere	60
3.3.3.2 Gonosomen und PAR	60
3.3.4 Assoziation mit genomischen Regionen/Charakteristika/Merkmalen	61
3.3.4.1 Annotation	61

3.3.4.2 Gene Ontology: PRDM9 und Krebs bzw. Ras	62
3.3.4.3 Repetitive Elemente	63
3.3.5 Motive.....	65
3.3.6 H3K4me3.....	66
3.3.7 Abschließender Vergleich zwischen Mm und Mf	67
3.3.8 PRDM9 und Speziation	69
4. FAZIT.....	70
5. ANHANG	71
6. LITERATURVERZEICHNIS	92
7.INTERNETQUELLEN.....	105
8. LEBENSLAUF	106

1. Einleitung und Zielsetzung

1.1 Einführung zu PRDM9

1.1.1 Entdeckung

Das erste Hybridsterilitätsgen bzw. Speziationsgen in Vertebraten wurde bereits 1974 von Forejt und Iványi beschrieben, auf Chromosom 17 von Maus lokalisiert und als *Hst1* (*hybrid sterility 1*) bezeichnet. Dafür wurde die Fertilität von wildlebenden Mäusen und Labormäusen und deren Hybride bestimmt. Die männlichen Nachkommen, die aus der Kombination wildlebender Väter mit Mütter verschiedener Laborstämme entstanden sind, zeigten einen Stillstand der Spermatogenese und kleinere Hoden. Die weibliche F1-Generation hingegen war fertil, was mit Haldanes Regel – nämlich, dass in Speziesshybriden häufiger das heterogametische Geschlecht steril ist – übereinstimmt (Haldane 1922). Die Region mit *Hst1* wurde mittels zweier Laborstämme bestimmt, die sich lediglich hier unterschieden und deren Nachkommen die gleichen Resultate aufwiesen. Weitere Kreuzungen aus wildlebenden Vätern mit Müttern aus anderen Laborstämmen führten zu fertilen Nachkommen, unabhängig vom Geschlecht. Hayashi et al. beschrieben 2005 eine Histonmethyltransferase, die spezifisch ist für die männliche und weibliche Meiose, die sie *Meisetz* (*Meiosis-induced factor containing PR/SET domain and zinc-finger motif*) nannten. Die Gruppe um Forejt konnte 2009 *Prdm9*, bei Mäusen als *Meisetz* bekannt, als *Hst1* identifizieren (Mihola et al. 2009), indem vier verschiedene, sich überlappende BAC-Klone (*bacterial artificial chromosome*) aus der Region mit dem gesuchten Gen in embryonale Stammzellen transfiziert wurden. Der Nachweis von fertilen Nachkommen war nur bei jenen Tieren mit dem Substitut, das ein komplettes PRDM9 kodiert, vorhanden. Der Name *PRDM9* steht für *PR domain containing protein 9*, wobei PR wiederum aus den Namen von *PRDM1* und *PRDM2* zusammengesetzt ist, die auch als *PRDI-BF1* (*positive regulatory domain I binding factor 1*) bzw. *RIZ1* (*retinoblastoma protein-interacting zinc finger gene 1*) bekannt sind (Buyse et al. 1995).

1.1.2 Aufbau

Von N- zu C-Terminus besteht PRDM9 aus einer KRAB-, einer SSXR-, einer PR/SET-Domäne mit einem anschließenden, einzelnen Zinkfinger (Znf) und einem Zinkfingerarray (Hayashi et al. 2005, Birtle und Ponting 2006).

Die KRAB-Domäne (*Krüppel-associated box*) spielt wahrscheinlich eine Rolle in der Proteininteraktion; jene von PRDM9 interagiert jedoch nicht wie die KRAB-Domänen anderer Proteine mit TRIM28 (*Tripartite motif-containing 28*), auch KAP1 genannt (*KRAB-associated protein-1*; Friedmann et al 1996, Patel et al. 2016, Imai et al. 2017), sondern unter anderem mit CXXC1, das dem COMPASS-Komplex zuzuschreiben ist und ein Ortholog von Spp1 von *Saccharomyces cerevisiae* darstellt (Parvanov et al. 2017, Imai et al. 2017). Spp1 (Secreted Phosphoprotein 1) ist in der Hefe notwendig, um potentielle Positionen von Doppelstrangbrüchen (DSB) mit der Chromosomenachse zu verbinden (Acquaviva et al. 2013, Sommermeyer et al. 2013). Auch weitere Interaktionspartner

wurden von Imai et al. (2017) mit *yeast two-hybrid assay* nachgewiesen: MCRS1, PIH1D1, CHAF1A, CEP70, FKBP6, IFT88.

In SSX-Proteinen fungiert die SSXRD-Domäne (*synovial sarcoma, X breakpoint repression domain*) als Transkriptionsrepressor (Lim et al. 1998). Bei PRDM9 scheint sie wie die KRAB-Domäne bei der Proteininteraktion beteiligt zu sein, eine schwache Interaktion mit CXXC1 und CHAF1A wurde von Imai et al. (2017) nachgewiesen.

Die PR/SET-Domäne, die 20-30% Übereinstimmung mit SET-Domänen (*Su(var)3-9, Enhancer-of-zeste and Trithorax*) aufweist (Huang et al. 1998), ist für die Trimethylierung der Lysine 4 und 36 des Histons 3 verantwortlich (Hayashi et al. 2005, Wu et al. 2013, Eram et al. 2014, Powers et al. 2016).

Die Zinkfinger sind mit ihrem Aufbau C-X(2)-C-X(12)-H-X(3)-H-TGEKPYX, X als beliebige Aminosäure, dem klassischen C₂H₂-Zinkfinger zuzuordnen, auch Krüppel-Typ Zinkfinger (KZNF) genannt (Miller et al. 1985, Bellefroid et al. 1991). Die zwei Cysteine und zwei Histidine binden dabei das Zinkion, die sieben Aminosäuren TGEKPYX bilden den Linker zwischen den einzelnen Wiederholungseinheiten und die DNA-bindenden Aminosäuren befinden sich an den Positionen -1, 2, 3 und 6 bezüglich der α -Helix (Miller et al. 1985, Pavletich und Pabo 1991, Fairall et al. 1993). Die Bindung erfolgt in der großen Furche des DNA-Stranges (Pavletich und Pabo 1991). Abgesehen von der Aminosäure an Position 2 unterliegen die restlichen drei DNA-bindenden Aminosäuren einem positiven Selektionsdruck, der bei verschiedenen Spezies für die PRDM9-Znf nachgewiesen wurde (Oliver et al. 2009, Thomas et al. 2009). Der einzelne Zinkfinger nach der PR/SET-Domäne weist als einziger über verschiedene Taxa hinweg die gleichen Aminosäuren an den bindenden Positionen auf (Myers et al. 2010). Der degenerierte Zinkfinger am Anfang des Arrays ist im Vergleich zu den restlichen Znf des Arrays konservierter, wie bei Schwarz et al. (2014) und Heerschop et al. (2016) gezeigt wird, da er über Spezies hinweg in der Phylogenie der einzelnen PRDM9-Zinkfinger von verschiedenen Primaten zusammenfällt.

Jedes einzelne Element ist für die Aufrechterhaltung der Funktion PRDM9s essentiell, was aus dem Vergleich von Orthologen (Baker et al. 2017) oder aus Funktionsstudien (Imai et al. 2017, Thibault-Sennett et al. 2018) hervorgeht.

1.1.3 Funktion

PRDM9 wird ausschließlich während der Prophase I der Meiose (Leptotän bis Zygotän) in der weiblichen und männlichen Keimbahn exprimiert, Knockouts können zu einem Stillstand im Pachytän-Stadium bei Mäusen führen, da die Reparatur der DSB, die Paarung der homologen Chromosomen und das Ausbilden des sogenannten *sex body*-Komplexes fehlschlagen (Hayashi et al. 2005, Sun et al. 2015, Parvanov et al. 2017). Dabei soll PRDM9 die Loci der Rekombinationshotspots determinieren (Baudat et al. 2010, Myers et al. 2010, Parvanov et al. 2010). Diese Hotspots sind 1-2 kb große Loci, an denen Crossing-over, der Austausch zwischen homologen Chromosomen, gehäuft stattfindet (Jeffreys et al. 2001). Das führt zu einer Erhöhung von Diversität, da die Zusammenstellung von Allelen verändert wird. Die kurzzeitige Verbindung der Chromosomen ist auch notwendig, damit sie sich bei der ersten meiotischen Teilung

fehlerfrei auftrennen (Stearns und Hoekstra 2003). Hotspots sind meist GC-reich und gehäuft in Telomernähe zu finden (siehe dazu beispielsweise Ikemura und Wada 1991, Williams und Hurst 2000, Yu et al. 2001, Kong et al. 2002, Paigen et al. 2008). Die Anzahl der Hotspots von etwa 30000 beim Menschen und 15000 bei der Maus übersteigen die Anzahl der Doppelstrangbrüche, etwa 200-400, welche zu 20-25 Crossing-over führen (The International HapMap Consortium 2005, Myers et al. 2005, Brick et al. 2012). Pro Chromosom ist mindestens ein Crossing-over zu beobachten (Hultén 1974). Auch in der pseudoautosomalen Region (PAR) ist mindestens ein Crossing-over notwendig (Burgoyne 1982), wobei PRDM9 beim Menschen für die dortigen Hotspots verantwortlich ist, jedoch bei der Maus noch Unsicherheiten hierzu bestehen (Brick et al. 2012, Hinch et al. 2014, Pratto et al. 2014). Laut Hinch et al. (2019) werden dort, wo sich die homologen Strukturen am schnellsten finden, aus DSB Crossing-over, was durch drei Faktoren beeinflusst wird: ein hoher Anteil an Guanin und Cytosin, eine Position an den Chromosomenenden und die Bindung PRDM9s an homologen Loci.

Der Ablauf, der von Imai et al. (2017) vorgeschlagen wird, sieht folgendermaßen aus:

PRDM9 bindet mittels der Zinkfinger die DNA meist in einer Loop-Region (Parvanov et al. 2017, Grey et al. 2017) und benachbarte Nukleosomen werden H3K4- und H3K36-trimethyliert und verschoben (Hayashi et al. 2005, Baker et al. 2014, Powers et al. 2016). Die Interaktion zu CXXC1 entsteht durch die KRAB- und SSXR-Domäne oder bzw. zugleich durch die Histonmodifikation, da CXXC1 methylierte H3K4s bindet (Eberl et al. 2013, Parvanov et al. 2017, Imai et al. 2017). Diese Verknüpfung sorgt dafür, dass die potentiellen DSB-Loci mit der Chromosomenachse verbunden werden, da CXXC1 auch IHO1 bindet. IHO1 bildet wiederum mit REC114 und MEI4 einen Komplex (Kumar et al. 2010, Stanzione et al. 2016, Kumar et al. 2018). Ob PRDM9 oder der IHO-REC114-MEI4-Komplex den Doppelstrangbruch induzierenden Komplex SPO11-TOPOVIBL herbeiführt, ist unklar (Robert et al. 2016, Imai et al. 2017). Dabei bindet PRDM9 mittels seiner Zinkfinger ein spezifisches Motiv, das bei Menschen die Konsensussequenz CCNCCNTNCCNC aufweist (Myers et al. 2008). Nicht alle Zinkfinger der humanen Arrays sind beteiligt, sondern nur vier bis sechs (Patel et al. 2016, Patel et al. 2017).

Ohne *PRDM9* wird auf einen wahrscheinlich ancestralen Mechanismus zurückgegriffen. Verschiedene Beispiele von Spezies ohne *PRDM9* zeigen, dass H3K4me3-Loci oder CpG-Inseln der DSB-Initiierung dienen und diese Hotspots, anders als PRDM9-induzierte, über evolutionäre Zeiträume konstant sind: Bei Hefe, die kein *PRDM9*-Ortholog besitzt, entstehen die Doppelstrangbrüche in Regionen ohne Nukleosomen (*nucleosome-depleted region*, NDR), die vor allem bei Promotoren und H3K4me3-Markierungen vorkommen (Borde et al. 2009, Pan et al. 2011). *PRDM9*-Knockout-Mäuse weisen DSB auf, allerdings an H3K4me3-Loci und nicht an den Positionen, die PRDM9 sonst vorgibt (Brick et al. 2012). Hunde besitzen ein pseudogenisiertes *PRDM9*, das mehrere Mutationen beinhaltet, und eine Verschiebung der Hotspots Richtung GC-reiche Regionen, vor allem hin zu CpG-Inseln, ist zu beobachten (Oliver et al. 2009, Muñoz-Fuentes et al. 2011, Axelsson et al. 2012, Auton et al. 2013). Bei Vögeln und Krokodilen ist *PRDM9* nicht vorhanden und die Hotspots sind auch bei CpG-Inseln und Transkriptionsstartsites zu finden (Singhal et al. 2015). Bei einer Mutter mit zwei nicht funktionalen *PRDM9*-Allelen konnte diese Verschiebung hin zu CpG-Inseln und Promotorregionen nicht nachgewiesen werden.

Deshalb vermuten die Autor*innen um Narasimhan (2016) bei Menschen einen weiteren Alternativmechanismus für die Bestimmung der Hotspots.

Imai et al. (2017) berichten, dass CXXC1 mit der KRAB-Domäne PRDM9s interagiert und den gebundenen Locus für den DSB an die Chromosomenachse heranführt. Auch als Alternativmechanismus schlagen sie CXXC1 vor, da es auch H3K4me3 und unmethylierte CpG-Loci bindet (Lee und Skalnik 2005, Thomson et al. 2010), was zu den berichteten Alternativhotspots, z.B. bei Hefe, Hunden und *PRDM9*-Knockout-Mäusen, passt. Tian et al. (2018) zeigen, dass CXXC1 nicht essentiell für die Interaktion zwischen PRDM9 und der Doppelstrangbruchmaschinerie bei Mäusen ist.

Zudem soll PRDM9 durch die schnell evolvierenden Zinkfinger neue Loci für die Doppelstrangbrüche generieren, da Hotspots mit der Zeit abgebaut werden (Boulton 1997 Lesecque et al. 2014, Baker et al. 2015a, Davies et al. 2016, Smagulova et al. 2016), Weiteres dazu im nächsten Kapitel.

1.1.4 Evolution

PRDM9 gehört der PRDM-Familie an, die sich durch ihre PR/SET-Domäne auszeichnet. Bis auf PRDM11 besitzen alle 17 Paraloge des Menschen Zinkfinger (Fumasoni et al. 2007, Hohenauer und Moore 2012). Bei Primaten ist die Anzahl der Mitglieder unterschiedlich, so weisen *Microcebus murinus* und *Macaca mulatta* 15, *Callithrix jacchus* und *Pan troglodytes* 17 und *Carlito syrichta* 12 Paraloge auf (Vervoort et al. 2016). Während der Primatenevolution soll *PRDM7* durch Duplikation aus *PRDM9* entstanden sein (Fumasoni et al. 2007). Vervoort et al. (2016) datieren die Duplikation entweder auf die Zeit vor der Auftrennung der Haplorrhini oder der Anthrozoidea. PRDM7 und PRDM9 besitzen, anders als die restlichen Mitglieder, eine KRAB-Domäne, die möglicherweise Vorgänger aller KRAB-Domänen ist (Birtle und Ponting 2006, Fumasoni et al. 2007, Hohenauer und Moore 2012). Die Aktivität der Histonmethyltransferasen PRDM7s ist für H3K4me3 gegenüber PRDM9 eingeschränkt und jene für H3K36me3 nicht vorhanden (Blazer et al. 2016).

Die Funktionen der verschiedenen PRDM-Mitglieder sind vielfältig (siehe dazu die Reviews: Fog et al. 2012, Di Zazzo et al. 2013, Mzoughi et al. 2016). Manche sind als Onkogene bekannt, wie zum Beispiel PRDM14 (Dettman und Justice 2008, Dettman et al. 2011, Nishikawa et al. 2007), andere als Tumorsuppressorgene, wie PRDM5 (Deng und Huang 2004, Watanabe et al. 2007, Cheng et al. 2010, Watanabe et al. 2009, Shu et al. 2011, Duan et al. 2007, Review: Fog et al. 2012, Mzoughi et al. 2016). Außerdem sind PRDM-Mitglieder in der Differenzierung involviert, wie PRDM16 bei braunem Fett (Seale et al. 2007, Kajimura et al. 2008, Seale et al. 2008, Seale et al. 2011), PRDM14 in embryonischen Stammzellen (Chia et al. 2010, Ma et al. 2011) und PRDM6 in vaskulären Zellen (Davis et al. 2006, Wu et al. 2008). PRDM2 und PRDM4 sind an Signaltransduktionswegen beteiligt (Medici et al. 1999, Abbondanza et al. 2000, Gazzero et al. 2006, Chittka et al. 1999, 2004, 2012).

PRDM9 ist in vielen Metazoen vorhanden und womöglich vor 700 Millionen Jahren entstanden (Oliver et al. 2009). Beispielhafte Ausnahmen sind im vorherigen Kapitel

erwähnt worden. Baker et al. (2017) haben eine Übersicht über 225 Spezies erstellt, die zeigt, dass *PRDM9* mehrfach und unabhängig aus Genomen verloren ging bzw. eine oder mehrere Domänen fehlen, was zum Funktionsverlust führt.

In Spezies mit vollständigem *PRDM9* ist eine schnelle Evolution im Zinkfingerarray zu verzeichnen (Baker et al. 2017). Dies bezieht sich auf die Anzahl der Zinkfinger, wie auch auf die drei der vier DNA-bindenden Aminosäuren (Oliver et al. 2009, Thomas et al. 2009, Myers et al. 2010). Bereits zwischen Schimpansen und Menschen sind keine übereinstimmenden Hotspots mehr vorhanden und auch das Zinkfingerarray unterscheidet sich komplett (Ptak et al. 2005, Winckler et al. 2005, Myers et al. 2010, Auton et al. 2012). Innerhalb der Primaten, aber auch innerhalb einzelner Spezies, ist eine große Diversität zu sehen und positive Selektion wirksam (Oliver et al. 2009, Parvanov et al. 2010, Auton et al. 2012, Groeneveld et al. 2012, Schwartz et al. 2014, Heerschop et al. 2016, Stevison et al. 2016). Abgesehen von *Microcebus murinus* sind bisher noch keine weiteren Zinkfingerarrays in Strepsirrhini bekannt.

Auch außerhalb der Primatenevolution wurde eine große Variabilität in Anzahl und Sequenz der Zinkfinger entdeckt: bei Mäusen sind über 80 Allele bekannt (Buard et al. 2014, Capilla et al. 2014, Kono et al. 2014) und sowohl Wiederkäuer als auch Vertreter der Pferde besitzen viele Varianten des Zinkfingerarrays (Steiner und Ryder 2013, Ahlawat et al. 2016a, 2016b).

Die Evolution der *PRDM9*-Zinkfinger gilt auch als Lösungsansatz für das *hotspot conversion paradox* (Boulton 1997): Wenn bei der Reparatur der Doppelstrangbrüche das homologe Chromosom als Vorlage dient, werden die darin auftretenden Mutationen in das zu reparierende Chromosom übernommen. Durch doppelstrangbruch-induzierte Genkonversion werden Motive kontinuierlich abgebaut, sodass sie nicht mehr als Hotspots erkannt werden (siehe auch Pineda-Krch und Redfield 2005, Coop und Myers 2007). Die Studien von Lesecque et al. (2014), Baker et al. (2015a), Davies et al. (2016), Smagulova et al. (2016) zeigen, dass die *PRDM9*-Bindungsstellen mit der Zeit verschwinden und neue *PRDM9*-Allele für neue Hotspots bei Mensch und Maus sorgen.

1.1.5 *PRDM9* und Speziation

Nimmt man das biologische Artkonzept (Dobzhansky 1937, Mayr 1963) als Grundlage, geht man bei zwei Taxa von unterschiedlichen Spezies aus, wenn die Nachkommen, deren Eltern aus verschiedenen Gruppen stammen, steril sind.

Mihola et al. (2009) konnten *PRDM9* als das erste Hybridsterilitätsgen in Wirbeltieren nachweisen. Forejt und Iványi (1974) zeigten, konsistent mit Haldanes Regel (Haldane 1922), dass bei einer Kreuzung verschiedener Mausstämmen männliche Nachkommen infertil waren, weibliche hingegen fertil. Dies wurde auf eine Region zurückgeführt, die *Hst1* beinhaltet, das später auch als *Meisetz* und *PRDM9* bekannt wurde, siehe auch Kapitel 1.1.1. Die Fertilität hybrider, eigentlich steriler, (PWD X B6)-Mäuse kann zum Teil durch verschiedene Interventionen aufrechterhalten werden: Durch das Entfernen des B6/*Prdm9*-Allels oder durch Überexpression mittels transgenen Allelen (Mihola et al.

2009, Flachs et al. 2012). Davies et al. (2016) stellten die Fertilität wieder her, indem sie das Zinkfingerarray des murinen B6-Allels durch das menschliche B-Allel ersetzten.

Hayashi et al. (2005) wiesen nach, dass bei Mäusen ein Knockout von *PRDM9* zu Sterilität bei beiden Geschlechtern führt, da die Meiose im Pachytän zum Erliegen kommt. Genauso verhält es sich in den Studien von Sun et al. (2015) mit verkürztem *PRDM9* und Parvanov et al. (2017). In Brick et al. (2012) wird bei *PRDM9*-Knockout-Mäusen eine Verschiebung der Hotspots hin zu H3K4me3-Loci, vor allem in Richtung von Promotoren, gezeigt. Erst kürzlich wiesen Mihola et al. (2019) nach, dass ein solcher Knockout bei Mäusen nicht allgemein zu Sterilität führt. Auch bei Menschen führen homozygote *loss of function* Varianten, genauer zwei verkürzte, nicht funktionale *PRDM9*-Allele nicht zu Sterilität, was anhand einer Mutter aus Großbritannien gezeigt wurde (Narasimhan et al. 2016). Zudem dienen verschiedene Spezies als Beleg, siehe Kapitel 1.1.3, dass *PRDM9* nicht für die Aufrechterhaltung der Fertilität notwendig ist. Bei Menschen sind jedoch zwei Fälle bekannt, bei denen SNPs (*single nucleotide polymorphisms*) in *PRDM9* mit Azoospermie im Zusammenhang standen (Miyamoto et al. 2008, Irie et al. 2009).

Davies et al. (2016) gehen davon aus, dass homologe Chromosomen aufgrund von asymmetrischen *PRDM9*-Bindungsstellen länger brauchen, um sich während der ersten meiotischen Teilungsphase zu finden, was zu Sterilität oder Subfertilität führen kann. Ausführlicher beschrieben, erodieren die Bindungsstellen der jeweiligen *PRDM9*-Allele innerhalb der einzelnen Gruppen, wie im letzten Kapitel zum *hotspot conversion paradox* beschrieben. Bei der hybriden F1-Generation bindet das *PRDM9*-Allel der einen Spezies die Chromosomen der anderen Spezies und vice versa, da dort die präferierten Bindungsstellen noch vorhanden sind. Somit werden seltener die gleichen Bindungsstellen auf beiden homologen Chromosomen gebunden, was die Autor*innen als Asymmetrie der Hotspots bezeichnen. Dies führt dazu, dass die Chromosomen sich langsamer finden, was problematisch für das Voranschreiten der Meiose ist und in Subfertilität oder Sterilität resultieren kann. Dagegen sprechen jedoch verschiedene Studien: Baker et al. (2015a) berichten, dass *PRDM9* neue Loci vorwiegend auf einem Homolog bindet statt auf beiden Homologen.

Eine hochrepetitive, 4,7 Mb (Megabase) große Region auf dem X-Chromosom bei Mäusen ist wohl mitverantwortlich, ob es bei Kreuzungen zu Subfertilität oder Sterilität kommt: Flachs et al. (2012) wiesen nach, dass die männlichen Nachkommen von PWD-Müttern und B6-Vätern steril sind, während die reziproke Kreuzung zu Subfertilität führt. Bhattacharyya et al. (2014) zeigen, dass hierfür die Region auf dem X-Chromosom verantwortlich sein könnte, die *Hstx2* (*hybrid sterility X 2*) genannt wird. Lustyk et al. (2019) können den Bereich auf 2,7 Mb eingrenzen.

Insgesamt ist daher der Mechanismus hinter *PRDM9* und Hybridsterilität noch nicht vollständig aufgeklärt. Die vielen Hinweise, die *PRDM9* für die Meiose unerlässlich erscheinen lassen, sowie diejenigen, die dagegen sprechen, weisen auf einen komplexen Hintergrund hin.

1.1.6 *PRDM9*-assoziierte Krankheiten

Nicht Allelische Homologe Rekombination (NAHR) ist meist das Resultat von inäqualem Crossing-over zwischen nicht homologen *low copy repeats* (LCR, siehe dazu Lupski 1998, Stankiewicz und Lupski 2002). Solche NAHRs können zu Erkrankungen wie Morbus Charcot-Marie-Tooth oder dem Potocki-Lupski- und Smith-Magenis-Syndrom führen (Pentao et al. 1992, Chance et al. 1994, Reiter et al. 1996, Potocki et al. 2000, Potocki et al. 2007, Chen et al. 1997, Shaw et al. 2002). In Regionen, die solche ektopischen Rekombinationen aufweisen, finden sich bei Menschen *PRDM9*-Loci bzw. *PRDM9*-Motive (Myers et al. 2008, Berg et al. 2010, Zhang et al. 2010, Pratto et al. 2014, Altemose et al. 2017).

Seltene *PRDM9*-Allele, die sich in ihrer Zinkfingerregion vom menschlichen A-Allel unterscheiden, könnten in manchen Fällen für die Trisomie 21 verantwortlich sein: In der Meiose der Mütter werden im Vergleich zu Müttern mit A-Allelen weniger Bindungsstellen auf dem langen Arm des Chromosoms 21 vorhergesagt. Als Folge kommt es häufiger zu einem Fehlschlagen der Rekombination dieses Chromosomenpaares, sodass es sich nicht trennt (Oliver et al. 2016).

Obwohl die Zusammenhänge noch nicht klar sind, ist eine Assoziation von *PRDM9* bzw. bestimmte Varianten von *PRDM9* und akuter lymphatischer Leukämie (ALL) bei Kindern vorhanden (Hussin et al. 2013, Woodward et al. 2014).

1.2 *Macaca mulatta* und *Macaca fascicularis*

Rhesusaffen, *Macaca mulatta*, und Javaneraffen, *Macaca fascicularis* gehören der Fascicularis-Gruppe innerhalb der Makaken an und sind nah verwandt (siehe Abbildung 1; Fooden 1976, Morales et al. 1998, Tosi et al. 2003).

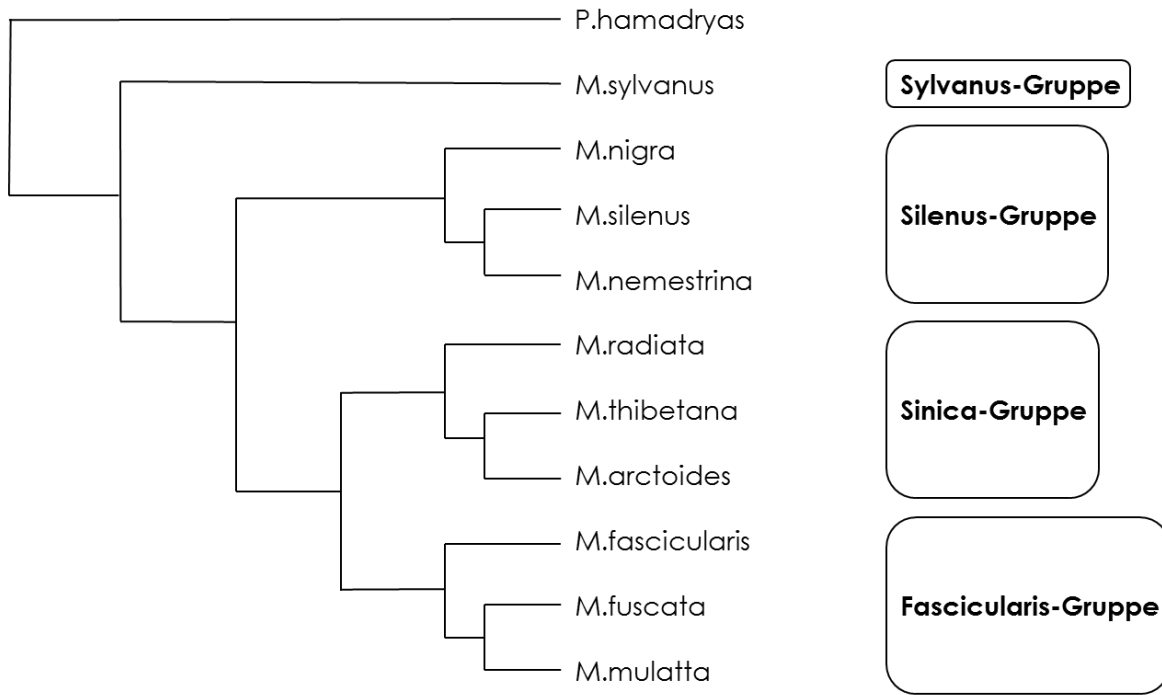


Abbildung 1: Schematische Phylogenie der Makaken mit *Papio hamadryas* als Außengruppe, nach Li et al. (2009b).

Der letzte gemeinsame Vorfahre wird von Stevison und Kohn (2009) auf 1,3 bzw. 3,5 Millionen Jahre von Perelman et al. (2011) zurückdatiert. Ihre Habitate erstrecken sich über große Flächen. *M.mulatta* ist in Indien, China, Afghanistan, Pakistan, Nepal, Bhutan, Bangladesch, Laos, Thailand, Myanmar und Vietnam beheimatet (Fooden 2000). *M.fascicularis* ist auf dem Festland in Laos, Thailand, Myanmar, Vietnam, Kambodscha und Malaysia zu finden, sowie auf den indonesischen Inseln Java, Borneo bis Osttimor und außerdem auf den Philippinen, jedoch nicht auf Sulawesi (Fooden 1995). Die Zone mit der stärksten Hybridisierung der beiden Spezies befindet sich zwischen dem 15. und 20. nördlichen Breitengrad in den Ländern Laos, Vietnam, Myanmar und Thailand, wobei die Vermischung darüber hinausgeht (Fooden 1995 und 2000, Hamada et al. 2006, Bunlungsup et al. 2017). Äußerlich unterscheiden sie sich durch Körpergröße und –gewicht, Fellfarbe und Schwanzlänge: während *M.mulatta* größer und schwerer als *M.fascicularis* ist, ist der Schwanz kürzer. Der Rhesusaffe hat ein braunes Fell und ein rötliches Gesicht; die Fellfarbe von *M.fascicularis* ist grau bis braun und das Gesicht ist pink-braun (Rowe et al. 1996). Die Schwanzlänge bzw. die Rate von Schwanzlänge zu Kopf- und Rumpflänge sowie die Fellfarbe von Hybriden bewegen sich hinsichtlich dieser Eigenschaften zwischen den beiden Arten (Fooden 1997).

Die Introgression, eine Übertragung genetischer Informationen auf eine andere Spezies, erfolgt vermehrt von *M.mulatta* auf *M.fascicularis* (Tosi et al. 2002, Hamada et al. 2006, Bonhomme et al. 2009, Bunlungsup et al. 2017). Das könnte laut Bunlungsup et al. (2017) folgende Gründe haben: zum einen sind *M.mulatta*-Männchen größer als ihre *M.fascicularis*-Konkurrenten, weshalb sie entweder attraktiver auf die *M.fascicularis*-Weibchen wirken oder sie sich ihnen aufdrängen (Hamada et al. 2006). Ein anderer Aspekt betrifft die Fortpflanzungszeiten. Während weibliche Rhesusaffen saisonal gebären, sind die Geburten der *M.fascicularis*-Weibchen über das ganze Jahr verteilt (Herndon 1983, Weinbauer et al. 2008, Kavanagh und Laursen 1984, van Schaik und Noordwijk 1985).

Ein Vergleich der Genome zweier Rhesusaffen und eines Javaneraffen ergibt, dass sich die größte Divergenz bei indischen *M.mulatta* und *M.fascicularis* zeigt; die Divergenz zwischen dem chinesischen und dem indischen *M.mulatta* ist zwar niedriger als bei dem chinesischen *M.mulatta* und *M.fascicularis*, aber vergleichbar hoch (Yan et al. 2011). Ein weiteres Zeichen für Introgression ist, dass ein Fünftel der Orthologen mehr Gemeinsamkeiten zwischen dem chinesischen *M.mulatta* und *M.fascicularis* aufweist als zwischen den beiden *M.mulatta* (Yan et al. 2011). *PRDM9* ist bei Makaken auf Chromosom 6, *PRDM7* auf Chromosom 20 lokalisiert (Fumasoni et al. 2007). Das *PRDM9*-Zinkfingerarray von *M.mulatta* und *M.nemestrina* war Teil verschiedener Studien (Oliver et al. 2009, Schwartz et al. 2014, Patel et al. 2017).

1.3 Zielsetzung

Die vorliegende Arbeit erweitert die Analyse der *PRDM9*-Merkmalsevolution für die Strepsirrhini. Die Sequenzierung des Zinkfingerarrays dreier madagassischer Lemuren mit *Daubentonia madagascariensis* als tiefste Verzweigung komplementiert meine Arbeiten zu Tarsiern und der *PRDM9*-Merkmalsevolution innerhalb der Haplorrhinen. Zudem werden die bisher unbeschriebenen Zinkfingersequenzen zwischen den drei strepsirrhinen Arten und *Microcebus murinus* verglichen.

Im Rahmen einer Merkmalsevolutionsanalyse zweier Mitglieder der *PRDM*-Genfamilie werde ich eine funktionell putativ wichtige Paralogenisierung innerhalb der Primatendivergenz analysieren. Die Datierung der Genduplikation, die zur Entstehung *PRDM7s* führte, soll mithilfe von phylogenetischen Rekonstruktionen paraloger und orthologer *PRDM*-Sequenzen näher bestimmt werden. Unterstützend sollen BLAST-Analysen und Vergleiche der Divergenz zwischen Sequenzen fungieren. Eine 89 bp-Insertion innerhalb eines *PRDM7*-Exons wird außerdem auf ihre Anwesenheit/Abwesenheit innerhalb der Primatendivergenz geprüft und darauf, ob die entsprechenden Muster somit als Unterscheidungsmerkmal zwischen *PRDM7* und *PRDM9* dienen kann. Um die Fixation der Insertion festzustellen, wird diese Sequenz bei mehreren nicht verwandten Schimpansen und Gorillas überprüft. Mit einer genaueren Datierung sollen mögliche Funktionen *PRDM7s* diskutiert werden, wie jene, dass das duplizierte Gen die Aufgaben *PRDM9s* übernehmen könnte.

Das Hauptprojekt stellt die Charakterisierung der PRDM9-Bindungsstellen zweier Makakenarten mittels eines explorativen Analyseansatzes der ersten *in vivo* ChIP-Daten nicht-humaner Primaten dar. Unterstützen sollen die gleichzeitig erstellten Daten von H3K4me3 und DMC1, die mit Rekombinationshotspots bzw. Doppelstrangbrüchen assoziiert werden. Dabei sollen die Loci auf bekannte Eigenschaften von Rekombinationshotspots wie GC-Gehalt und ihrer Position entlang der Chromosomen überprüft werden. Des Weiteren sollen Übereinstimmungen mit genomischen Loci wie repetitive Elemente oder regulatorischen Einheiten untersucht werden sowie die Bindungsstellen in der pseudoautosomalen Region in Hinblick auf unterschiedlichen rekombinatorischen Beitrag von PRDM9 bei Maus und Mensch in diesem Bereich. Hinweisen auf Beteiligung an Erkrankungen und auf mögliche weitere Funktionen werden überdies nachgegangen. Auch Motive innerhalb der von PRDM9 gebundenen Loci sollen erstellt und zwischen den beiden Arten *Macaca mulatta* und *Macaca fascicularis* verglichen werden. Weitere Unterschiede und Gemeinsamkeiten zwischen den Spezies in den genannten Eigenschaften werden dahingehend überprüft, inwieweit die Erosion der Rekombinationshotspots fortgeschritten ist. Dies soll insgesamt auch in Hinblick auf die entsprechenden Implikationen für den Prozess einer Speziation diskutiert werden. Zuletzt sollen mögliche Protein-Interaktionspartner von PRDM9 mit einer Immunopräzipitation und anschließender Massenspektrometrie ermittelt werden.

Ziel dieses Hauptprojekts ist es daher, mittels eines komparativen Ansatzes die Unterschiede und Gemeinsamkeiten der *PRDM9*-Bindungsstellen, die für putative Rekombinationshotspots stehen, zweier rezenter und nahverwandter Primatenarten, die Hybride hervorbringen können, zu beschreiben.

2. PRDM9 in der Primatenevolution

2.1 Die PRDM9-Zinkfinger der Strepsirrhini

2.1.1 Material und Methoden

2.1.1.1 Labor


Die DNAs der beiden Strepsirrhini *Cheirogaleus medius* und *Daubentonia madagascariensis* lagen in der Arbeitsgruppe Zischler vor. Für *Eulemur coronatus* war Milzgewebe vorhanden, aus dem mittels des QIAamp® DNA Mini Kit von Qiagen DNA extrahiert wurde. Die PRDM9-Sequenzen von *Microcebus murinus* und *Mus musculus* entstammen der Datenbank Ensembl, Version 91 (Dezember 2017), erstere wurde unter PRDM7 geführt.

Für die Sequenzierung des letzten Exons des Gens PRDM9 wurde die Sequenz zunächst über eine Polymerasekettenreaktion, kurz PCR, mithilfe des Taq PCR Core Kit von Qiagen vervielfältigt. Ansatz und Programm sind den Tabellen 1 und 2 zu entnehmen. Die Oligonukleotide 5'-GCCCAGAACAGGCCAGA-3' und 5'-TTCTGGTCTCTTTACTCTTTGG-3' (Tm: 50°C) wurden für alle PCRs verwendet außer für die der Spezies *Cheirogaleus medius*. Die dafür verwendeten Primer waren 5'-GGGATCAGAATCAGGAGCAG-3' und 5'-CTCTAGTCATGAAAGTAGAGGATTTG-3' (Tm: 54°C). Die Auftrennung nach Größe der PCR-Produkte erfolgte mit einer Agarosegelelektrophorese. Eine 100 bp (Basenpaar) plus-Leiter von Fermentas und die PCR-Produkte wurden dafür mit Loading Dye auf das 1,5%-ige Agarosegel aufgetragen, das in einer Endkonzentration von 0,5 µg/ml mit Ethidiumbromid versetzt war. Das jeweils kürzere PCR-Produkt von *Daubentonia madagascariensis* und *Eulemur coronatus* wurde mit dem QIAquick® Gel Extraction Kit von Qiagen aus dem Agarosegel isoliert. Die Aufreinigung des Produkts von *Cheirogaleus medius* fand mit den Enzymen Exonuklease I und Fast Alkaline Phosphatase statt. Die Sequenzierung übernahm die Firma Macrogen.

Tabelle 1: PCR-Ansatz, 30 µl (Mengenangaben in µl).

Mix, unten		Trennung durch Wachsschicht	Mix, oben	
ddH ₂ O	6,4		ddH ₂ O	16,85
10xPuffer	1		10xPuffer	2
Primer, forward	1		Taq- Polymerase	0,15
Primer, reverse	1			
dNTPs	0,6			
10			19 + 1 µl DNA oder WGA	

Tabelle 2: Thermocycler-Programm.

PCR- Schritt	Temperatur in °C	Zeit	Zyklen
Initiale Denaturierung	94	3 min	
Denaturierung	94	40 sec	
Annealing	T _m *	1 min	
Elongation	72	1 min	
Endelongation	72	5 min	
Kühlen	4	∞	

*Primerspezifische Schmelztemperatur

Jessica Ries führte im Rahmen ihrer Bachelorarbeit und unter meiner Betreuung die DNA-Extraktion für *Eulemur coronatus* durch sowie die PCR und weitere Vorbereitungen für die Sequenzierung von *Cheirogaleus medius*.

2.1.1.2 Datenanalysen

In BioEdit 7.2.5 (Hall 1999) wurden die Sequenzen von *Daubentonia madagascariensis*, *Cheirogaleus medius* und *Eulemur coronatus* überprüft und die Zinkfinger (Znf) der drei genannten Strepsirrhini und von *Microcebus murinus* vereinzelt. Zusammen mit dem degenerierten Znf von *Mus musculus* wurde ein Alignment der C2H2- und der degenerierten Znf erstellt. Dabei fungierte das Kodon für das erste Cystein als Startpunkt erstgenannter Znf. Da die zweitgenannten Znf durch den Verlust des ersten Cystein degeneriert waren, war deren erstes Kodon dasjenige für das zweite Cystein. Das Alignment stellte zusammen mit den Sequenzdaten aus Schwartz et al. 2014 und Heerschop et al. 2016 die Grundlage für die Analysen dar. Dazu wurden zudem die vier Triplets entfernt, die für die bindenden Aminosäuren kodieren. Zinkfinger, die mehrfach innerhalb einer Art vorkamen, wurden bis auf ein Exemplar aussortiert. Das Substitutionsmodell für die Baumrekonstruktion wurde mit jModelTest 2.1.6 (Guindon und Gascuel 2003, Darriba et al. 2012) bestimmt. Die Erstellung der Phylogenie erfolgte nach dem Optimalitätskriterium der Maximum Likelihood in PhyML 3.0 (Guindon und Gascuel 2003). Die Unterstützungswerte wurden mit dem Shimodaira-Hasegawa *approximate likelihood-ratio test* (SH-aLRT) berechnet. Abgesehen vom eingegebenen Substitutionsmodell, SYM+G als Ergebnis des jModelTests, wurden die Standardeinstellungen belassen. Der degenerierte Znf von *Mus musculus* diente als Außengruppe. Knoten mit einem Unterstützungswert unter 0,5 wurden in TreeGraph 2.2.0-407 beta (Stöver und Müller 2010) entfernt. In FigTree 1.3.1 (Rambaut 2009) wurde die kreisförmige Darstellung der Phylogenie erstellt.

2.1.2 Ergebnisse

Sowohl *Daubentonia madagascariensis* (*Dm*) als auch *Microcebus murinus* (*Mm*) weisen vier C2H2-Znf auf zusätzlich zum degenerierten. *Cheirogaleus medius* (*Cm*) und *Eulemur coronatus* (*Ec*) verfügen über acht bzw. zehn der klassischen Znf. Nach der Entfernung der bindenden Kodons und der mehrfachen Znf, reduzieren sich die Zahlen auf zwei (*Dm* und *Mm*), drei (*Cm*) und sieben (*Ec*).

Die Phylogenie in Abbildung 2 zeigt, dass die degenerierten und C2H2-Znf sowohl der Tarsier als auch der Anthropoidea distinkte Cluster bilden. Im Gegensatz dazu gruppieren sich die strepsirrhinen Znf nicht als Klade, sondern meist als Schwestergruppen, allerdings getrennt von Tarsiern und Anthropoidea. Drei Ausnahmen bei den C2H2-Znf sind zu erwähnen. Ein Znf von *Daubentonia madagascariensis* ist im Cluster der Anthropoidea zu finden, wohingegen ein Znf der Hominoidea im Abschnitt der Strepsirrhini verortet ist. Ein Znf von *Cheirogaleus medius* bildet eine Außengruppe zu den übrigen C2H2-Znf.

Im Vergleich zur Baumrekonstruktion mit dem degenerierten Zinkfinger ausschließlich eines Strepsirrhinen als Außengruppe in Abbildung 3, die eine Variante der Phylogenie aus Heerschoop et al. 2016 zeigt, ist die Auflösung der Haplorrhini akkurater im Hinblick auf den allgemeingültigen Artenbaum der Primaten: Die klassischen Znf der Tarsier fallen zum einen außerhalb der Gruppe der Anthropoidea und nicht mehr zusammen mit platyrrhinen Znf.

Bei der Berücksichtigung der Aufspaltungszeiten der Taxa, vgl. Tabelle 3, zeigt sich, dass erst eine Zeitspanne von über 80 Millionen Jahren (MY) zu distinkten Gruppen führt, wie bei der Aufspaltung von Strepsi- und Haplorrhini (87 MY) sowie von Tarsier und Anthropoidea (81 MY).

Tabelle 3: Aufspaltungszeiten der in der Phylogenie verwendeten Taxa. Die Daten sind entnommen aus Perelman et al. 2011.

Knotenpunkt	Abspaltungszeit in Millionen Jahren
Cercopithecoidea/Hominoidea	32
Platyrrhini/Catarrhini	43
Tarsier/Anthropoidea	81
Haplorrhini/Strepsirrhini	87
Dm/EcCmMm	59
Ec/CmMm	39
Cm/Mm	25

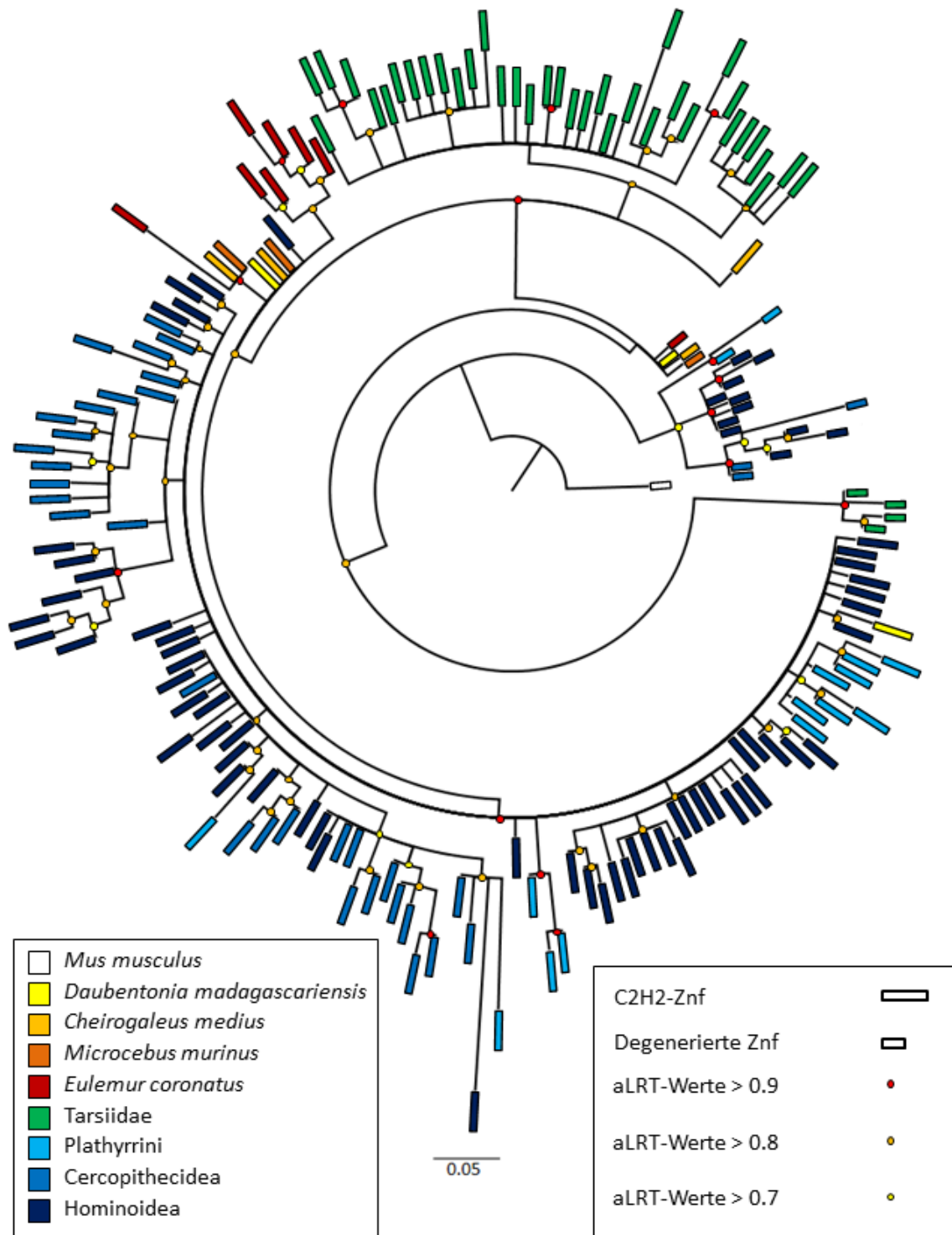


Abbildung 2: Phylogenie mit degenerierten (kurz) und C2H2-Znf (lang) von Vertretern der Primaten mit dem degenerierten *Mus musculus* Zinkfinger als Außengruppe, erweiterte Variante der Baumrekonstruktion aus Heerschop et al. (2016). Die Taxa sind farblich gekennzeichnet und die Knotenpunkte mit aLRT-Werten über 0.7, 0.8 und 0.9 sind gelb, orange bzw. rot dargestellt, siehe dazu die Legenden.

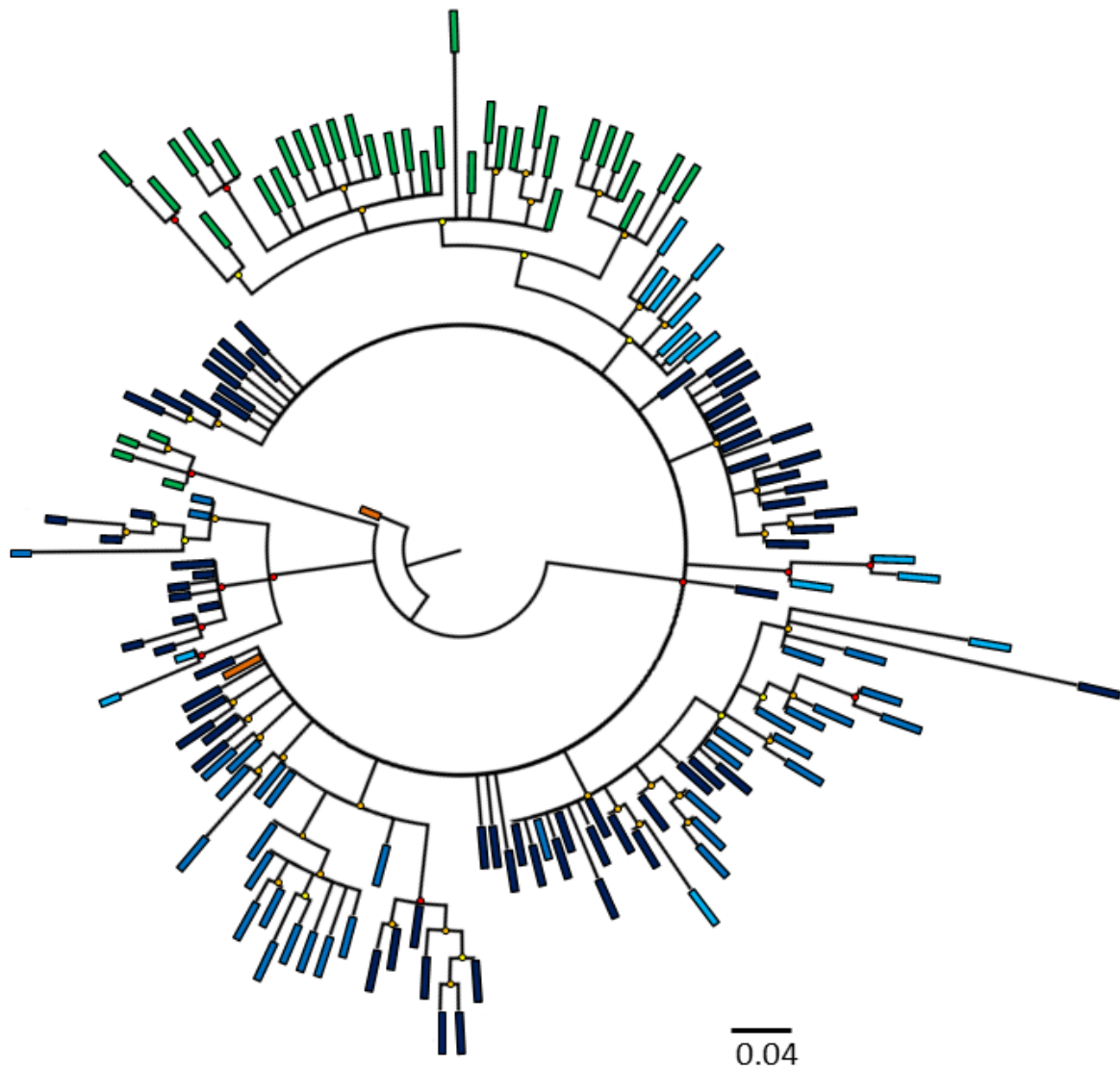


Abbildung 3: Phylogenie mit degenerierten (kurz) und C2H2-Znf (lang) von Haplorrhini mit *Microcebus murinus* Zinkfinger als Außengruppe, abgeändert entnommen aus Heerschop et al. (2016). Die Taxa sind farblich gekennzeichnet und die Knotenpunkte mit aLRT-Werten über 0.7, 0.8 und 0.9 sind gelb, orange bzw. rot dargestellt, vergleiche dazu die Legenden von Abbildung 2.

2.1.3 Diskussion

Die Auflösung der Phylogenie der *PRDM9*-Zinkfinger hat sich durch das Hinzufügen weiterer strepsirrhiner Znf erhöht.

Es ist anzunehmen, dass die konservierte Zinkfingerstruktur nicht dazu geeignet ist rezenterere Aufspaltungen als 80 MY abzubilden. In diesem Projekt wurden zusätzlich die schnell evolvierenden, für die Bindung der DNA verantwortlichen Kodons entfernt. Da diese unter positivem Selektionsdruck stehen (Oliver et al 2009, Thomas et al. 2009, Myers et al. 2010), ist anzunehmen, dass sie die Phylogenie verzerren könnten. Schwartz et al. 2014, auf deren Baumrekonstruktion die vorliegenden Phylogenien beruhen, haben die bindenden Kodons aus diesem Grund ausgeschlossen.

2.2 Datierung des Duplikationsevents PRDM7/PRDM9

2.2.1 Material und Methoden



2.2.1.1 Labor

Die WGAs (*whole genome amplification*) von *Daubentonia madagascariensis*, *Cheirogaleus medius*, *Carlito syrichta*, *Callithrix jacchus*, *Macaca mulatta*, *Hylobates lar*, *Pongo abelii*, *Gorilla gorilla gorilla*, *Pan troglodytes* und *Homo sapiens* lagen in der Arbeitsgruppe vor. Weitere DNA-Proben von *Pan troglodytes* und *Gorilla gorilla gorilla* stellte Dr. Ernst Verschoor vom *Biomedical Primate Research Centre* in Rijswijk bereit.

Es wurden PCRs durchgeführt, um festzustellen, wann die für *PRDM7* typische Insertion im vorletzten Exon auftrat. Sowohl Ansatz als auch Cyclerprogramm sind im Kapitel 2.1.1.1 nachzuschlagen. Die verwendeten Primer waren 5'-AACTGTGCCCGGGATGAT-3' und 5'-ACTTGCTGCCCCACTTGAT-3' mit einer Schmelztemperatur von 51°C. Ausgewählte Vertreter der Primaten (*D.madagascariensis*, *C.medius*, *C.syrichta*, *C.jacchus*, *H.lar*, *P.abelii*, *G.g.gorilla*, *P.troglodytes* und *H.sapiens*) und mehrere Schimpansen (Frits, Nana, Orlando, Regina) und Gorillas (Miliki, N'diki, Viringika, Vinoto, Vizuri) standen als Material zur Verfügung.

Weitere PCRs wurden für das letzte Exon inkl. Zinkfingerarray angesetzt, um das Vorhandensein von *PRDM7* nachzuweisen. Dazu wurden die Spezies *M.mulatta*, *H.lar*, *P.abelii*, *G.g.gorilla*, *P.troglodytes* und *H.sapiens* herangezogen. Die Primer hierfür waren 5'-ACATCTACCCTGACCAAAAAC-3' und 5'-CGGATTTGTTTAATCAGTTATTTTC-3' mit einer Schmelztemperatur von 50 °C für *H. lar*, 54°C für *G.g.gorilla* und 52 °C für die weiteren Primaten. Bei dem Ansatz handelt es sich um den Standardansatz, siehe Kapitel 2.1.1.1. Anstelle des üblichen Cyclerprogramms wurde eine Touch down-PCR durchgeführt, siehe Tabelle 4. In den ersten 20 Zyklen wird die Schmelztemperatur um jeweils 0,5°C verringert, um eine spezifischere Bindung der Primer an das *template* zu erhalten.

Tabelle 4: Thermocycler-Programm, Touch down-PCR.

PCR- Schritt	Temperatur in °C	Zeit	Zyklen
Initiale Denaturierung	94	3 min	
Denaturierung	94	40 sec	 20, -0,5°C
Annealing	Tm*+10	1 min	
Elongation	72	1 min	
Denaturierung	94	40 sec	 20
Annealing	Tm*	1 min	
Elongation	72	1 min	
Endelongation	72	5 min	
Kühlen	4	∞	

*Primerspezifische Schmelztemperatur

Für eine anschließende, stichprobenartige Sequenzierung bei Macrogen wurden folgende PCR-Produkte aus dem Gel extrahiert, vgl. auch Kapitel 2.1.1.1: drei Banden von *H. lar* und *M. mulatta*, sowie die kleinste Bande von *P. troglodytes*. Außerdem wurde genauso mit beiden Banden von *P. troglodytes* und *H. sapiens* aus der PCR mit dem vorletzten Exon verfahren, wie auch mit einer „Extra“-Bande von *C. jacchus*.

2.2.1.2 Datenanalysen

2.2.1.2.1 Sequenzen aus dem 1000-Genome-Projekt

Als Ergänzung zu den PCRs, die sich der Insertion in *PRDM7* annehmen, wurden menschliche Daten vom 1000-Genome-Projekt (The 1000 Genomes Project Consortium 2015) herangezogen. Sowohl die Regionen mit *PRDM7* als auch jene mit *PRDM9* wurden mittels eines *shell*-Skriptes gezogen. Sequenzen, die verkürzt waren, wurden entfernt. Dies traf ausschließlich auf einige von *PRDM7* zu, wobei Anfang- und Endregionen betroffen waren. Selbsterstellte Python-Skripts entnahmen alle Exons, konkatenierten sie und stellten Sets aus nicht identischen Sequenzen zusammen. Dabei war das Zinkfingerarray im letzten Exon nicht enthalten. Ein Alignment dieser Sequenzen wurde in BioEdit 7.2.5 (Hall 1999) angefertigt.

2.2.1.2.2 Erstellung der Phylogenien

Als weitere Überprüfungsmöglichkeit, wann *PRDM7* sich abgespalten hat, wurden vier Phylogenien mit verschiedenen Sequenzen und diversen Primatenspezies als Grundlage erstellt. Zum einen wurden 1002 Basen der cDNA-Sequenz (*complementary DNA*) ab dem Startkodon mit KRAB-Domäne und dem Großteil der SET-Domäne herangezogen, zum anderen wurden die intronischen Regionen eins, zwei, sieben und neun zusammengeschlossen. Weitere Introns wurden aufgrund unspezifischer Basen, „Ns“, ausgeschlossen oder auch, wenn die Anzahl der Basen der kürzesten Sequenz weniger als 40% der längsten Sequenz aufwies. Außerdem wurden jeweils 2000 Basen vor dem Startkodon und nach dem Stoppkodon für zwei Phylogenien herausgesucht. Dabei wurden „Ns“ herausgenommen. Die Daten entstammten alle der ensembl-Datenbank, Version 91 (Dezember 2017), vgl. Tabelle 5. Spezies und Homologe, die in den jeweiligen Baumrekonstruktionen enthalten sind, sind in Tabelle 5 mit Abkürzungen gekennzeichnet. Die Generierung der Phylogenien erfolgte über phylogeny.fr mit folgenden Einstellungen: Das Alignment wurde mit MUSCLE (*full mode*) erstellt und es wurden 100 Bootstraps gezogen. Die newick-Dateien wurden heruntergeladen und in Treegraph 2.2 (Stöver und Müller 2010) geöffnet, um die Außengruppe (Strepsirrhini) zu bestimmen und Knoten unter 0,5 kollabieren zu lassen. Um eine Skala zu erhalten, wurden die abschließenden Bilder in Figtree 1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>, Rambaut 2009) erstellt.

Tabelle 5:Ensembl Gene IDs.

Primat	<i>PRDM7</i>		<i>PRDM9</i>	
<i>Homo sapiens</i>	ENSG00000126856	IK53	ENSG00000164256	IK53
<i>Pan troglodytes</i>			ENSPTRG00000028674	IK53
<i>Pan paniscus</i>			ENSPPAG00000031671	IK53
<i>Gorilla gorilla gorilla</i>	101135995 **	IK53	ENSGGOG00000024584	IK53
<i>Pongo abelii</i>			ENSPPYG00000015366	IK53
<i>Nomascus leucogenys</i>	100583589 **	53	ENSNLEG00000031321	53
<i>Macaca mulatta</i>	100427538**	53	ENSMMUG00000003772	IK53
<i>Macaca fascicularis</i>			ENSMFAG00000003001	IK53
<i>Papio anubis</i>	101024444 **	53	ENSPANG00000013143	IK53
<i>Rhinopithecus roxellana*</i>			ENSRROG000000031264	53
<i>Rhinopithecus bieti*</i>			ENSRBIG000000036346	53
<i>Chlorocebus angolensis palliatus</i>			ENSCSAG00000017229	IK53
<i>Cercocebus atys*</i>			ENSCATG000000041590	53
<i>Cebus capucinus imitator *</i>			ENSCCAG000000023876	IK53
<i>Callithrix jacchus</i>			ENSCJAG00000010419	IK53
<i>Carlito syrichta *</i>			ENSTSYG000000031191	IK
<i>Propithecus coquereli *</i>			ENSPCOG000000013284	IK
<i>Microcebus murinus</i>			ENSMICG000000037613	IK53

*: scaffold; **: Entrez gene ID, I: intron, K: KRAB/SET, 5:5', 3: 3'

2.2.1.2.3 Sequenzvergleich mit needle

Um die Sequenzidentität zu prüfen, wurden die cDNA-Sequenzen von *M.murinus* (ENSMICT000000031671.2), *C.jacchus* (ENSCJAT00000020315.2), *M.mulatta* (ENSMMUT00000005338.3) und *H.sapiens* (ENST00000296682.3, ENST00000449207.6), jeweils *PRDM9* und für letztere Spezies auch *PRDM7*, aus ensembl Version 91 zusammengestellt. Untersucht wurde die Teilsequenz ab dem Startkodon bis zur Base 1098, gewählt aufgrund der großen Übereinstimmung. *M.murinus* wies im Alignment eine Deletion auf, dieser Abschnitt wurde bei allen Spezies entfernt, sodass letztendlich 1047 Basen zur Verfügung standen. Auf https://www.ebi.ac.uk/Tools/psa/emboss_needle/ wurden die Sequenzen paarweise verglichen.

2.2.1.2.4 BLAST der PRDM7-Sequenzen in Cercopithecoidea

Die sequenzierten Proben wurden über das BLAST-Programm von NCBI gesucht.

Das Chromosom 20 von *M. mulatta* (Mmul_8.0.1) und *P. anubis* (Panu_3.0) wurde von NCBI heruntergeladen. Die orthologen Sequenzen der humanen *PRDM7*-Exons (Ensembl

Version 91; Zerbino et al. 2018) wurden mit blast-2.2.28+ (<https://www.ncbi.nlm.nih.gov/books/NBK131777/>, Camacho 2013) extrahiert.

2.2.1.2.5 ensembl-Suche

Über die Suchfunktion bei ensembl wurden *PRDM7* und *PRDM9* bei allen verfügbaren Primaten gesucht.

2.2.2 Ergebnisse

2.2.2.1 PCR der internen *PRDM7*-Duplikation

Die in Fumasoni et al. (2007) beschriebene Insertion im vorletzten Exon *PRDM7s* tritt lediglich bei Schimpanse und Mensch auf. In Abbildung 4 (links) ist der exonische Bereich, in dem die 89 Basen lange Insertion bei *PRDM7* liegt, bei verschiedenen Vertretern der Primaten dargestellt, wobei weitere, größere Banden nur bei den bereits genannten Spezies vorliegt. Die Sequenz der Bande über 500 bp von *C. jacchus* ergibt bei einem Blast über NCBI die Ausgabe: *PREDICTED: Callithrix jacchus glyceraldehyde-3-phosphate dehydrogenase pseudogene (LOC103789031), misc_RNA* und stellt somit kein weiteres PRDM-Pendant dar. Der Blast der sequenzierten Produkte von Mensch und Schimpanse, kleine und große Banden, stimmen mit Fumasoni et al. (2007) überein, die kleinere Bande stellt *PRDM9* und die größere *PRDM7* dar. Das Gelbild in Abbildung 4 (rechts) bestätigt, dass bei mehreren Vertretern der Spezies *G.g.gorilla* und *P.troglodytes*, die erste Art weiterhin nur eine Bande aufweist und Schimpansen durchgehend mindestens zwei bis drei Produkte zeigen. Das Alignment, das mit allen nicht identischen *PRDM7* und *PRDM9* Sequenzen aus dem 1000-Genome-Projekt erstellt wurde, weist bei allen 246 *PRDM7* Sequenzen die Insertion von 89 Basen auf im Gegensatz zu den 74 *PRDM9* Sequenzen.

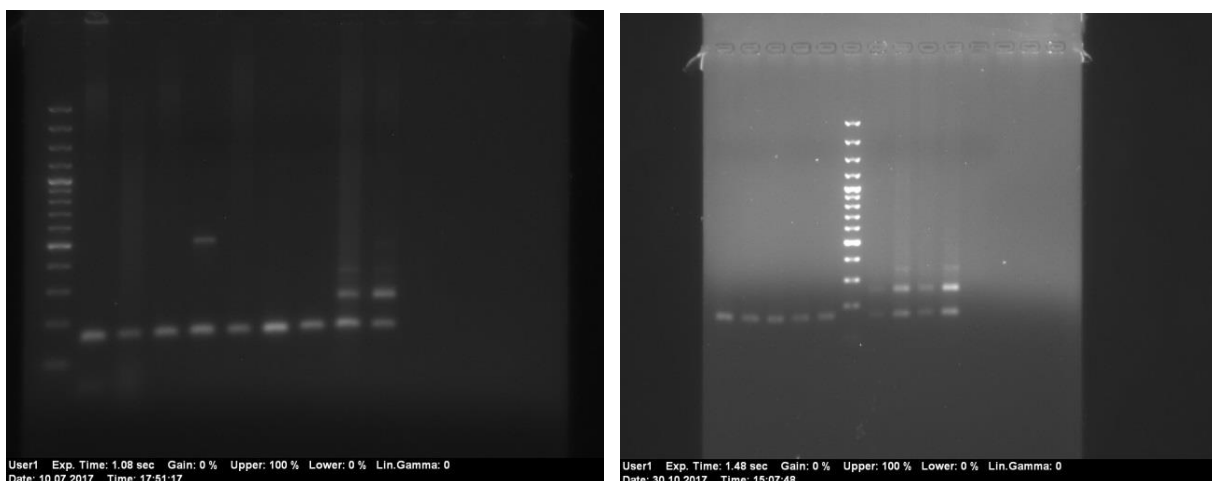


Abbildung 4: Agarosegelbilder mit einem Ausschnitt aus dem vorletzten Exon von *PRDM7* und *PRDM9*. links: *D.madagascariensis* - *C.medius* - *C.syrichta* - *C.jacchus* - *H.lar* - *P.abelii* - *G.g.gorilla* - *P.troglodytes* - *H.spaniens*; rechts: fünf Vertreter der Spezies *G.g.gorilla* (Miliki, N'diki, Viringika, Vinoto, Vizuri) und vier Vertreter von *P.troglodytes* (Frits, Nana, Orlando, Regina). Der Marker ist jeweils die 100 bp+ Leiter von Fermentas.

2.2.2.2 PCR des Zinkfingerarrays

Die PCRs zum Exon mit dem Zinkfingerarray zeigen bei den Hominoidea nach der Auftrennung der PCR-Produkte auf dem Agarosegel mindestens zwei Banden, nur *H. lar* weist eine dritte, hellere Bande auf. Die kleineren Produkte entsprechen dabei *PRDM7*, die größeren *PRDM9*, siehe Abbildung 5 (links). In der Abbildung rechts sind die drei Banden, die das Exon bei *M.mulatta* bilden, erkennbar, dabei handelt es sich um die erste Probe. Die Sequenzierung der jeweils drei Banden von *H. lar* und *M.mulatta* sowie des kurzen Produkts von *P.troglodytes* und das Blasten zeigten, dass die kurze Sequenz *PRDM7* darstellt und die beiden größeren zwei Allele von *PRDM9*.

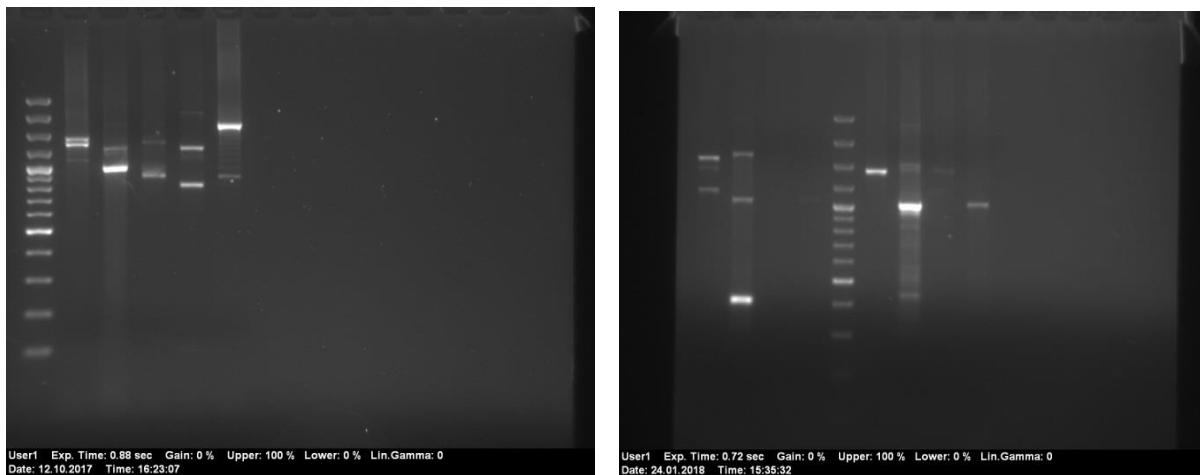


Abbildung 5: Agarosegelbilder mit dem letzten Exon von *PRDM7* und *PRDM9*. links: *H. lar* - *P. abelii* – *G.g. gorilla* - *P. troglodytes* - *H. sapiens*; rechts: das PCR-Produkt ganz links zeigt *M. mulatta*. Der Marker ist jeweils die 100 bp+ Leiter von Fermentas.

2.2.2.3 *PRDM7/PRDM9*-Phylogenien

Die vier erstellten Phylogenien zeigen weitestgehend eine übereinstimmende Topologie auf. Die *PRDM7*-Sequenzen spalten sich dabei jeweils vor den *PRDM9*-Sequenzen der Catarrhini ab und nach den *PRDM9*-Sequenzen der Platyrrhini. Beispielhaft ist die Baumrekonstruktion mit den 2000 Basen nach dem Stoppkodon als Grundlage in Abbildung 6 zu sehen. Die weiteren Phylogenien sind im Anhang zu finden (S. 71-73). In der Rekonstruktion mit den 2000 Basen vor dem Startkodon fallen die *PRDM7*-Sequenzen nicht komplett zusammen, jedoch zeigen sie als Schwestergruppe das gleiche Schema wie oben beschrieben.

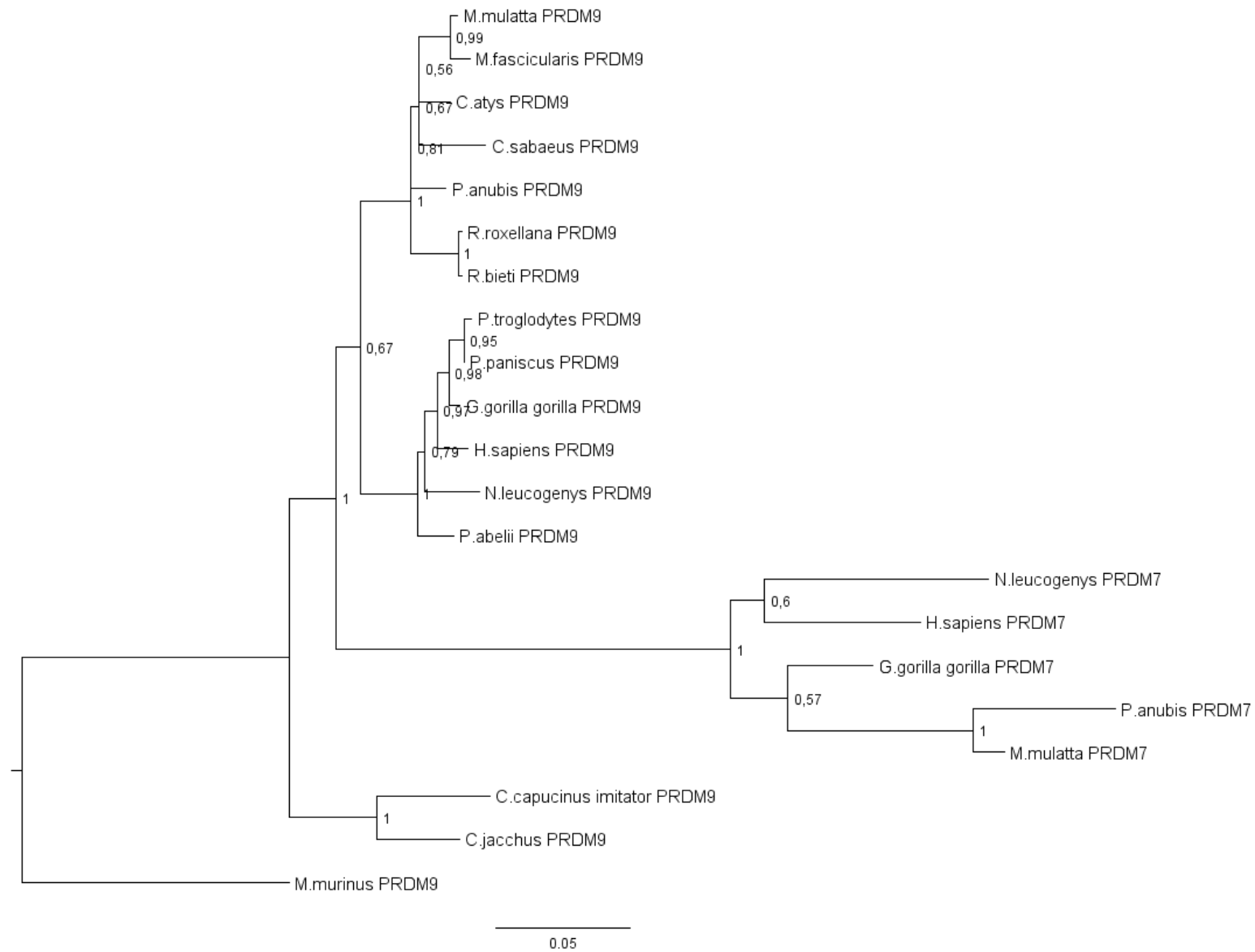


Abbildung 6: Phylogenie basierend auf den 2000 Basen nach dem Stoppkodon (3').

2.2.2.4 Vergleich von Teilsequenzen mit needle

Die Übereinstimmung der Sequenzen nimmt bei *PRDM9* mit höherem Verwandtschaftsgrad der Spezies zu. Bei Betrachtung aller Sequenzidentitäten im paarweisen Vergleich ist das humane *PRDM7* bezüglich der Prozentzahl zwischen *C.jacchus* und *M.mulatta* (*PRDM9*) anzusiedeln, siehe Tabelle 6. Außerdem zeigen die Sequenzidentitäten von Strepsirrhini zu den Haplorrhini, von Platyrrhini zu Catarrhini, sowie *PRDM7* (*H.sapiens*) zu *PRDM9* (*M.mulatta*, *H.sapiens*) jeweils ähnliche Werte.

Tabelle 6: Sequenzidentität zwischen den verschiedenen Primatensequenzen in Prozent.

	HOM- <i>PRDM9</i>	MAC- <i>PRDM9</i>	HOM- <i>PRDM7</i>	CAL- <i>PRDM9</i>	MIC- <i>PRDM9</i>
HOM- <i>PRDM9</i>	-	98,6	96,8	94,7	90,3
MAC- <i>PRDM9</i>	98,6	-	96,8	94,2	89,7
HOM- <i>PRDM7</i>	96,8	96,8	-	94,8	90,1
CAL- <i>PRDM9</i>	94,7	94,2	94,8	-	89,7
MIC- <i>PRDM9</i>	90,3	89,7	90,1	89,7	-

2.2.2.5 BLAST gegen Chromosom 20

Die Suche nach den orthologen Sequenzen der humanen *PRDM7*-Exons im Chromosom 20 zweier Papionini als Vertreter der Cercopithecoidea ergab für die ersten fünf Exons und für das vorletzte Exon keine Treffer. Für die restlichen war je ein Treffer und für das letzte Exon zwei sich überlappende Regionen, die zusammen das komplette Exon abbilden, zu verzeichnen (siehe Tabelle 7). Dies gilt sowohl für *M.mulatta* als auch für *P.anubis*.

Tabelle 7: Übereinstimmungen der humanen *PRDM7*-Exons mit Sequenzen auf Chromosom 20 von *M.mulatta* und *P.anubis*.

Exons HSA <i>PRDM7_202</i>	bp	Chromosom 20					
		<i>Macaca mulatta</i>	query	sbjct	<i>Papio anubis</i>	query	sbjct
1	98	no hit			no hit		
2	154	no hit			no hit		
3	124	no hit			no hit		
4	108	no hit			no hit		
5	50	no hit			no hit		
6	157	1	3-157	74956129-74955975	1	3-157	72025469-72025315
7	102	1	3-102	74954865-74954766	1	3-102	72024205-72024106
8	272	1	1-271	74954561-74954291	1	5-271	72023898-72023634
9	68	1	1-68	74953889-74953822	1	1-68	72023232-72023165
10	283	no hit			no hit		
11	755	2	1-606	74950315-74949713	2	1-606	72021446-72020844
			520-755	74949882-74949644		520-755	72021013-72020775

2.2.2.6 ensembl-Suche

Beim Durchsuchen der ensembl-Datenbank nach den Genen *PRDM7* und *PRDM9*, wurde letzteres Gen mit Ausnahme von *Saimiri boliviensis boliviensis* bei allen zur Verfügung stehenden Primaten gefunden. *PRDM7* wird ausschließlich bei *H.sapiens*, *G.g.gorilla*, *N.leucogenys*, *M.mulatta*, *P.anubis*, *R.bieti* und *C.jacchus* von ensembl ausgegeben, wobei die Gen-ID bei *C.jacchus* identisch ist mit der von *PRDM9*, siehe Tabelle 8. *PRDM7* ist damit nur bei Vertretern der Catarrhini zu finden.

Tabelle 8: Gen-IDs von *PRDM7* und *PRDM9*.

Spezies	Taxon	<i>PRDM7</i>	<i>PRDM9</i>
<i>Homo sapiens</i>	H	ENSG00000126856	ENSG00000164256
<i>Pan troglodytes</i>	H		ENSPTRG00000028674
<i>Pan paniscus</i>	H		ENSPPAG00000031671
<i>Gorilla gorilla gorilla</i>	H	101135995 (Chr16)**	ENSGGOG00000024584
<i>Pongo abelii</i>	H		ENSPPYG00000015366
<i>Nomascus leucogenys</i>	H	100583589 (Chr2)**	ENSNLEG00000031321
<i>Macaca mulatta</i>	C	100427538 (Chr20)**	ENSMMUG00000003772
<i>Macaca fascicularis</i>	C		ENSMFAG00000003001
<i>Macaca nemestrina*</i>	C		ENSMNEG00000044045
<i>Papio anubis</i>	C	101024444 (Chr20)**	ENSPANG00000013143
<i>Mandrillus leucophaeus*</i>	C		ENSMLEG00000032078
<i>Colobus angolensis palliatus</i>	C		ENSCANG00000038646
<i>Rhinopithecus roxellana*</i>	C		ENSRROG00000031264
<i>Rhinopithecus bieti*</i>	C	ENSRBIG00000043258	ENSRBIG00000036346
<i>Chlorocebus sabaeus</i>	C		ENSCSAG00000017229
<i>Cercocebus atys*</i>	C		ENSCATG00000041590
<i>Cebus capucinus imitator*</i>	P		ENSCCAG00000023876
<i>Aotus nancymae*</i>	P		ENSANAG00000027642
<i>Saimiri boliviensis boliviensis*</i>	P		x
<i>Callithrix jacchus</i>	P	ENSCJAG00000010419	ENSCJAG00000010419
<i>Carlito syrichta*</i>	T		ENSTSYG00000031191
<i>Otolemur garnettii*</i>	S		ENSOGAG00000025273 ENSOGAG00000028747
<i>Propithecus coquereli*</i>	S		ENSPCOG00000013284
<i>Microcebus murinus</i>	S		ENSMICG00000037613

*: scaffold; **:Entrez gene ID; H: Hominoidea, C: Cercopithecoidea, P: Platyrrhini, T: Tarsius, S: Strepsirrhini

2.2.3 Diskussion

Eine Einordnung des Duplikationsevents, das zu *PRDM7* führte, ist bislang nur grob erfolgt und zwar in der Primatenevolution (Fumasoni et al. 2007). Vervoort et al. (2016) vermuten, dass die Duplikation *PRDM9s* entweder vor der Aufspaltung der Haplorrhini oder der Anthrooidea stattfand, können dies aber nicht anhand einer phylogenetischen Rekonstruktion nachweisen. In dieser Arbeit konnte anhand verschiedener Methoden der putative Abspaltungszeitraum festgestellt werden. Sowohl die Sequenzidentitäten als auch die erstellten Baumrekonstruktionen und die ensembl-Suche nach *PRDM7* und *PRDM9* weisen darauf hin, dass die Duplikation erst kurz vor der Aufspaltung der Catarrhini geschah. Somit wird hier festgestellt, dass die Duplikation rezenter als von Vervoort et al. (2016) angenommen ist und zwischen 43 und 32 Millionen Jahren, nämlich vor der Verzweigung der Cercopithecoidea/Hominoidea stattfand (Perelman et al. 2011). Dies stimmt mit einer Zeitspanne (35 MYA) überein, in der eine große Anzahl segmentaler Duplikationen auftrat, aus der 5% des menschlichen Genoms resultieren sollen (International Human Genome Sequencing Consortium 2001, Review: Samonte und Eichler 2001). Die Duplikationen sind nicht gleich verteilt über die Chromosomen (Chr): Zum einen sind manche Chromosomen, wie Chr 16, häufiger betroffen. Außerdem finden sich die Duplikationen vor allem in zentromer- und telomernahen Regionen (International Human Genome Sequencing Consortium 2001, Bailey et al. 2001, Cheung et al. 2001). *PRDM7s* Position ist an einem Ende des langen Arms (q) des Chromosoms 16 im menschlichen Genom. Die genaue Lage nach NCBI (National Center for Biotechnology Information) ist 90,054,579 - 90,079,511, wobei die Gesamtlänge 90338345 Basen beträgt. In Zusammenhang mit den Informationen zu segmentaler Duplikation liegt die Vermutung nahe, dass *PRDM9* eher das ancestrale Gen darstellt. Fumasoni et al. (2007) sehen *PRDM9*, aufgrund der Konservierung der genomischen Umgebung und der größeren Ähnlichkeit auf Gen- und Proteinebene mit den Orthologen von Taxa außerhalb der Primaten, ebenso als das Vorgängergen an.

Da nicht alle Exons von *PRDM7* bei *M.mulatta* und *P.anubis* auf Chromosom 20 zu finden sind, ist es möglich, dass das Gen auf den Linien zu Hominoidea und Cercopithecoidea deutlich unterschiedlich evolviert ist. Dies könnte darauf hinweisen, dass die Entstehung *PRDM7s* kurz vor der Aufspaltung der Catarrhini stattfand. Da hier Papionini untersucht wurden, wären die *PRDM7*-Sequenzen weiterer Cercopithecoidea, wie Vertreter der Cercopithecini und Colobinae, interessant. Die bei Fumasoni et al. (2007) beschriebene Insertion bei *PRDM7* ist allerdings nur bei Mensch und Schimpanse ein Merkmal des duplizierten Gens, was in der vorliegenden Arbeit per PCR nachgewiesen wurde, und kann somit nicht als Unterscheidungsmerkmal der beiden Gene dienen. Bei einer möglichen weiteren Funktion *PRDM9s* als Transkriptionsaktivator fehlt H3K36me3, somit könnte dies auch *PRDM7s* Funktion vor allem in Testis sein (Hayashi et al. 2005, Blazer et al. 2016, Cano-Rodriguez et al. 2016, Altemose et al. 2017, <https://www.proteinatlas.org/ENSG00000126856-PRDM7/tissue>).

Obwohl Ahlawat et al. (2016b) von *PRDM7* bei Wiederkäuern berichten, ist es aufgrund der Duplikation während der Primatenevolution und der hohen Sequenzidentität wahrscheinlicher, dass ein weiteres Homolog in diesem Taxon entstanden ist.

3. Charakterisierung der PRDM9-Bindungsstellen als putative Rekombinationshotspots bei zwei Makakenarten

3.1 Material und Methoden

Als Material für die folgenden Versuche wurde Hodengewebe von *Macaca mulatta* und *Macaca fascicularis* genutzt, das aus dem Primatenzentrum Göttingen stammt.

3.1.1 Labor

3.1.1.1 RNA-Extraktion und cDNA-Synthese

Mittels RNA-Extraktion, cDNA-Synthese und anschließender PCR wurde überprüft, ob *PRDM9* bei den beiden Makakenarten transkribiert wurde. Für ersteres wurde TRI Reagent® und das Sigma-Aldrich Protokoll verwendet, für zweiteres SuperScript™ III Reverse Transcriptase und das Protokoll von Invitrogen™. Die Primer für die PCR wurden in zwei unterschiedliche Exons gelegt (5'-GGCAGCCAACAATGGATAC-3', 5'-GAGATTCTCTGGTTGGAAGAGTC-3', Tm: 54°C). Durch den Längenunterschied, der durch das Vorhandensein (genomische DNA) bzw. Nichtvorhandensein (cDNA) des Introns entsteht, war es möglich zu bestimmen, welche Art der DNA vorlag. Der Ansatz sowie das Touch down-Programm sind Kapitel 2.1.1.1 und 2.2.1.1 zu entnehmen.

3.1.1.2 Sequenzierung der Zinkfingerarrays

Weitere PCRs wurden durchgeführt, um die Zinkfingerarrays der beiden Makaken zu erhalten. Dabei wurden PCR-Ansatz und Cyclerprogramm wie in Kapitel 2.1.1.1 verwendet. Die Primersequenzen waren wie folgt gewählt: 5'-GGCCAGAAAGTGAATCCAGA-3' und 5'-TGGGCAAGTGTGTGGTGACT-3'. Da beim PCR-Produkt von *Macaca mulatta* zwei, sich durch zwei Zinkfinger (168 bp) unterscheidende, Allele vorlagen, wurden diese mittels pGEM T-Vektor in elektrokompente TOP10 Zellen transfiziert. Mittels Gelelektrophorese, siehe Kapitel 2.1.1.1, wurden die Größen bestimmt und somit auch die getrennten Allele unterschieden. Nach enzymatischer Aufreinigung wurden die PCR-Produkte bei Macrogen von beiden Richtungen Sanger-sequenziert.

3.1.1.3 Western Blot

Ein nativer Western Blot wurde angefertigt, um den verwendeten PRDM9-Antikörper (ThermoFisher, AB-730062; für native Anwendungen) auf seine Spezifität zu testen, vor allem in Hinsicht auf das nahverwandte PRDM7. Die beiden Proteine unterscheiden sich durch ihr Molekulargewicht. Das ausführliche Protokoll ist dem Anhang (S. 76-78) zu entnehmen und basiert auf einem von abcam. Es wurde ein Gradient 4-15% PAA-Gel von Biorad genutzt.

3.1.1.4 Chromatinimmunopräzipitation

Die Chromatinimmunopräzipitation mit anschließender Hochdurchsatzsequenzierung, im Folgenden als ChIP-Seq abgekürzt, wurde nach einer überarbeiteten Variante des abcam-Protokolls durchgeführt (siehe Anhang, S. 74-76). Ziel der ChIP-Seq war es, die DNA-Bindungsstellen der Proteine PRDM9 und DMC1 sowie der Histonmodifikation H3K4me3 zu extrahieren, indem der DNA-Protein-Komplex durch Einsatz eines Antikörpers gegen das Zielprotein herausgefiltert wurde.

Das Protokoll sah eine native Herangehensweise vor, somit wurde von einer Fixierung der Protein-DNA-Bindung mittels Formaldehyd (*cross-linking*) abgesehen. Der Großteil der Arbeiten geschah bei 4°C bzw. auf Eis, um das native System aufrechtzuerhalten.

Die einzelnen Schritte im groben Überblick waren die Zellyse, der enzymatische Verdau, die Immunopräzipitation, die Waschungen und die DNA-Isolation mit anschließender Hochdurchsatzsequenzierung.

Die Zellyse wurde mit einem nichtionischen Detergens, Tween 20, und einem Zellsieb durchgeführt. Die verschiedenen Puffer wurden vor der Anwendung gekühlt. Darauf folgte der enzymatische Verdau mittels *micrococcal nuclease* (MNase), die zwischen den Nukleosomen schnitt und vor allem zu Mono-, Di- und Trinukleosomen führte, welche in etwa 150, 300 und 450 Basenpaare groß waren. Die Immunopräzipitation erfolgte über Nacht mit nachstehenden Antikörpern: Maus anti-human PRDM9 (Thermo Fisher Scientific, 730062), Kaninchen anti-human H3K4me3 (ActiveMotif, 39159; Cell Signaling, 9751, auch spezifisch gegen Maus, Ratte, Primaten, *D. melanogaster*, *S. cerevisiae*), Maus anti-human DMC1 (Santa Cruz, sc-373862; auch spezifisch gegen Pferde, Hunde, Rinder, Schweine und Vögel), Kaninchen anti-human IgG (Abcam, ab133470; auch spezifisch gegen Maus und Ratte). Außerdem wurde bei beiden Replikatdurchführungen eine Kontrolle ohne Antikörper angefertigt (*input*). Für die Herauslösung des DNA-Protein-Antikörper-Komplexes aus der Gesamtmasse wurden Protein G *magnetic beads* eingesetzt, die den Antikörper banden und mit Hilfe von magnetischen Racks zur Auftrennung führten. Danach wurden verschiedene Waschpuffer mit unterschiedlichem Kochsalzgehalt zur Aufreinigung verwendet. Die Trennung der DNA-Fragmente von dem beads-Antikörper-Protein-Komplex erfolgte mittels einer Waschung mit einem mit SDS versetzten Puffer, einer Phenol-Chloroform-Aufreinigung und einer Ethanolpräzipitation.

Da beim ersten Replikatdurchgang hochmolekulare Banden, möglicherweise Proteinrückstände, im Agarosegel zu sehen waren, wurde mit dem Pippin Prep (sage science) und dem SSQ-Set jeweils die Region zwischen 50 bp und 500 bp extrahiert.

Die Erstellung der *libraries* und die Hochdurchsatzsequenzierung wurden zum einen von der IMB Core facility (Illumina HiSeq; Input-1, IgG-1, PRDM9-1, H3K4me3-1), zum anderen von BGI, Hongkong (BGI SEQ 500; Input-2, PRDM9-2, H3K4me3-2, DMC1-2) mit je 50 bp, *single end* ausgeführt. Die biologischen Replikate entstammen aus dem gleichen Gewebe und waren somit isogenisch (genetisch identisch), siehe dazu das Encode Projekt (<https://www.encodeproject.org/data-standards/terms/>).

3.1.1.5 Immunopräzipitation und Massenspektrometrie

Für die Bestimmung möglicher Proteininteraktionspartner wurde mit dem Pierce Classic IP Kit von Thermo Fisher eine Immunopräzipitation durchgeführt. Mittels eines SDS-PAGE-Gels wurden die relevanten Banden zweier Proben und einer Kontrolle pro Spezies extrahiert und an die AG Schlüter in Hamburg zur massenspektrometrischen Bestimmung geschickt.

3.1.2 Datenanalyse

3.1.2.1 Aufbereitung der Zinkfingersequenzen

Die *PRDM9*-Zinkfingersequenzen von *M.mulatta* und *M.fascicularis* wurden in BioEdit 7.2.5 (Hall 1999) kontrolliert und aufbereitet.

3.1.2.2 Ermittlung der DNA-Bindungssequenzen anhand der Zinkfingerarrays

Die Vorhersage der Bindungssequenzen erfolgte mit dem *Interactive PWM predictor* (Position Weight Matrix) unter dem *expanded linear SMV (Support Vector Machines)*-Model auf Basis der Zinkfingerarrays (Persikov et al. 2009, Persikov und Singh 2014).

3.1.2.3 Analyse der CHIP-Sequenzen

3.1.2.3.1 Qualitätskontrolle der Rohdaten

Mit FastQC (Version 0.11.5, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, Andrews 2010) wurden die erhaltenen Sequenzen auf ihre Qualität getestet.

3.1.2.3.2 Galaxy

In Galaxy 15.07 (lokal installiert, Giardine et al. 2005, Blankenberg et al. 2010b, Goecks et al. 2010) wurden die folgenden Programme verwendet: FASTQ Groomer (Galaxy Tool Version 1.0.4, Blankenberg et al. 2010a), bowtie (Galaxy Tool Version 1.1.3, Langmead et al. 2009) und sam-to-bam converter (Galaxy Tool Version 2.1, Li et al. 2009a, Li 2011a, Li 2011b).

Mit dem FASTQ Groomer wurden die Sequenzdaten vom Format FASTQ zu Sanger FASTQ konvertiert. Dazu musste der jeweilige FASTQ *quality score*-Typ angegeben werden. Für die ersten Replikate war das Sanger & Illumina 1.8+ (IMB) und für die zweiten Illumina 1.3-1.7 (BGI).

Bowtie wurde zum Lokalisieren der Sequenzen bzw. Reads in den Referenzgenomen von *Macaca mulatta* (rheMac8/JSUE03, rheMac3/AEHK01) und *Macaca fascicularis*

(macFas5/AQIA01, AEHL01) verwendet. Dabei wurden die Sequenzen von *M.mulatta* gegen rheMac8, rheMac3 und macFas5, jene von *M.fascicularis* gegen macFas5, AEHL01 und rheMac8. Die Genome bestehen jeweils aus den Autosomen und das X-Chromosom, scaffolds und contigs wurden nicht mit aufgenommen. Bei rheMac8 ist außerdem das Y-Chromosom enthalten. Die gewählten Änderungen der Soforteinstellungen sind: ‚doBest‘, ‚use strata option‘ und Alignments, die mehr als einmal vorkamen, sollten unterdrückt werden. Außerdem sollten Sequenzen, die nicht oder mehr als einmal aligniert wurden, separat abgespeichert werden.

Für den Download der Daten und als Vorbereitung für das *peak calling* wurden die Sequenzen konvertiert von sam-to-bam.

3.1.2.3.3 Qualitätswerte

Um Duplikate, Sequenzen mit identischem 5'-Ende, zu entfernen, wurde das Softwarepaket samtools (Li et al. 2009a) eingesetzt. Duplikate könnten durch Überamplifikation während der PCR entstanden sein. Diese zu entfernen war das konservativere Vorgehen. Phantompeakqualtools (Kharchenko et al. 2008, Landt et al. 2012) berechnete die Qualitätsmerkmale RSC (Relative Strand Cross-correlation coefficient), NSC (Normalized Strand Cross-correlation coefficient) und einen zu letzterem gehörenden *quality tag*. Zudem wurden der NRF (Non-Redundant Fraction) und der PBC (PCR Bottleneck Coefficient) in Excel ermittelt (Landt et al. 2012, <https://www.encodeproject.org/data-standards/terms/>).

3.1.2.3.4 Peak Calling

Mit MACS2 (Model-based Analysis of ChIP-Seq, Zhang et al. 2008) wurden die Peaks, also mit Reads angereicherte Regionen im Vergleich zu den Kontrollsequenzen, bestimmt. Die Kommandozeile lautete: `macs2 callpeak -t [ChIP-Datei] -c [Kontrolldatei] -n [output-Name] --bw 150 -p 0.05 -m 5,50`. Um mehr Übereinstimmungen zwischen den Replikaten zu erhalten, rechnete MACS2 mit gelockerten Bedingungen ($p=0.05$).

Mit bedtools wurden die Peaks nach Chromosom und Region sortiert. Danach wurde intersect, das auch im bedtools-Paket enthalten ist, genutzt, um die in beiden Replikaten enthaltenen Peaks zu bestimmen. Die Länge der Peaks wurden jeweils vor intersect auf 500 bp (± 250 bp um das Zentrum der Peaks) gesetzt. Falls der Start- oder Stopppunkt der Peaks damit über die Chromosomenenden hinausragte, wurden sie entsprechend getrimmt. Außerdem wurden Übereinstimmungen zwischen den Bindungsstellen der verschiedenen Antikörper (anti-PRDM9, anti-H3K4me3, anti-DMC1) gesucht, indem die Peaks diesmal auf 5 kb vergrößert wurden, da es weniger um eine exakte Übereinstimmung geht als um die Nähe der Peaks zueinander. Peaks, die mittels beider PRDM9-Replikate gefunden wurden, werden im Folgenden mit P9 abgekürzt, jene, die in PRDM9- und H3K4me3-Replikaten vorhanden sind mit P9H3 und solche die zudem im DMC1-Replikat präsent sind mit P9H3D1. Eine Darstellung der H3K4me3- und DMC1-Peaks, die sich ± 5 kb um PRDM9-Peaks befinden, wurde in Excel erstellt. Des Weiteren

wurde intersect genutzt, um übereinstimmende Peaks zwischen *M.mulatta* und *M.fascicularis* auf jeweils einem Genom (rheMac8 oder macFas5) zu finden, dazu wurden die Peaks auf 2 kb vergrößert, was der Größe von Rekombinationshotspots entspricht.

3.1.2.3.5 Verteilung der Peaks im Genom

Die Verteilung der P9-Peaks in den verschiedenen Genomen wurde zum einen dargestellt, indem die Prozente der Peaks pro Chromosom gegen den prozentualen Anteil der Chromosomen im Genom aufgetragen wurden. Zu diesem Aspekt wurden Rangkorrelationen, sowohl Pearson als auch Spearman, gerechnet, um zu überprüfen, wie der Anteil der Peaks pro Chromosom mit dessen GC-Gehalt in Zusammenhang steht (<https://www.socscistatistics.com/>). Dabei wurden die Anzahl der Peaks eines Chromosoms durch die Basenpaare des Chromosoms dividiert, um sie mit dem prozentualen GC-Gehalt zu vergleichen.

Außerdem wurde die prozentuale Verteilung der Peaks (P9 und P9H3D1) von *M.mulatta* auf rheMac8 und *M.fascicularis* auf macFas5 entlang aller Autosomen (Mittelwert), die in 10%-Schritten aufgeteilt wurden, mittels Pivot-Tabelle und -Diagramm dargestellt. Ein genaueres Diagramm wurde mit Chromosom 20 erstellt, das in 1 Mb Schritten aufgeteilt wurde, einmal für P9, P9H3 und P9H3D1 mit *M.mulatta* auf rheMac8, sowie für P9 mit *M.mulatta* und *M.fascicularis* auf jeweils beide Referenzgenome, rheMac8 und macFas5. Beispielhaft wurden zudem die Verteilung der P9-Peaks von *M.mulatta* auf rheMac8, macFas5 und rheMac3, sowie jene von *M.fascicularis* auf rheMac8 und macFas5 mittels UCSC für Chromosom 10, 11 und X aufgezeigt.

3.1.2.3.6 Annotation und Genfunktion

Für die Annotation der Peaks wurde HOMER (Hypergeometric Optimization of Motif EnRichment, Version 4.10; Heinz et al. 2010) verwendet mit dem Zusatz, dass eine statistische Übersicht erstellt wurde. Dies wurde für die PRDM9- und H3K4me3-Peaks durchgeführt.

Die Annotation diente auch als Grundlagen für weitere Analysen. Zum einen wurden die H3K4me3-Peaks, die +/- 2000 bp um die *transcription start site* gelegen waren, grafisch dargestellt, sowohl für *M.mulatta* als auch *M.fascicularis* auf beiden Genomen: rheMac8 und macFas5. Die dazugehörigen Gene wurden mit der Gene Functional Classification von DAVID (The Database for Annotation, Visualization and Integrated Discovery, Version 6.8, Huang et al. 2009a, Huang et al. 2009b) untersucht, ebenso jene Gene, in die PRDM9-Peaks fielen. Diese Peaks lagen entweder in Exons, Introns, TSS (*transcription starting site*), TTS (*transcription termination site*), 5'- oder 3'-UTRs (*untranslated regions*). Zudem wurden die Gene, mit H3K4me3-Peaks +/- 2kb um ihre TSS, mit dem Testisproteom von *Macaca mulatta* (Zhou et al. 2015) und den Haushaltsgenen von *Homo sapiens* (Eisenberg, Levanon 2013) verglichen. Dies geschah auch für die mit PRDM9-Peaks überlappenden Gene. Dabei wurden die Peaks von *M.mulatta* und *M.fascicularis* untersucht, die sich auf rheMac8 befanden. Um einen Vergleichswert zu erhalten, wurde

der prozentuale Anteil der Haushaltsgene im Testisproteom berechnet. Außerdem wurden jeweils Chi-Quadrat-Tests durchgeführt, um nachzuweisen, ob die aufgezeigten Unterschiede signifikant sind (<http://quantpsy.org>, Preacher 2001).

3.1.2.3.7 Motive

Die Suche nach Motiven wurde mit der MEME Suite 5.0.2 durchgeführt. Dazu wurden die Peaks absteigend nach ihrem p-value sortiert und die 1000 bestunterstützten Peaks herausgezogen. Diese wurden zudem auf +/-75 bp um den Peak-Mittelpunkt erweitert. Die dazugehörigen Fasta-Sequenzen wurden mit bedtools getfasta generiert. Mit MEME (Multiple Em for Motif Elicitation; Bailey und Elkan 1994, Bailey et al. 2009) wurde nach drei Motiven gesucht, die in allen vorgegebenen Sequenzen vorkamen. Für *M.mulatta* wurden die Peaks, die auf den Genomen rheMac8 lagen, untersucht, für *M.fascicularis* diejenigen auf macFas5. Über FIMO (Find Individual Motif Occurrences; Grant et al. 2011) wurde überprüft, wie oft das jeweils einzig signifikante Motiv in den PRDM9-Peaks der beiden Spezies zu finden sind. Diese Peaks wurden auf ihre Verteilung auf den Chromosomen ihrer jeweiligen Referenzgenome mittels Excel geprüft.

Mit STAMP wurden die mit MEME erstellten Motive von *M.mulatta* und *M.fascicularis* mit den vorhergesagten Motiven verglichen, siehe Kapitel 3.1.2.2 (Mahony et al. 2007, Mahony und Benos 2007). Dabei wurden die Soforteinstellungen genutzt.

3.1.2.3.8 piRNA Cluster

PRDM9-Peaks, die mit piRNA-Cluster überlappten, wurden mit intersect nachgewiesen. Um zu überprüfen, ob sie signifikant häufiger auftreten als durch Zufall, wurde 10 Mal über bedtools random die gleiche Anzahl an beliebigen Sequenzabschnitten mit den gleichen Längen gezogen und die Übereinstimmungen mit den Clustern per intersect überprüft. Der Durchschnitt dieser Übereinstimmungen wurde mit den Überlappungen der tatsächlichen PRDM9-Peaks mit einem Chi-Quadrat-Test überprüft (<http://quantpsy.org>, Preacher 2001).

3.1.2.3.9 Auswertung der Daten aus der Massenspektrometrie

Die Namen der Proteine wurden zu uniprot IDs umgewandelt (<https://www.uniprot.org/uploadlists/>) und in CRAPome (Contaminant Repository for Affinity Purification, <http://crapome.org/>) darauf getestet, welche häufig als Kontaminant bzw. unspezifische Interaktionen erscheinen. Proteine, die maximal dreimal in den 411 Experimenten auftraten, wurden in die Ergebnisse aufgenommen.

3.2 Ergebnisse

3.2.1 Überprüfung PRDM9s auf RNA- und Proteinebene

Die vorbereitenden Methoden und Analysen, wie RNA-Extraktion mit anschließender cDNA-Synthese und der Western Blot, zeigen zum einen, dass *PRDM9* im Hoden der beiden Makakenarten transkribiert wird, dass die resultierende mRNA stabil translatiert werden kann, und, dass der gewählte Maus anti-human PRDM9-Antikörper, der die gefaltete, humane PRDM9-SET-Domäne bindet, Produkte mit erwartetem Molekulargewicht bei Makaken spezifisch erkennt. Wie in Abbildung 7a beispielhaft zu sehen, stimmt die Größe des PCR-Produkts (*M. mulatta*), links neben dem Marker, mit der erwarteten Fragmentgröße (454bp) der gespleißten PRDM9-mRNA bzw. dem entsprechenden cDNA-Produkt überein. Inklusive Introns würde die Länge 3532 bp betragen. Rechts neben dem Marker sind die beiden Allele des Zinkfingerarrays von *M. mulatta* abgebildet, wobei sich die größere Bande etwas schwächer abzeichnet. Der Western Blot in Abbildung 7b zeigt für beide Makakenarten jeweils eine Bande, nachdem dieser mit Maus anti-human PRDM9 und Sekundärantikörper behandelt wurde.

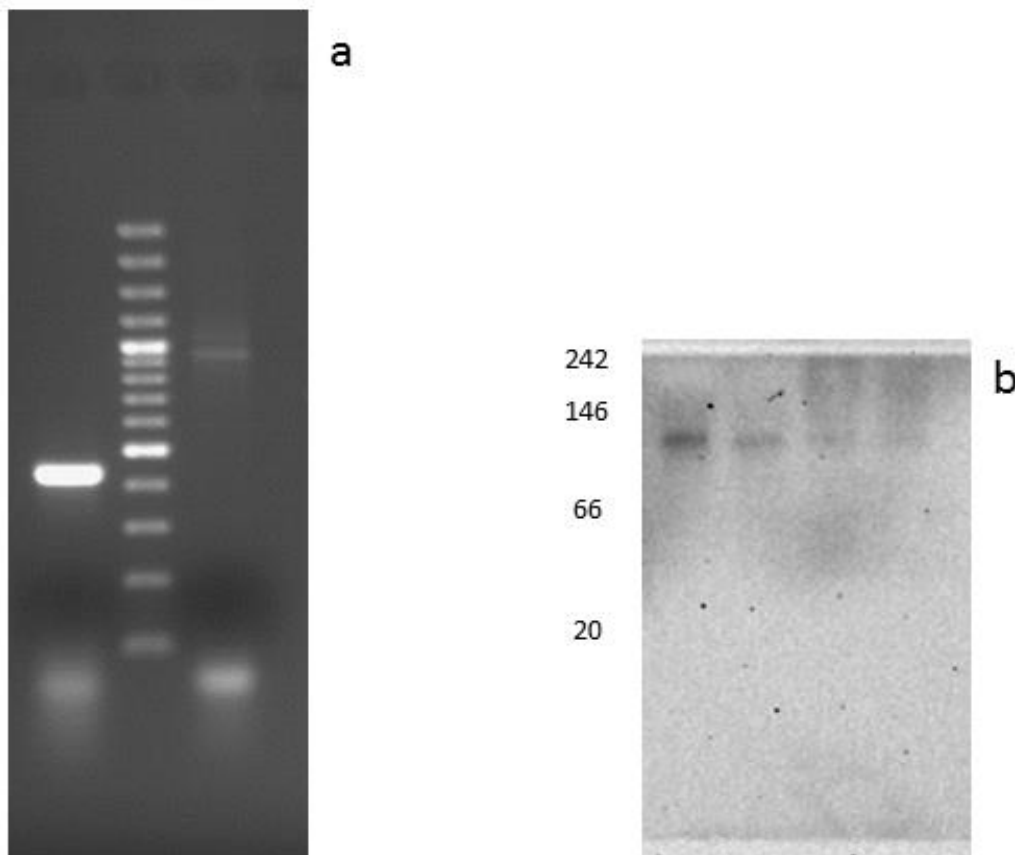


Abbildung 7: a) Agarosegelbild mit PCR-Produkten aus zwei PCRs inkl. 100 bp+ Leiter von Fermentas; links ist das Produkt zu sehen, das das Vorhandensein der mRNA (in cDNA synthetisiert) von *PRDM9* bestätigt, rechts sind die unterschiedlich großen Zinkfingerarrays *PRDM9*s zu sehen, als Grundlage diente jeweils Gewebematerial von *M. mulatta*. b) Western Blot mit je zwei Banden von *M. mulatta* (links) und *M. fascicularis* (rechts), die jeweils linke Bande hat jeweils die doppelte Proteinmenge Untersuchungsmaterial als Basis. Behandelt wurde der Blot mit Maus anti-human PRDM9.

3.2.2 Die Sequenzen der Zinkfingerarrays

Die Sequenzierung der Zinkfingerarrays hat ergeben, dass *M. mulatta* am *PRDM9*-Locus heterozygot und *M. fascicularis* homozygot ist. Die drei Allele unterscheiden sich in ihrer Länge und in ihrer Sequenz, wie auch in Abbildung 8 ersichtlich. *M. mulattas PRDM9* besitzt acht bzw. zehn Zinkfinger, *M. fascicularis* neun. Dabei entspricht das kürzere Allel von *M. mulatta* in der Struktur der repetitiven Einheiten der Sequenz des rheMac8 Referenzgenoms. Bis auf einen zusätzlichen Zinkfinger trifft das auch auf das *M. fascicularis*-Allel zu. Das längere Allel von *M. mulatta* ist allenfalls mit dem Array des Genoms macFas5 vergleichbar. Die Aussagen beziehen sich vor allem auf die bindenden Aminosäuren, die auch in Abbildung 8 zu sehen sind.

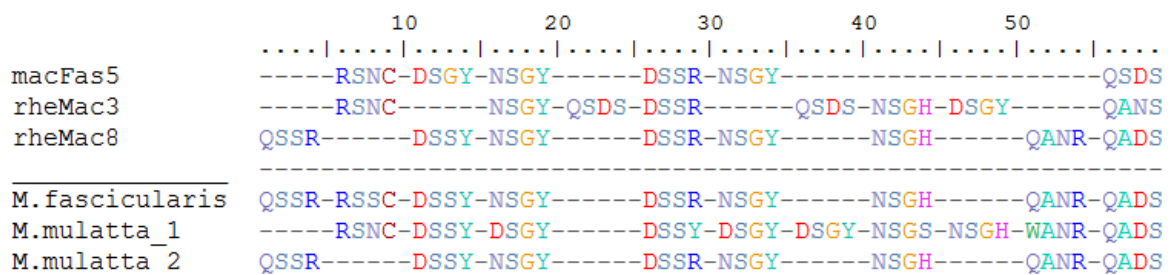


Abbildung 8: Darstellung der bindenden Aminosäuren (-1, 2, 3, 6 bezüglich der Alpha-Helix; abgekürzt mit dem Ein-Buchstaben-Code für Aminosäuren) im Zinkfingerarray von PRDM9. Je vier Aminosäuren stellen einen Zinkfinger dar. Die oberen drei Arrays sind aus drei verschiedenen Makakengenomen (macFas5: *M. fascicularis*, rheMac3 und rheMac8: *Macaca mulatta*). Die unteren drei stellen die Allele dar, die in den beiden untersuchten Individuen gefunden wurden.

3.2.3 Ergebnisse der Chromatinimmunopräzipitation

Ein Agarosegelbild, Abbildung 9, der fertigen ChIP-Produkte zeigt das typische Muster nach einer enzymatischen Behandlung mit Micrococcal Nuclease. Das Enzym Micrococcal Nuclease schneidet im Histon H1-Linkerbereich zwischen den Nukleosomen, was – je nach Effektivität und Dauer der enzymatischen Behandlung – zu Mono- (ca. 150 bp), Di-, Trinukleosomen und weiteren Vielfachen führt.

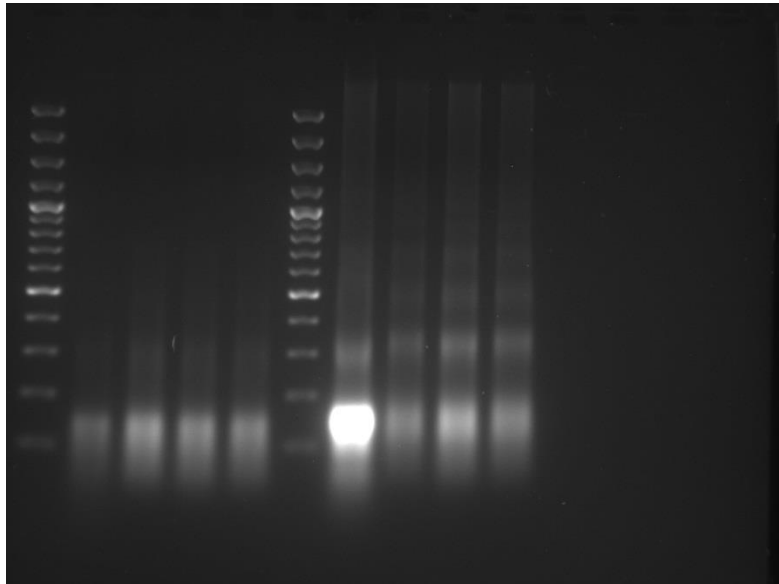


Abbildung 9: Agarosebild mit den ChIP-Proben nach der DNA-Isolation, die das typische Nukleosommuster von 150 bp und Vielfachen aufweist. Reihenfolge: 100 bp+ Leiter von Fermentas – *M.mulatta*: input – PRDM9 – H3K4me3 – DMC1 – 100 bp+ Leiter – *M.fascicularis*: input – PRDM9 – H3K4me3 – DMC1.

3.2.3.1 Übersicht über alle Reads

Die Anzahl der Rohsequenzen beläuft sich auf ca. 17 bis 105 Millionen. Davon lassen sich jeweils zum Großteil mehr als 65 % genomisch lokalisieren, d.h. „mappen“, mindestens jedoch 50 %. Die weggefallenen Sequenzen sind entweder nicht im jeweiligen Genom zu finden (ca. 5-24%) oder gleich an mehreren Loci (ca. 9-30%) und werden somit aussortiert. Von jenen Sequenzen, die mehrfach an derselben 5'-Position starten, verbleibt zur Vermeidung von Redundanzen nur je eine im Datensatz. Der NRF-Wert (*non redundant fraction*), der die Rate der nicht redundanten Reads zu der Gesamtzahl der Reads darstellt, liegt, abgesehen von zwei Datensätzen, bei über 0,5 und ist somit akzeptabel (siehe <https://www.encodeproject.org/data-standards/terms/>). Die nicht-identischen Sequenzen stellen die Grundlagen weiterer Analysen dar. Tabelle 9 sind die einzelnen Werte zu entnehmen, die sich auch je nach Referenzgenom unterscheiden. Im Anhang (S. 79-80) sind Tabellen mit Qualitätswerten zu finden, die sich vorwiegend im Rahmen befinden und die sich auf die Reads beziehen, sowohl auf die Anzahl vor und nach dem Mappen als auch nach dem Entfernen der Duplikate.

Tabelle 9: Übersicht über die Anzahl der Reads pro Replikat (*M.mulatta*: Mm1 und Mm2, *M.fascicularis*: Mf1 und Mf2) und über die Anzahl der im Genom lokalisierten Reads mit und ohne Duplikate und je nach Referenzgenom (rheMac3 und rheMac8 für *M.mulatta*, macFas5 und AEHL für *M.fascicularis*).

			rheMac8		macFas5		rheMac3	
		alle Reads	Reads im Genom	ohne Duplikate	Reads im Genom	ohne Duplikate	Reads im Genom	ohne Duplikate
Mm1	input	21851998	16776268	16319634	16607997	16265479	16747234	15524341
	PRDM9	17865381	13850535	13520432	13725464	13474009	13643402	12829812
	H3K4me3	21302496	16353890	15914178	16201029	15858641	16145592	15072318
Mm2	input	105148033	72513562	67332286	71727113	66951484	72631171	64337052
	PRDM9	77578623	56454730	45482942	55750708	45103326	56454730	43272028
	H3K4me3	72198659	50973278	47033314	50319297	46754843	50973278	44999735
	DMC1	68410957	38647317	34257436	37765449	34003483	41329450	32509459
							AEHL	
Mf1	input	20330700	15545569	15155387	15788206	15438832	14612103	13895929
	PRDM9	15587633	11964584	11711614	12166530	11939419	11184699	10707953
	H3K4me3	18362226	14045012	13716293	14284094	13982035	13117348	12532041
Mf2	input	65356120	49735370	46881957	50575489	47675434	46654191	42963086
	PRDM9	82028731	58086247	53478057	59113070	54381451	54644600	48585588
	H3K4me3	58442991	38418425	35490571	38881488	36141175	36828231	32091599
	DMC1	48263437	31405148	29176632	31834409	29713180	29901590	26267998

3.2.3.2 Anzahl der Peaks

MACS2 hat je nach Datensatz mit den gewählten Einstellungen zwischen 45679 und 551294 Peaks gefunden. Tabelle 10 zeigt, neben der Anzahl der Peaks pro Replikat, die Peaks, die in beiden Replikaten pro verwendetem Antikörper zu finden sind und solche, die in den Replikaten aller drei Antikörper (PRDM9, H3K4me3, DMC1; kurz: P9H3D1) vorkommen. Die Anzahl der Peaks variiert je nach Referenzgenom, wobei beim Vergleich der einzelnen Replikate fast immer die meisten sowohl bei *M.mulatta* als auch bei *M.fascicularis* mit rheMac8 auftreten. Die wenigsten Peaks sind immer bei PRDM9_R2 von *M.fascicularis* zu sehen, unabhängig vom Referenzgenom. Die vergleichsweise niedrigere Gesamtzahl an Peaks findet sich in beiden *M.fascicularis* PRDM9-Replikaten, in den PRDM9- und den P9H3D1-Peaks.

Legt man die Referenzgenomsequenz rheMac8 zugrunde, sind 617 der PRDM9-Peaks von *M.fascicularis* auch bei *M.mulatta* zu finden, was 14,6% der *M.fascicularis* und 1,8% der *M.mulatta* PRDM9-Peaks entspricht. Mit macFas5 als Referenz sind es 490 Peaks, was 11,6% (*M.fascicularis*) bzw. 1,9% (*M.mulatta*) gleichkommt.

Tabelle 10: Übersicht über die Anzahl der Peaks pro Genom, Spezies, Replikat und zudem über die Anzahl der Peaks in mehreren Replikaten.

Genom	Spezies	Antikörper	Peaks	intersect Replikate*	P9H3D1-Peaks
rheMac8	<i>M.mulatta</i>	PRDM9_R1	231613	34526	4636
		PRDM9_R2	551294		
		H3K4me3_R1	311949	22541	
		H3K4me3_R2	194079		
		DMC1_R2	520270		
	<i>M.fascicularis</i>	PRDM9_R1	201225	4233	631
		PRDM9_R2	48700		
		H3K4me3_R1	260473	19138	
		H3K4me3_R2	157434		
		DMC1_R2	248571		
macFas5	<i>M.mulatta</i>	PRDM9_R1	232031	25309	3865
		PRDM9_R2	377974		
		H3K4me3_R1	318174	23522	
		H3K4me3_R2	192714		
		DMC1_R2	337196		
	<i>M.fascicularis</i>	PRDM9_R1	204414	4230	431
		PRDM9_R2	48910		
		H3K4me3_R1	164155	9852	
		H3K4me3_R2	126212		
		DMC1_R2	138512		
rheMac3	<i>M.mulatta</i>	PRDM9_R1	182407	17870	2503
		PRDM9_R2	380429		
		H3K4me3_R1	247226	17189	
		H3K4me3_R2	185769		
		DMC1_R2	341626		
AEHL	<i>M.fascicularis</i>	PRDM9_R1	160174	3032	329
		PRDM9_R2	45679		
		H3K4me3_R1	198871	10723	
		H3K4me3_R2	101052		
		DMC1_R2	123558		

*ohne Duplikate

3.2.3.3 Verteilung der Peaks

In den zwei folgenden Diagrammen (Abbildungen 10 und 11) ist die prozentuale Verteilung der PRDM9-Peaks (P9) auf die Chromosomen der Referenzgenome *rheMac8* und *macFas5* abgebildet, zusammen mit dem Verhältnis der Größe (bp) einzelner Chromosomen zur Genomgröße (1-20 und X, inkl. Y bei *M.mulatta*). Zu sehen ist, dass die Peaks sich mit Ausnahmen relativ gleichmäßig je nach Größe der Chromosomen verteilen, vor allem jene von *M.fascicularis*. Auf den Chromosomen 10, 16, 19 und 20 sind die Anteile der Peaks erhöht, zum Teil um das doppelte als die prozentualen Größenanteile der Chromosomen an der Genomgröße. Der gegenteilige Fall liegt beim X-Chromosom vor mit vermutlich signifikantem Unterschied zwischen den prozentualen Anteilen der Peaks und der relativen Chromosomengröße. Noch deutlicher dokumentiert dies Abbildung 12, die den Log₂-Wert der Rate zwischen den beiden Prozenten abbildet. Die meisten Werte liegen sehr nah an null, was darauf hinweist, dass die korrigierte Peakzahl und relative Chromosomenlänge sich nur gering unterscheiden. Die Tendenzen und deren Stärke der oben erwähnten Ausnahmen sind daraus gut abzulesen. Auffällig ist zudem, dass die Peaks von *M.mulatta* bei fünf Chromosomen (2, 5, 6, 12, 17) in ihren Tendenzen erkennbar stärker sind als bei *M.fascicularis*.

Die Referenzgenome *rheMac8* und *macFas5* wirken sich nur geringfügig auf die Verteilung der Peaks von *M.mulatta* und *M.fascicularis* aus.

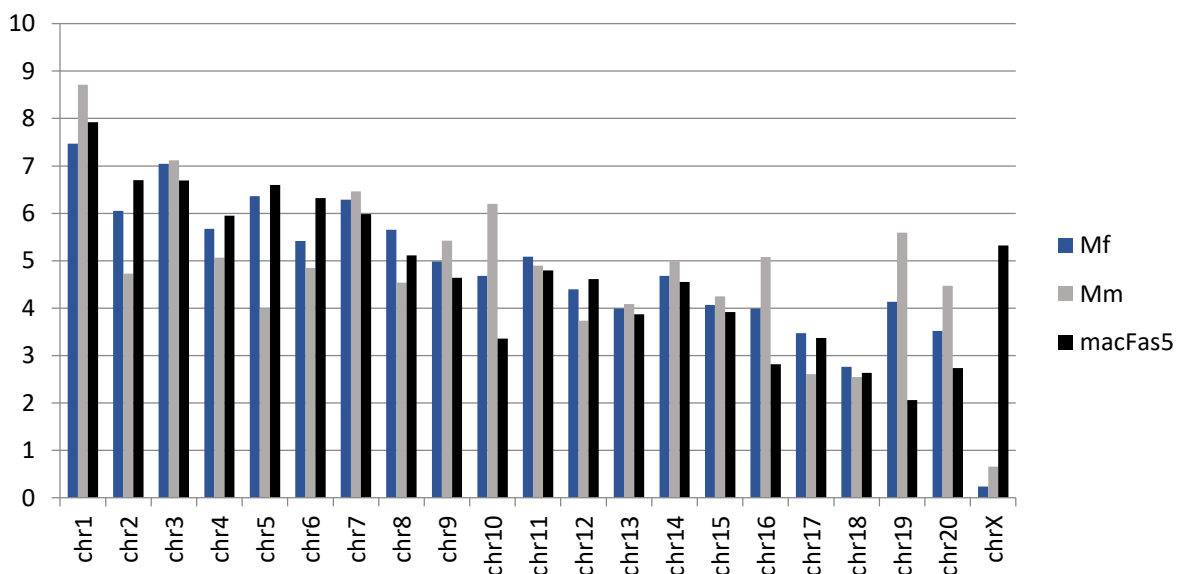


Abbildung 10: Prozentualer Anteil der P9-Peaks pro Chromosom für *M.fascicularis* (Mf, blau) und *M.mulatta* (Mm, grau) auf dem Referenzgenom *macFas5* und der Anteil der einzelnen Chromosomen im Verhältnis zu allen Chromosomen (schwarz).

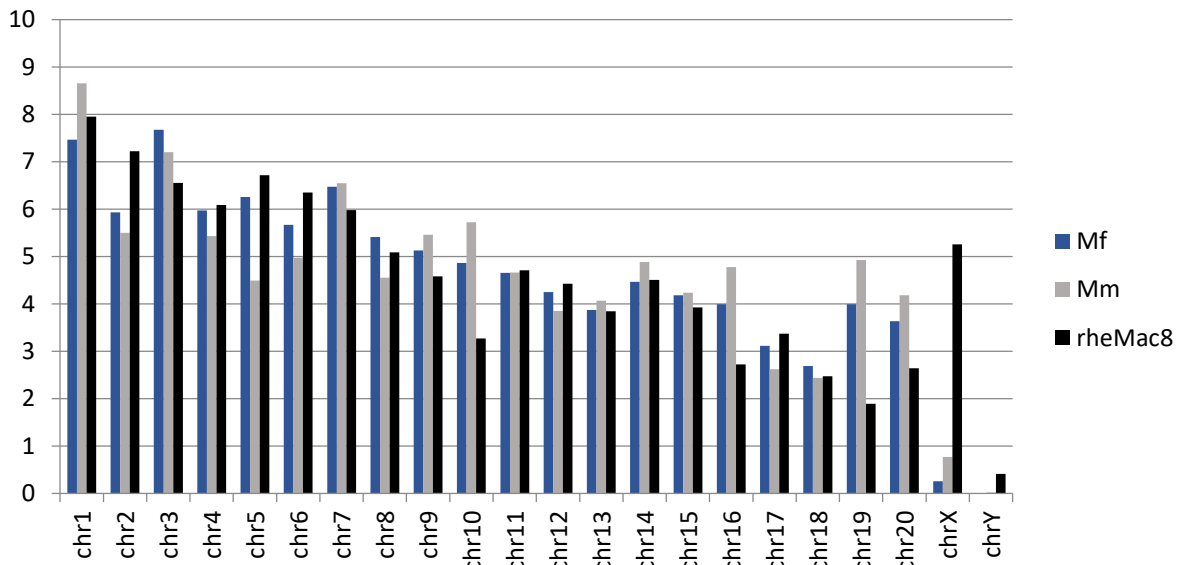


Abbildung 11: Prozentualer Anteil der P9-Peaks pro Chromosom für *M.fascicularis* (Mf, blau) und *M.mulatta* (Mm, grau) auf dem Referenzgenom rheMac8 und der Anteil der einzelnen Chromosomen im Verhältnis zu allen Chromosomen (schwarz). Der Anteil der Peaks auf dem Y-Chromosom liegt bei *M.fascicularis* bei 0% und bei *M.mulatta* bei 0,023%.

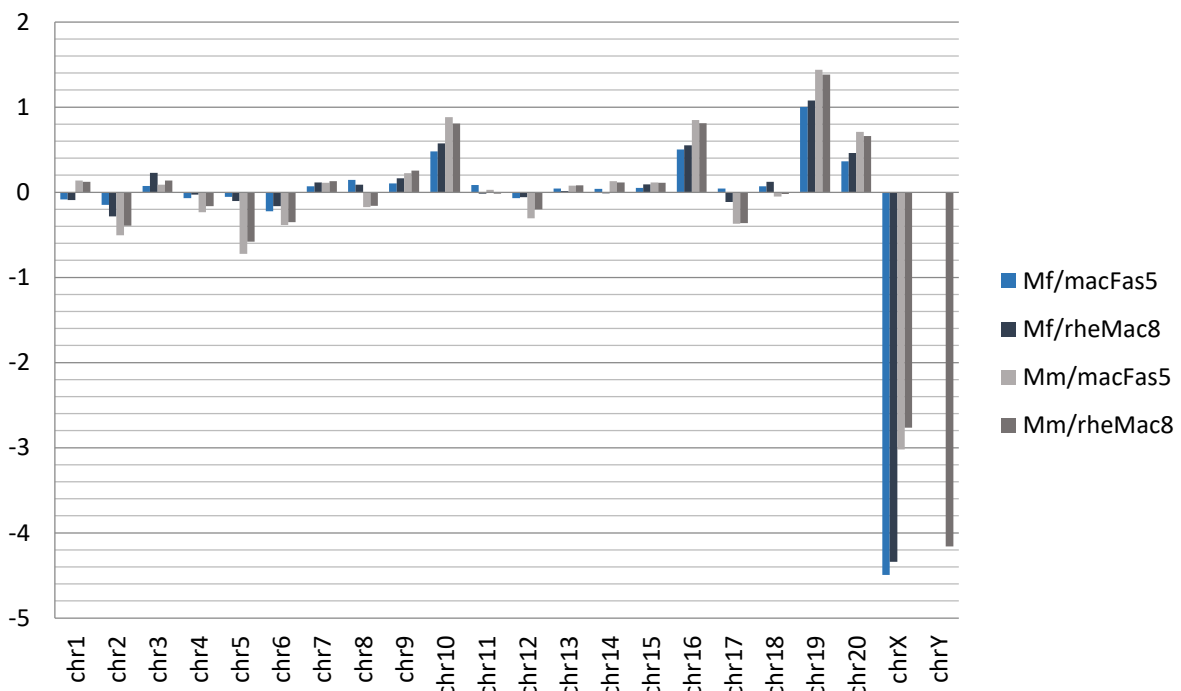


Abbildung 12: Logarithmus zur Basis 2 des Verhältnisses des prozentualen Anteils der P9-Peaks zum Anteil der einzelnen Chromosomen im Vergleich zu allen Chromosomen für *M.fascicularis* (blau) und *M.mulatta* (grau) zu den Referenzgenomen macFas5 (hell) und rheMac8 (dunkel). Für Chromosom Y ist nur der Wert von *M.mulatta* auf rheMac8 abgebildet, da keine *M.fascicularis* keine Peaks auf dem Y-Chromosom von rheMac8 aufweist.

3.2.3.4 GC-Gehalt

Die Chromosomen mit mehr Peaks als über die relative Chromosomenlänge erwartet sind jene mit hohem GC-Gehalt (10, 16, 19, 20). Jene mit weniger Peaks weisen analog den geringsten GC-Gehalt auf (2, 5, 6, 12, 17, X, Y). Die Rangkorrelationen, die den Zusammenhang zwischen Peaks (normalisiert durch Größe der Chromosomen) und GC-Gehalt pro Chromosom untersuchen, weisen R-Werte auf, die für eine starke positive Korrelation sprechen, siehe Tabellen 11 und 12.

Tabelle 11: Rangkorrelationen zu Anzahl der *M. fascicularis*-Peaks im Verhältnis zur Chromosomgröße und dem jeweiligen GC-Gehalt des Chromosoms in macFas5. Die Korrelationen wurden zum einen für Peaks in beiden PRDM9-Replikaten (P9) und in allen erstellten Replikaten (P9H3D1) berechnet, sowohl mit und ohne Peaks in CpG-Inseln.

<i>M.fascicularis/macFas5</i>					
				Peaks in CpG-Inseln entfernt	
		P9	P9H3D1	P9	P9H3D1
Spearman	R	0,706763	0,771132	0,659298	0,637192
	p-Wert (two-tailed)	0,00034	0,00004	0,00115	0,00189
Pearson	R	0,7611	0,8621	0,7488	0,7972
	R ²	0,5793	0,7432	0,5607	0,6355
	p-Wert	0, 000062	<0,00001	0, 000094	0,000015

Dies gilt sowohl, wenn man die Peaks von PRDM9 als auch jene von P9H3D1 betrachtet. Um einen Effekt durch sich in CpG-Inseln befindenden Peaks auszuschließen, wurden diese entfernt und weitere Korrelationen berechnet, die wiederum die positive Korrelation in ähnlicher Stärke bestätigen. Die Signifikanzwerte befinden sich alle unter dem 5%-Niveau.

Tabelle 12: Rangkorrelationen zu Anzahl der *M.mulatta*-Peaks pro Chromosomgröße und dem jeweiligen GC-Gehalt des Chromosoms in rheMac8. Die Korrelationen wurden zum einen für Peaks in beiden PRDM9-Replikaten (P9) und in allen erstellten Replikaten (P9H3D1) berechnet, sowohl mit und ohne Peaks in CpG-Inseln.

<i>M.mulatta/rheMac8</i>					
				Peaks in CpG-Inseln entfernt	
		P9	P9H3D1	P9	P9H3D1
Spearman	R	0,91156	0,818876	0,899124	0,779316
	p-Wert (two-tailed)	0	0	0	0,00002
Pearson	R	0,9042	0,8951	0,8985	0,8946
	R ²	0,8176	0,8012	0,8073	0,8003
	p-Wert	<0,00001	<0,00001	<0,00001	<0,00001

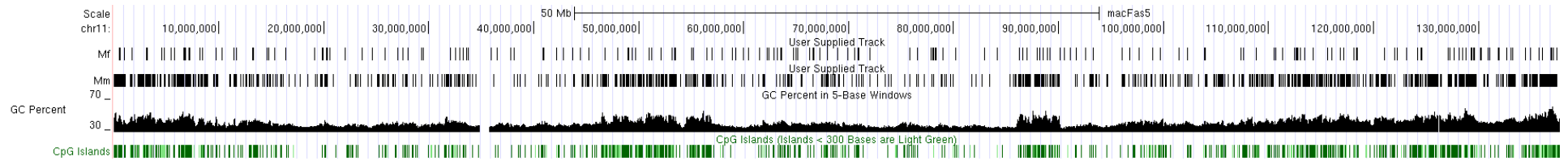
Im Vergleich von *M.mulatta* und *M.fascicularis* fällt auf, dass die R-Werte bei ersterem höher ausfallen.

3.2.3.5 Zusammenhang zwischen Verteilung und GC-Gehalt

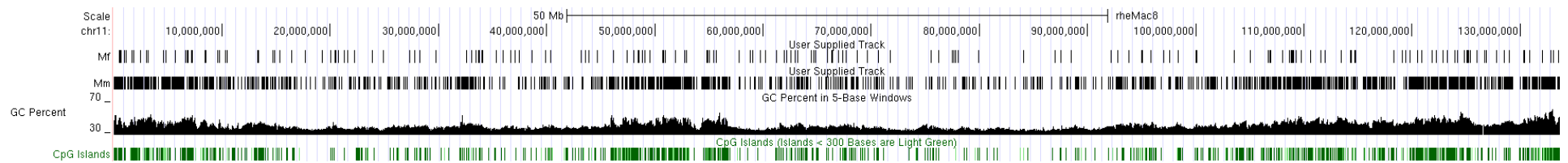
Die Verteilung der Peaks über einzelne Chromosomen wird hier am Beispiel von Chromosom 11, das den durchschnittlichen GC-Gehalt der Genome aufweist, in Abbildung 13 gezeigt. Dabei dienen rheMac8, macFas5 und rheMac3 als Referenzgenome. Aufgrund der höheren Anzahl an *M.mulatta*-Peaks ist der Zusammenhang zwischen Peaks und GC-Gehalt deutlicher zu sehen als bei *M.fascicularis*. Vor allem an den p- und q-terminalen Enden des Chromosoms und bei 50 Mb ist bei allen drei Teilabbildungen ein erhöhtes Aufkommen von Peaks bei hohem GC-Gehalt zu beobachten. Die CpG-Inseln zeigen ein ähnliches Muster. Im Gegensatz dazu liegen nur wenige Peaks um 80 Mb, was mit niedrigem GC-Gehalt zusammenfällt.

Weitere Abbildungen in der Art für die Chromosomen 10 und X sind im Anhang zu finden (S. 81-82). Sie zeigen exemplarisch die Verteilung bei GC-reichen und -armen Chromosomen.

a



b



c

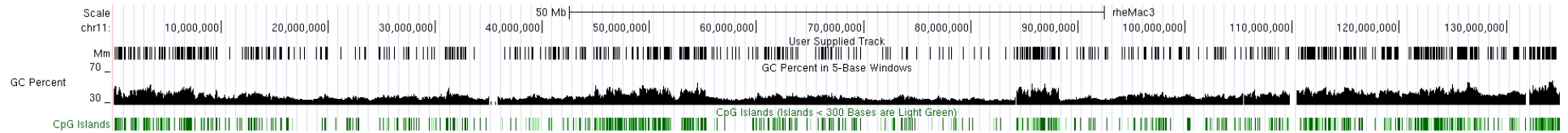


Abbildung 13: Verteilung der PRDM9-Peaks von *M.fascicularis* auf Chr11 von macFas5 (a) und rheMac8 (b) und von *M.mulatta* auf Chr11 von macFas5 (a), rheMac8 (b) und rheMac3 (c). Erstellt mit dem UCSC genome browser inkl. Darstellung des GC-Gehalts und Positionen der CpG-Inseln.

3.2.3.6 Prozentuale Verteilung auf einzelne Chromosomen

Der Mittelwert der prozentualen Verteilung aller P9- und P9H3D1-Peaks entlang der Autosomen ist in 10%-Abschnitten aufgeteilt (siehe Abbildung 14). Abgesehen von den p- und q-terminalen Abschnitten liegen die Anteile der Peaks pro Einheit überwiegend zwischen 5% und 10%. An beiden Enden steigt der prozentuale Anteil auf bis 14% bis 20%, wobei der Anteil der P9H3D1-Peaks im Vergleich zu den P9-Peaks höher liegt, sowohl bei *M.mulatta* als auch bei *M.fascicularis*. Bei wenigen Autosomen fallen die prozentualen Anteile in den telomernahen Regionen unter 10%.

Betrachtet man die prozentuale Verteilung der *M.mulatta*-Peaks entlang des Chromosoms 20 von der Referenzsequenz rheMac8, Abbildung 15, ist ebenfalls eine Häufung an den Chromosomenenden erkennbar. Vergleicht man die Peaks, die in beiden PRDM9-Replikaten vorkommen, mit jenen, die in PRDM9- und H3K4me3-Replikaten bzw. in diesen und DMC1-Replikaten zu finden sind, ist festzustellen, dass die beschriebene Verteilung der Peaks bei den letzten beiden noch stärker hervortritt. Etwa 30% der Peaks fallen hierbei in die p- und q-terminalen drei Millionen Basenpaare. Diese sechs Mb entsprechen acht Prozent des Chromosoms 20. Die weiteren 70% der Peaks verteilen sich nicht gleichmäßig über das restliche Chromosom, sondern sind an manchen Positionen gar nicht bzw. kaum vorhanden (siehe 47 Mb) oder stärker vertreten (siehe 26-29 Mb). In den ersten und letzten 10% des Chromosoms befinden sich etwa 36% aller PRDM9-Peaks und knapp 44% der P9H3D1-Peaks.

In Abbildung 16 ist Chromosom 20 von rheMac8 und macFas5 zu sehen und die Verteilung der PRDM9-Peaks beider Makakenarten. Das Muster der Peaks bei rheMac8 und macFas5 ist durchaus ähnlich, wenn auch etwas verschoben, da Chr20 von macFas5 ca. vier Mb größer ist als das von rheMac8. Vergleicht man die Peaks von *M.mulatta* und *M.fascicularis* jeweils innerhalb des Chromosoms, ist festzustellen, dass Peaks von *M.mulatta* an den Chromosomenenden stärker vertreten sind und solche von *M.fascicularis* entlang des Chromosoms immer wieder regionspezifisch erhöht sind (siehe 31 Mb bei Chr20 von macFas5).

Die Verteilung der PRDM9-Peaks auf den Geschlechtschromosomen unterscheidet sich im Vergleich zu jener auf den Autosomen. Das X-Chromosom zeigt deutlich mehr Peaks, sowohl P9 als auch P9H3D1, an den Enden. Bei *M.mulatta* sind 44% der P9- und 78% der P9H3D1-Peaks in den ersten und letzten 10% zu finden (rheMac8), bei *M.fascicularis* sind es vier von den insgesamt zehn P9-Peaks, die zwei P9H3D1-Peaks liegen in der Mitte des Gonosoms (macFas5). Von den acht P9-Peaks auf dem Y-Chromosom liegen drei in den ersten und letzten 10%, die weiteren fünf liegen separat entlang des Gonosoms. Die pseudoautosomale Region ist am Ende des kurzen Arms des X- und des langen Arms des Y-Chromosom zu finden und beträgt 2,7 Mb (Ellis et al. 1990, Hughes et al. 2012). Bei *M.mulatta* liegen 24% der P9- und 64% der P9H3D1-Peaks in der X-chromosomalen und 25% der P9-Peaks in der Y-chromosomalen PAR, bei *M.fascicularis* ist es ein P9-Peak in der X-chromosomalen PAR.

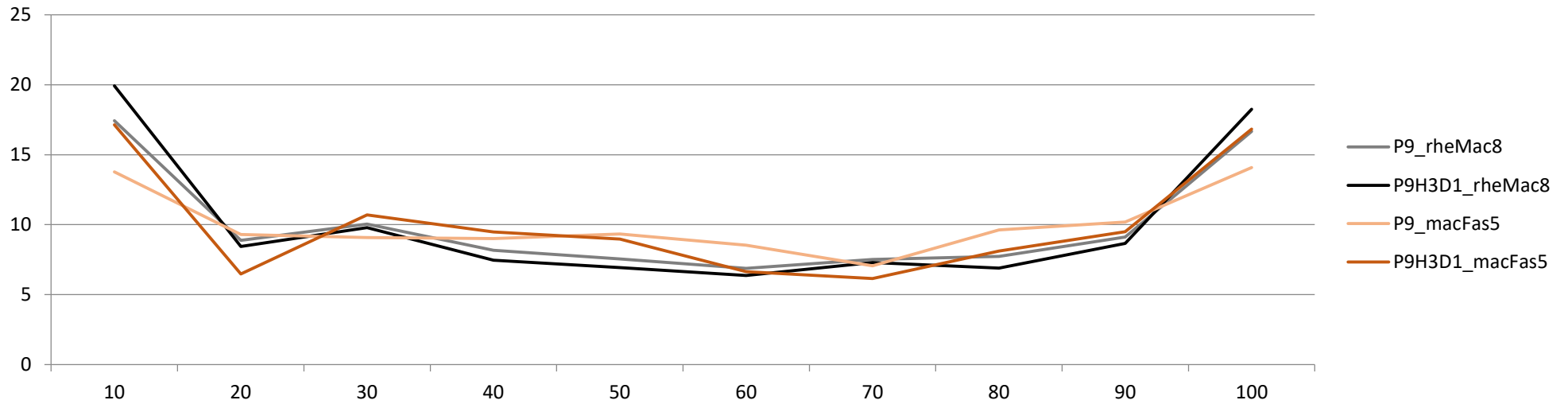


Abbildung 14: Prozentuale Verteilung der PRDM9- (P9, in hell) und PRDM9-, H3K4me3- und DMC1-Peaks (P9H3D1, in dunkel) pro 10% der Chromosomenlänge für *Macaca mulatta* auf rheMac8 (schwarz) und *Macaca fascicularis* auf macFas5 (orange), jeweils nur für die Autosomen.

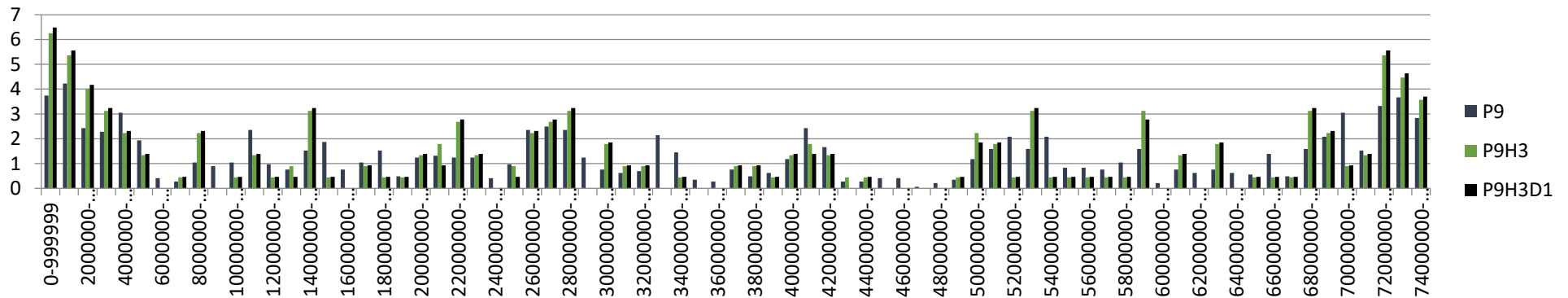
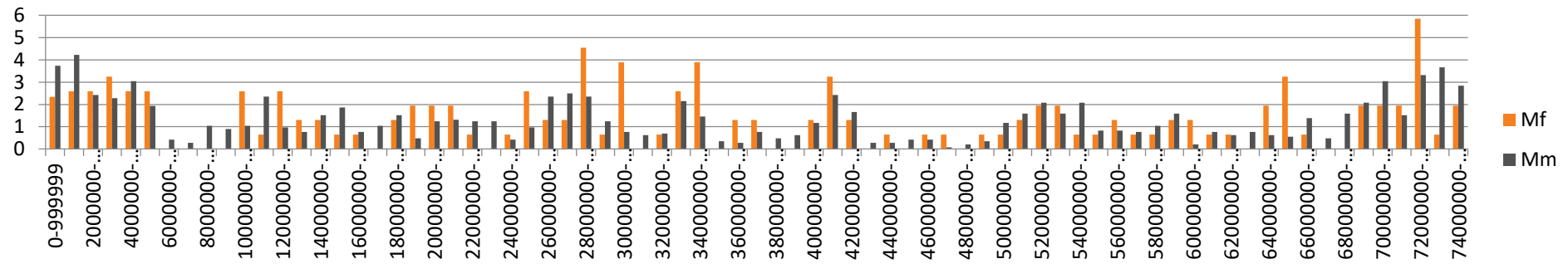


Abbildung 15: Prozentuale Verteilung der PRDM9- (P9), PRDM9- und H3K4me3- (P9H3), PRDM9-, H3K4me3- und DMC1-Peaks (P9H3D1) pro 1 Mb von *M.mulatta* auf Chromosom 20 des Referenzgenoms rheMac8.

a



b

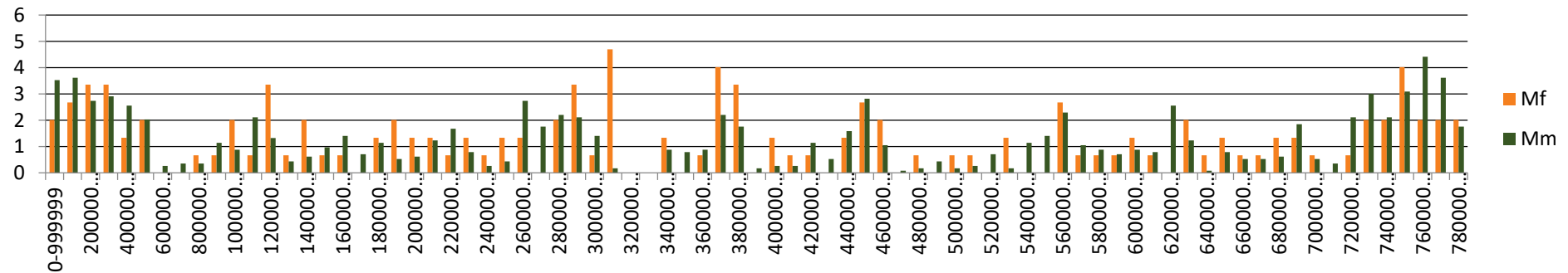


Abbildung 16: a) Prozentuale Verteilung der PRDM9-Peaks pro 1 Mb von *M.fascicularis* und *M.mulatta* auf Chromosom 20 des Referenzgenoms rheMac8. b) Prozentuale Verteilung der PRDM9-Peaks pro 1 Mb von *M.fascicularis* und *M.mulatta* auf Chromosom 20 des Referenzgenoms macFas5.

3.2.3.7 Verteilung der H3K4me3- und DMC1-Peaks PRDM9-Peaks

Der prozentuale Anteil der H3K4me3- und DMC1-Peaks ist sowohl bei *M.mulatta* als auch bei *M.fascicularis* etwa 1 kb 5' und 3' von den Zentren der PRDM9-Peaks in einem Bereich von +/- 5 kb erhöht. Die höchsten Werte liegen dabei 100 bp vor und nach den Zentren, was vor allem bei H3K4me3 zu sehen ist, da sich hier die zwei Werte deutlich von der Kurve hervorheben, siehe Abbildung 17.

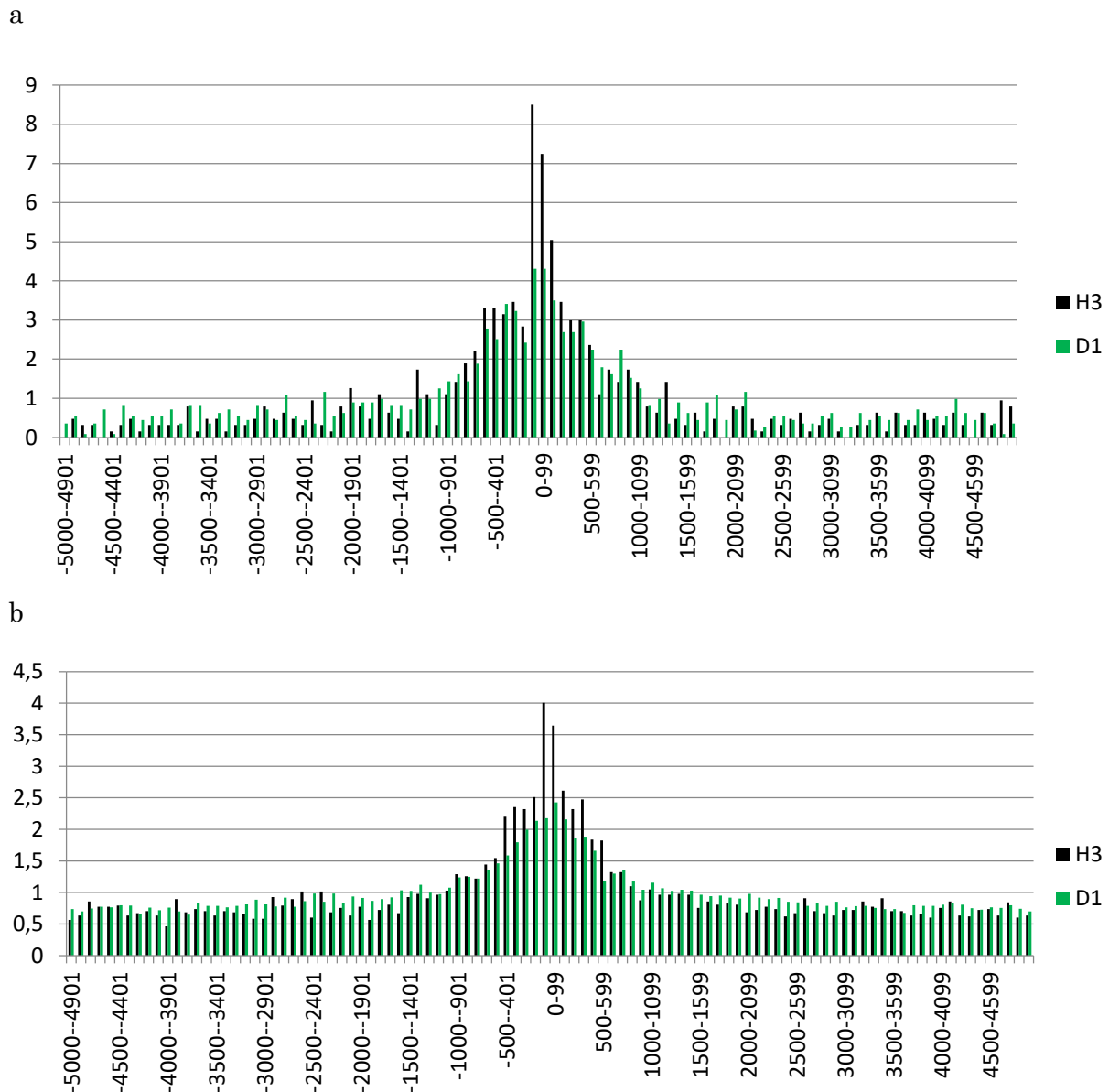


Abbildung 17: Prozentuale Verteilung der H3K4me3- (H3 in schwarz) und DMC1-Peaks (D1 in grün) +/- 5 kb um das Zentrum der PRDM9-Peaks (Position 0 im Diagramm) in 100 bp Abschnitten. In a) sind die Peaks von *M. fascicularis* auf macFas5 abgebildet, in b) jene von *M.mulatta* auf rheMac8.

3.2.3.8 Annotation PRDM9s

Der prozentuale Anteil der PRDM9-Peaks, die in 5'UTRs, Promotoren, Retroposons (SVAs, SINE-VNTR/variable number tandem repeats/-Alus) und CpG-Inseln fallen, übersteigt den Anteil dieser Bereiche im Genom um etwa ein Acht- bis Sechzehnfaches (Abbildung 18). Für Peaks in exonischen Bereichen, 3'-UTRs und TTSs (*transcription termination sites*) ist eine zwei- bis vierfache Erhöhung zu erkennen. In LINEs, *long interspersed nuclear elements*, und Satelliten-DNA liegt der Anteil der Peaks bei der Hälfte bzw. bei einem Viertel im Verhältnis zum Anteil dieser beiden Bereiche im Genom. Der prozentuale Anteil der Peaks in Introns, intergenischen Bereichen, SINEs (*short interspersed nuclear elements*) und LTRs (*long terminal repeats*) gleicht dem dieser Bereiche im Gesamtgenom. Beim Vergleich der Referenzgenome ist kein großer Unterschied zu vermerken. Die Peaks in Retroposons und Satelliten-DNA sind die Ausnahmen, da deren Anteile sich bei *M.mulatta* und *M.fascicularis* unterscheiden. Während keine Peaks von *M.fascicularis* in Retroposons fallen, sind sie bei *M.mulatta* bis zu sechzehnfach erhöht. Bei den Minisatelliten ist der Anteil der *M.mulatta*-Peaks geringer, bei *M.fascicularis* sind sie fast identisch bzw. höher.

Die hohen Balken im negativen Bereich bei Retroposons und RNA in Abbildung 18, kleiner gleich -12, können nicht in Betracht gezogen werden, da der Logarithmus von 0 nicht definiert ist und sie auf diesen basieren.

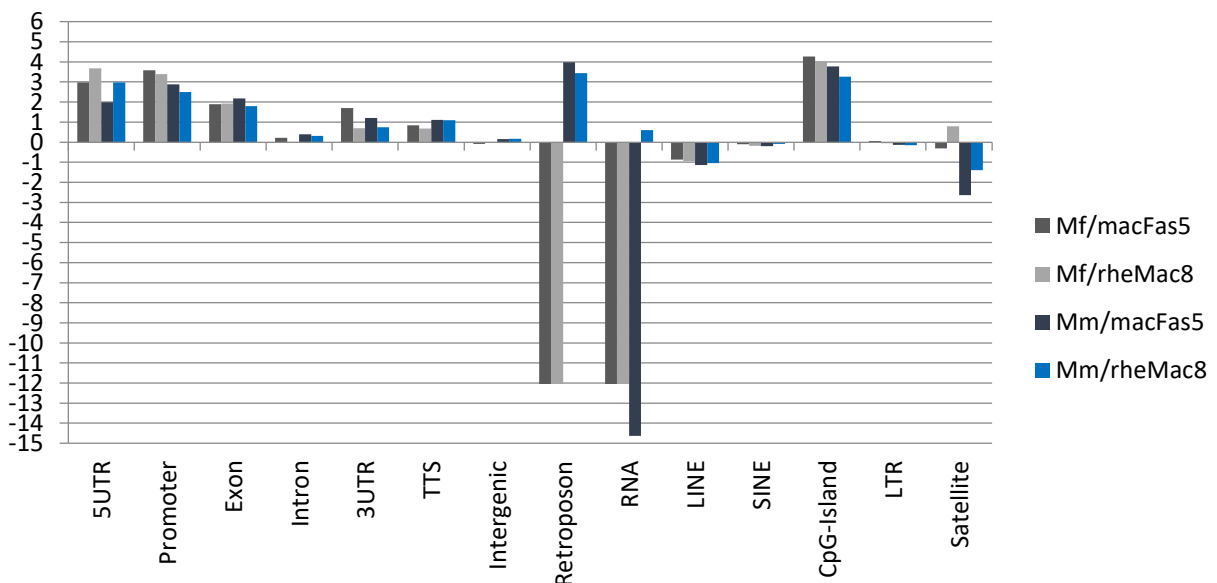


Abbildung 18: Logarithmus zur Basis 2 des Verhältnisses des prozentualen Anteils der PRDM9-Peaks in verschiedenen, nachfolgenden Bereichen zu deren Anteil im Genom. Die Bereiche umfassen: 5'UTR, Promotor, Exon, Intron, 3'UTR, TTS, intergenische Regionen, Retroposon, RNA, LINE, SINE, CpG-Inseln, LTR und Satelliten-DNA für *M. fascicularis* (grau) und *M. mulatta* (blau) für jeweils macFas5 (dunkel) und rheMac8 (hell).

Obwohl Peaks beispielsweise häufiger in Promotorbereichen zu finden sind, als diese Regionen wiederum im Genom, beträgt der prozentuale Anteil der Peaks im Promotorbereich für *M.mulatta* 1,35% und für *M.fascicularis* 0,94% (vgl. Tabelle 13). Der

Großteil der Peaks, über 87% bei *M.mulatta* und 95% bei *M.fascicularis*, liegt in intergenischen Regionen.

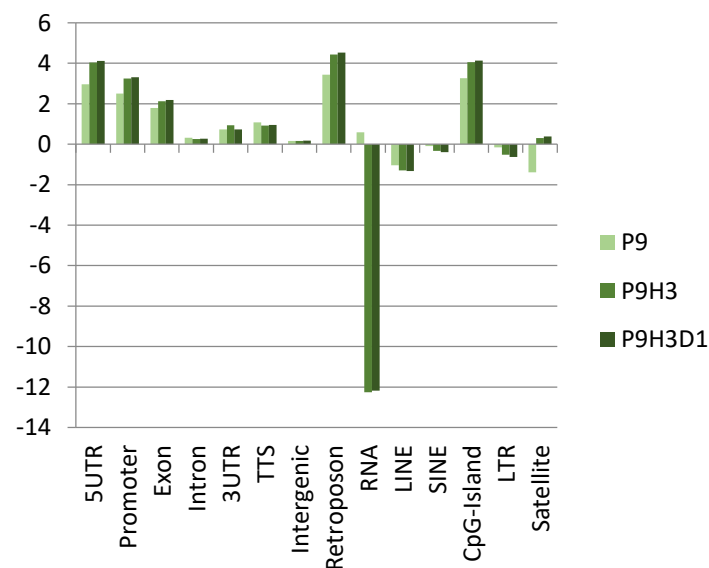
Tabelle 13: Anzahl und prozentualer Anteil der P9-Peaks pro Annotation für *M.mulatta* (Mm) auf rheMac 8 und *M.fascicularis* (Mf) auf macFas5. Die CpG-Inseln sind separat aufgeführt, da sie Teil der anderen Kategorien sind.

	Mm/rheMac8		Mf/macFas5	
Gesamtzahl an Peaks	34525		4229	
Annotation	Anzahl an Peaks	prozentualer Anteil	Anzahl an Peaks	prozentualer Anteil
3UTR	67	0,19406227	4	0,094585008
TTS	172	0,49818972	6	0,141877512
Exon	285	0,82548878	12	0,283755025
Intron	3251	9,41636495	131	3,097659021
Intergenic	30241	87,5916003	4034	95,38898085
Promoter	469	1,35843592	40	0,945850083
5UTR	40	0,11585807	2	0,047292504
CpG-Island	2053	5,94641564	498	11,77583353

Die Verteilung der P9H3- und P9H3D1-Peaks in den verschiedenen Bereichen des Genoms weist zum Großteil die gleiche Tendenz auf wie PRDM9-Peaks, meist nochmals stärker, vor allem bei *M.fascicularis*. Die Peaks dieser beiden Konstellationen in den Minisatelliten kommen bei beiden Makaken nun prozentual häufiger vor als Minisatelliten im Genom, siehe Abbildung 19.

Die negativen Werte unter -8 in Abbildung 19, bei beiden Makakenarten, basieren, wie oben, auf dem nicht definierten Logarithmus von 0.

a



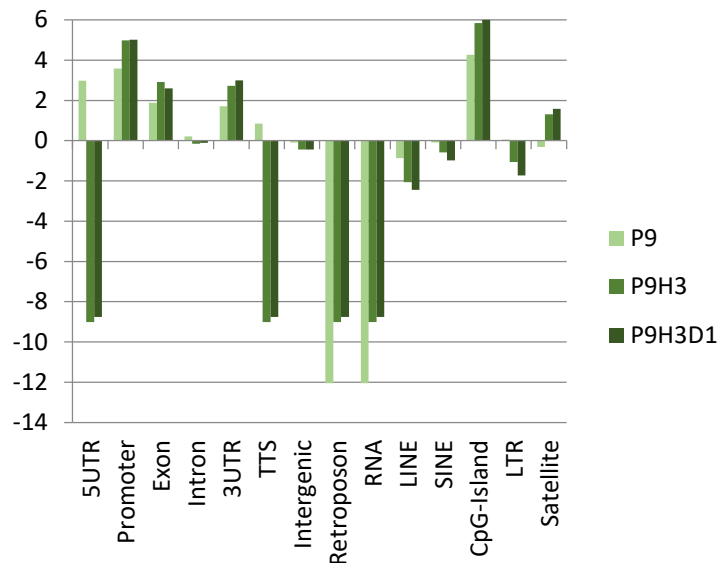
b

Abbildung 19: Logarithmus zur Basis 2 des Verhältnisses des prozentualen Anteils der P9-, P9H3- und P9H3D1-Peaks in verschiedenen, nachfolgenden Bereichen zu deren Anteil im Genom. Die Bereiche umfassen: 5'UTR, Promotor, Exon, Intron, 3'UTR, TTS, intergenische Regionen, Retroposon, RNA, LINE, SINE, CpG-Inseln, LTR und Satelliten-DNA für *M. fascicularis* auf macFas5 (b) und *M. mulatta* auf rheMac8 (a).

3.2.3.9 Gene Ontology

Die Gene Functional Classification der PRDM9-Peaks, die sich in Genen befanden, ergibt für *M. mulatta* auf rheMac8, rheMac3 und macFas5 jeweils 25, 8 und 10 Gruppen mit signifikantem *enrichment score* über eins (Huang et al. 2009b). Bei *M. fascicularis* finden sich drei signifikante Gruppen mit rheMac8 als Grundlagen, bei macFas5 als Referenzgenom waren keine vorhanden. Eine Gengruppe, die bei allen vier Konstellation auftritt, besteht aus Mitgliedern der Ras-Superfamilie. Ein Überblick zu allen Gruppen steht im Anhang zur Verfügung (S. 83-91). Von den 61 RAS-Genen, die mit PRDM9-Peaks assoziiert sind, sind 11 mit P9H3D1-Peaks zu finden

Die Anzahl der Gene, in denen PRDM9-Peaks zu finden sind, beläuft sich bei *M. mulatta* auf 2396 von 34526 und 484 von 4233 bei *M. fascicularis* mit rheMac8 als Referenzgenom. Das entspricht etwa 7 bzw. 11 %, siehe Tabelle 14. Von diesen Genen finden sich bei beiden Spezies 279 gemeinsame. Über 50% der Gene mit PRDM9-Peaks sind im Testisproteom (Zhou et al. 2015) vertreten und grob 30% sind Haushaltsgene (Eisenberg, Levanon 2013, Tabelle 14). Der prozentuale Anteil der Haushaltsgene im Testisproteom beträgt etwa 34% und ist somit höher als der Anteil der Gene mit PRDM9-Peaks in Haushaltsgenen. Der Unterschied ist bei *M. mulatta* statistisch signifikant, bei *M. fascicularis* jedoch nicht, vgl. Chi-Quadrat-Tests in Tabellen 15 und 16.

Tabelle 14: Anzahl und Anteil der PRDM9-Peaks von *M.mulatta* und *M.fascicularis* in Genen, die Übereinstimmung dieser Gene in beiden Spezies, sowie Anzahl und Anteil der Gene im Testisproteom und der Gene, die zu den Haushaltsgenen gehören. Das Referenzgenom ist jeweils rheMac8.

	Gene mit PRDM9-Peaks	%	gemeinsame Gene	%
<i>M.mulatta</i>	2396	6,93969762	279	11,6444073
<i>M.fascicularis</i>	484	11,4339712		57,6446281
	im Testisproteom	%	in Haushaltsgene	%
<i>M.mulatta</i>	1288	53,7562604	682	28,4641068
<i>M.fascicularis</i>	270	55,785124	153	31,6115702

Tabelle 15: Chi-Quadrat-Test zu Genen mit PRDM9-Peaks von *M.mulatta* bzw. Genen im Testisproteom, die zu den Haushaltsgenen gehören oder nicht, inkl. Chi-Quadrat- und p-Wert. Das Referenzgenom ist rheMac8.

	Haushaltsgene	keine Haushaltsgene	total
Gene mit PRDM9-Peaks	682	1714	2396
Gene im Testisproteom	2918	5722	8640
total	3600	7436	11036
Chi-Quadrat	24,055	p-Wert	0,00000094

Tabelle 16: Chi-Quadrat-Test zu Genen mit PRDM9-Peaks von *M.fascicularis* bzw. Genen im Testisproteom, die zu den Haushaltsgenen gehören oder nicht, inkl. Chi-Quadrat- und p-Wert. Das Referenzgenom ist rheMac8.

	Haushaltsgene	keine Haushaltsgene	total
Gene mit PRDM9-Peaks	153	331	484
Gene im Testisproteom	2918	5722	8640
total	3071	6053	9124
Chi-Quadrat	0,959	p-Wert	0,32743896

3.2.3.10 piRNA Cluster

Von den 4230 PRDM9-Peaks von *M.fascicularis* auf macFas5 liegen 28 in piRNA-Cluster, bei *M.mulatta* auf rheMac8 sind es 152 von 34526. Dabei kommt es auch vor, dass Peaks in das gleiche Cluster fallen. Zufällig erstellte Peaks sind jeweils nur halb so oft in diesen Loci zu finden. Die Unterschiede sind, wie in den Tabellen 17 und 18 zu sehen ist, signifikant.

Tabelle 17: Chi-Quadrat-Test zu Peaks von *M.fascicularis* auf macFas5 und zufällig gezogenen Loci inner- und außerhalb von piRNA-Clustern, inkl. Chi-Quadrat- und p-Wert.

Peaks	in Cluster	außerhalb von Cluster	total
beobachtet	28	4202	4230
zufällig gezogen	10	4220	4230
total	38	8422	8460
Chi-Quadrat	8,565	p-Wert	0,00342687

Tabelle 18: Chi-Quadrat-Test zu Peaks von *M.mulatta* auf rheMac8 und zufällig gezogenen Loci inner- und außerhalb von piRNA-Clustern, inkl. Chi-Quadrat- und p-Wert.

Peaks	in Cluster	außerhalb von Cluster	total
beobachtet	152	34374	34526
zufällig gezogen	67	34459	34526
total	219	68833	69052
Chi-Quadrat	33,096	p-Wert	0,00000001

Die Prozentzahlen der Cluster mit PRDM9-Peaks, sowohl bei *M.fascicularis* als auch bei *M.mulatta*, unterscheiden sich im Verhältnis zum Anteil der jeweiligen Gruppen an piRNA-Clustern. Gruppe 1 besteht aus Clustern mit speziesübergreifender Expression, Gruppe 2 aus Clustern mit speziesübergreifendem Vorkommen und Gruppe 3 aus Clustern mit speziespezifischem Vorkommen. Während bei beiden Spezies der Anteil in Gruppe 1 höher ist, ist er bei *M.fascicularis* in Gruppe 3 und bei *M.mulatta* in Gruppe 2 ähnlich, bei *M.fascicularis* in Gruppe 2 und *M.mulatta* in Gruppe 3 niedriger als der jeweilige Anteil der Gruppe zur Gesamtanzahl der Cluster.

Die im Vergleich zu oben genannten Zahlen abweichenden absoluten Gesamtzahlen in Tabelle 19 sind durch die Anwesenheit mehrerer Peaks pro Cluster zu erklären. Bei *M.fascicularis* sind es drei Peaks, die wegfallen, davon gehört einer zu Gruppe 1 und zwei zu Gruppe 3. Ein deutlicherer Unterschied ist bei *M.mulatta* vorzufinden. Hier verteilen sich 49 Peaks auf Cluster, die bereits einen Peak aufweisen. Die Mehrheit, 30 Peaks, gehören zu Gruppe 1, 12 zu Gruppe 2 und sieben zu Gruppe 3.

Tabelle 19: Anzahl und Anteil aller piRNA-Cluster und piRNA-Cluster (aus Gebert et al. 2019) mit PRDM9-Peaks in den verschiedenen Gruppen (siehe unten), sowohl für *M.fascicularis* auf macFas5 als auch für *M.mulatta* auf rheMac8.

Cluster*	Anteil der Cluster macFas5		Cluster mit PRDM9-Peaks (Mf/macFas5)		Anteil der Cluster rheMac8		Cluster mit PRDM9-Peaks (Mm/rheMac8)	
	abs.	%**	abs.	%	abs.	%**	abs.	%**
1	146	24	9	36	106	24	35	34
2	246	41	7	28	175	39	39	38
3	210	35	9	36	164	37	29	28
gesamt	602	100	25	100	445	100	103	100

*1: Cluster mit speziesübergreifender Expression, 2: Cluster mit speziesübergreifendem Vorkommen, 3: Cluster mit speziespezifischem Vorkommen. Speziesübergreifend bezieht sich dabei auf die Primatenarten *Microcebus murinus*, *Nycticebus coucang*, *Loris tardigradus*, *Callithrix jacchus*, *Macaca mulatta*, *Macaca fascicularis* und *Homo sapiens*.

** : gerundet, abs. = absolute Zahlen

3.2.3.11 Motiv-Konsensussequenz

Die Motiv-Konsensussequenz (kurz Motiv), die jeweils in allen 1000 höchsten nach p-Wert sortierten Peaks bei *M.fascicularis* und *M.mulatta* vorkommt, ist bei beiden Arten 15 Basenpaare lang und wird von Guanin bzw. Cytosin dominiert. Die zwei Motive sind nahezu identisch. Vergleicht man das *M.fascicularis*-Motiv mit dem reversen Komplement von *M.mulatta* stimmt es von Position eins bis vierzehn mit dem Bereich zwei bis fünfzehn von *M.mulatta* überein, siehe Abbildung 20. Abweichungen stellen die Anteile der Basen an den jeweiligen Positionen dar.

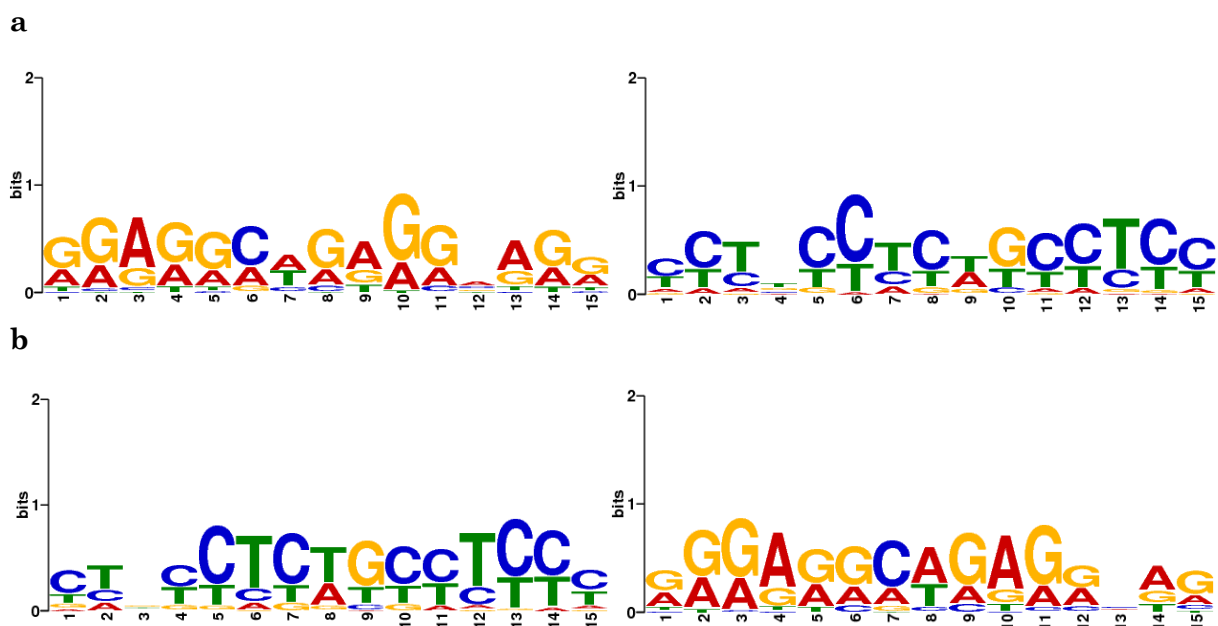


Abbildung 20: Motive, die mit MEME suite ermittelt wurden, von *M.fascicularis* (a) und *M.mulatta* (b) mit dem revers komplementären Motiv auf der jeweils rechten Seite. Die Größe der Basen steht für deren Häufigkeit, die Gesamthöhe pro Position stellt den Informationsgehalt der Position in Bit dar.

Beide Motive sind zu knapp 40% in den PRDM9-Peaks von *M.mulatta* und zu etwa 35% in den PRDM9-Peaks von *M.fascicularis* vorhanden, siehe Tabelle 20.

Tabelle 20: Übersicht zur Anzahl und zum prozentualen Anteil der Motive in den PRDM9-Peaks der beiden Makaken. Die dazugehörige Standardeinstellung für den p-Wert ist kleiner als 0,0001.

FIMO				
Motiv	Mm_PRDM9-Peaks (auf rheMac8)	Prozentualer Anteil in Peaks	Mf_PRDM9-Peaks (auf macFas5)	Prozentualer Anteil in Peaks
Mf	13781	39,91484678	1521	35,957447
Mm	13212	38,26681342	1417	33,498818

*ohne scaffolds

Die Verteilung der Peaks mit Motiv ist in Abbildung 21 zu sehen. Zudem sind auch alle Peaks von *M.fascicularis* auf macFas5 und jene von *M.mulatta* auf rheMac8 dargestellt. Bei beiden Konstellationen, sowohl für macFas5 als auch für rheMac8, sind die prozentualen Anteile der Peaks im Vergleich zum prozentualen Anteil der Chromosomen 10, 16, 19 und 20 zur Gesamtheit der Chromosomen um etwa das Doppelte erhöht. Ähnliches gilt für die entgegengesetzte Richtung und in abgeschwächter Form für die Chromosomen 2, 5, 6, 12 und 17. Die Tendenz der Peaks mit Motiv ist meist ausgeprägter.

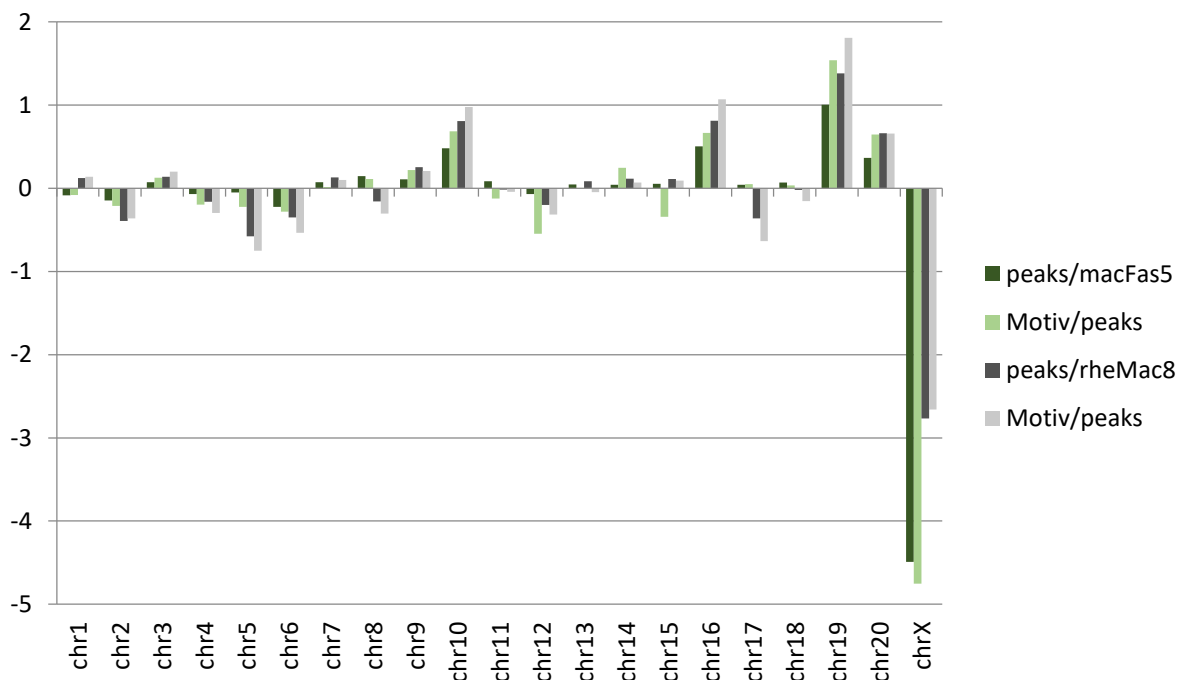


Abbildung 21: Logarithmus zur Basis 2 des Verhältnisses des prozentualen Anteils der Peaks (Peaks/macFas5 bzw. Peaks/rheMac8) und Peaks mit Motiv (Motiv/Peaks) zum Anteil der einzelnen Chromosomen im Vergleich zu allen Chromosomen für *M.fascicularis* (grün) und *M.mulatta* (grau).

3.2.3.11.1 Zusammenhang zwischen Motiv und Alu-Elementen

Der Anteil von Alu-Elementen an allen Peaks beträgt bei beiden Makaken etwa 9,5% und ist somit niedriger als die ca. 11%, die Alus im Genom beider Arten ausmachen. Betrachtet man die Peaks mit Motiv, erhöht sich der prozentuale Anteil an Alus auf mehr als 20%, was dem Doppelten an Alu-SINEs im Genom entspricht. Die Alu-Subgruppen mit unterschiedlich tief koaleszierenden Retropositionsmaxima nehmen ungefähr 24% (AluJ), 55-57% (AluS) und 19-21% (AluY) in Anspruch, siehe Tabelle 21. Im Vergleich zu diesen Zahlen ist der prozentuale Anteil an „jungen“ AluY in beiden Peak-Kategorien geringer, 5-7%. Bei „alten“ AluJ hingegen sind die Anteile bei den beiden Peak-Kategorien erhöht. Das Gegenteil ist bei AluS der Fall, bei dem die Werte bei den beiden Peaks niedriger sind.

Tabelle 21: Anzahl und prozentualer Anteil der Alus in PRDM9-Peaks und in PRDM9-Peaks mit Motiv, sowie beides für die drei Subgruppen AluJ, AluS und AluY und ein Überblick über Anzahl und prozentualen Anteil der Subgruppen aller Alus in den Genomen macFas5 und rheMac2 (ersetzt hier rheMac8).

<i>M.fascicularis</i>						
	Peaks		Peaks mit Motiv		Alus in macFas5	
	Alus	Peaks	Alus	Peaks		
	399	4230	327	1522	1136139	
prozentualer Anteil der Alus	9,432624113		21,4848883			
		%		%		%
AluJ	191	47,8696742	141	43,1192661	267228	23,5207136
AluS	187	46,8671679	162	49,5412844	646921	56,940304
AluY	21	5,26315789	24	7,33944954	221990	19,5389825
<i>M.mulatta</i>						
	Peaks		Peaks mit Motiv		Alus in rheMac2	
	Alus	Peaks	Alus	Peaks		
	3274	34526	3093	13212	1046142	
prozentualer Anteil der Alus	9,482708683		23,41053588			
		%		%		%
AluJ	1310	40,0122175	1101	35,5965082	252581	24,1440455
AluS	1790	54,6731827	1824	58,971872	576614	55,1181388
AluY	174	5,31459988	168	5,43161979	216947	20,7378157

3.2.3.11.2 Vergleich der Motive: MEME vs. Binding Site Prediction

STAMP vergleicht die mit MEME ermittelten Motive mit jenen, die anhand des Zinkfingerarrays vorhergesagt wurden. Die MEME-Motive sowohl von *M.mulatta* als auch von *M.fascicularis* alignieren jeweils in dem Bereich, in dem sich die vorhergesagten Bindungssequenzen unterscheiden. Der direkte Vergleich von *M.mulatta* und *M.fascicularis* zeigt, dass ihre MEME-Motive an der jeweils identischen Position in den drei Allelen alignieren, und, dass der e-Wert am geringsten beim längeren *M.mulatta*-Allel ist und am höchsten beim kürzeren, siehe Abbildungen 22 und 23.

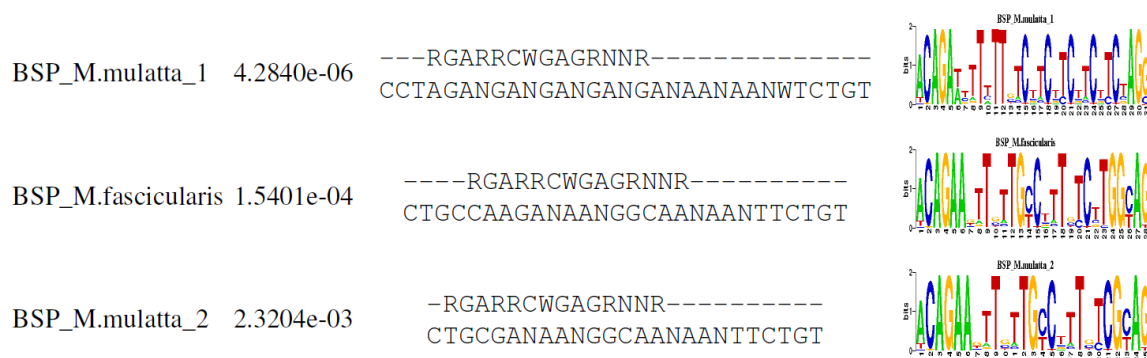


Abbildung 22: Von links nach rechts sind der Name des vorhergesagten Motivs (BSP = *binding site prediction*), der e-Wert, die Alignierung der Sequenzen der Motive mit dem MEME-Motiv von *M.mulatta* oben und die Abbildung des Motivs angegeben.

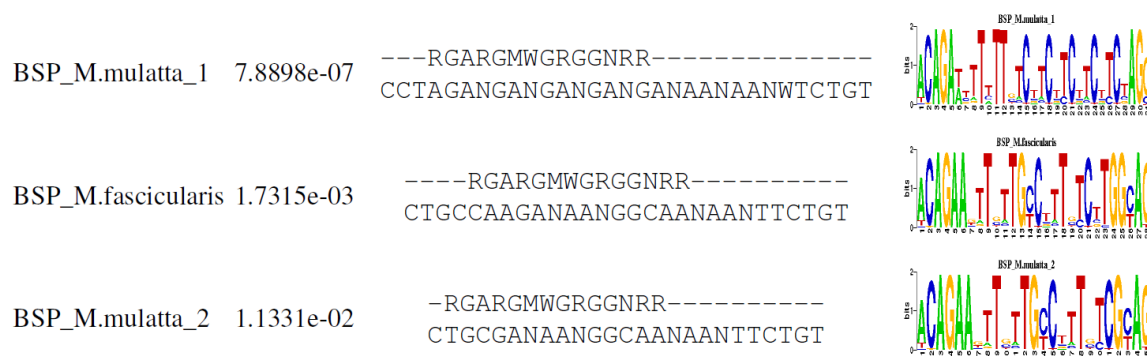


Abbildung 23: Von links nach rechts sind der Name des vorhergesagten Motivs (BSP = *binding site prediction*), der e-Wert, die Alignierung der Sequenzen der Motive mit dem MEME-Motiv von *M.fascicularis* oben und die Abbildung des Motivs angegeben.

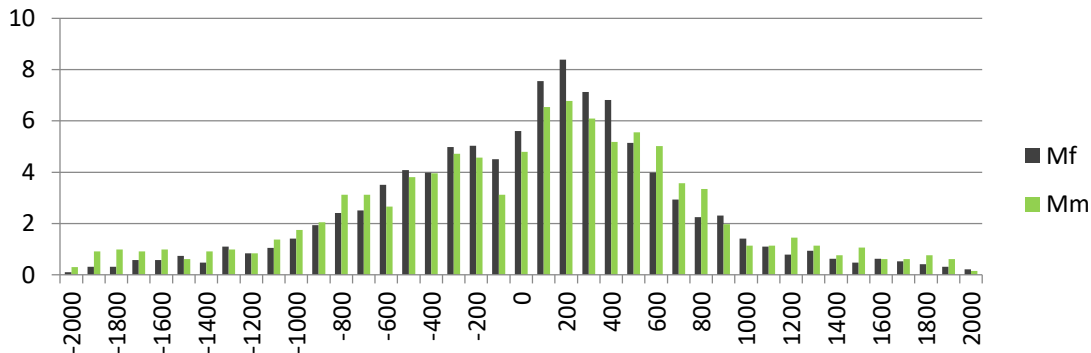
3.2.3.12 Interaktionspartner

Die mittels Massenspektrometrie und CRAPome ermittelten Interaktionspartner für PRDM9 von *M.fascicularis* sind Acrosin (ACR), Acrosin Binding Protein (ACRBP) und KIAA1210. Bei *M.mulatta* finden sich Suppressor Of Variegation 3-9 Homolog 2 (SUV39H2), Toll interacting protein (TOLLIP) und Vav Guanine Nucleotide Exchange Factor 1 (VAV1). Die Ergebnisse sind jeweils nur in einem Replikat zu finden.

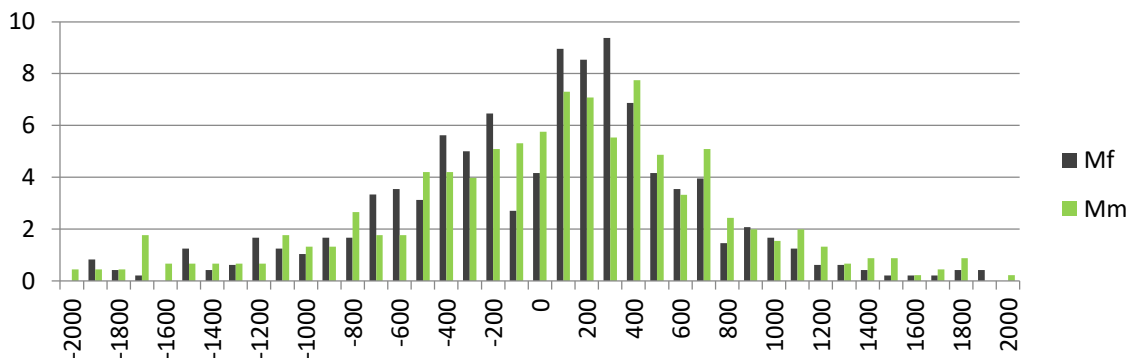
3.2.3.13 H3K4me3

Der prozentuale Anteil der H3K4me3-Peaks, die sich um die *transcription starting site* (TSS) befinden, steigt unabhängig vom Referenzgenom und der Spezies ab 1 kb 5' der TSS an und ebbt 1 kb 3' ab, mit einem Höchstwert bei etwa 300 bp auf der 3'-Seite der TSS, siehe Teilabbildungen a und b in Abbildung 24.

a



b



c

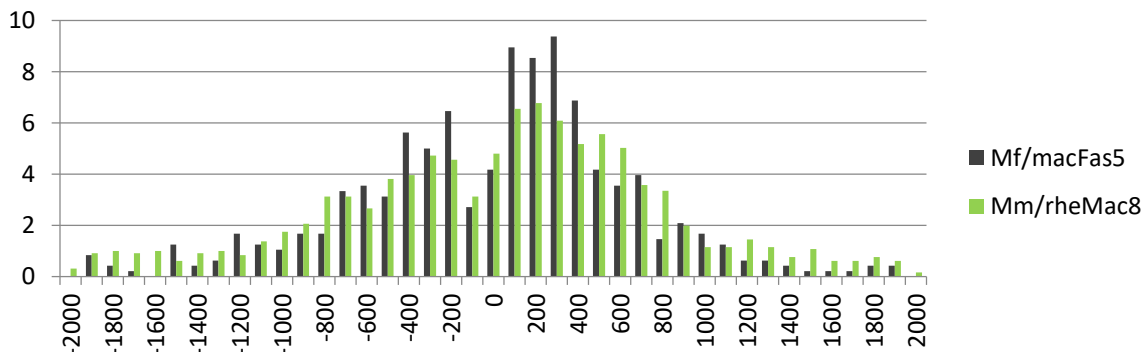


Abbildung 24: Prozentuale Verteilung der H3K4me3-Peaks um die TSS (*transcription starting site*), was 0 in den Diagrammen entspricht, in 100 bp-Abschnitten. a) zeigt die Verteilung von *M. fascicularis* (Mf) und *M. mulatta* (Mm) auf rheMac8, b) zeigt ebensolche auf macFas5 und in c) ist die Verteilung von *M. fascicularis* auf macFas5 und *M. mulatta* auf rheMac8 graphisch zusammengefasst.

Sieht man sich nun Teilabbildung c in Abbildung 24 an, fällt auf, dass ein Abfall des prozentualen Anteils der Peaks 100 bp 5' und 3' der TSS zu sehen ist. Das ist nur der Fall, wenn *M.mulatta*-Peaks auf rheMac8 und *M.fascicularis*-Peaks auf macFas5 abgebildet werden. In den Teilabbildungen a und b sind die Peaks von *M.mulatta* auf macFas5 und *M.fascicularis* auf rheMac8 zu sehen, die diese Abnahme nicht (*M.mulatta* auf macFas5) oder kaum aufweisen (*M.fascicularis* auf rheMac8).

3.2.3.13.1 Testis, Haushalt

Die Anzahl der Gene, um deren TSS H3K4me3-Peaks liegen (+/-2kb), beträgt bei *M.mulatta* 903 von 22541, bei *M.fascicularis* 1441 von 19138 mit je rheMac8 als Referenzgenom, was 4% bzw. 7,5% entspricht, vgl. Tabelle 22. Die beiden Makaken teilen sich davon 307 gemeinsame Gene. Etwa 60% der Gene mit H3K4me3-peak sind im Testisproteom zu finden und 38% gehören zu den Haushaltsgenen. Der prozentuale Anteil der Haushaltsgene im Testisproteom beträgt etwa 34% und ist somit geringer als der Anteil der Gene mit H3K4me3-Peaks in Haushaltsgenen. Die Unterschiede sind zudem statistisch signifikant, siehe Tabellen 23 und 24.

Tabelle 22: Anzahl und prozentualer Anteil der H3K4me3-Peaks von *M.mulatta* und *M.fascicularis* +/-2kb um die TSSs herum, die Übereinstimmung dieser Gene in beiden Spezies; Anzahl und Anteil der Gene (mit H3K4me3 um TSS) im Testisproteom und der Gene (mit H3K4me3 um TSS), die zu den Haushaltsgenen gehören. Das Referenzgenom ist jeweils rheMac8.

	Gene mit H3K4me3-Peaks um TSS	%	gemeinsame Gene	%
<i>M.mulatta</i>	903	4,00603345	307	33,9977852
<i>M.fascicularis</i>	1441	7,52952242		21,3046495
	im Testisproteom	%	in Haushaltsgene	%
<i>M.mulatta</i>	512	56,6998893	344	38,0952381
<i>M.fascicularis</i>	892	61,9014573	537	37,2657876

Tabelle 23: Chi-Quadrat-Test zu Genen mit H3K4me3-Peaks +/-2kb um die TSSs herum von *M.mulatta* bzw. Genen im Testisproteom, die zu den Haushaltsgenen gehören oder nicht, inkl. Chi-Quadrat- und p-Wert. Das Referenzgenom ist rheMac8.

	Haushaltsgene	keine Haushaltsgene	total
Gene mit H3K4me3-Peaks	344	559	903
Gene im Testisproteom	2918	5722	8640
total	3262	6281	9543
Chi-Quadrat	6,788	p-Wert	0,00917727

Tabelle 24: Chi-Quadrat-Test zu Genen mit H3K4me3-Peaks +/-2kb um die TSSs herum von *M.fascicularis* bzw. Genen im Testisproteom, die zu den Haushaltgenen gehören oder nicht, inkl. Chi-Quadrat- und p-Wert. Das Referenzgenom ist rheMac8.

	Haushaltsgene	keine Haushaltsgene	total
Gene mit H3K4me3-Peaks	537	904	1441
Gene im Testisproteom	2918	5722	8640
total	3455	6626	10081
Chi-Quadrat	6,688	p-Wert	0,0097064

3.2.3.13.2 GO mit DAVID

Die Gene, um deren TSS (+/-2kb) herum sich die H3K4me3-Peaks finden, kodieren bei den *M.mulatta*-Peaks auf rheMac8 und rheMac3 für Untereinheiten der RNA-Polymerasen 2 und 3 sowie RNA-Bindungsmotivproteine und Splicingfaktoren. Bei den *M.mulatta*-Peaks auf den beiden *M.mulatta*-Genomen und macFas5 finden sich (mitochondrial-) ribosomale Proteine. Bei den *M.fascicularis*-Peaks, sowohl auf rheMac8 als auch macFas5, gehören Gene, die für (mitochondrial-) ribosomale Proteine kodieren, und Mitglieder der Ras-Superfamilie.

3.2.3.13.3 Annotation

Eine deskriptiv statistische Auswertung der genomischen Loci der H3K4me3-Peaks zeigt, dass die Peaks in 5'-UTRs, Promotoren, Retroposons und CpG-Inseln um das acht- bis 32-fache häufiger vorkommen als dies der prozentuale Anteil dieser Loci im Genom erwarten lässt. Bei Exons und RNA-Wiederholungseinheiten sind sie zum Teil zwei- bis viermal so oft vertreten. In LINEs und Satelliten-DNA kommen sie im Vergleich zu diesen Loci im Genom nur halb so oft vor. Im Vergleich liegt bei den beiden Makakenarten unabhängig vom Referenzgenom meist eine gleiche Tendenz vor, die bei *M.fascicularis* etwas ausgeprägter ist, siehe Abbildung 25.

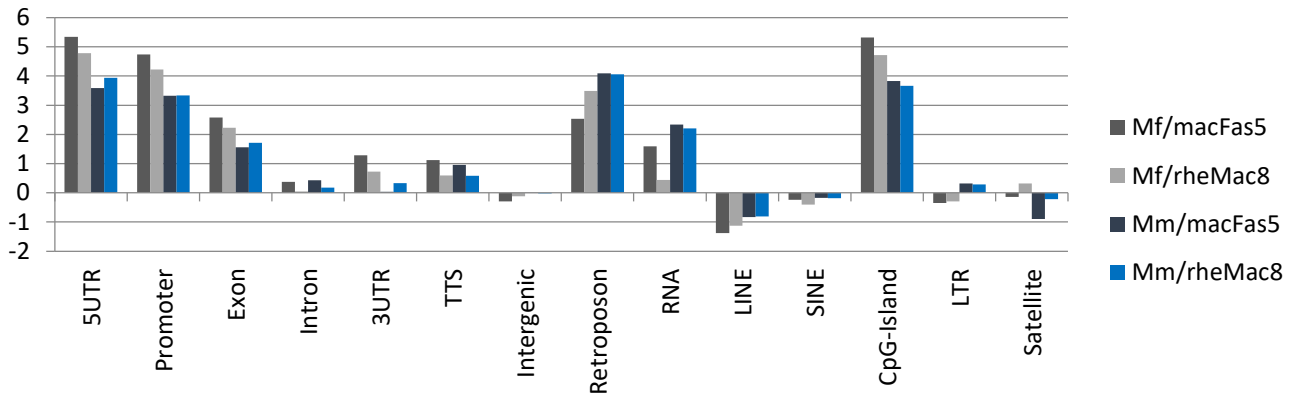


Abbildung 25: Logarithmus zur Basis 2 des Verhältnisses des prozentualen Anteils der H3K4me3-Peaks in verschiedenen, nachfolgenden Bereichen zu deren Anteil im Genom. Die Bereiche umfassen: 5'UTR, Promotor, Exon, Intron, 3'UTR, TTS, intergenische Regionen, Retroposon, RNA, LINE, SINE, CpG-Inseln, LTR und Satelliten-DNA für *M. fascicularis* (Mf, grau) und *M. mulatta* (Mm, blau) für jeweils macFas5 (dunkel) und rheMac8 (hell).

3.3 Diskussion

Die hier charakterisierten Eigenschaften der PRDM9-Bindungsstellen bei *Macaca mulatta* und *Macaca fascicularis* und die Zusammenhänge mit genomischen Regionen und Interaktionspartner werden im Folgenden ebenso diskutiert wie die Eigenschaften der Histonmodifikation H3K4me3. Final werden die Unterschiede zwischen den beiden Makakenarten bezüglich der PRDM9-Peaks bzw. Rekombinationshotspots aufgegriffen und die möglichen Auswirkungen der unterschiedlichen Zinkfingerarrays auf die Artentstehung beleuchtet.

3.3.1 Reads, Peaks und Qualitätsmerkmale

Der Anteil der Reads, der ausgeschlossen wurde, da sie mehrfach oder gar nicht zu Mappen bzw. redundant waren, hält sich in dem Rahmen, den das encode-Projekt für zulässig beschreibt. Die Qualitätswerte sind auch positiv zu bewerten.

Auffällig ist, dass ausschließlich bei den PRDM9-Peaks von *Macaca fascicularis* die Anzahl deutlich geringer ausfällt. Der Grund für die wenigen PRDM9-Peaks liegt wohl beim zweiten Replikat, das bei BGI sequenziert wurde. Trotz ähnlicher Anzahl an Reads, auch nach den oben erwähnten Reduktionsschritten, erreichten die Peaks nach dem Zusammenführen der Replikate eine vierstellige Summe, während die Peaks aller anderen Konstellationen fünfstellig sind. Das ist bei allen drei Referenzgenomen der Fall.

3.3.2 Verteilung der PRDM9-Peaks und GC-Gehalt

Betrachtet man die Verteilung der PRDM9-Peaks auf allen Chromosomen und auf den einzelnen Chromosomen, fällt eine Häufung auf GC-reichen Chromosomen, auch unter Ausschluss von CpG-Inseln, auf. Rangkorrelationen unterstützen diese Aussage und zeigen einen stark positiven Zusammenhang. Zudem sind die Peaks vermehrt in GC-reichen Regionen der Chromosomen zu finden. Aus früheren Studien zu verschiedensten Organismen ist eine Verbindung zwischen GC-Gehalt und Rekombinationshotspots bekannt: Mensch (Ikemura und Wada 1991, Eyre-Walker 1993, Eisenbarth et al. 2000 und 2001; Fullerton et al. 2001, Galtier et al. 2001), Nager (Williams und Hurst 2000, Jensen-Seaman et al. 2004), Vogel (Hurst et al. 1999, Galtier et al. 2001), Fliege (Marais et al. 2001, Takano-Shimizu 2001), Wurm (Marais et al. 2001) und Hefe (Gerton et al. 2000) beschrieben, siehe dazu Birdsall et al. (2002). Ob GC-reiche Regionen eine Voraussetzung für Rekombinationshotspots sind (Gerton et al. 2000, Petes 2001, Blat et al. 2002, Petes and Merker 2002, Marsolier-Kergoat und Yeramian 2009) oder Rekombinationshotspots GC-reiche Regionen durch Genkonversion entstehen lassen (Bernardi 1993, Montoya-Burgos et al. 2003, Meunier and Duret 2004, Clément und Arndt 2013), wird zwar in Jensen-Seaman et al. (2004) sowie in Popa et al. (2012) aufgegriffen, bleibt jedoch ungeklärt.

3.3.3 Verteilung auf einzelne Chromosomen

3.3.3.1 Autosome und Telomere

Die PRDM9-Peaks sind nicht gleichmäßig auf den einzelnen Chromosomen verteilt. Eine Häufung an den p- und q-terminalen Enden ist sowohl am Beispiel von Chromosom 20 als auch bei den Mittelwerten aller Chromosomen zu sehen. Diese Bereiche sind GC-reich (siehe Abbildung 13, Costantini et al. 2006). Das steht im Einklang mit den Aussagen im vorherigen Kapitel, dass GC-reiche Regionen und Rekombination zusammenfallen. Außerdem ist die Rekombinationsrate höher in Richtung Telomere und niedriger in Zentromernähe; bei männlichen Individuen, bei Menschen und Mäusen, ist dieser Zusammenhang deutlicher als bei weiblichen (Broman et al. 1998, Yu et al. 2001, Kong et al. 2002, Paigen et al. 2008). Da Zentromere wichtig sind für die Segregation der Chromosomen und Crossing-over hier bzw. in ihrer Nähe zu Aneuploidien führen können (Lamb et al. 1996, Thomas et al. 2001), sind sie durch RNAi (RNA-Interferenz) und Heterochromatin vor Crossing-over geschützt (Ellermeier et al. 2010).

Noch deutlicher wird dieses Verteilungsmuster bei den Peaks, die bei PRDM9, H3K4me3 und DMC1 zu finden sind. Diese Peaks stehen aufgrund der Funktion DMC1s bei der Alignierung der homologen Chromosomen (Li et al. 1997, Pittman et al. 1998, Yoshida et al. 1998) für Loci mit Doppelstrangbrüchen (DSB). Somit sind bereits die potentiellen Hotspots der Rekombination (PRDM9-Peaks) eher an den Chromosomenenden anzutreffen, aber die Auswahl für die DSB verschiebt die Verteilung nochmals weiter Richtung Chromosomenenden. Wie die Loci für DSB und die Crossing-over-Ereignisse letztlich gewählt werden, ist bisher nicht bekannt. Die H3K4me3-Peaks sind vor allem stark in den 100 bp-Bereichen vor und nach dem PRDM9-Zentrum vertreten, während die DMC1-Peaks zwar auch Richtung Zentrum ansteigen und dann wieder abfallen, aber ohne einen deutlichen Ausschlag direkt vor und nach dem Mittelpunkt. Dieses Muster ist auch in Paigen und Petkov (2018) mit Hinweis auf Brick et al. (2012) und Baker et al. (2014) beschrieben worden.

Bei den Chromosomen 1, 2 und 19 von *M.mulatta* liegen jeweils an einem Chromosomenende weniger als 10% der Peaks. Möglicherweise sind die gegenüberliegenden Enden häufiger im Rekombinationsprozess involviert und ein einziges Rekombinationsereignis kann ausreichend sein (Hultén 1974). *M.fascicularis* weist zwar mehr Chromosomen deren Enden weniger Peaks besitzen, auf, allerdings kann dies auf der insgesamt geringeren Anzahl an Peaks beruhen.

3.3.3.2 Gonosomen und PAR

Die Geschlechtschromosomen gehören zu jenen mit niedrigem GC-Gehalt von etwa 40 %. Aber im Vergleich mit anderen Chromosomen mit ähnlichem bzw. geringerem GC-Gehalt, wie Chr 6 und 17, ist die Anzahl der PRDM9-Peaks im Verhältnis zur Chromosomengröße deutlich kleiner. Somit ist der GC-Gehalt nicht ausschlaggebend für die vergleichsweise wenigen Peaks. Chromosom X hat, abgesehen von den zwei P9H3D1-Peaks auf Chr X von macFas5, zudem den höchsten prozentualen Wert in den ersten 10% des Chromosoms,

sowohl bei P9- als auch bei P9H3D1-Peaks. Die wenigen Peaks und deren starke Präsenz am Ende des kurzen Arms sind höchstwahrscheinlich darauf zurückzuführen, dass die Rekombination der Geschlechtschromosomen bei männlichen Individuen in einem bestimmten Bereich stattfindet, der als pseudoautosomalen Region (PAR) bezeichnet wird. Diese befindet sich bei *M.mulatta* im Gegensatz zum Menschen, der zwei PARs besitzt, nur am Ende des kurzen Arms bei Chromosom X und am Ende des langen Arms des Y-Chromosoms (Ellis et al. 1990, Hughes et al. 2012). Ob PRDM9 überhaupt für die Rekombinationshotspots in der pseudoautosomalen Region verantwortlich ist, ist je nach Spezies unterschiedlich: Bei der Maus soll ein unbekannter Mechanismus die Hotspots und somit das Crossing-over zwischen X und Y-Chromosom garantieren (Brick et al. 2012). Beim Menschen (Pratto et al. 2014, Hinch et al. 2014) soll PRDM9 allerdings die Loci in der PAR bestimmen. Bei *Macaca mulatta* ist der Großteil der DMC1-Peaks, die mit H3K4me3- und PRDM9-Peaks zusammenfallen, in der PAR des X-Chromosoms zu finden. Das lässt vermuten, dass PRDM9 hier durchaus eine Rolle in der Vorbereitung der Rekombination spielt. Die wenigen Peaks bei *Macaca fascicularis* könnten dafür verantwortlich sein, dass für diese Spezies nicht ähnliches berichtet werden kann. Anders liegt der Fall bei *M.mulatta* und dem Y-Chromosom. Hier sind zwar PRDM9-, aber keine P9H3D1-Peaks zu verzeichnen. Ob die Rekombination somit ausschließlich vom X-Chromosom ausgeht oder doch ein Alternativprogramm zum Zuge kommt, bleibt fraglich.

Die Rekombination bei Geschlechtschromosomen ist beim heterogametischen Geschlecht räumlich und zeitlich von jener der Autosomen getrennt. Das X- und Y-Chromosom formen einen Komplex, der *sex body* oder *XY-body* genannt wird, und werden für die Zeit inaktiviert, während derer sich die Autosomen rekombinieren; der Austausch zwischen den Gonosomen findet erst danach statt (siehe dazu z.B. Hoyer-Fender 2003, Handel 2004, Kauppi et al. 2012). SUV39H2, das laut Immunopräzipitation und Massenspektrometrie mit PRDM9 von *M.mulatta* interagieren könnte, ist im *XY-body*-Komplex angereichert (O'Carroll et al. 2000).

3.3.4 Assoziation mit genomischen Regionen/Charakteristika/Merkmalen

3.3.4.1 Annotation

Der Großteil der PRDM9-Peaks ist in intergenischen Regionen (87-95%) zu finden, die nächstgrößeren Anteile in den Introns (3-9%) und CpG-Inseln (CGIs, 6-12%). Bei den ersten beiden stimmt dies in etwa mit dem Anteil dieser Loci im Genom überein. Der Anteil der Peaks in den CpG-Inseln ist höher als der prozentuale Anteil der CGIs am Genom. Eine mögliche Erklärung ist, dass diese Bereiche GC-reich sind und somit der in Kapitel 3.3.2 dargestellte Zusammenhang zwischen GC-Gehalt und Rekombinationshotspots als Grundlage dient. Eine andere ist jene, dass CGIs häufig in Promotoren vorkommen (Strachan und Read 2005): die geringe Prozentzahl der Peaks von etwa 0,95-1,36% in Promotoren stimmt mit den Ergebnissen von Brick et al. (2012) überein, die besagen, dass PRDM9 die Doppelstrangbrüche von den Transkriptionsstartsites wegdirigiert, da, im Gegensatz zu 3% beim Wildtyp, 44% der Rekombinationshotspots bei

PRDM9-Knockout-Mäusen in Promotorregionen zu finden sind. Dennoch sind in genspezifischen Regionen wie Promotoren, UTRs, Exons und TTs häufiger Peaks vorhanden als diese Regionen im Genom anteilig sind. Dabei ist eine ähnliche Erhöhung um etwa das Sechzehnfache sowohl bei 5'UTRs, Promotoren und CGIs zu beobachten. Circa 7-11% der *PRDM9*-Peaks kommen in Genen (Promotor, Exon, Intron, TTS, 5'- und 3'UTR) vor. Mehr als die Hälfte dieser Gene von *M.fascicularis* sind auch in *M.mulatta* zu finden. Rekombinationsreiche Regionen weisen beim Menschen eine erhöhte Mutationsrate auf (Lercher und Hurst 2002), was vor allem in Genen und Regulationsbereichen von Nachteil sein kann. *PRDM9*-Bindungsstellen finden sich beim Menschen zwar auch in Promotorbereichen, allerdings sollen sie nicht zu Rekombination führen, möglicherweise aufgrund einer fehlenden H3K36me3-Markierung, und stattdessen diese Gene regulieren bzw. aktivieren (Hayashi et al. 2005, Cano-Rodriguez et al. 2016, Altemose et al. 2017). Dazu passend fällt mehr als die Hälfte der Gene mit *PRDM9*-Peaks bei Makaken mit jenen Genen zusammen, die das Testisproteom ausmachen. Der Anteil der Gene mit *PRDM9*-Peak in den Haushaltsgenen (ca. 30%) ist geringer als der Anteil der Testisproteom-Gene in den Haushaltsgenen. Somit sind die „*PRDM9*-Gene“ im Vergleich nicht gehäuft in den Haushaltsgenen anzutreffen. Zhou et al. (2013) untersuchten Gene, die mit Hotspots zusammenfallen und stellten dabei fest, dass diese Gene in Haushaltsgenen signifikant unterrepräsentiert sind.

3.3.4.2 Gene Ontology: *PRDM9* und Krebs bzw. Ras

Mitglieder der *PRDM*-Familie sind dafür bekannt, dass sie als Onkogene und Tumorsuppressorgene agieren (siehe dazu die Reviews: Fog et al. 2012, Mzoughi et al. 2016, Hohenauer und Moore 2012). *PRDM9* selbst wird in Hussin et al. (2013) und Woodward et al. (2014) mit akuter lymphatischer Leukämie (ALL) in Verbindung gebracht. Seltene *PRDM9*-Allele sind häufiger bei Kindern mit ALL und ihren Eltern zu finden. Ob die versetzten Bindungsstellen während der parental Meiose eine nicht vorgesehene Expression *PRDM9*s außerhalb der Meiose und in den Krebszellen oder ein weiterer Mechanismus Grundlage für die Verbindung darstellen, bleibt unklar. Die Expression *PRDM9*s in Krebszellen wurde in Feichtinger et al. (2012) und Houle et al. (2018) untersucht und in 5 bzw. 32 Arten nachgewiesen. Houle et al. (2018) berichten, dass es an den somatischen Bindungsstellen von *PRDM9* zu Doppelstrangbrüchen kommt. Wenn diese nicht umgehend repariert werden, kann dies zu strukturellen Veränderungen führen (Jackson 2002) und somit zu genomischer Instabilität, was wiederum ein Kennzeichen von Krebserkrankungen darstellt (Hanahan und Weinberg 2011).

Sowohl bei *M.fascicularis* als auch bei *M.mulatta* überschneiden sich *PRDM9*-Peaks mit Genen, die Vertreter der Ras-Superfamilie, vor allem der dazugehörigen Rab- und Ras-Subfamilien, sind. Sie kodieren für GTPasen. Diese Enzyme sind aktiv, wenn sie GTP (Guanosintriphosphat) binden. Wird GTP zu GDP (Guanosindiphosphat) durch Abspaltung eines Phosphats mithilfe von GAP (GTPase-aktivierendes Protein), sind die Enzyme inaktiv. GTP-Austauschfaktoren (GEF, Guanosine triphosphate Exchange Factor) bewirken, dass GDP durch GTP ersetzt wird. Die Funktionen der GTPase-Subfamilien erstrecken sich von Signalwegen, Zellproliferation und -differenzierung und

Apoptose (Ras) zu intrazellulärem Transport von Proteinen (Rab) oder Vesikel (Arf) und Genexpression (Rho, Ras), siehe dazu Goitre et al. (2014). Außerdem sind Zusammenhänge zwischen den GTPasen und Krebs bekannt, wie z.B. die Hoch- und Herunterregulierungen von Rab in Krebszellen (Hendrix et al. 2010, Nam et al. 2010, Ho et al. 2012, Li et al. 2015, siehe dazu Banworth und Li 2018) oder, dass Mutationen in KRas, NRas und HRas in 30% der Tumore beim Menschen zu finden sind (Fernández-Medarde und Santos 2011). Einige Mitglieder der Ras-Superfamilie sind in der akuten lymphatischen Leukämie involviert: Rab1A, Rab7A, Rab9B und Rab14 sind hoch-, während Rab22B, Rab27A und Rab32 herunterreguliert sind (Rodrigues und Pereira-Leal 2014, Banworth und Li 2018). Villagomez et al. (2017) sehen Rab35 als potentiell Onkogen in der ALL. Mutationen in NRas und KRas sind häufig bei der ALL zu finden, vor allem auch bei Kindern, die rückfällig werden (Shu et al. 2004, Liang et al. 2006, Yamamoto et al. 2006, Fernández-Medarde und Santos 2011, Irving et al. 2014). Rab5C soll als Regulator des B-ALL-Zellwachstums agieren (Tan et al. 2014). Von diesen Ras-Genen sind Rab1A, Rab5C, Rab27A, Rab32 und Rab35 jene, die mit PRDM9-Peaks zusammenfallen, siehe die Listen zur Gene Ontology im Anhang (S. 83-91).

Somit bergen DSB und eine damit verbundene, möglich fehlerhafte Reparatur (Strathern et al. 1995, Lupski 1998, Lercher und Hurst 2002, Hicks et al. 2010, Deem et al. 2011, siehe auch Rodgers und McVey 2016) ein hohes Risiko. Allerdings sind weniger Übereinstimmungen von P9H3D1-Peaks mit diesen Genen zu verzeichnen. Dies könnte darauf hindeuten, dass hier wie bei anderen Genen regulär kaum Doppelstranbrüche bzw. keine Rekombination stattfinden (siehe Altemose et al. 2017) und PRDM9 an diesen Loci die Transkription der Gene aktiviert bzw. eine noch unbekannt Funktion ausübt (Hayashi et al. 2005, Cano-Rodriguez et al. 2016).

Zum Cluster der Ras-Gene gehören auch zwei Septine, SEPT6 und SEPT7, die zusammen mit SEPT12, SEPT2 und SEPT4 eine Filamentstruktur am Schlussring, der sich zwischen dem Mittelstück und dem Schwanz des Spermiums befindet, bilden (Kuo et al. 2015).

3.3.4.3 Repetitive Elemente

Einige repetitive Elemente, wie Alus, SVAs und piRNAs sind GC-reich (International Human Genome Sequencing Consortium 2001, Gebert et al. 2019), was sie als mögliche Rekombinationshotspots prädestinieren könnte, siehe Kapitel 3.3.2. Yamada et al. (2017) zeigen, dass bei Mäusen Doppelstrangbrüche zu etwa einem Drittel in repetitiven Einheiten stattfinden, was einem hohen prozentualen Anteil entspricht. Da repetitive Regionen allerdings 50 % des Genoms ausmachen, sind DSB in diesen Bereichen unterrepräsentiert.

Nicht Allelische Homologe Rekombinationen (NAHR) entstehen meist durch inäquales Crossing-over zwischen zwei nicht homologen *low copy repeats* (LCR), auch segmentale Duplikationen genannt (siehe dazu Lupski 1998, Stankiewicz und Lupski 2002). Allerdings kann eine solche ektopische Rekombination auch zwischen repetitiven Elementen stattfinden, wie bei Alus (siehe dazu Deininger und Batzer 1999, Bailey et al. 2003, Sen et al. 2006), LINEs (Burwinkel und Kilimann 1998, Han et al. 2008) oder

vermutlich bei SVAs (siehe dazu Ostertag et al. 2003, Hancks und Kazazian 2010, Lee et al. 2012). Somit wäre zu erwarten, dass PRDM9 hier weniger oft bindet. Abgesehen von den SVA genannten Retroposons, sind die PRDM9-Peaks seltener oder erwartungsgemäß oft in repetitiven Einheiten bei *M.mulatta* zu finden. Die SVAs in den *Cercopithecidae* (MacSVAs) sind Vorgänger der SVAs in den Hominiden (Han et al. 2007). Erstere sind kürzer, etwa 350 bp, und seltener, 100-200 Einheiten, im Vergleich zu 900 bp großen SVAs, die 2000 bis 5500 Mal in den Hominiden vorkommen (Tang und Liang 2019). Insgesamt sind sieben PRDM9-Peaks von *M.mulatta* in SVAs zu finden. Im Gegensatz dazu sind bei *M.fascicularis* keine in SVAs vorhanden. Beim Menschen sind SVAs eher in coldspots zu finden (Myers et al. 2008). NAHR führt zu Umgestaltungen wie Insertionen, Deletionen als auch Inversionen (siehe dazu Lupski 1998, Stankiewicz und Lupski 2002) und kann zu Krankheiten führen; Charcot-Marie-Tooth, HNPP (*Hereditary Neuropathy with liability to Pressure Palsies* bzw. Hereditäre Neuropathie mit Neigung zu Druckläsionen), X-chromosomale Ichthyose, das Potocki-Lupski- und das Smith-Magenis-Syndrom entstehen durch NAHR (Pentao et al. 1992, Chance et al. 1994, Reiter et al. 1996, Chen et al. 1997, Potocki et al. 2000, Shaw et al. 2002, van Esch et al. 2005, Potocki et al. 2007) und weisen bei Menschen zudem von PRDM9 determinierte Hotspots bzw. dessen Bindungssequenz auf (Myers et al. 2008, Berg et al. 2010, Zhang et al. 2010, Pratto et al. 2014, Altemose et al. 2017).

KRAB-Zinkfingerproteine können diverse Funktionen ausüben. Sie sind beispielsweise an Imprinting und Embryogenese beteiligt (Li et al. 2008, Quenneville et al. 2012). Ihre Hauptfunktion ist jedoch die Expression verschiedener transposabler Elemente (TEs) in embryonischen Stammzellen (ESC) zu unterdrücken (Wolf und Goff 2009, Nishitsuji et al. 2012, Tan et al. 2013, Jacobs et al. 2014, siehe auch Wolf et al. 2015). Dafür interagieren sie unter anderem mit KAP1 (Friedman et al. 1996, Peng et al. 2000). Da die Anzahl der TEs und der KRAB-Zinkfingerproteine korrelieren, scheint eine Koevolution der beiden wahrscheinlich (Thomas und Schneider 2011). Obwohl PRDM9 als Vorgänger aller KRAB-Zinkfingerproteine gehandelt wird (Birtle und Ponting 2006), interagiert dessen KRAB-Domäne nicht mit KAP1, sondern mit CXXC1 (Parvanov et al. 2017). Jedoch ist es möglich, dass PRDM9 über die gehäuften Bindungsstellen in den piRNA-Cluster und in Verbindung mit nicht allelischer, homologer Rekombination (NAHR) zur Verbreitung der Cluster und somit zur Kontrolle transposabler Elemente beiträgt (Assis und Kondrashov 2009). Dies wäre allerdings nur der Fall, wenn die Cluster den präpachytenen piRNAs zugeordnet werden könnten, da diese TEs unterdrücken (Aravin et al. 2007, Aravin et al. 2008, De Fazio et al. 2011). Die pachytenen piRNAs sind mit der Degradation von mRNAs und lncRNA, *long non-coding RNA*, verknüpft (Yamamoto et al. 2013, Goh et al. 2015, Gou et al. 2015, Watanabe et al. 2015). Es wäre möglich, dass PRDM9 eine aktivierende Rolle spielt, beispielweise indem die H3K4me3-Markierung das Chromatin öffnet. Außerdem sind piRNA-Cluster GC-reich und mit offenem Chromatin bzw. H3K4me3 verknüpft (Gebert et al. 2019), was sie als Bindungsstelle von PRDM9 prädestiniert, siehe Kapitel 3.3.2 und 3.3.6. Bei den verschiedenen Gruppen an piRNA-Cluster fällt auf, dass die Gruppe mit speziesspezifischer Expression bei den piRNA-Clustern mit PRDM9-Peaks prozentual bis zu 1,5-fach häufiger vorkommt, als diese Gruppe innerhalb der Cluster einnimmt. Im Gegensatz dazu ist die Gruppe mit speziesspezifischem Vorkommen bei *M.fascicularis* und mit speziesspezifischem Vorkommen bei *M.mulatta* niedriger. Dass bei beiden die speziesspezifische Expression von piRNA-Clustern erhöht ist, passt

wiederum zu einer möglichen aktivierenden Funktion PRDM9s. Während *PRDM9* im Leptotän und Zygotän exprimiert wird, sind die PIWI-Proteine bei Maus nicht aktiv (Di Giacomo et al. 2013).

Nach den ENCODE-Richtlinien sollen Reads, die mit Sequenzen an multiplen Loci im Genom übereinstimmen, entfernt werden (Landt et al. 2012). Damit wird sichergestellt, dass nur Reads in die Analysen eingehen, die eindeutig zu verorten sind. Somit sind möglicherweise nicht alle repetitive Einheiten, die gebunden wurden, in den Ergebnissen vertreten.

3.3.5 Motive

Unter der Annahme, dass PRDM9-Peaks den Hotspots der Rekombination gleichzusetzen sind, ist der Anteil der Hotspots mit Motiv bei Makaken und Menschen vergleichbar. Myers et al. (2008) zeigen, dass etwa 40% der Hotspots ein 13 bp-langes Motiv aufweisen. Bei beiden Makaken ist das jeweilige Motiv in 38% (*M.mulatta*) bzw. 36% (*M.fascicularis*) der PRDM9-Peaks vorhanden. Im Gegensatz dazu weisen die zwei Mausstämme in fast allen Hotspots ihr jeweils eigenes Motiv auf (Brick et al. 2012).

Beim Menschen sind vier bis sechs Zinkfinger für die DNA-Bindung zuständig. Diese befinden sich, abgesehen vom letzten Zinkfinger, am C-Terminus des Zinkfingerarrays (Patel et al. 2016, Patel et al. 2017). Mit einer Länge von 15 bp sind fünf Zinkfinger, die jeweils drei Basen erkennen, ausreichend, um die Motive von Makaken zu binden. Betrachtet man die Überlappung der vorhergesagten Motive mit jenen, die durch MEME bestimmt wurden, ist es wahrscheinlich, dass die N-terminalen Zinkfinger für die Motivbindung verantwortlich sind. Pratto et al. (2014) vergleichen die Hotspots von Männern mit verschiedenen *PRDM9*-Allelstatus. Sowohl zwischen zwei Individuen, die homozygot für das häufigste Allel A sind, als auch zwischen homozygot-A und heterozygot-A|B stimmen 89% bzw. 88% der Hotspots überein. Das sich unterscheidende Aminosäuretriplett zwischen A- und B-Allel ist zudem nicht unter den 4-6 Zinkfingern, die für die DNA-Bindung zuständig sind. Anders verhält es sich zwischen homozygot-A und heterozygot-A|C. Die Übereinstimmung der Hotspots beläuft sich auf 43% und das C-Allel weist mehrere Unterschiede im C-Terminus auf. Somit könnte davon ausgegangen werden, dass das *M.fascicularis*-Allel und das kürzere *M.mulatta*-Allel für die übereinstimmenden PRDM9-Peaks (ca. 14%, siehe Kapitel 3.2.3.2) verantwortlich sind, da sie bis auf einen Zinkfinger identisch sind und vier der N-terminalen Zinkfinger übereinstimmen. Für eine genauere Beurteilung wären weitere Vergleiche zwischen diversen Allel-Status nötig.

Alu-Elemente sind in den PRDM9-Peaks nur zu etwa 9% vorhanden und in Peaks mit Motiv zu ca. 22%. Außerdem sind die PRDM9-Peaks in Alus zu 95% in den meist nicht aktiven AluJ- und AluS-Gruppen (siehe dazu Comeaux et al. 2009) zu finden, siehe Kapitel 3.2.3.11.1. Auch in Peaks mit Motiv sind die aktiven AluY unterrepräsentiert. Eine mögliche Erklärung wäre, dass somit NAHRs zwischen Alus und daraus entstehende Erkrankungen entgegengewirkt werden soll, siehe Kapitel 3.3.4.3. Auch das mit PRDM9 in Verbindung stehende 13mer-Motiv des Menschen ist in Alu-Elementen zu finden und

die Rekombinationsrate ist in manchen Alu-Elementen erhöht, wobei Alus etwa 5% der Hotspots darstellen (Myers et al. 2008, Myers et al. 2010). Bei Gorillas hingegen wird im Gegensatz zu anderen Menschenaffen vermutet, dass Alus als ehemalige Hotspots dienten (Wacholder und Pollock 2017). Auch wenn nur wenige PRDM9-Peaks in Alu-Elementen zu finden sind, schließt das nicht aus, dass die Rekombinationsrate hier erhöht sein könnte.

3.3.6 H3K4me3

Die Loci der Rekombination wird bei Säugetieren meist, mit wenigen Ausnahmen, durch PRDM9, das über eine SET-Domäne verfügt, die für eine H3K4-Trimethyltransferase kodiert, bestimmt. Bei Hunden, Vögeln und Krokodilen, die kein funktionierendes PRDM9 besitzen, bei manchen *PRDM9*-Knockout-Mäusen und bei *Saccharomyces cerevisiae* sind die Hotspots an H3K4me3-Loci zu finden (Munoz-Fuentes et al. 2011, Axelsson et al. 2012, Auton et al. 2013, Singhal et al. 2015, Brick et al. 2012, Borde et al. 2009).

Die Trimethylierung des vierten Lysins des Histons 3 steht allgemein für Transkriptionsaktivierung (Bernstein et al. 2002, Santos-Rosa et al. 2002, Schneider et al. 2004, Bernstein et al. 2005). Eine Häufung von H3K4me3 in der Promotorregion aktiver Gene geht mit einer Verteilung der H3K4-Trimethylierung einher, die durch eine deutliche Reduktion der Methylierung ca. -200 bp bis +50 bp um die TSS gekennzeichnet ist (vgl. hierzu Cui et al. 2012). Dies ist durch eine nukleosom-freie Region (NFR) *upstream* der TSS zu erklären (Boeger et al. 2003, Lee et al. 2004, Ozsolak et al. 2007, Radman-Livaja und Rando 2010). Das Muster ist auch in dieser Arbeit nachgewiesen worden. Bei der Verteilung der H3K4me3-Peaks von *M.mulatta* und *M.fascicularis* fällt auf, dass das jeweilige, zugrundeliegende Referenzgenom sogar bei nahverwandten Spezies eine bedeutende Rolle spielt. Das Muster ist, wenn *M.mulatta*-Daten auf das *M.fascicularis*-Genom abgebildet werden und vice versa, zwar zu erkennen, jedoch in deutlich abgeschwächter Form. Zudem passen die Ergebnisse der Annotation der H3K4me3-Loci mit der transkriptionsaktivierenden Funktion überein, da sie vor allem in 5'-UTRs, Promotoren und CpG-Inseln häufiger als erwartet vorkommen.

Jene Gene, um deren TSS herum H3K4me3-Markierung je zwei kb in 5'- und 3'-Richtung zu finden sind, gehören zu 57-62% zum Testisproteom und zu 37-38% zu den Haushaltsgenen, wobei letztere als Annäherung herhalten, da die Daten spezifisch für den Mensch sind. Barski et al. 2007 zeigen, dass 59% der H3K4me3-Loci beim Menschen in Promotoren inaktiver Gene zu finden sind und Cui et al. (2012) weisen bei Maus nach, dass 15% der aktiven Gene keine H3K4-Trimethylierung aufweisen. Somit wird deutlich, dass nicht alle testis-spezifischen Gene und Haushaltsgene eine H3K4-Trimethylierung zur Aktivierung benötigen. Ob H3K4me3 beispielsweise nicht eher als Elongationsfaktor für die Transkription dient (Chen et al. 2015) bzw. die Stabilität der Expression gewährleistet (Mosesson et al. 2014) oder den Status des aktiven Zustands weitergibt (Muramoto et al. 2010), wird von Howe et al. (2017) diskutiert. Neben der Funktion während der Transkription und der Rolle in der Meiose ist H3K4me3 auch ein Teil der DNA-Reparatur (Peña et al. 2008, Faucher und Wellinger 2010).

Im Vergleich zum prozentualen Anteil der Testisproteom-Gene in den Haushaltsgenen ist der Anteil der Haushaltsgene mit H3K4-Trimethylierung erhöht, woraus geschlossen werden könnte, dass H3K4me3 vor allem bei den Haushaltsgenen eine wichtige Funktion innehat. Zu den Prozessen, in denen Haushaltsgene involviert sind, gehören unter anderem Translation, metabolische RNA-Prozesse inkl. Splicing, Zellzyklus, Biogenese der Ribosomen und der Transport (Chang et al. 2011). Die Gene Ontology der Gene mit H3K4me3 um ihre TSS zeigt, dass sie für Untereinheiten von RNA-Polymerasen, RNA-Bindungsmotivproteinen, Splicingfaktoren, ribosomale Proteinen und Mitgliedern der Ras-Superfamilie, die zum Teil für Transport zuständig ist, kodieren. Zu den Übereinstimmungen zwischen der Gene Ontology der H3K4me3-Gene und den Prozessen der Haushaltsgene passt, dass der Anteil der Haushaltsgene mit H3K4-Trimethylierung erhöht ist.

Insgesamt befindet sich nur ein Bruchteil (4-7%) aller H3K4me3-Positionen in TSS-Nähe, was im Vergleich zu den beschriebenen 59% in Maus testis gering erscheint (Cui et al. 2012). Zudem gehen Cui et al. (2012) davon aus, dass ein Großteil der restlichen H3K4me3-Loci entweder Loci markieren, an denen sich noch unbekannte Gene befinden bzw. 10 kb up- und downstream der TSS und TTS von Genen vorkommen. Das könnte auch bei den Makaken der Fall sein, womit sich der Anteil der H3K4me3-Loci in Gennähe in zukünftigen Analysen erhöhen könnte.

3.3.7 Abschließender Vergleich zwischen Mm und Mf

Der letzte gemeinsame Vorfahre von *Macaca mulatta* und *Macaca fascicularis* bzw. die Zeit seit ihrer Auftrennung wird auf 1,3 Millionen Jahre von Stevison und Kohn (2009) bzw. 3,53 Millionen Jahre von Perelman et al. (2011) datiert. Somit sind die beiden Arten nah verwandt. Außerdem befindet sich in Indochina zwischen dem 15. und 20. nördlichen Breitengrad eine Zone mit Hybriden, die fertil sind (Gray 1954, Fooden 1964, 1995 und 2000, Hamada et al. 2006, Bunlungsup et al. 2017). Inwieweit PRDM9 sich bereits in seiner Zinkfingersequenz und in seinen Bindungseigenschaften bei den Makakenarten unterscheidet, wird im Folgenden aufgezeigt.

Die Zinkfingersequenz von *M.fascicularis* ist bis auf eine Wiederholungseinheit identisch mit der des kürzeren Allels von *M.mulatta*. Da dieser Unterschied innerhalb der bindenden Zinkfinger vorkommt (siehe Kapitel 3.2.3.11.2 und 3.3.5), könnte dies bereits das Bindungsverhalten beeinflussen. Das größere Allel ist deutlich anders aufgebaut, was sich wohl bereits auf die DNA-Bindungsstellen auswirken könnte, wobei die Motive dennoch stark übereinzustimmen scheinen. Trotzdem ist der Anteil der sich überlappenden PRDM9-Loci von *M.mulatta* und *M.fascicularis* gering. In welchem Maß die zwei Allele von *M.mulatta* die Bindungsloci bestimmen, kann nicht festgestellt werden. Man könnte aufgrund der Ähnlichkeit der Zinkfingerarrays meinen, dass das kurze Allel von *M.mulatta* für die PRDM9-Peaks verantwortlich ist, die mit den Peaks von *M.fascicularis* übereinstimmen. Da diese nur einen geringen Anteil ausmachen, spräche dies für eine Dominanz des längeren Allels. Sowohl beim Menschen als auch bei der Maus sind solche dominanten Allele bekannt (Ségurel et al. 2011, Brick et al. 2012, Pratto et al. 2014, Baker et al. 2015b).

Letztendlich sorgen wohl bereits minimale Abweichungen für Veränderungen im Bindungsverhalten, welche, wie im nächsten Kapitel aufgegriffen wird, zu Speziation führen könnten.

Die beschriebenen Eigenschaften, wie GC-Gehalt und Telomernähe, sind zum Teil auch ähnlich in anderen Spezies beschrieben, wie auch im jeweiligen Kapitel erwähnt. Somit ist es nicht außergewöhnlich, dass diese aufrechterhalten werden, womöglich anzestral sind und sie sich in den Makakenspezies nicht unterscheiden.

Die beiden Makakenarten zeigen in den untersuchten Eigenschaften eine ähnliche Tendenz, allerdings in unterschiedlich starker Ausprägung. Ein Grund dafür, wie auch für weitere Differenzen, kann die geringere Anzahl an Peaks bei *M.fascicularis* sein. Dabei ist das zweite Replikat der begrenzende Faktor. Trotz einer Anzahl an Reads, die vergleichbar ist mit der anderer Proben, ist die Anzahl der Peaks verhältnismäßig gering. Eine andere Möglichkeit ist, dass die Veränderungen im Motiv sich auch auf den GC-Gehalt und die Verteilung der Peaks auswirken und somit Auslöser sind für die wenigen Übereinstimmungen der Peaks.

Zu den Eigenschaften, die gleich ausgeprägt sind, gehört zum einen die Korrelation zwischen GC-Gehalt und Peaks pro Chromosom, zum anderen das Schema der Verteilung der H3K4me3- und DMC1-Peaks um PRDM9-Peaks. Bei der Gene Ontology sind bei allen Kombinationen Mitglieder der Ras-Superfamilie und eine vergleichsweise hohe Übereinstimmung der Gene im Vergleich zu den Überlappungen der Peaks zu verzeichnen. Außerdem finden sich die Peaks bei beiden häufiger als zu erwarten in piRNA-Cluster und die Motive haben eine Verknüpfung zu Alu-Elementen.

Der prozentuale Anteil der Peaks mit Motiv und aller Peaks, die mit Alu-Elementen zusammenfallen, unterscheidet sich kaum bei den beiden Makaken. Bei den Annotationen wie auch beim GC-Gehalt und Peaks pro Chromosom ist jeweils die Tendenz ähnlich, jedoch bei *M.mulatta* stärker ausgeprägt. Eine Ausnahme ist bei GC-Gehalt und Peaks pro Chromosom das X-Chromosom, bei dem die Tendenz deutlicher bei *M.fascicularis* abgebildet wird. Der Zusammenhang zwischen Peaks entlang eines Chromosoms und dem dazugehörigen GC-Gehalt ist bei *M.mulatta* eindeutiger, höchstwahrscheinlich aufgrund der höheren Anzahl an Peaks. Der Anteil der Gene im Testisproteom bzw. in Haushaltsgenen ist bei beiden Arten ähnlich. Der Unterschied zwischen PRDM9-Peaks und Genen im Testisproteom, die mit Haushaltsgenen übereinstimmen, ist bei *M.fascicularis*, womöglich aufgrund der geringen Anzahl an Peaks, nicht signifikant.

Die größten Unterschiede sind im Folgenden zu verzeichnen. Zum einen ist die Verteilung in den verschiedenen Gruppen der piRNA-Cluster zu erwähnen, bei der die Cluster mit PRDM9-Peaks von *M.mulatta* seltener in speziesspezifisch vorkommenden Clustern und jene von *M.fascicularis* seltener in speziessübergreifend vorkommenden Clustern festzuhalten sind. Zum anderen ist zwar das Schema, wie sich die DMC1- und H3K4me3-Peaks um PRDM9-Peaks verteilen, identisch, jedoch ist der prozentuale Anteil im Vergleich zu *M.mulatta* bei *M.fascicularis* doppelt so hoch. Zuletzt ist die geringe Übereinstimmung der Peaks nochmals zu erwähnen, die trotz aller Gemeinsamkeiten darauf hinweist, dass die Hotspotloci deutlich unterschiedlich evolviert sind.

3.3.8 PRDM9 und Speziation

Rekombinationshotspots verändern mit der Zeit ihre Lage im Genom, da das Allel, das als Hotspot erkannt wird, mit dem mutierten Allel, das nicht erkannt wird, als Vorlage repariert wird (Boulton 1997). Dieses Phänomen wird als *hotspot conversion paradox* bezeichnet. PRDM9 mit seinen schnell evolvierenden Zinkfingern soll dem Verschwinden der Hotspots entgegenwirken, indem es alternative Loci bindet, die als neue Hotspots dienen sollen (Lesecque et al. 2014, Baker et al. 2015a, Davies et al. 2016, Smagulova et al. 2016). Inkompatibilitäten zwischen neuentstehenden Spezies werden zum einen dadurch erklärt, dass die *PRDM9*-Allele vornehmlich die Chromosomen der jeweils anderen Art binden, da dort die Bindungssequenz häufiger vorhanden ist und somit PRDM9 seltener an beiden Homologen bindet, was als Asymmetrie der Bindungsstellen beschrieben wird (Davies et al. 2016). Eine weitere Theorie steht in Zusammenhang mit einem Locus auf dem X-Chromosom, dem *Hstx2*, mit dem PRDM9 möglicherweise interagiert (Bhattacharyya et al. 2014, Lustyk et al. 2019).

Bei Mäusen sind Hybridsterilitäten bereits nach 500000 Jahren nach der Auftrennung von *M.m.domesticus* und *M.m.musculus* nachzuweisen (Forejt und Iványi 1974, Boursot et al. 1993). Jedoch stammt mindestens ein parentales Tier aus Laborstämmen, die das Resultat von Inzucht sind (Beck et al. 2000). Daher ist es schwierig bzw. nicht möglich, davon ausgehend Rückschlüsse auf die Konstellation der hier bearbeiteten Makaken zu ziehen.

Dass nur wenige PRDM9-Peaks bei *M.mulatta* und *M.fascicularis* überlappen, könnte jedoch darauf hinweisen, dass sich das Bindungsverhalten bereits verändert hat bzw. dass der Abbau der Bindungsstellen bereits im Gange ist. Obwohl die Motive, die sich in jeweils 40% der Peaks wiederfinden, ähnlich sind, führen sie nicht zu mehr übereinstimmenden Loci. Da *M.mulatta* zwei Allele vorzuweisen hat, die sich in dem Aufbau des Zinkfingerarrays unterscheiden, kann nicht eindeutig bestimmt werden, welches Allel für welche Loci verantwortlich ist. Das kürzere Allel ist nahezu identisch zum *M.fascicularis*-Allel. Der zusätzliche Zinkfinger bei *M.fascicularis* gehört wahrscheinlich zu den DNA-bindenden Zinkfingern. Trotzdem sind diese beiden Allele aufgrund der Ähnlichkeit auf einen gemeinsamen Ursprung zurückzuführen und damit könnte das längere Allel *M.mulattas* das neuere darstellen, das für die Bereitstellung neuer Hotspots zuständig ist. Passend dazu wird die Übereinstimmung des vorhergesagten Motivs des längeren Allels mit dem von MEME erstellten Motivs besser unterstützt im Vergleich zum kürzeren. Anhand der Zinkfingerarrays allein sind keine Aussagen über die Speziation zu treffen, da diese zum Beispiel auch innerhalb von Arten hohe Heterozygotieraten aufweisen (Berg et al. 2010, Kono et al. 2014, Schwartz et al. 2014, Heerschop et al. 2016). Studien in der Hybridzone zwischen den Verbreitungsgebieten der beiden Makakenarten, die sich zwischen dem 15. und 20. nördlichen Breitengrad in Südostasien befindet, zeigen, dass die hybriden Nachkommen fertil sind (Gray 1954, Fooden 1964, Hamada et al. 2006, Satkoski Trask et al. 2013, Bunlungsup et al. 2017). Somit sind jedenfalls in den parapatrischen Gebieten die Barrieren zwischen den Arten noch durchlässig, wobei eine Introgression verstärkt von *M.mulatta* zu *M.fascicularis* zu beobachten ist (siehe dazu Tosi et al. 2002, Hamada et al. 2006, Bonhomme et al. 2009, Bunlungsup et al. 2017). Das Verbreitungsgebiet *M.mulattas* erstreckt sich zwischen dem 15. und 40. nördlichen

Breitengrad und jenes von *M.fascicularis* zwischen dem 10. südlichen und dem 20. nördlichen Breitengrad (Fooden 1997). Individuen der beiden Spezies, die weiter entfernt voneinander zu verorten sind, könnten bereits deutlichere Barrieren aufweisen. Da nicht bekannt ist, woher die beiden hier getesteten Makaken bzw. deren Vorfahren stammen, fehlt die Ausgangslage, um dahingehend eine Aussage zu treffen.

Transposable Elemente liegen in Hybriden von Heringsmaränen und Wallabies überexprimiert vor und stehen somit in Verdacht bei der Speziation involviert zu sein (O'Neill et al 1998, Dion-Côté et al. 2014). Ob dahingehend eine Verbindung mit den von PRDM9 gebundenen piRNA-Cluster besteht, bleibt offen.

4. Fazit

Mit den verschiedenen Projekten konnte die Evolution *PRDM9s* innerhalb der Primaten auf verschiedenen Ebenen beleuchtet werden:

Die Zinkfinger der strepsirrhinen Spezies bilden eine distinkte Gruppe wie auch die der Tarsier und der Anthropeidea. Innerhalb der letztgenannten ist die Auflösung gering, weshalb die konservierte Zinkfingerstruktur ohne die Basen, die für die bindenden Aminosäuren kodieren, eher für ältere Aufspaltungen interessant ist.

Das Duplikationsereignis, das zu *PRDM7* führte, wird in dieser Arbeit auf die Zeit vor der Auftrennung der Catarrhini in Hominoidea und Cercopithecoidea datiert. Die Evolution *PRDM7s* ist in den beiden Taxa unterschiedlich verlaufen. Zudem sind die Duplikationen innerhalb *PRDM7s* nur bei Menschen und Schimpansen zu finden, wobei Menschen nur zwei Wiederholungen aufzeigen und Schimpansen acht.

Mittels der Chromatinimmunopräzipitation mit anschließender Hochdurchsatzsequenzierung konnten bekannte Eigenschaften von Hotspots, wie Telomernähe und hoher GC-Gehalt, bei den zwei Makakenarten bestätigt werden. Zudem sind Zusammenhänge mit der Ras-Superfamilie und mit piRNA-Cluster festzustellen. Ersteres könnte Anhaltspunkte für den Zusammenhang von *PRDM9* und akuter lymphatischer Leukämie (ALL) liefern.

Obwohl viele der beleuchteten Merkmale bei *Macaca mulatta* und *Macaca fascicularis* übereinstimmen, fällt nur ein Bruchteil der *PRDM9*-Peaks zusammen. Das könnte auf die sich verändernden Bindungsstellen zurückzuführen sein. Andererseits stimmt ein Allel von *M.mulatta* bis auf einen Zinkfinger mit dem von *M.fascicularis* überein. Ob das möglicherweise auf ein Hybridisierungsevent zurückzuführen ist, ist leider nicht nachvollziehbar. Für mehr Einsichten in die Zusammenhänge von *PRDM9* und Speziation bei Makaken wäre es von Vorteil, die Zusammensetzung und Verteilung der Zinkfingerarrays beider Arten in ihren Verbreitungsräumen und in der Hybridzone zu kennen. Gerade die Hybridzone wäre ein interessantes Untersuchungsgebiet. Auch eine Miteinbeziehung weiterer Makakenarten, entweder als Außengruppen oder für die Evolution *PRDM9s* innerhalb der Makaken, könnte informativ sein.

5. Anhang

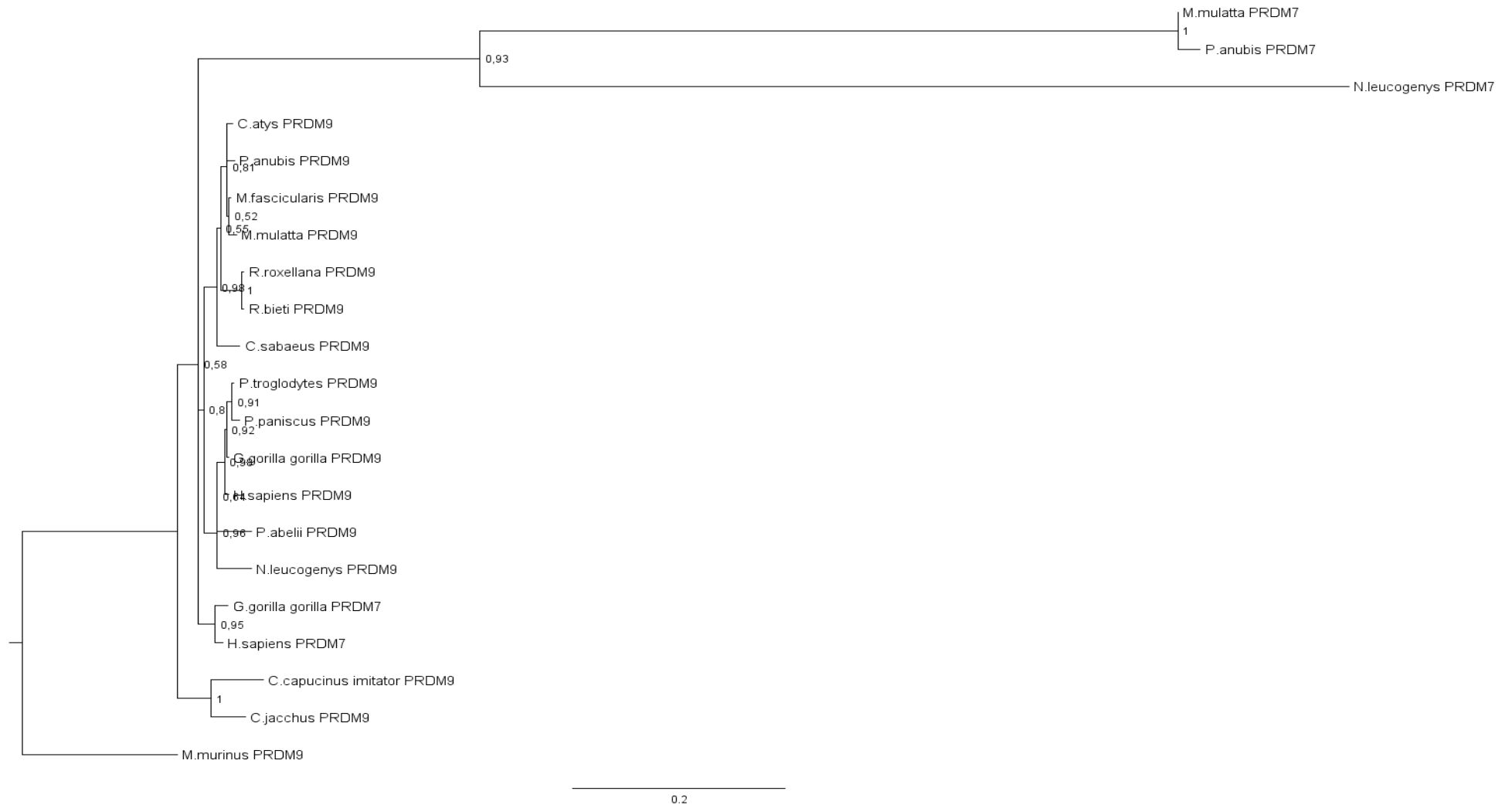


Abbildung I: Phylogenie basierend auf den 2000 Basen vor dem Startkodon (5').

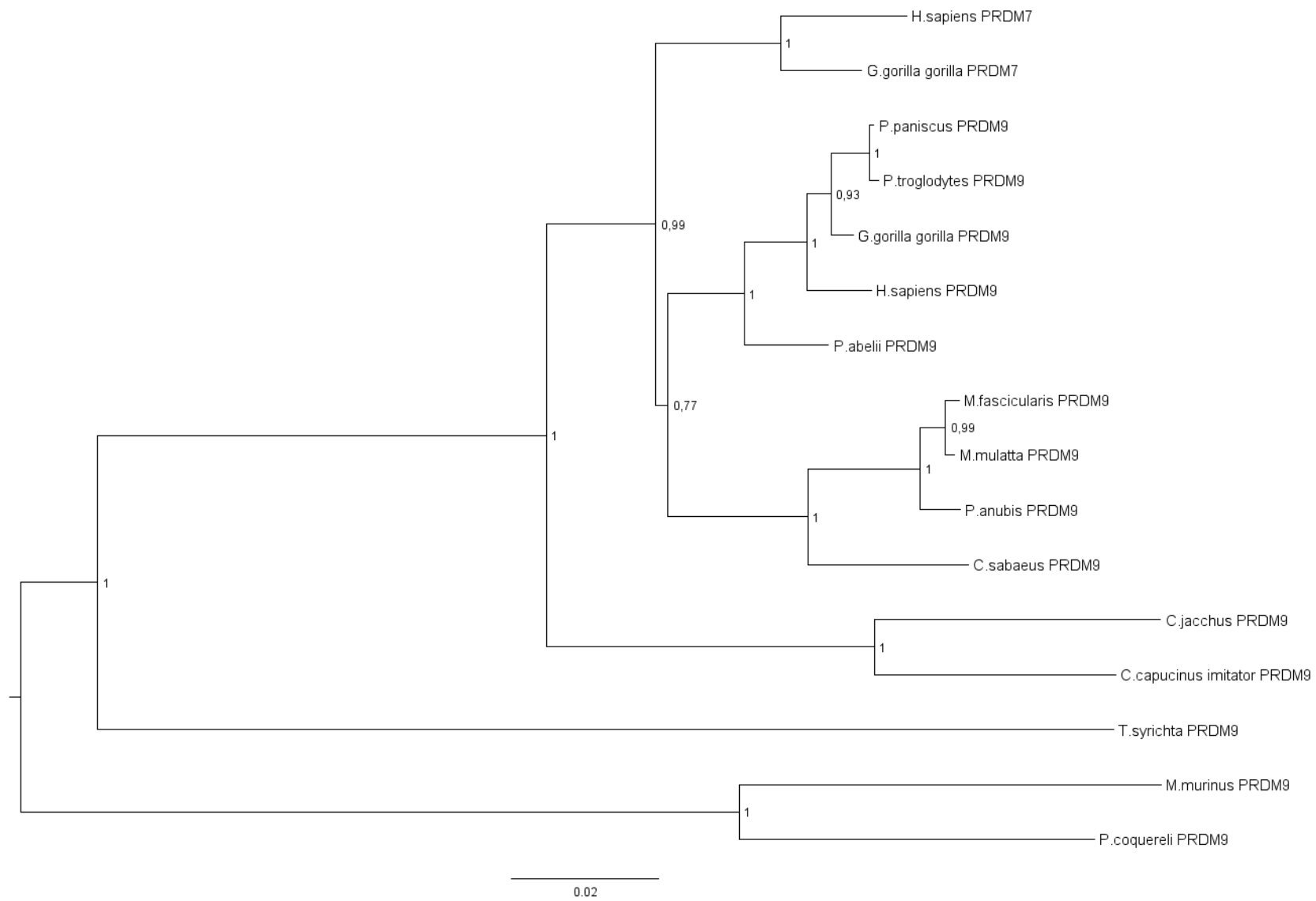


Abbildung II: Phylogenie basierend auf den Introns 1, 2, 7 und 9.

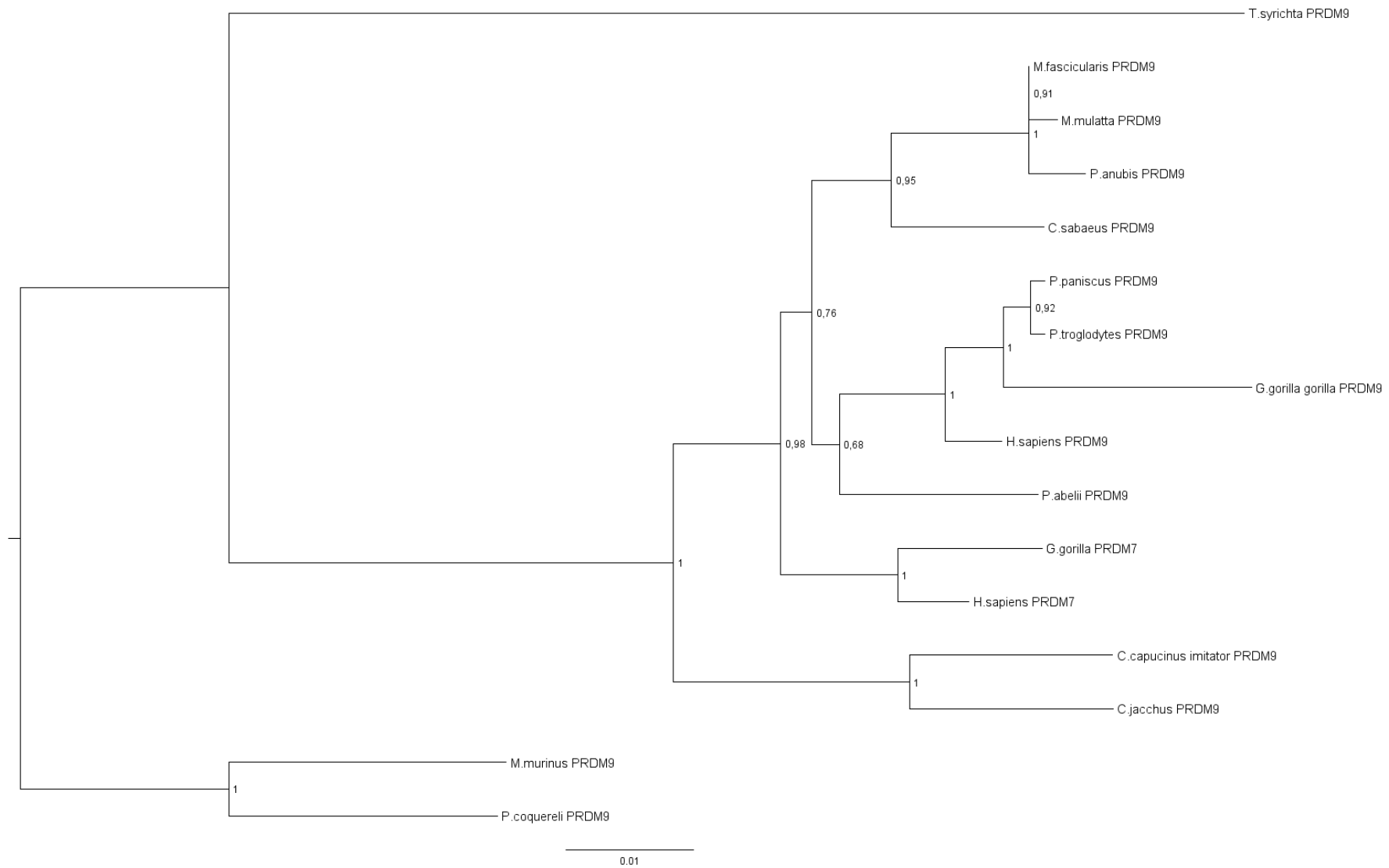


Abbildung III: Phylogenie basierend auf den ersten 1002 Basen der cDNA ab dem Startkodon.

N-ChIP Protokoll

(abgeändert nach <https://www.abcam.com/protocols/native-chromatin-immunoprecipitation-protocol>)

Gewebevorbereitung

- 500µl kaltes PBS (OHNE PMSF) auf 70 µm Strainer vorlegen und Gewebe(leicht angetaut) darauf verteilen
- nochmals 500µl kaltes PBS (OHNE PMSF) dazugeben, mit gummiertem Spritzenstößel verstreichen
- Durchfluss in Eppi geben, Schälchen mit 500µl kaltem PBS (OHNE PMSF) auswaschen und zum Eppi pipettieren
- Davon 5 µl mit 5 µl Coomassie-Blau anfärben und Zellen überprüfen
- bei 4°C 300g 5 min und 600g 2 min zentrifugieren
- 2x Waschen mit 1ml PBS inkl. 50 µl PMSF und 5 µl 20xPI, jeweils bei bei 4°C 300g 5 min und 600g 2 min zentrifugieren
- Pellet in 500 µl TBS (inkl. Tween, PMSF, 20xPI Mix) resuspendieren
- 10 min bei 4 °C rotieren lassen (turntable auf 20)
- Zentrifugieren 6,2rpm 20 sec 4°C
- Pellets in 500 µl TBS (inkl. Tween, PMSF, 20xPI Mix) resuspendieren (nicht zu viel pipettieren)
- 10 min bei 4 °C rotieren lassen (turntable auf 20)
- Zentrifugieren: gelöste Pellets bei 4°C 15 min 14.000g
- pro tube Pellet in 100 µl 25% sucrose resuspendieren und zusammenführen
- weitere 200 µl hinzufügen
- 200 µl 50% sucrose unterlegen
- zentrifugieren 14.000 g 4°C 25 min
- Pellet mit 400 µl 25% sucrose waschen, zentrifugieren 14.000 g 4°C 25 min

MNase-Verdau

- Pellets in je 25 µl MNase Puffer lösen und poolen + 10 µl MNase Puffer zum Ausspülen des zweiten Eppis und zu erstem geben
- 3,6 µl MNase pro Proben
- 10 min 37°C inkubieren
- 5,86 µl 50mM EDTA zu Proben für Abbruch des Verdaus
- 5 min auf Eis, auf Agarosegel testen

PBS: 1 ml PBS + 50µl PMSF

TBS: 1 ml TBS + 50µl PMSF + 5µl PI-Mix + 50µl 20% Tween

Digestion buffer (10ml): ohne PMSF aliquotieren und einfrieren

0,32 M sucrose ≈	1,09 g sucrose
50 mM Tris (pH 8) ≈	500 µl 1M Tris (pH 8)
4 mM MgCl ₂ ≈	40 µl 1 M MgCl ₂
1 mM CaCl ₂ ≈	10 µl 1 M CaCl ₂
0,1 mM PMSF ≈	10 µl 100 mM PMSF
Auffüllen mit H ₂ O	

Inkubationspuffer: 50 mM NaCl

20 mM Tris (pH 8)

5 mM Na₂EDTA

0,1 mM PMSF

Pre-clearing

- Proben mit Inkubationspuffer auf 800 µl auffüllen
- Protein-G beads 50% v/v (20 µl pro Probe) 2x mit PBS waschen (etwa 750 µl)
- 1x mit Inkubationspuffer(inkl. PI: 100 µl in 5 ml Ink.puffer) waschen
- Pellet in 66 µl Inkubationspuffer lösen und je 20 µl zu Probe geben
- 15 min bei 4°C rotieren lassen
- Mittels magnetic rack Überstand abpipettieren und in neues Eppi geben
- Nochmals in magnetic rack und Überstand wiederum in neues tube

Immunoprecipitation

- 800 µl Überstand auf 4x 2ml Eppis verteilen (input, Anti-IgG, Anti-PRDM9, Anti-H3K27)
- Je 2 µg Antikörper zu letzten dreien:
- 2µl IgG, 4 µl Anti-PRDM9, 2 µl Anti-H3K27
- Über Nacht bei 4 °C auf rotierendem turntable(10, sehr schräg), auch input
- Slurry mit 3X Inkubationspuffer waschen und in Inkubationspuffer aufnehmen 50% v/v
- Pro Probe 15 µg geben und 2 h bei 4°C inkubieren/rotieren lassen
- Kurz an zentrifugieren, 1min in magnetic rack stellen
- 400 µl Waschpuffer A dazugeben, Pellet lösen, 10 min auf Eis, in magnetic rack 1 min
- 400 µl Waschpuffer B dazugeben, Pellet lösen, 10 min auf Eis, in magnetic rack 1 min
- 400 µl Waschpuffer C dazugeben, Pellet lösen, 10 min auf Eis, in magnetic rack 1 min
- 400 µl Waschpuffer C dazugeben, Pellet lösen, 10 min auf Eis, in magnetic rack 1 min
- Pellet in 100 µl Inkubationspuffer (inkl. 1% SDS) lösen, 15 min auf turntable (20) bei Raumtemperatur
- 1 min in magnetic rack, Überstand abpipettieren -> Bound 1 bzw. Elu 1
- 100 µl Inkubationspuffer inkl. 1 % SDS, 10 min auf Eis
- 1 min in magnetic rack, Überstand abpipettieren -> Bound 2 bzw. Elu 2
- Beide Elus poolen

Waschpuffer A:	50 mM Tris (pH 8)	[≈500 µl 1M Tris (pH 8),
	10 mM EDTA	200 µl 0,5M EDTA
	50 mM NaCl -> 100 mM	100 µl 5M NaCl auf 10 ml H2O]

Waschpuffer B:	50 mM Tris (pH 8)
	10 mM EDTA
	100 mM NaCl -> 200 mM

Waschpuffer C:	50 mM Tris (pH 8)
	10 mM EDTA
	150 mM NaCl -> 300 mM

DNA Isolation

- Zu Input 20 µl 10% SDS und 185 µl Inkubationspuffer -> insgesamt 400 µl
- Zu anderen Proben 205 µl Inkubationspuffer (ohne SDS) -> 0,5% SDS in Gesamtlösung (400 µl)
- 132 µl Phenol-Chloroform zu allen samples, 10 sec vortexen
- Zentrifugieren 13.000 rpm 10 min 4°C
- Überstand zu 400 µl Chloroform überführen, 10 sec vortexen
- Zentrifugieren 13.000 rpm 10 min 4°C
- Überstand (etwa 350 µl) auf 2 Eppis aufteilen: 200 µl (1) und 150 µl (2)
- Zu (1): 800 µl EtOH, 20 µl 4M LiCl, 0.5 µl Glycogen
- Zu (2): 600 µl EtOH, 15 µl 4M LiCl, 0.5 µl Glycogen
- 10 sec vortexen

- Über Nacht bei -20°C präzipitieren
- Zentrifugieren: 13.000 g 15 min 4°C, Überstand abkippen
- (2): Pellets mit 2x 150 µl kaltem 70% EtOH spülen und zu (1) pipettieren
- (1): Pellets mit 700 µl 70% EtOH waschen
- bei 4°C 13.000 g 15 min zentrifugieren
- Überstand vorsichtig abpipettieren, Pellet löst sich sehr leicht
- Nochmals zentrifugieren: 13.000 g 2 min 4°C
- Überstand abpipettieren und Pellet trocknen bei Raumtemperatur
- Pellet in 50 µl 1xTE Puffer lösen, vortexen und leicht zentrifugieren
- Auf Agarosegel testen

Protein Isolation

- Zur organischen Phase aus erstem Phenol/Chloroform-Schritt 5 µl BSA- Lösung (1 mg/ml), 0,01 vol (4 µl) 10 M H₂SO₄ und 12 vol Aceton geben
- Precipitation bei -20°C. Pellet mit H₂SO₄-Aceton Lösung (1:6 -> 100 mM H₂SO₄:Aceton) waschen, danach 3x mit Aceton.

Western Blot Protokoll

Zellyse

- Gewebe mit flüssigem Stickstoff und Mörser zerkleinern -> kleines Stück in Eppi überführen und wiegen
- Auf etwa 80 mg 500 µl eiskalten Lysepuffer (2xRIPA) geben, Gewebe mit Pipettenspitze zerkleinern
- Bei 4°C ca. 1-max. 2h schütteln lassen

10xRIPA (250ml): 25 ml NP-40 detergent
 125 ml of 1M Tris-HCl pH 7,4
 75 ml 5M NaCl
 12,5 g Sodium Deoxycholate
 5 ml 0,5M EDTA
 20 ml milliQ water

2xRIPA: 10xRIPA verdünnen und mit PI (Roche) versetzen –
 200 µl 10xRIPA, 750 µl H₂O, 50 µl PI

- Zentrifugieren bei 4°C, 12000 rpm, 20 min; Überstand in neues Eppi überführen
- Pellet kann resuspendiert und nochmals geschüttelt werden, falls Proteinkonzentration niedrig; Pellet einfrieren

Messung der Proteinkonzentration

- 1 ml Bradford assay und 1 µl Lysat in Küvette pipettieren und mischen, sollte sich dabei verfärben = Indikator für genügend hohe Proteinkonzentration
- Je zwei Küvetten pro Probe vorbereiten (Duplikat) zur Überprüfung der Messergebnisse
- Als Standard eine Küvette mit 1 ml Bradford assay bereitstellen
- Nach Messung den Mittelwert der beiden Ergebnisse berechnen und in Formel einsetzen:
$$y = (-0,0162)/0,069$$
- Bei mehreren Proben gleiche Konzentration herstellen durch Verdünnen
- 300 µg für WB bereitstellen, weiteres Lysat aliquotieren (200 µg)

Puffer für Kathode (innere Kammer):

6,4 g Tris + 5,1 g Glycin pro Liter Wasser, pH 8,9

Puffer für Anode (äußere Kammer):

12,2 g Tris pro Liter Wasser, pH 7,8

Beladen der Gele

- Proben mit 4xLaemmli (mit und ohne DTT) versetzen: 3 Teile Probe mit 1 Teil 4xLaemmli
- Ohne DTT: 27 µl Probe + 9 µl 4xLaemmli, davon einmal 20 µl und einmal 10 µl auftragen
- Mit DTT: 27 µl Probe + 9 µl 4xLaemmli (1000 µl Laemmli inkl. 4 µl DTT)
- Proteinstandard: native Mark Protein Std 57030
- Bei 30 mA laufen lassen bis Marker am unteren Ende angekommen ist (Strom niedrig halten für niedrige Temperatur -> verhindert Denaturierung)

4xLaemmli-Puffer: 4,4 ml 0,5M Tris pH 6,8
 4,4 ml Glycerol
 0,5 ml 1% Bromophenol Blue
 2,2 ml milliQ water
 pH 6,8
 bei -20°C lagern

Transfer

- PVDF-Membran und Whatman-Papier auf Größe des Gels zuschneiden
- Membran 1-2 min in Methanol einlegen, danach 5 min in eiskaltem Transferpuffer inkubieren (auf Schüttler)
- Gel 3-5 min in eiskaltem Transferpuffer inkubieren (auf Schüttler)
- Membran und Gel inkl. Transferpuffer für ein paar Minuten in den Kühlschrank
- Schwämme waschen mit Wasser: warm-kalt-destilliert
- Wet-transfer-Anrichtung: Schwamm/Papier/Gel/Membran/Papier/Schwamm [in eiskaltem Transferpuffer]
- Eventuelle Luftblasen entfernen und zusammenklemmen
- Blot-Aufbau mit Elektrotransferpuffer (eiskalt) in Kammer setzen: über Nacht, bei 4°C, 30 V, 0.05A
- Kammer auf Rührer stellen und Rührfisch in den Puffer geben für einheitliche Temperatur

10xTowbin's Electrotransfer Puffer:

30.3 g (0,25M) Tris Base

144 g (1,92 M) Glycine

Auffüllen auf 1000 ml ddH₂O

- Membran markieren, um Blotting-Seite später zu erkennen

Ponceau-S Färbung

- Membran mit ddH₂O waschen
- Membran in Ponceau (1:10 Verdünnung) 5 min inkubieren
- Membran mit ddH₂O waschen bis Banden gut sichtbar sind
- Bild mit BioRad erstellen -> Membran zwischen Plastikfolie legen
- Membran mit TBS waschen
- Membran mit TBST waschen (TBS mit 0.2% Tween); Tween in kleiner Menge TBS lösen, dann erst zu TBS geben
- Ponceau ist reversibel und kann mehrere Male verwendet werden

10xPonceau S: Ponceaus S 2% (w/v)
TCA 30% (w/v)

TBS 10x (concentrated TBS): 24.23 g Trizma HCl
80.06 g NaCl
Mix in 800 ml ultra pure water.
pH to 7.6 with pure HCl.
Top up to 1 L.

- TBST: 100 ml of TBS 10x + 900 ml ultra pure water + 1ml Tween20

Blockieren der Membran

- Für 2-3h bei Raumtemperatur, auf Schüttler, in 5% skim milk TBST
- Membran mit TBST waschen

Inkubation mit primärem Antikörper

- AK-Lösung (1:1000) : 10 ml TBST inkl. 10 µl Antikörper und 3% BSA
- Über Nacht bei 4°C auf Schüttler
- Waschen: 6x10 min mit TBST bei Raumtemperatur

Inkubation mit sekundärem Antikörper (Mausspezifisch)

- AK-Lösung (1:10000): 10 ml TBST inkl. 1 µl Antikörper und 1% BSA
- 1-2h bei Raumtemperatur auf Schüttler
- Waschen: 6x 10 min mit TBST bei Raumtemperatur
- Weiterer Waschschrift: 1x10 min mit TBS bei Raumtemp. (ohne Detergenzien)

Detektion

- ECL-Reagenzien 1:1 mischen (etwa 8ml für 2 Membrane)
- Membran darin 1 min inkubieren, abtropfen lassen und zwischen Folie legen für Visualisierung mit BioRad -> addiert mehrere Bilder

ECL-Reagenzien:

- 200 µl 250 mM Luminol (0,44g/10ml DMSO)*
88 µl 90 mM p-Coumarsiure (0,15g/10ml DMSO)*
2 ml 1M Tris-HCl pH 8,5
17,7 ml H₂O (bidest o. HPLC)
- 12 µl 30% H₂O₂
2 ml M Tris-HCl pH 8,5
18 ml H₂O (bidest)

* Luminol ist lichtempfindlich -> Alufolie um Gefäß wickeln!

Tabelle I: Übersicht über die Anzahl der Reads pro Replikat (M.mulatta: Mm1 und Mm2, M.fascicularis: Mf1 und Mf2), über die Anzahl der im Genom lokalisierten Reads mit und ohne Duplikate für die Referenzgenome von *M.mulatta* und die dazugehörigen Qualitätswerte (siehe dazu Landt et al. 2012 und <https://www.encodeproject.org/data-standards/terms/>). Quality Tag-Werte beziehen sich auf den NSC und sind unter 0 negativ, von 0 neutral und über 0 positiv zu bewerten.

rheMac8									
		alle reads	mapped reads	ohne duplikate	NSC	RSC	Quality Tag	NRF	PBC _{min}
Mm1	i	21851998	16776268	16319634	1,012992	0,604973	0	0,746826	0,9455619
	P9	17865381	13850535	13520432	1,01337	0,549035	0	0,756795	0,9523335
	H3	21302496	16353890	15914178	1,013428	0,582824	0	0,747057	0,9462254
Mf1	i	20330700	15545569	15155387	1,013122	0,51931	0	0,745443	0,9498015
	P9	15587633	11964584	11711614	1,014868	0,496619	-1	0,75134	0,9577135
	H3	18362226	14045012	13716293	1,013979	0,49369	-1	0,746984	0,9531906
Mm2	i	105148033	72513562	67332286	1,025203	1,029846	1	0,640357	0,857095
	P9	82028731	56454730	45482942	1,07972	1,266625	1	0,554476	0,6113067
	H3	72198659	50973278	47033314	1,040346	1,187458	1	0,651443	0,8454106
	D1	68410957	38647317	34257436	1,079849	1,189971	1	0,500759	0,7728235
Mf2	i	65356120	49735370	46881957	1,024851	1,227258	1	0,717331	0,8852562
	P9	82028731	58086247	53478057	1,029028	1,041786	1	0,651943	0,8413328
	H3	58442991	38418425	35490571	1,034917	0,94626	0	0,607268	0,8475807
	D1	48263437	31405148	29176632	1,045858	0,94456	0	0,604529	0,8580796
rheMac3									
Mm1	i	21851998	16747234	15524341	1,015362	0,7399036	0	0,7104312	0,8539588
	P9	17865381	13643402	12829812	1,015754	0,6754398	0	0,7181382	0,8807350
	H3	21302496	16145592	15072318	1,016033	0,7088068	0	0,7075377	0,8670505
Mm2	i	105148033	72631171	64337052	1,029531	1,024875	1	0,6118712	0,7716099
	P9	82028731	56454730	43272028	1,087361	1,260093	1	0,5275228	0,5329815
	H3	72198659	50973278	44999735	1,044665	1,174984	1	0,6232766	0,7656206
	D1	68410957	41329450	32509459	1,086693	1,239103	1	0,4752084	0,5731861

Tabelle II: Übersicht über die Anzahl der Reads pro Replikat (M.mulatta: Mm1 und Mm2, M.fascicularis: Mf1 und Mf2), über die Anzahl der im Genom lokalisierten Reads mit und ohne Duplikate für die Referenzgenome von *M.fascicularis* und die dazugehörigen Qualitätswerte (siehe dazu Landt et al. 2012 und <https://www.encodeproject.org/data-standards/terms/>). Quality Tag-Werte beziehen sich auf den NSC und sind unter 0 negativ, von 0 neutral und über 0 positiv zu bewerten.

macFas5									
		alle reads	mapped reads	ohne duplikate	NSC	RSC	Quality Tag	NRF	PBC _{min}
Mm1	i	21851998	16607997	16265479	1,012398	0,7074829	0	0,7443475	0,9587526
	P9	17865381	13725464	13474009	1,011882	0,6232015	0	0,7541966	0,9633593
	H3	21302496	16201029	15858641	1,012127	0,6552906	0	0,7444499	0,9577326
Mf1	i	20330700	15788206	15438832	1,013194	0,6150467	0	0,7593852	0,9557424
	P9	15587633	12166530	11939419	1,014891	0,6129757	0	0,7659546	0,9626663
	H3	18362226	14284094	13982035	1,014467	0,6067066	0	0,7614564	0,9577069
Mm2	i	105148033	71727113	66951484	1,025791	1,021323	1	0,6367355	0,8668389
	P9	82028731	55750708	45103326	1,081831	1,267744	1	0,5498479	0,618036
	H3	72198659	50319297	46754843	1,040914	1,190805	1	0,647586	0,8583266
	D1	68410957	37765449	34003483	1,081145	1,186726	1	0,4970473	0,8007721
Mf2	i	65356120	50575489	47675434	1,02525	1,313238	1	0,7294716	0,8853178
	P9	82028731	59113070	54381451	1,029443	1,08956	1	0,6629561	0,8399129
	H3	58442991	38881488	36141175	1,034845	0,9961234	0	0,6184005	0,8590428
	D1	48263437	31834409	29713180	1,045719	0,988114	0	0,6156458	0,8667336
AEHL									
Mf1	i	20330700	14612103	13895929	1,019466	0,7336845	0	0,6834949	0,9019752
	P9	15587633	11184699	10707953	1,020978	0,6927532	0	0,6869518	0,9147503
	H3	18362226	13117348	12532041	1,02043	0,6938909	0	0,6824903	0,9107583
Mf2	i	65356120	46654191	42963086	1,030696	1,209413	1	0,657369	0,8417675
	P9	82028731	54644600	48585588	1,036112	1,082897	1	0,5922996	0,7782393
	H3	58442991	36828231	32091599	1,042721	1,059995	1	0,5491095	0,7427717
	D1	48263437	29901590	26267998	1,055581	1,056165	1	0,5442629	0,7569633

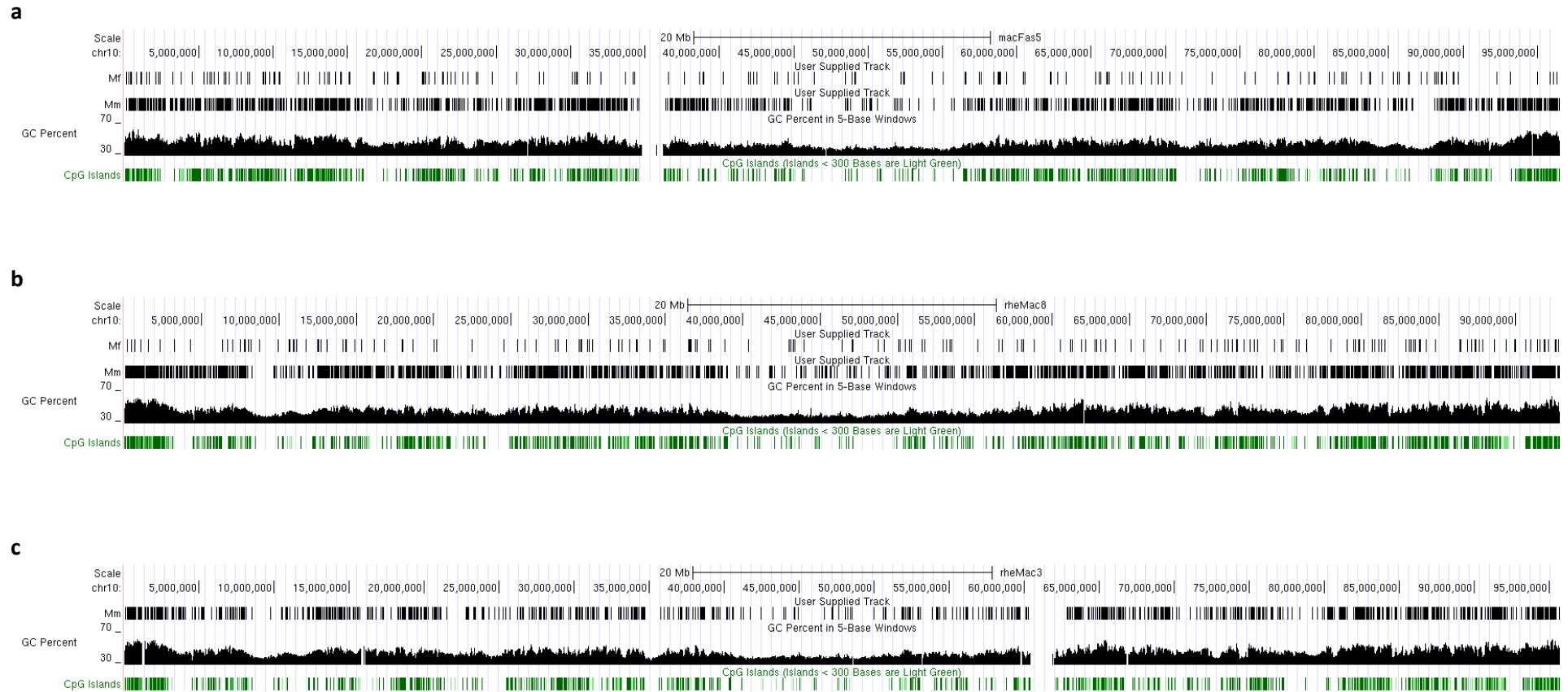


Abbildung IV: Verteilung der PRDM9-Peaks von *M.fascicularis* auf Chr10 von macFas5 (a) und rheMac8 (b) und von *M.mulatta* auf Chr10 von macFas5 (a), rheMac8 (b) und rheMac3 (c). Erstellt mit dem UCSC genome browser inkl. Darstellung des GC-Gehalts und Positionen der CpG-Inseln.

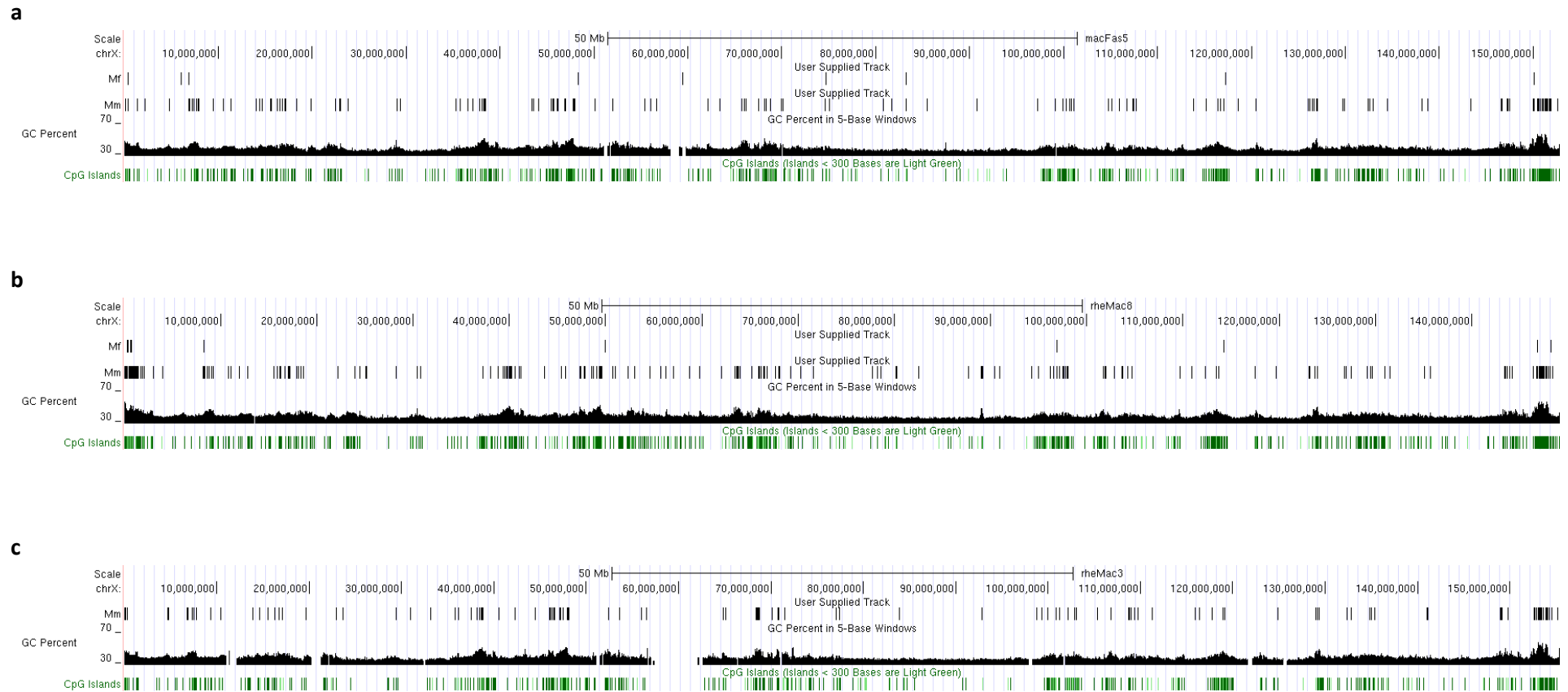


Abbildung V: Verteilung der PRDM9-Peaks von *M.fascicularis* auf ChrX von macFas5 (a) und rheMac8 (b) und von *M.mulatta* auf ChrX von macFas5 (a), rheMac8 (b) und rheMac3 (c). Erstellt mit dem UCSC genome browser inkl. Darstellung des GC-Gehalts und Positionen der CpG-Inseln.

Gene Functional Classification (DAVID) der in Genen befindlichen PRDM9-Peaks:

M.mulatta/rheMac8

Gene Group 1 Enrichment Score: 3.546231138219674

OFFICIAL_GENE_SYMBOL	Gene Name
COQ8B	coenzyme Q8B(COQ8B)
ETNK2	ethanolamine kinase 2(ETNK2)
MAPK4	mitogen-activated protein kinase 4(MAPK4)
TNNI3K	TNNI3 interacting kinase(TNNI3K)
ADCK1	aarF domain containing kinase 1(ADCK1)
FN3KRP	fructosamine 3 kinase related protein(FN3KRP)

Gene Group 2 Enrichment Score: 2.961122905463588

OFFICIAL_GENE_SYMBOL	Gene Name
SOS2	SOS Ras/Rho guanine nucleotide exchange factor 2(SOS2)
AKT3	AKT serine/threonine kinase 3(AKT3)
AKT1	AKT serine/threonine kinase 1(AKT1)
PIK3R2	phosphoinositide-3-kinase regulatory subunit 2(PIK3R2)
PRKACB	protein kinase cAMP-activated catalytic subunit beta(PRKACB)
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha(PIK3CA)
RAF1	Raf-1 proto-oncogene, serine/threonine kinase(RAF1)
PRKCA	protein kinase C alpha(PRKCA)
NFKBIA	NFKB inhibitor alpha(NFKBIA)
PIK3CG	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma(PIK3CG)
NFKB1	nuclear factor kappa B subunit 1(NFKB1)
IKBKB	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta(IKBKB)
MAP2K1	mitogen-activated protein kinase kinase 1(MAP2K1)
AKT2	AKT serine/threonine kinase 2(AKT2)
PIK3R3	phosphoinositide-3-kinase regulatory subunit 3(PIK3R3)
PLCB2	phospholipase C beta 2(PLCB2)
PRKCB	protein kinase C beta(PRKCB)
MAPK10	mitogen-activated protein kinase 10(MAPK10)

Gene Group 3 Enrichment Score: 2.7161951616037756

OFFICIAL_GENE_SYMBOL	Gene Name
RASL10A	RAS like family 10 member A(RASL10A)
EEF2	eukaryotic translation elongation factor 2(EEF2)
EEF1A1	eukaryotic translation elongation factor 1 alpha 1(EEF1A1)
ARL15	ADP ribosylation factor like GTPase 15(ARL15)
RAB3B	RAB3B, member RAS oncogene family(RAB3B)
RAB2A	RAB2A, member RAS oncogene family(RAB2A)
DRG2	developmentally regulated GTP binding protein 2(DRG2)
TUFM	Tu translation elongation factor, mitochondrial(TUFM)
ARF1	ADP ribosylation factor 1(ARF1)
RABL3	RAB, member of RAS oncogene family like 3(RABL3)
RIT2	Ras like without CAAX 2(RIT2)
RAP1A	RAP1A, member of RAS oncogene family(RAP1A)
MTG1	mitochondrial ribosome associated GTPase 1(MTG1)
RAP1B	RAP1B, member of RAS oncogene family(RAP1B)
SRPRB	SRP receptor beta subunit(SRPRB)
DNAJC27	DnaJ heat shock protein family (Hsp40) member C27(DNAJC27)
RAB27A	RAB27A, member RAS oncogene family(RAB27A)
RASL10B	RAS like family 10 member B(RASL10B)
RALA	RAS like proto-oncogene A(RALA)
AGAP2	ArfGAP with GTPase domain, ankyrin repeat and PH domain 2(AGAP2)
AGAP1	ArfGAP with GTPase domain, ankyrin repeat and PH domain 1(AGAP1)
RAB35	RAB35, member RAS oncogene family(RAB35)
RAB3C	RAB3C, member RAS oncogene family(RAB3C)
MRAS	muscle RAS oncogene homolog(MRAS)
ARF4	ADP ribosylation factor 4(ARF4)
RAB18	RAB18, member RAS oncogene family(RAB18)
SEPT6	septin 6(SEPT6)
RAB6A	RAB6A, member RAS oncogene family(RAB6A)
RAB1A	RAB1A, member RAS oncogene family(RAB1A)
DIRAS1	DIRAS family GTPase 1(DIRAS1)
RAB5C	RAB5C, member RAS oncogene family(RAB5C)
DIRAS2	DIRAS family GTPase 2(DIRAS2)
RERG	RAS like estrogen regulated growth inhibitor(RERG)
RAB8B	RAB8B, member RAS oncogene family(RAB8B)
RAB5A	RAB5A, member RAS oncogene family(RAB5A)
SEPT7	septin 7(SEPT7)
RHOBTB3	Rho related BTB domain containing 3(RHOBTB3)

RAB4A RAB4A, member RAS oncogene family(RAB4A)
RRAS2 related RAS viral (r-ras) oncogene homolog 2(RRAS2)
RAB28 RAB28, member RAS oncogene family(RAB28)
ARL5B ADP ribosylation factor like GTPase 5B(ARL5B)
RASD1 ras related dexamethasone induced 1(RASD1)
RAB33B RAB33B, member RAS oncogene family(RAB33B)
RASL12 RAS like family 12(RASL12)
SAR1B secretion associated Ras related GTPase 1B(SAR1B)
RAB32 RAB32, member RAS oncogene family(RAB32)

Gene Group 4 Enrichment Score: 2.309890617254016

OFFICIAL_GENE_SYMBOL	Gene Name
PRKACB	protein kinase cAMP-activated catalytic subunit beta(PRKACB)
PRKCB	protein kinase C beta(PRKCB)
GNAQ	G protein subunit alpha q(GNAQ)
ADCY2	adenylate cyclase 2(ADCY2)
GNAO1	G protein subunit alpha o1(GNAO1)
PLCB2	phospholipase C beta 2(PLCB2)
PLCB1	phospholipase C beta 1(PLCB1)
ADCY3	adenylate cyclase 3(ADCY3)
GNAI2	G protein subunit alpha i2(GNAI2)

Gene Group 5 Enrichment Score: 2.261333113164759

OFFICIAL_GENE_SYMBOL	Gene Name
GNG7	G protein subunit gamma 7(GNG7)
GNAO1	G protein subunit alpha o1(GNAO1)
GNG12	G protein subunit gamma 12(GNG12)
GNG11	G protein subunit gamma 11(GNG11)
GNG2	G protein subunit gamma 2(GNG2)
GNB1	G protein subunit beta 1(GNB1)
GNG3	G protein subunit gamma 3(GNG3)
GNAI2	G protein subunit alpha i2(GNAI2)

Gene Group 6 Enrichment Score: 2.195548802545862

OFFICIAL_GENE_SYMBOL	Gene Name
PRKD3	protein kinase D3(PRKD3)
TGFBR2	transforming growth factor beta receptor 2(TGFBR2)
PRKCQ	protein kinase C theta(PRKCQ)
CSNK1A1	casein kinase 1 alpha 1(CSNK1A1)
BMPR1A	bone morphogenetic protein receptor type 1A(BMPR1A)
STK38L	serine/threonine kinase 38 like(STK38L)
STK24	serine/threonine kinase 24(STK24)
CAMK1	calcium/calmodulin dependent protein kinase I(CAMK1)
CSNK2A1	casein kinase 2 alpha 1(CSNK2A1)
MAPK4	mitogen-activated protein kinase 4(MAPK4)
TESK2	testis-specific kinase 2(TESK2)
CDK6	cyclin dependent kinase 6(CDK6)
DAPK1	death associated protein kinase 1(DAPK1)
NEK7	NIMA related kinase 7(NEK7)
MAP3K14	mitogen-activated protein kinase kinase kinase 14(MAP3K14)
ACVR1C	activin A receptor type 1C(ACVR1C)
SRPK2	SRSF protein kinase 2(SRPK2)
ACVR1B	activin A receptor type 1B(ACVR1B)
TNNI3K	TNNI3 interacting kinase(TNNI3K)
CDK5	cyclin dependent kinase 5(CDK5)
BMPR2	bone morphogenetic protein receptor type 2(BMPR2)
CDK20	cyclin dependent kinase 20(CDK20)
RPS6KA4	ribosomal protein S6 kinase A4(RPS6KA4)
CSNK1E	casein kinase 1, epsilon(CSNK1E)
CAMK2G	calcium/calmodulin dependent protein kinase II gamma(CAMK2G)
PSKH1	protein serine kinase H1(PSKH1)
MAPKAPK3	mitogen-activated protein kinase-activated protein kinase 3(MAPKAPK3)
CDK17	cyclin dependent kinase 17(CDK17)
DYRK1B	dual specificity tyrosine phosphorylation regulated kinase 1B(DYRK1B)
PAK5	p21 (RAC1) activated kinase 5(PAK5)
EIF2AK2	eukaryotic translation initiation factor 2 alpha kinase 2(EIF2AK2)
CDC7	cell division cycle 7(CDC7)
CDK8	cyclin dependent kinase 8(CDK8)
TNIK	TRAF2 and NCK interacting kinase(TNIK)
PAK6	p21 (RAC1) activated kinase 6(PAK6)
OXS1	oxidative stress responsive 1(OXS1)
DAPK3	death associated protein kinase 3(DAPK3)
MAP2K3	mitogen-activated protein kinase kinase 3(MAP2K3)

SBK2 SH3 domain binding kinase family member 2(SBK2)
CDK7 cyclin dependent kinase 7(CDK7)
BMPR1B bone morphogenetic protein receptor type 1B(BMPR1B)
CDK12 cyclin dependent kinase 12(CDK12)
STK16 serine/threonine kinase 16(STK16)

Gene Group 7 Enrichment Score: 2.159181341832219

OFFICIAL_GENE_SYMBOL Gene Name
EGLN3 egl-9 family hypoxia inducible factor 3(EGLN3)
EGLN2 egl-9 family hypoxia inducible factor 2(EGLN2)
OGFOD3 2-oxoglutarate and iron dependent oxygenase domain containing 3(OGFOD3)
P4HA2 prolyl 4-hydroxylase subunit alpha 2(P4HA2)

Gene Group 8 Enrichment Score: 1.9392450041395672

OFFICIAL_GENE_SYMBOL Gene Name
FGF12 fibroblast growth factor 12(FGF12)
FGF22 fibroblast growth factor 22(FGF22)
FGF3 fibroblast growth factor 3(FGF3)
FGF1 fibroblast growth factor 1(FGF1)

Gene Group 9 Enrichment Score: 1.8476976748404568

OFFICIAL_GENE_SYMBOL Gene Name
UCKL1 uridine-cytidine kinase 1 like 1(UCKL1)
UCK2 uridine-cytidine kinase 2(UCK2)
NME1 non-metastatic cells 1, protein (NM23A) expressed in(NME1)
DCK deoxycytidine kinase(DCK)
NMRK1 nicotinamide riboside kinase 1(NMRK1)

Gene Group 10 Enrichment Score: 1.6904531901251496

OFFICIAL_GENE_SYMBOL Gene Name
ARFGAP3 ADP ribosylation factor GTPase activating protein 3(ARFGAP3)
SMAP2 small ArfGAP2(SMAP2)
ACAP2 ArfGAP with coiled-coil, ankyrin repeat and PH domains 2(ACAP2)
ARFGAP1 ADP ribosylation factor GTPase activating protein 1(ARFGAP1)

Gene Group 11 Enrichment Score: 1.6716037258540446

OFFICIAL_GENE_SYMBOL Gene Name
EIF4EBP2 eukaryotic translation initiation factor 4E binding protein 2(EIF4EBP2)
EIF1B eukaryotic translation initiation factor 1B(EIF1B)
EIF2S2 eukaryotic translation initiation factor 2 subunit beta(EIF2S2)
EIF4E3 eukaryotic translation initiation factor 4E family member 3(EIF4E3)

Gene Group 12 Enrichment Score: 1.6507160015023878

OFFICIAL_GENE_SYMBOL Gene Name
PPP2R5E protein phosphatase 2 regulatory subunit B'epsilon(PPP2R5E)
PPP2R2C protein phosphatase 2 regulatory subunit Bgamma(PPP2R2C)
PPP2R5A protein phosphatase 2 regulatory subunit B'alpha(PPP2R5A)
PPP2R1A protein phosphatase 2 scaffold subunit Aalpha(PPP2R1A)

Gene Group 13 Enrichment Score: 1.5886615993974884

OFFICIAL_GENE_SYMBOL Gene Name
PSMB2 proteasome subunit beta 2(PSMB2)
PSMB3 proteasome subunit beta 3(PSMB3)
PSMB7 proteasome subunit beta 7(PSMB7)
PSMA4 proteasome subunit alpha 4(PSMA4)
PSMA2 proteasome subunit alpha 2(PSMA2)
PSMA3 proteasome subunit alpha 3(PSMA3)
PSMA5 proteasome subunit alpha 5(PSMA5)
PSMB9 proteasome subunit beta 9(PSMB9)

Gene Group 14 Enrichment Score: 1.55389281447378

OFFICIAL_GENE_SYMBOL Gene Name
GMPPA GDP-mannose pyrophosphorylase A(GMPPA)
GMDS GDP-mannose 4,6-dehydratase(GMDS)
GMPPB GDP-mannose pyrophosphorylase B(GMPPB)
TSTA3 tissue specific transplantation antigen P35B(TSTA3)

Gene Group 15 Enrichment Score: 1.5474044930053577

OFFICIAL_GENE_SYMBOL Gene Name
ZMYM6 zinc finger MYM-type containing 6(ZMYM6)
CXXC5 CXXC finger protein 5(CXXC5)
ZMYM2 zinc finger MYM-type containing 2(ZMYM2)
ZMYM4 zinc finger MYM-type containing 4(ZMYM4)

KIAA1958 KIAA1958 ortholog(KIAA1958)
GTF2I general transcription factor Ii(GTF2I)
GTF2IRD1 GTF2I repeat domain containing 1(GTF2IRD1)

Gene Group 16 Enrichment Score: 1.5448569881137681

OFFICIAL_GENE_SYMBOL	Gene Name
FGR	FGR proto-oncogene, Src family tyrosine kinase(FGR)
JAK2	Janus kinase 2(JAK2)
YES1	YES proto-oncogene 1, Src family tyrosine kinase(YES1)
SYK	spleen associated tyrosine kinase(SYK)
CSK	c-src tyrosine kinase(CSK)
JAK1	Janus kinase 1(JAK1)

Gene Group 17 Enrichment Score: 1.4063159782793917

OFFICIAL_GENE_SYMBOL	Gene Name
UBE3A	ubiquitin protein ligase E3A(UBE3A)
HECTD1	HECT domain E3 ubiquitin protein ligase 1(HECTD1)
HERC3	HECT and RLD domain containing E3 ubiquitin protein ligase 3(HERC3)
SMURF1	SMAD specific E3 ubiquitin protein ligase 1(SMURF1)
HECW1	HECT, C2 and WW domain containing E3 ubiquitin protein ligase 1(HECW1)
WWP2	WW domain containing E3 ubiquitin protein ligase 2(WWP2)

Gene Group 18 Enrichment Score: 1.3992082559502574

OFFICIAL_GENE_SYMBOL	Gene Name
GSTT1	glutathione S-transferase theta 1(GSTT1)
MGST3	microsomal glutathione S-transferase 3(MGST3)
GSTK1	glutathione S-transferase kappa 1(GSTK1)
GSTA3	glutathione S-transferase alpha 3(GSTA3)
GSTP1	glutathione S-transferase pi 1(GSTP1)
GSTO2	glutathione S-transferase omega 2(GSTO2)

Gene Group 19 Enrichment Score: 1.3696582467836456

OFFICIAL_GENE_SYMBOL	Gene Name
UBE2V2	ubiquitin conjugating enzyme E2 V2(UBE2V2)
UBE2N	ubiquitin conjugating enzyme E2 N(UBE2N)
UBE2E2	ubiquitin-conjugating enzyme E2E 2(UBE2E2)
UBE2F	ubiquitin conjugating enzyme E2 F (putative)(UBE2F)
UBE2H	ubiquitin conjugating enzyme E2 H(UBE2H)
UBE2S	ubiquitin conjugating enzyme E2 S(UBE2S)
UBE2K	ubiquitin conjugating enzyme E2 K(UBE2K)

Gene Group 20 Enrichment Score: 1.3569719266263176

OFFICIAL_GENE_SYMBOL	Gene Name
NLRP6	NLR family pyrin domain containing 6(NLRP6)
NLRP8	NLR family pyrin domain containing 8(NLRP8)
NLRP11	NLR family pyrin domain containing 11(NLRP11)
NLRP2	NLR family pyrin domain containing 2(NLRP2)
NLRP1	NLR family pyrin domain containing 1(NLRP1)
NLRP9	NLR family pyrin domain containing 9(NLRP9)
NLRP5	NLR family pyrin domain containing 5(NLRP5)

Gene Group 21 Enrichment Score: 1.3355655654160672

OFFICIAL_GENE_SYMBOL	Gene Name
PRKCE	protein kinase C epsilon(PRKCE)
PRKD3	protein kinase D3(PRKD3)
PRKCQ	protein kinase C theta(PRKCQ)
PRKCA	protein kinase C alpha(PRKCA)
PRKCB	protein kinase C beta(PRKCB)

Gene Group 22 Enrichment Score: 1.3332416235750113

OFFICIAL_GENE_SYMBOL	Gene Name
POLR3F	RNA polymerase III subunit F(POLR3F)
POLR3B	RNA polymerase III subunit B(POLR3B)
POLR2D	RNA polymerase II subunit D(POLR2D)
POLR3H	RNA polymerase III subunit H(POLR3H)
POLR2B	RNA polymerase II subunit B(POLR2B)
POLR1B	RNA polymerase I subunit B(POLR1B)

Gene Group 23 Enrichment Score: 1.2312650028430385

OFFICIAL_GENE_SYMBOL	Gene Name
LYNX1	Ly6/neurotoxin 1(LYNX1)
DPEP1	dipeptidase 1 (renal)(DPEP1)
LYNX1	Ly6/neurotoxin 1(LYNX1)

GPC1 glypican 1(GPC1)

Gene Group 24 Enrichment Score: 1.1553678855207619

OFFICIAL_GENE_SYMBOL	Gene Name
DCTD	dCMP deaminase(DCTD)
APOBEC3D	apolipoprotein B mRNA editing enzyme catalytic subunit 3D(APOBEC3D)
APOBEC3A	apolipoprotein B mRNA editing enzyme catalytic subunit 3A(APOBEC3A)
APOBEC3F	apolipoprotein B mRNA editing enzyme catalytic subunit 3F(APOBEC3F)

Gene Group 25 Enrichment Score: 1.0383978293687017

OFFICIAL_GENE_SYMBOL	Gene Name
PDE9A	phosphodiesterase 9A(PDE9A)
PDE4B	phosphodiesterase 4B(PDE4B)
PDE1B	phosphodiesterase 1B(PDE1B)
PDE1A	phosphodiesterase 1A(PDE1A)
PDE2A	phosphodiesterase 2A(PDE2A)
PDE6A	phosphodiesterase 6A(PDE6A)

M.mulatta/rheMac3

Gene Group 1 Enrichment Score: 1.9057340001944152

OFFICIAL_GENE_SYMBOL	Gene Name
RASL10A	RAS like family 10 member A(RASL10A)
RAB1A	RAB1A, member RAS oncogene family(RAB1A)
RAB3C	RAB3C, member RAS oncogene family(RAB3C)
RAB5A	RAB5A, member RAS oncogene family(RAB5A)
RASD1	ras related dexamethasone induced 1(RASD1)
MTG1	mitochondrial ribosome associated GTPase 1(MTG1)
RERG	RAS like estrogen regulated growth inhibitor(RERG)
RASL10B	RAS like family 10 member B(RASL10B)
ARF1	ADP ribosylation factor 1(ARF1)
RAB35	RAB35, member RAS oncogene family(RAB35)
ARL5B	ADP ribosylation factor like GTPase 5B(ARL5B)
AGAP1	ArfGAP with GTPase domain, ankyrin repeat and PH domain 1(AGAP1)
RAB31	RAB31, member RAS oncogene family(RAB31)
RABL3	RAB, member of RAS oncogene family like 3(RABL3)
ARL15	ADP ribosylation factor like GTPase 15(ARL15)
SEPT6	septin 6(SEPT6)
RAB11B	RAB11B, member RAS oncogene family(RAB11B)
NKIRAS2	NFKB inhibitor interacting Ras like 2(NKIRAS2)
RAP1A	RAP1A, member of RAS oncogene family(RAP1A)
RAB28	RAB28, member RAS oncogene family(RAB28)
EEF1A1	eukaryotic translation elongation factor 1 alpha 1(EEF1A1)
RRAS2	related RAS viral (r-ras) oncogene homolog 2(RRAS2)
RAB23	RAB23, member RAS oncogene family(RAB23)
RAB4A	RAB4A, member RAS oncogene family(RAB4A)
NKIRAS1	NFKB inhibitor interacting Ras like 1(NKIRAS1)
RASL12	RAS like family 12(RASL12)
RIT2	Ras like without CAAX 2(RIT2)

Gene Group 2 Enrichment Score: 1.4318416356102754

OFFICIAL_GENE_SYMBOL	Gene Name
HSDL2	hydroxysteroid dehydrogenase like 2(HSDL2)
CBR3	carbonyl reductase 3(CBR3)
DHRS7	dehydrogenase/reductase 7(DHRS7)
CBR4	carbonyl reductase 4(CBR4)
RDH16	retinol dehydrogenase 16 (all-trans)(RDH16)
HSD17B6	hydroxysteroid 17-beta dehydrogenase 6(HSD17B6)

Gene Group 3 Enrichment Score: 1.4105056243183727

OFFICIAL_GENE_SYMBOL	Gene Name
SYVN1	synoviolin 1(SYVN1)
RNF10	ring finger protein 10(RNF10)
TRIM62	tripartite motif containing 62(TRIM62)
TRIM34	tripartite motif containing 34(TRIM34)
RNF26	ring finger protein 26(RNF26)
MARCH4	membrane associated ring-CH-type finger 4(MARCH4)
RNF113A	ring finger protein 113A(RNF113A)
ZNRF1	zinc and ring finger 1(ZNRF1)

BRAP BRCA1 associated protein(BRAP)
RNF40 ring finger protein 40(RNF40)
BFAR bifunctional apoptosis regulator(BFAR)
RNF167 ring finger protein 167(RNF167)
RNF152 ring finger protein 152(RNF152)
TRIM37 tripartite motif containing 37(TRIM37)
TRIM35 tripartite motif containing 35(TRIM35)
RCHY1 ring finger and CHY zinc finger domain containing 1(RCHY1)
RNF115 ring finger protein 115(RNF115)
RNF121 ring finger protein 121(RNF121)
TRIM8 tripartite motif containing 8(TRIM8)
RNF144B ring finger protein 144B(RNF144B)
RNF126 ring finger protein 126(RNF126)
TRIM27 tripartite motif containing 27(TRIM27)
TRIM52 tripartite motif containing 52(TRIM52)
PCGF3 polycomb group ring finger 3(PCGF3)

Gene Group 4 Enrichment Score: 1.2873475604050748

OFFICIAL_GENE_SYMBOL	Gene Name
NEK9	NIMA related kinase 9(NEK9)
PRKD3	protein kinase D3(PRKD3)
CDK7	cyclin dependent kinase 7(CDK7)
MAP3K11	mitogen-activated protein kinase kinase kinase 11(MAP3K11)
MAP3K3	mitogen-activated protein kinase kinase kinase 3(MAP3K3)
MAPKAPK3	mitogen-activated protein kinase-activated protein kinase 3(MAPKAPK3)
DAPK1	death associated protein kinase 1(DAPK1)
STK38L	serine/threonine kinase 38 like(STK38L)
CSNK1E	casein kinase 1, epsilon(CSNK1E)
PSKH1	protein serine kinase H1(PSKH1)
CDK20	cyclin dependent kinase 20(CDK20)
PAK6	p21 (RAC1) activated kinase 6(PAK6)
SBK2	SH3 domain binding kinase family member 2(SBK2)
CDK6	cyclin dependent kinase 6(CDK6)
CSNK2A1	casein kinase 2 alpha 1(CSNK2A1)
PHKG1	phosphorylase kinase catalytic subunit gamma 1(PHKG1)
STK16	serine/threonine kinase 16(STK16)
CDK8	cyclin dependent kinase 8(CDK8)
TESK2	testis-specific kinase 2(TESK2)
MAPKAPK5	mitogen-activated protein kinase-activated protein kinase 5(MAPKAPK5)
SRPK2	SRSF protein kinase 2(SRPK2)
MAPK4	mitogen-activated protein kinase 4(MAPK4)
OXSRI	oxidative stress responsive 1(OXSRI)
DAPK3	death associated protein kinase 3(DAPK3)
MAP2K3	mitogen-activated protein kinase kinase 3(MAP2K3)
CDK12	cyclin dependent kinase 12(CDK12)
DYRK1B	dual specificity tyrosine phosphorylation regulated kinase 1B(DYRK1B)

Gene Group 5 Enrichment Score: 1.1507033151546646

OFFICIAL_GENE_SYMBOL	Gene Name
GNAO1	G protein subunit alpha o1(GNAO1)
GNG12	G protein subunit gamma 12(GNG12)
GNG2	G protein subunit gamma 2(GNG2)
GNB1	G protein subunit beta 1(GNB1)
GNG3	G protein subunit gamma 3(GNG3)

Gene Group 6 Enrichment Score: 1.1191690163448005

OFFICIAL_GENE_SYMBOL	Gene Name
WSB1	WD repeat and SOCS box containing 1(WSB1)
HIRA	HIR histone cell cycle regulation defective homolog A (S. cerevisiae)(HIRA)
WDR37	WD repeat domain 37(WDR37)
WDR77	WD repeat domain 77(WDR77)
WDR61	WD repeat domain 61(WDR61)
LRWD1	leucine rich repeats and WD repeat domain containing 1(LRWD1)
TLE1	transducin like enhancer of split 1(TLE1)
DTL	denticleless E3 ubiquitin protein ligase homolog(DTL)
GNB1L	G protein subunit beta 1 like(GNB1L)
WDR83	WD repeat domain 83(WDR83)
BUB3	BUB3, mitotic checkpoint protein(BUB3)
DDB2	damage specific DNA binding protein 2(DDB2)
WDR5B	WD repeat domain 5B(WDR5B)
DCAF4	DDB1 and CUL4 associated factor 4(DCAF4)
POC1B	POC1 centriolar protein B(POC1B)
PWP2	PWP2 periodic tryptophan protein homolog (yeast)(PWP2)

SNRNP40 small nuclear ribonucleoprotein U5 subunit 40(SNRNP40)
 THOC3 THO complex 3(THOC3)
 TAF5L TATA-box binding protein associated factor 5 like(TAF5L)
 WDSUB1 WD repeat, sterile alpha motif and U-box domain containing 1(WDSUB1)
 PAK1IP1 PAK1 interacting protein 1(PAK1IP1)
 WDR55 WD repeat domain 55(WDR55)

Gene Group 7 Enrichment Score: 1.0826571781369951

OFFICIAL_GENE_SYMBOL Gene Name
 TIA1 TIA1 cytotoxic granule associated RNA binding protein(TIA1)
 SRSF4 serine and arginine rich splicing factor 4(SRSF4)
 HNRNPDL heterogeneous nuclear ribonucleoprotein D like(HNRNPDL)
 SNRNP70 small nuclear ribonucleoprotein U1 subunit 70(SNRNP70)
 SNRPA small nuclear ribonucleoprotein polypeptide A(SNRPA)
 PSPC1 paraspeckle component 1(PSPC1)
 SNRNP48 small nuclear ribonucleoprotein U11/U12 subunit 48(SNRNP48)
 EWSR1 EWS RNA binding protein 1(EWSR1)
 SF3B6 splicing factor 3b subunit 6(SF3B6)
 POLDIP3 DNA polymerase delta interacting protein 3(POLDIP3)
 BOLL boule homolog, RNA binding protein(BOLL)
 LARP4B La ribonucleoprotein domain family member 4B(LARP4B)
 HNRNPF heterogeneous nuclear ribonucleoprotein F(HNRNPF)
 SRSF7 serine and arginine rich splicing factor 7(SRSF7)
 PABPC1L poly(A) binding protein cytoplasmic 1 like(PABPC1L)
 MATR3 matrin 3(MATR3)
 SNRNP40 small nuclear ribonucleoprotein U5 subunit 40(SNRNP40)
 RBM42 RNA binding motif protein 42(RBM42)
 SRSF12 serine and arginine rich splicing factor 12(SRSF12)
 RBM5 RNA binding motif protein 5(RBM5)
 RBM34 RNA binding motif protein 34(RBM34)
 RBM17 RNA binding motif protein 17(RBM17)
 HNRNPH3 heterogeneous nuclear ribonucleoprotein H3(HNRNPH3)
 SRSF2 serine and arginine rich splicing factor 2(SRSF2)
 G3BP1 G3BP stress granule assembly factor 1(G3BP1)

Gene Group 8 Enrichment Score: 1.0735742795283763

OFFICIAL_GENE_SYMBOL Gene Name
 PIK3R3 phosphoinositide-3-kinase regulatory subunit 3(PIK3R3)
 RAF1 Raf-1 proto-oncogene, serine/threonine kinase(RAF1)
 PRKCA protein kinase C alpha(PRKCA)
 MAP2K1 mitogen-activated protein kinase kinase 1(MAP2K1)
 AKT2 AKT serine/threonine kinase 2(AKT2)
 PIK3R2 phosphoinositide-3-kinase regulatory subunit 2(PIK3R2)

M.mulatta/macFas5

Gene Group 1 Enrichment Score: 3.9106954039016424

REFSEQ_MRNA Gene Name
 NM_001319527 microcephalin 1(MCPH1)
 NM_001319560 protein phosphatase 1 regulatory subunit 42(PPP1R42)
 NM_001319406 dynein regulatory complex subunit 3(DRC3)
 NM_001284767 centrosomal protein 70(CEP70)

Gene Group 2 Enrichment Score: 3.599236171576447

REFSEQ_MRNA Gene Name
 NM_001284918 eukaryotic translation initiation factor 4A3(EIF4A3)
 NM_001285017 eukaryotic translation initiation factor 4A1(EIF4A1)
 NM_001284712 eukaryotic translation initiation factor 4A2(EIF4A2)
 NM_001319400 DEAD-box helicase 25(DDX25)
 NM_001283451 DEAD-box helicase 1(DDX1)
 NM_001287668 DEAD (Asp-Glu-Ala-Asp) box polypeptide 19B(DDX19B)

Gene Group 3 Enrichment Score: 2.501933969047725

REFSEQ_MRNA Gene Name
 NM_001283220 annexin A11(ANXA11)
 NM_001283231 annexin A5(ANXA5)
 NM_001284044 annexin A6(ANXA6)
 NM_001285067 annexin A7(ANXA7)

Gene Group 4 Enrichment Score: 2.040562807608741
REFSEQ_MRNA Gene Name
NM_001287637 cytochrome P450, family 2, subfamily B, polypeptide 6(CYP2B6)
NM_001319483 cytochrome P450 1A2(LOC102130662)
NM_001319491 steroid 17-alpha-hydroxylase/17,20 lyase(LOC102133235)
NM_001319492 phylloquinone omega-hydroxylase CYP4F2(LOC102134301)
NM_001284679 thromboxane A synthase 1(TBXAS1)
NM_001284028 cytochrome P450 2A13-like(CYP2A26)
NM_001287732 cytochrome P450 19A1-like(LOC102117886)

Gene Group 5 Enrichment Score: 1.8268936510440261
REFSEQ_MRNA Gene Name
NM_001283671 muscle RAS oncogene homolog(MRAS)
NM_001285006 RAB8A, member RAS oncogene family(RAB8A)
NM_001284847 NFkB inhibitor interacting Ras like 2(NKIRAS2)
NM_001283979 RAB3A, member RAS oncogene family(RAB3A)
NM_001319567 RAS like proto-oncogene B(RALB)
NM_001283973 RAB5A, member RAS oncogene family(RAB5A)
NM_001283736 RAB18, member RAS oncogene family(RAB18)
NM_001284098 RAB28, member RAS oncogene family(RAB28)
NM_001283926 ADP ribosylation factor 5(ARF5)
NM_001285405 RAP1B, member of RAS oncogene family(RAP1B)
NM_001283838 secretion associated Ras related GTPase 1A(SAR1A)
NM_001283692 ADP ribosylation factor 1(ARF1)

Gene Group 6 Enrichment Score: 1.770184618612254
REFSEQ_MRNA Gene Name
NM_001285355 SRSF protein kinase 2(SRPK2)
NM_001285112 MOK protein kinase(MOK)
NM_001284139 3-phosphoinositide dependent protein kinase 1(PDPK1)
NM_001283976 casein kinase 1 delta(CSNK1D)
NM_001283843 mitogen-activated protein kinase kinase 5(MAP2K5)
NM_001285266 serum/glucocorticoid regulated kinase family member 3(SGK3)
NM_001283662 NIMA related kinase 11(NEK11)
NM_001283945 TP53 regulating kinase(TP53RK)
NM_001284722 serine/threonine kinase 32C(STK32C)
NM_001319386 cyclin dependent kinase 19(CDK19)
NM_001285247 dual specificity tyrosine phosphorylation regulated kinase 3(DYRK3)
NM_001283991 protein-O-mannose kinase(POMK)
NM_001283246 calcium/calmodulin dependent protein kinase kinase 2(CAMKK2)
NM_001319385 cyclin dependent kinase 14(CDK14)
NM_001283450 RIO kinase 1(RIOK1)
NM_001284745 ribosomal protein S6 kinase A2(RPS6KA2)
NM_001283970 serine/threonine kinase like domain containing 1(STKLD1)
NM_001284072 serine/threonine kinase 40(STK40)
NM_001284874 ribosomal protein S6 kinase B1(RPS6KB1)

Gene Group 7 Enrichment Score: 1.4630746384406808
REFSEQ_MRNA Gene Name
NM_001289964 glutathione S-transferase pi 1(GSTP1)
NM_001287669 glutathione S-transferase zeta 1(GSTZ1)
NM_001285257 glutathione S-transferase M5(GSTM5)
NM_001319441 glutathione S-transferase omega 2(GSTO2)

Gene Group 8 Enrichment Score: 1.36376933131626
REFSEQ_MRNA Gene Name
NM_001283747 mediator complex subunit 20(MED20)
NM_001284829 Kruppel like factor 9(KLF9)
NM_001319638 zinc finger MYM-type containing 5(ZMYM5)
NM_001285037 zinc finger protein 35(ZNF35)
NM_001285306 thyroid hormone receptor beta(THRB)
NM_001284600 distal-less homeobox 1(DLX1)
NM_001284997 nuclear factor I C(NFIC)
NM_001284987 paired box 6(PAX6)
NM_001287638 nuclear receptor subfamily 2 group C member 1(NR2C1)
NM_001287719 SRY-box 14(SOX14)
NM_001319642 zinc finger protein 614(ZNF614)
NM_001287653 zinc finger protein 512(ZNF512)
NM_001283230 TATA-box binding protein like 1(TBPL1)
NM_001283994 histone deacetylase 11(HDAC11)
NM_001285372 TATA-box binding protein associated factor, RNA polymerase I subunit B(TAF1B)
NM_001285209 zinc finger matrin-type 4(ZMAT4)
NM_001319551 progesterone receptor(PGR)

NM_001284865 zinc finger protein 576(ZNF576)
 NM_001283775 CCR4-NOT transcription complex subunit 9(RQCD1)
 NM_001285110 small nuclear RNA activating complex polypeptide 1(SNAPC1)
 NM_001283923 ZFP30 zinc finger protein(ZFP30)
 NM_001285338 Kruppel like factor 4(KLF4)
 NM_001319515 zinc finger protein 665(LOC107126545)
 NM_001284982 eukaryotic translation elongation factor 1 delta(EEF1D)
 NM_001285109 zinc finger protein 385A(ZNF385A)
 NM_001284102 mediator complex subunit 21(MED21)

Gene Group 9 Enrichment Score: 1.3500321538071285
 REFSEQ_MRNA Gene Name
 NM_001283175 leucine carboxyl methyltransferase 1(LCMT1)
 NM_001284964 protein arginine methyltransferase 1(PRMT1)
 NM_001283361 protein arginine methyltransferase 2(PRMT2)
 NM_001285350 methyltransferase like 16(METTL16)

Gene Group 10 Enrichment Score: 1.247052670428277
 REFSEQ_MRNA Gene Name
 NM_001319631 tetratricopeptide repeat domain 12(TTC12)
 NM_001284938 G-protein signaling modulator 2(GPSM2)
 NM_001319633 tetratricopeptide repeat domain 29(TTC29)
 NM_001319632 tetratricopeptide repeat domain 16(TTC16)
 NM_001283287 tetratricopeptide repeat domain 34(TTC34)
 NM_001285224 FK506 binding protein like(FKBPL)
 NM_001283694 DnaJ heat shock protein family (Hsp40) member C7(DNAJC7)
 NM_001283216 tetratricopeptide repeat domain 9B(TTC9B)
 NM_001283126 protein phosphatase 5 catalytic subunit(PPP5C)
 NM_001319588 small glutamine rich tetratricopeptide repeat containing alpha(SGTA)

M.fascicularis/rheMac8

Gene Group 1 Enrichment Score: 2.0721395443385964
 OFFICIAL_GENE_SYMBOL Gene Name
 RAB23 RAB23, member RAS oncogene family(RAB23)
 RAP1B RAP1B, member of RAS oncogene family(RAP1B)
 SEPT7 septin 7(SEPT7)
 RRAS2 related RAS viral (r-ras) oncogene homolog 2(RRAS2)
 ARL15 ADP ribosylation factor like GTPase 15(ARL15)
 RHOH ras homolog family member H(RHOH)
 RAP2B RAP2B, member of RAS oncogene family(RAP2B)
 RHEB Ras homolog enriched in brain(RHEB)
 RAB5A RAB5A, member RAS oncogene family(RAB5A)
 RAB8B RAB8B, member RAS oncogene family(RAB8B)
 RHOU ras homolog family member U(RHOU)
 RIT2 Ras like without CAAX 2(RIT2)
 ARF4 ADP ribosylation factor 4(ARF4)

Gene Group 2 Enrichment Score: 1.515282141638127
 OFFICIAL_GENE_SYMBOL Gene Name
 USP38 ubiquitin specific peptidase 38(USP38)
 USP46 ubiquitin specific peptidase 46(USP46)
 USP14 ubiquitin specific peptidase 14(USP14)
 USP15 ubiquitin specific peptidase 15(USP15)

Gene Group 3 Enrichment Score: 1.1373094571879139
 OFFICIAL_GENE_SYMBOL Gene Name
 DYRK1A dual specificity tyrosine phosphorylation regulated kinase 1A(DYRK1A)
 TBK1 TANK binding kinase 1(TBK1)
 STK38L serine/threonine kinase 38 like(STK38L)
 MAPKAPK3 mitogen-activated protein kinase-activated protein kinase 3(MAPKAPK3)
 DYRK2 dual specificity tyrosine phosphorylation regulated kinase 2(DYRK2)
 CDC7 cell division cycle 7(CDC7)
 STK24 serine/threonine kinase 24(STK24)
 PAK2 p21 (RAC1) activated kinase 2(PAK2)

6. Literaturverzeichnis

- Abbondanza C, Medici N, Nigro V, et al. 2000. The retinoblastoma-interacting zinc-finger protein RIZ is a downstream effector of estrogen action. *Proceedings of the National Academy of Sciences of the United States of America*. 97(7):3130–3135.
- Acquaviva L, Székvölgyi L, Dichtl B, Dichtl BS, La Roche Saint André C de, Nicolas A, Géli V. 2013. The COMPASS subunit Spp1 links histone methylation to initiation of meiotic recombination. *Science (New York, N.Y.)*. 339(6116):215–218.
- Ahlawat S, Sharma P, Sharma R, Arora R, Verma NK, Brahma B, Mishra P, De S. 2016a. Evidence of positive selection and concerted evolution in the rapidly evolving PRDM9 zinc finger domain in goats and sheep. *Animal genetics*. 47(6):740–751.
- Ahlawat S, Sharma P, Sharma R, Arora R, De S. 2016b. Zinc Finger Domain of the PRDM9 Gene on Chromosome 1 Exhibits High Diversity in Ruminants but Its Paralog PRDM7 Contains Multiple Disruptive Mutations. *PloS one*. 11(5):e0156159.
- Altemose N, Noor N, Bitoun E, Tumian A, Imbeault M, Chapman JR, Aricescu AR, Myers SR. 2017. A map of human PRDM9 binding provides evidence for novel behaviors of PRDM9 and other zinc-finger proteins in meiosis. *eLife*. 6.
- Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, Bestor T, Hannon GJ. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Molecular cell*. 31(6):785–799.
- Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. 2007. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science (New York, N.Y.)*. 316(5825):744–747.
- Assis R, Kondrashov AS. 2009. Rapid repetitive element-mediated expansion of piRNA clusters in mammalian evolution. *Proceedings of the National Academy of Sciences of the United States of America*. 106(17):7079–7082.
- Auton A, Fledel-Alon A, Pfeifer S, et al. 2012. A fine-scale chimpanzee genetic map from population sequencing. *Science (New York, N.Y.)*. 336(6078):193–198.
- Auton A, Rui Li Y, Kidd J, et al. 2013. Genetic recombination is targeted towards gene promoter regions in dogs. *PLoS genetics*. 9(12):e1003984.
- Axelsson E, Webster MT, Ratnakumar A, Ponting CP, Lindblad-Toh K. 2012. Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome. *Genome research*. 22(1):51–63.
- Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE. 2001. Segmental duplications. Organization and impact within the current human genome project assembly. *Genome research*. 11(6):1005–1017.
- Bailey JA, Liu G, Eichler EE. 2003. An Alu transposition model for the origin and expansion of human segmental duplications. *American journal of human genetics*. 73(4):823–834.
- Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings. International Conference on Intelligent Systems for Molecular Biology*. 2:28–36.
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE. Tools for motif discovery and searching. *Nucleic acids research*. 37(Web Server issue):W202-8.
- Baker CL, Kajita S, Walker M, Saxl RL, Raghupathy N, Choi K, Petkov PM, Paigen K. 2015a. PRDM9 drives evolutionary erosion of hotspots in *Mus musculus* through haplotype-specific initiation of meiotic recombination. *PLoS genetics*. 11(1):e1004916.
- Baker CL, Petkova P, Walker M, Flachs P, Mihola O, Trachtulec Z, Petkov PM, Paigen K. 2015b. Multimer Formation Explains Allelic Suppression of PRDM9 Recombination Hotspots. *PLoS genetics*. 11(9):e1005512.
- Baker CL, Walker M, Kajita S, Petkov PM, Paigen K. 2014. PRDM9 binding organizes hotspot nucleosomes and limits Holliday junction migration. *Genome research*. 24(5):724–732.
- Baker Z, Schumer M, Haba Y, Bashkirova L, Holland C, Rosenthal GG, Przeworski M. 2017. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *eLife*. 6.
- Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell*. 129(4):823–837.

- Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, Massy B de. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science (New York, N.Y.)*. 327(5967):836–840.
- Beck JA, Lloyd S, Hafezparast M, Lennon-Pierce M, Eppig JT, Festing MF, Fisher EM. 2000. Genealogies of mouse inbred strains. *Nature genetics*. 24(1):23–25.
- Bellefroid EJ, Poncelet DA, Lecocq PJ, Revelant O, Martial JA. 1991. The evolutionarily conserved Krüppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proceedings of the National Academy of Sciences of the United States of America*. 88(9):3608–3612.
- Berg IL, Neumann R, Lam K-WG, Sarbajna S, Odenthal-Hesse L, May CA, Jeffreys AJ. 2010. PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nature genetics*. 42(10):859–863.
- Bernardi G. 1993. The vertebrate genome. Isochores and evolution. *Molecular biology and evolution*. 10(1):186–204.
- Bernstein BE, Humphrey EL, Erlich RL, Schneider R, Bouman P, Liu JS, Kouzarides T, Schreiber SL. 2002. Methylation of histone H3 Lys 4 in coding regions of active genes. *Proceedings of the National Academy of Sciences of the United States of America*. 99(13):8695–8700.
- Bernstein BE, Kamal M, Lindblad-Toh K, et al. 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*. 120(2):169–181.
- Bhattacharyya T, Reifova R, Gregorova S, Simecek P, Gergelits V, Mistrik M, Martincova I, Pialek J, Forejt J. 2014. X chromosome control of meiotic chromosome synapsis in mouse inter-subspecific hybrids. *PLoS genetics*. 10(2):e1004088.
- Birdsell JA. 2002. Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. *Molecular biology and evolution*. 19(7):1181–1197.
- Birtle Z, Ponting CP. 2006. Meisetz and the birth of the KRAB motif. *Bioinformatics (Oxford, England)*. 22(23):2841–2845.
- Blankenberg D, Gordon A, Kuster G von, Coraor N, Taylor J, Nekrutenko A. 2010a. Manipulation of FASTQ data with Galaxy. *Bioinformatics (Oxford, England)*. 26(14):1783–1785.
- Blankenberg D, Kuster G von, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. 2010b. Galaxy. A web-based genome analysis tool for experimentalists. *Current protocols in molecular biology*. Chapter 19: Unit 19.10.1-21.
- Blat Y, Protacio RU, Hunter N, Kleckner N. 2002. Physical and Functional Interactions among Basic Chromosome Organizational Features Govern Early Steps of Meiotic Chiasma Formation. *Cell*. 111(6):791–802.
- Blazer LL, Lima-Fernandes E, Gibson E, Eram MS, Loppnau P, Arrowsmith CH, Schapira M, Vedadi M. 2016. PR Domain-containing Protein 7 (PRDM7) Is a Histone 3 Lysine 4 Trimethyltransferase. *The Journal of biological chemistry*. 291(26):13509–13519.
- Boeger H, Griesenbeck J, Strattan JS, Kornberg RD. 2003. Nucleosomes Unfold Completely at a Transcriptionally Active Promoter. *Molecular cell*. 11(6):1587–1598.
- Bonhomme M, Cuartero S, Blancher A, Crouau-Roy B. 2009. Assessing natural introgression in 2 biomedical model species, the rhesus macaque (*Macaca mulatta*) and the long-tailed macaque (*Macaca fascicularis*). *The Journal of heredity*. 100(2):158–169.
- Borde V, Robine N, Lin W, Bonfils S, Géli V, Nicolas A. 2009. Histone H3 lysine 4 trimethylation marks meiotic recombination initiation sites. *The EMBO journal*. 28(2):99–111.
- Boulton A, Myers RS, Redfield RJ. 1997. The hotspot conversion paradox and the evolution of meiotic recombination. *Proceedings of the National Academy of Sciences of the United States of America*. 94(15):8058–8063.
- Boursot P, Auffray J-C, Britton-Davidian J, Bonhomme F. 1993. The Evolution of House Mice. *Annu. Rev. Ecol. Syst.* 24(1):119–152.
- Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. 2012. Genetic recombination is directed away from functional genomic elements in mice. *Nature*. 485(7400):642–645.
- Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. 1998. Comprehensive human genetic maps. Individual and sex-specific variation in recombination. *American journal of human genetics*. 63(3):861–869.
- Buard J, Rivals E, Dunoyer de Segonzac D, Garres C, Caminade P, Massy B de, Boursot P. 2014. Diversity of Prdm9 zinc finger array in wild mice unravels new facets of the evolutionary turnover of this coding minisatellite. *PLoS one*. 9(1):e85021.

- Bunlungsup S, Kanthaswamy S, Oldt RF, Smith DG, Houghton P, Hamada Y, Malaivijitnond S. 2017. Genetic analysis of samples from wild populations opens new perspectives on hybridization between long-tailed (*Macaca fascicularis*) and rhesus macaques (*Macaca mulatta*). *Am. J. Primatol.* 79(12).
- Burgoyne PS. 1982. Genetic homology and crossing over in the X and Y chromosomes of Mammals. *Human genetics.* 61(2):85–90.
- Burwinkel B, Kilimann MW. 1998. Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. *Journal of molecular biology.* 277(3):513–517.
- Buyse IM, Shao G, Huang S. 1995. The retinoblastoma protein binds to RIZ, a zinc-finger protein that shares an epitope with the adenovirus E1A protein. *Proceedings of the National Academy of Sciences of the United States of America.* 92(10):4467–4471.
- Cano-Rodriguez D, Gjaltema RAF, Jilderda LJ, Jellema P, Dokter-Fokkens J, Ruiters MHJ, Rots MG. 2016. Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. *Nature communications.* 7:12284.
- Capilla L, Medarde N, Alemany-Schmidt A, Oliver-Bonet M, Ventura J, Ruiz-Herrera A. 2014. Genetic recombination variation in wild Robertsonian mice. On the role of chromosomal fusions and Prdm9 allelic background. *Proceedings. Biological sciences.* 281(1786).
- Chance PF, Abbas N, Lensch MW, Pentao L, Roa BB, Patel PI, Lupski JR. 1994. Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. *Human molecular genetics.* 3(2):223–228.
- Chang C-W, Cheng W-C, Chen C-R, Shu W-Y, Tsai M-L, Huang C-L, Hsu IC. 2011. Identification of human housekeeping genes and tissue-selective genes by microarray meta-analysis. *PloS one.* 6(7):e22859.
- Chen KS, Manian P, Koeuth T, Potocki L, Zhao Q, Chinault AC, Lee CC, Lupski JR. 1997. Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nature genetics.* 17(2):154–163.
- Chen K, Chen Z, Wu D, et al. 2015. Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nature genetics.* 47(10):1149–1157.
- Cheng H-Y, Chen X-W, Cheng L, Liu Y-D, Lou G. 2010. DNA methylation and carcinogenesis of PRDM5 in cervical cancer. *Journal of cancer research and clinical oncology.* 136(12):1821–1825.
- Cheung VG, Nowak N, Jang W, et al. 2001. Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature.* 409(6822):953–958.
- Chia N-Y, Chan Y-S, Feng B, et al. 2010. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature.* 468(7321):316–320.
- Chittka A, Chao MV. 1999. Identification of a zinc finger protein whose subcellular distribution is regulated by serum and nerve growth factor. *Proceedings of the National Academy of Sciences of the United States of America.* 96(19):10705–10710.
- Chittka A, Arevalo JC, Rodriguez-Guzman M, Pérez P, Chao MV, Sendtner M. 2004. The p75NTR-interacting protein SC1 inhibits cell cycle progression by transcriptional repression of cyclin E. *The Journal of cell biology.* 164(7):985–996.
- Chittka A, Nitarska J, Grazini U, Richardson WD. 2012. Transcription factor positive regulatory domain 4 (PRDM4) recruits protein arginine methyltransferase 5 (PRMT5) to mediate histone arginine methylation and control neural stem cell proliferation and differentiation. *The Journal of biological chemistry.* 287(51):42995–43006.
- Clément Y, Arndt PF. 2013. Meiotic recombination strongly influences GC-content evolution in short regions in the mouse genome. *Molecular biology and evolution.* 30(12):2612–2618.
- Comeaux MS, Roy-Engel AM, Hedges DJ, Deininger PL. 2009. Diverse cis factors controlling Alu retrotransposition. What causes Alu elements to die? *Genome research.* 19(4):545–555.
- Coop GM, Myers SR. 2005. Live hot, die young. Transmission distortion in recombination hotspots. *PLoS genetics.* preprint(2007):e35.
- Costantini M, Clay O, Auletta F, Bernardi G. 2006. An isochore map of human chromosomes. *Genome research.* 16(4):536–541.
- Cui P, Liu W, Zhao Y, et al. 2012. Comparative analyses of H3K4 and H3K27 trimethylations between the mouse cerebrum and testis. *Genomics, proteomics & bioinformatics.* 10(2):82–93.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2. More models, new heuristics and parallel computing. *Nature methods.* 9(8):772.

- Davies B, Hatton E, Altemose N, et al. 2016. Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. *Nature*. 530(7589):171–176.
- Davis CA, Haberland M, Arnold MA, et al. 2006. PRISM/PRDM6, a transcriptional repressor that promotes the proliferative gene program in smooth muscle cells. *Molecular and cellular biology*. 26(7):2626–2636.
- Deem A, Keszthelyi A, Blackgrove T, Vayl A, Coffey B, Mathur R, Chabes A, Malkova A. 2011. Break-induced replication is highly inaccurate. *PLoS biology*. 9(2):e1000594.
- De Fazio S, Bartonicek N, Di Giacomo M, et al. 2011. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature*. 480(7376):259–263.
- Deininger PL, Batzer MA. 1999. Alu repeats and human disease. *Molecular genetics and metabolism*. 67(3):183–193.
- Deng Q, Huang S. 2004. PRDM5 is silenced in human cancers and has growth suppressive activities. *Oncogene*. 23(28):4903–4910.
- Dettman EJ, Justice MJ. 2008. The zinc finger SET domain gene Prdm14 is overexpressed in lymphoblastic lymphomas with retroviral insertions at Evi32. *PloS one*. 3(11):e3823.
- Dettman EJ, Simko SJ, Ayanga B, Carofino BL, Margolin JF, Morse HC, Justice MJ. 2011. Prdm14 initiates lymphoblastic leukemia after expanding a population of cells resembling common lymphoid progenitors. *Oncogene*. 30(25):2859–2873.
- Di Giacomo M, Comazzetto S, Saini H, et al. 2013. Multiple epigenetic mechanisms and the piRNA pathway enforce LINE1 silencing during adult spermatogenesis. *Molecular cell*. 50(4):601–608.
- Di Zazzo E, Rosa C de, Abbondanza C, Moncharmont B. 2013. PRDM Proteins. *Molecular Mechanisms in Signal Transduction and Transcriptional Regulation*. *Biology*. 2(1):107–141.
- Dion-Côté A-M, Renaut S, Normandeau E, Bernatchez L. 2014. RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young lake whitefish species. *Molecular biology and evolution*. 31(5):1188–1199.
- Dobzhansky T, Gould SJ. 1982. *Genetics and the origin of species*, New York: Columbia University Press.
- Duan Z, Person RE, Lee H-H, Huang S, Donadieu J, Badolato R, Grimes HL, Papayannopoulou T, Horwitz MS. 2007. Epigenetic regulation of protein-coding and microRNA genes by the Gfi1-interacting tumor suppressor PRDM5. *Molecular and cellular biology*. 27(19):6889–6902.
- Eberl HC, Spruijt CG, Kelstrup CD, Vermeulen M, Mann M. 2013. A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. *Molecular cell*. 49(2):368–378.
- Eisenbarth I, Striebel AM, Moschgath E, Vogel W, Assum G. 2001. Long-range sequence composition mirrors linkage disequilibrium pattern in a 1.13 Mb region of human chromosome 22. *Human molecular genetics*. 10(24):2833–2839.
- Eisenbarth I, Vogel G, Krone W, Vogel W, Assum G. 2000. An isochore transition in the NF1 gene region coincides with a switch in the extent of linkage disequilibrium. *American journal of human genetics*. 67(4):873–880.
- Eisenberg E, Levanon EY. 2013. Human housekeeping genes, revisited. *Trends in genetics: TIG*. 29(10):569–574.
- Ellermeier C, Higuchi EC, Phadnis N, Holm L, Geelhood JL, Thon G, Smith GR. 2010. RNAi and heterochromatin repress centromeric meiotic recombination. *Proceedings of the National Academy of Sciences of the United States of America*. 107(19):8701–8705.
- Ellis N, Yen P, Neiswanger K, Shapiro LJ, Goodfellow PN. 1990. Evolution of the pseudoautosomal boundary in old world monkeys and great apes. *Cell*. 63(5):977–986.
- Eram MS, Bustos SP, Lima-Fernandes E, et al. 2014. Trimethylation of histone H3 lysine 36 by human methyltransferase PRDM9 protein. *The Journal of biological chemistry*. 289(17):12177–12188.
- Eyre-Walker A. 1993. Recombination and mammalian genome evolution. *Proceedings. Biological sciences*. 252(1335):237–243.
- Fairall L, Schwabe JW, Chapman L, Finch JT, Rhodes D. 1993. The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature*. 366(6454):483–487.
- Faucher D, Wellinger RJ. 2010. Methylated H3K4, a transcription-associated histone modification, is involved in the DNA damage response pathway. *PLoS genetics*. 6(8).

- Feichtinger J, Aldeaij I, Anderson R, et al. 2012. Meta-analysis of clinical data using human meiotic genes identifies a novel cohort of highly restricted cancer-specific marker genes. *Oncotarget*. 3(8):843–853.
- Flachs P, Mihola O, Simeček P, et al. 2012. Interallelic and intergenic incompatibilities of the Prdm9 (Hst1) gene in mouse hybrid sterility. *PLoS genetics*. 8(11):e1003044.
- Fog CK, Galli GG, Lund AH. 2012. PRDM proteins. Important players in differentiation and disease. *BioEssays: news and reviews in molecular, cellular and developmental biology*. 34(1):50–60.
- Fooden J. 1976. Provisional classifications and key to living species of macaques (primates. Macaca). *Folia primatologica; international journal of primatology*. 25(2-3):225–236.
- Fooden J. 1995. Systematic review of Southeast Asian longtail macaques, *Macaca fascicularis* (Raffles, [1821]) / Jack Fooden, [Chicago, Ill.]: Field Museum of Natural History.
- Fooden J. 1997. Tail length variation in *Macaca fascicularis* and *M. mulatta*. *Primates*. 38(3):221–231.
- Fooden J. 2000. Systematic review of the rhesus macaque, *Macaca mulatta* (Zimmermann, 1780) / Jack Fooden, Chicago, Ill.: Field Museum of Natural History.
- Forejt J, Iványi P. 1974. Genetic studies on male sterility of hybrids between laboratory and wild mice (*Mus musculus* L.). *Genetical research*. 24(2):189–206.
- Friedman JR, Fredericks WJ, Jensen DE, Speicher DW, Huang XP, Neilson EG, Rauscher FJ. 1996. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes & development*. 10(16):2067–2078.
- Fullerton SM, Bernardo Carvalho A, Clark AG. 2001. Local rates of recombination are positively correlated with GC content in the human genome. *Molecular biology and evolution*. 18(6):1139–1142.
- Fumasoni I, Meani N, Rambaldi D, Scafetta G, Alcalay M, Ciccarelli FD. 2007. Family expansion and gene rearrangements contributed to the functional specialization of PRDM genes in vertebrates. *BMC evolutionary biology*. 7:187.
- Galtier N, Piganeau G, Mouchiroud D, Duret L. 2001. GC-content evolution in mammalian genomes. The biased gene conversion hypothesis. *Genetics*. 159(2):907–911.
- Gazzerro P, Abbondanza C, D'Arcangelo A, Rossi M, Medici N, Moncharmont B, Puca GA. 2006. Modulation of RIZ gene expression is associated to estradiol control of MCF-7 breast cancer cell proliferation. *Experimental cell research*. 312(3):340–349.
- Gebert D, Zischler H, Rosenkranz D. 2019. Primate piRNA Cluster Evolution Suggests Limited Relevance of Pseudogenes in piRNA-Mediated Gene Regulation. *Genome biology and evolution*. 11(4):1088–1104.
- Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD. 2000. Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*. 97(21):11383–11390.
- Giardine B, Riemer C, Hardison RC, et al. 2005. Galaxy. A platform for interactive large-scale genome analysis. *Genome research*. 15(10):1451–1455.
- Goecks J, Nekrutenko A, Taylor J. 2010. Galaxy. A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome biology*. 11(8):R86.
- Goh WSS, Falciatori I, Tam OH, Burgess R, Meikar O, Kotaja N, Hammell M, Hannon GJ. 2015. piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. *Genes & development*. 29(10):1032–1044.
- Goitre L, Trapani E, Trabalzini L, Retta SF. 2014. The Ras superfamily of small GTPases. The unlocked secrets. *Methods in molecular biology (Clifton, N.J.)*. 1120:1–18.
- Gou L-T, Dai P, Yang J-H, et al. 2015. Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell research*. 25(2):266.
- Grant CE, Bailey TL, Noble WS. 2011. FIMO. Scanning for occurrences of a given motif. *Bioinformatics (Oxford, England)*. 27(7):1017–1018.
- Gray AP. 1954. Mammalian Hybrids. Farnham Royal, Bucks.: Commonwealth Agricultural Bureaux.
- Grey C, Clément JAJ, Buard J, Leblanc B, Gut I, Gut M, Duret L, Massy B de. 2017. In vivo binding of PRDM9 reveals interactions with noncanonical genomic sites. *Genome research*. 27(4):580–590.

- Groeneveld LF, Atencia R, Garriga RM, Vigilant L. 2012. High diversity at PRDM9 in chimpanzees and bonobos. *PloS one*. 7(7):e39064.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology*. 52(5):696–704.
- Haldane JBS. 1922. Sex ratio and unisexual sterility in hybrid animals. *Journ. of Gen.* 12(2):101–109.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98.
- Hamada Y, Urasopon N, Hadi I, Malaivijitnond S. 2006. Body Size and Proportions and Pelage Color of Free-Ranging *Macaca mulatta* from a Zone of Hybridization in Northeastern Thailand. *Int J Primatol*. 27(2):497–513.
- Han K, Konkel MK, Xing J, et al. 2007. Mobile DNA in Old World monkeys. A glimpse through the rhesus macaque genome. *Science (New York, N.Y.)*. 316(5822):238–240.
- Han K, Lee J, Meyer TJ, Remedios P, Goodwin L, Batzer MA. 2008. L1 recombination-associated deletions generate human genomic variation. *Proceedings of the National Academy of Sciences of the United States of America*. 105(49):19366–19371.
- Hanahan D, Weinberg RA. 2011. Hallmarks of cancer. The next generation. *Cell*. 144(5):646–674.
- Hancks DC, Kazazian HH. 2010. SVA retrotransposons. Evolution and genetic instability. *Seminars in cancer biology*. 20(4):234–245.
- Handel MA. 2004. The XY body. A specialized meiotic chromatin domain. *Experimental cell research*. 296(1):57–63.
- Hayashi K, Yoshida K, Matsui Y. 2005. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature*. 438(7066):374–378.
- Heerschoop S, Zischler H, Merker S, Perwitasari-Farajallah D, Driller C. 2016. The pioneering role of PRDM9 indel mutations in tarsier evolution. *Scientific reports*. 6:34618.
- Heinz S, Benner C, Spann N, et al. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*. 38(4):576–589.
- Hendrix A, Maynard D, Pauwels P, et al. 2010. Effect of the secretory small GTPase Rab27B on breast cancer growth, invasion, and metastasis. *Journal of the National Cancer Institute*. 102(12):866–880.
- Herndon JG. 1983. Seasonal breeding in rhesus monkeys. Influence of the behavioral environment. *Am. J. Primatol*. 5(3):197–204.
- Hicks WM, Kim M, Haber JE. 2010. Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science (New York, N.Y.)*. 329(5987):82–85.
- Hinch AG, Altemose N, Noor N, Donnelly P, Myers SR. 2014. Recombination in the human Pseudoautosomal region PAR1. *PLoS genetics*. 10(7):e1004503.
- Hinch AG, Zhang G, Becker PW, Moralli D, Hinch R, Davies B, Bowden R, Donnelly P. 2019. Factors influencing meiotic recombination revealed by whole-genome sequencing of single sperm. *Science (New York, N.Y.)*. 363(6433).
- Ho JR, Chapeaublanc E, Kirkwood L, et al. 2012. Deregulation of Rab and Rab effector genes in bladder cancer. *PloS one*. 7(6):e39469.
- Hohenauer T, Moore AW. 2012. The Prdm family. Expanding roles in stem cells and development. *Development (Cambridge, England)*. 139(13):2267–2282.
- Houle AA, Gibling H, Lamaze FC, et al. 2018. Aberrant PRDM9 expression impacts the pan-cancer genomic landscape. *Genome research*. 28(11):1611–1620.
- Howe FS, Fischl H, Murray SC, Mellor J. 2017. Is H3K4me3 instructive for transcription activation? *BioEssays: news and reviews in molecular, cellular and developmental biology*. 39(1):1–12.
- Hoyer-Fender S. 2003. Molecular aspects of XY body formation. *Cytogenetic and genome research*. 103(3-4):245–255.
- Huang DW, Sherman BT, Lempicki RA. 2009a. Bioinformatics enrichment tools. Paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research*. 37(1):1–13.
- Huang DW, Sherman BT, Lempicki RA. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*. 4(1):44–57.
- Huang S, Shao G, Liu L. 1998. The PR domain of the Rb-binding zinc finger protein RIZ1 is a protein binding interface and is related to the SET domain functioning in chromatin-mediated gene expression. *The Journal of biological chemistry*. 273(26):15933–15939.

- Hughes JF, Skaletsky H, Brown LG, et al. 2012. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature*. 483(7387):82–86.
- Hultén M. 1974. Chiasma distribution at diakinesis in the normal human male. *Hereditas*. 76(1):55–78.
- Hurst LD, Brunton CFA, Smith NGC. 1999. Small introns tend to occur in GC-rich regions in some but not all vertebrates. *Trends in genetics: TIG*. 15(11):437–439.
- Hussin J, Sinnott D, Casals F, et al. 2013. Rare allelic forms of PRDM9 associated with childhood leukemogenesis. *Genome research*. 23(3):419–430.
- Ikemura T, Wada K. 1991. Evident diversity of codon usage patterns of human genes with respect to chromosome banding patterns and chromosome numbers; relation between nucleotide sequence data and cytogenetic data. *Nucleic acids research*. 19(16):4333–4339.
- Imai Y, Baudat F, Taillepierre M, Stanzione M, Toth A, Massy B de. 2017. The PRDM9 KRAB domain is required for meiosis and involved in protein interactions. *Chromosoma*. 126(6):681–695.
- International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature*. 409(6822):860–921.
- Irie S, Tsujimura A, Miyagawa Y, Ueda T, Matsuoka Y, Matsui Y, Okuyama A, Nishimune Y, Tanaka H. 2009. Single-nucleotide polymorphisms of the PRDM9 (MEISETZ) gene in patients with nonobstructive azoospermia. *Journal of andrology*. 30(4):426–431.
- Irving J, Matheson E, Minto L, et al. 2014. Ras pathway mutations are prevalent in relapsed childhood acute lymphoblastic leukemia and confer sensitivity to MEK inhibition. *Blood*. 124(23):3420–3430.
- Jackson SP. 2002. Sensing and repairing DNA double-strand breaks. *Carcinogenesis*. 23(5):687–696.
- Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D. 2014. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature*. 516(7530):242–245.
- Jeffreys AJ, Kauppi L, Neumann R. 2001. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature genetics*. 29(2):217–222.
- Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen C-F, Thomas MA, Haussler D, Jacob HJ. 2004. Comparative recombination rates in the rat, mouse, and human genomes. *Genome research*. 14(4):528–538.
- Kajimura S, Seale P, Tomaru T, et al. 2008. Regulation of the brown and white fat gene programs through a PRDM16/CtBP transcriptional complex. *Genes & development*. 22(10):1397–1409.
- Kauppi L, Jasin M, Keeney S. 2012. The tricky path to recombining X and Y chromosomes in meiosis. *Annals of the New York Academy of Sciences*. 1267:18–23.
- Kavanagh M, Laursen E. 1984. Breeding seasonality among long-tailed macaques, *Macaca fascicularis*, in Peninsular Malaysia. *Int J Primatol*. 5(1):17–29.
- Kharchenko PV, Tolstorukov MY, Park PJ. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature biotechnology*. 26(12):1351–1359.
- Kong A, Gudbjartsson DF, Sainz J, et al. 2002. A high-resolution recombination map of the human genome. *Nature genetics*. 31(3):241–247.
- Kono H, Tamura M, Osada N, Suzuki H, Abe K, Moriwaki K, Ohta K, Shiroishi T. 2014. Prdm9 polymorphism unveils mouse evolutionary tracks. *DNA research: an international journal for rapid publication of reports on genes and genomes*. 21(3):315–326.
- Kumar R, Bourbon H-M, Massy B de. 2010. Functional conservation of Mei4 for meiotic DNA double-strand break formation from yeasts to mice. *Genes & development*. 24(12):1266–1280.
- Kumar R, Oliver C, Brun C, Juarez-Martinez AB, Tarabay Y, Kadlec J, Massy B de. 2018. Mouse REC114 is essential for meiotic DNA double-strand break formation and forms a complex with MEI4. *Life science alliance*. 1(6):e201800259.
- Kuo Y-C, Shen Y-R, Chen H-I, Lin Y-H, Wang Y-Y, Chen Y-R, Wang C-Y, Kuo P-L. 2015. SEPT12 orchestrates the formation of mammalian sperm annulus by organizing core octameric complexes with other SEPT proteins. *Journal of cell science*. 128(5):923–934.
- Lamb NE, Feingold E, Savage A, et al. 1997. Characterization of susceptible chiasma configurations that increase the risk for maternal nondisjunction of chromosome 21. *Human molecular genetics*. 6(9):1391–1399.
- Landt SG, Marinov GK, Kundaje A, et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome research*. 22(9):1813–1831.

- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*. 10(3):R25.
- Lee C-K, Shibata Y, Rao B, Strahl BD, Lieb JD. 2004. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nature genetics*. 36(8):900–905.
- Lee J-H, Skalnik DG. 2005. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *The Journal of biological chemistry*. 280(50):41725–41731.
- Lee J, Ha J, Son S-Y, Han K. 2012. Human Genomic Deletions Generated by SVA-Associated Events. *Comparative and functional genomics*. 2012:807270.
- Lercher MJ, Hurst LD. 2002. Human SNP variability and mutation rate are higher in regions of high recombination. *Trends in genetics: TIG*. 18(7):337–340.
- Lesecque Y, Glémin S, Lartillot N, Mouchiroud D, Duret L. 2014. The red queen model of recombination hotspots evolution in the light of archaic and modern human genomes. *PLoS genetics*. 10(11):e1004790.
- Li H. 2011a. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford, England)*. 27(21):2987–2993.
- Li H. 2011b. Improving SNP discovery by base alignment quality. *Bioinformatics (Oxford, England)*. 27(8):1157–1158.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009a. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*. 25(16):2078–2079.
- Li J, Han K, Xing J, Kim H-S, Rogers J, Ryder OA, Disotell T, Yue B, Batzer MA. 2009b. Phylogeny of the macaques (Cercopithecidae. *Macaca*) based on Alu elements. *Gene*. 448(2):242–249.
- Li X, Ito M, Zhou F, Youngson N, Zuo X, Leder P, Ferguson-Smith AC. 2008. A maternal-zygotic effect gene, *Zfp57*, maintains both maternal and paternal imprints. *Developmental cell*. 15(4):547–557.
- Li Y, Jia Q, Zhang Q, Wan Y. 2015. Rab25 upregulation correlates with the proliferation, migration, and invasion of renal cell carcinoma. *Biochemical and biophysical research communications*. 458(4):745–750.
- Li Z, Golub EI, Gupta R, Radding CM. 1997. Recombination activities of HsDmc1 protein, the meiotic human homolog of RecA protein. *Proceedings of the National Academy of Sciences of the United States of America*. 94(21):11221–11226.
- Liang D-C, Shih L-Y, Fu J-F, et al. 2006. K-Ras mutations and N-Ras mutations in childhood acute leukemias with or without mixed-lineage leukemia gene rearrangements. *Cancer*. 106(4):950–956.
- Lim FL, Soulez M, Koczan D, Thiesen HJ, Knight JC. 1998. A KRAB-related domain and a novel transcription repression domain in proteins encoded by SSX genes that are disrupted in human sarcomas. *Oncogene*. 17(15):2013–2018.
- Lupski JR. 1998. Genomic disorders. Structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends in genetics: TIG*. 14(10):417–422.
- Lustyk D, Kinský S, Ullrich KK, et al. 2019. Genomic structure of Hstx2 modifier of Prdm9-dependent hybrid male sterility in mice. bioRxiv 670422; doi: <https://doi.org/10.1101/670422>.
- Ma Z, Swigut T, Valouev A, Rada-Iglesias A, Wysocka J. 2011. Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. *Nature structural & molecular biology*. 18(2):120–127.
- Mahony S, Auron PE, Benos PV. 2007. DNA familial binding profiles made easy. Comparison of various motif alignment and clustering strategies. *PLoS computational biology*. 3(3):e61.
- Mahony S, Benos PV. 2007. STAMP. A web tool for exploring DNA-binding motif similarities. *Nucleic acids research*. 35(Web Server issue):W253-8.
- Marais G, Mouchiroud D, Duret L. 2001. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. *Proceedings of the National Academy of Sciences of the United States of America*. 98(10):5688–5692.
- Marsolier-Kergoat M-C, Yeramian E. 2009. GC content and recombination. Reassessing the causal effects for the *Saccharomyces cerevisiae* genome. *Genetics*. 183(1):31–38.
- Mayr E. 1969. Animal species and evolution. Cambridge: Harvard University Press.

- Medici N, Abbondanza C, Nigro V, et al. 1999. Identification of a DNA binding protein cooperating with estrogen receptor as RIZ (retinoblastoma interacting zinc finger protein). *Biochemical and biophysical research communications*. 264(3):983–989.
- Meunier J, Duret L. 2004. Recombination drives the evolution of GC-content in the human genome. *Molecular biology and evolution*. 21(6):984–990.
- Mihola O, Pratto F, Brick K, et al. 2019. Histone methyltransferase PRDM9 is not essential for meiosis in male mice. *Genome research*. 29(7):1078–1086.
- Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. 2009. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science (New York, N.Y.)*. 323(5912):373–375.
- Miller J, McLachlan AD, Klug A. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *The EMBO journal*. 4(6):1609–1614.
- Miyamoto T, Koh E, Sakugawa N, Sato H, Hayashi H, Namiki M, Sengoku K. 2008. Two single nucleotide polymorphisms in PRDM9 (MEISETZ) gene may be a genetic risk factor for Japanese patients with azoospermia by meiotic arrest. *Journal of assisted reproduction and genetics*. 25(11-12):553–557.
- Montoya-Burgos JI, Boursot P, Galtier N. 2003. Recombination explains isochores in mammalian genomes. *Trends in genetics: TIG*. 19(3):128–130.
- Morales JC, Melnick DJ. 1998. Phylogenetic relationships of the macaques (Cercopithecidae. *Macaca*), as revealed by high resolution restriction site mapping of mitochondrial ribosomal genes. *Journal of human evolution*. 34(1):1–23.
- Mosesson Y, Voichek Y, Barkai N. 2014. Divergence and selectivity of expression-coupled histone modifications in budding yeasts. *PloS one*. 9(7):e101538.
- Muñoz-Fuentes V, Di Rienzo A, Vilà C. 2011. Prdm9, a major determinant of meiotic recombination hotspots, is not functional in dogs and their wild relatives, wolves and coyotes. *PloS one*. 6(11):e25498.
- Muramoto T, Müller I, Thomas G, Melvin A, Chubb JR. 2010. Methylation of H3K4 is required for inheritance of active transcriptional states. *Current biology: CB*. 20(5):397–406.
- Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of recombination rates and hotspots across the human genome. *Science (New York, N.Y.)*. 310(5746):321–324.
- Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, McVean G, Donnelly P. 2010. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science (New York, N.Y.)*. 327(5967):876–879.
- Myers S, Freeman C, Auton A, Donnelly P, McVean G. 2008. A common sequence motif associated with recombination hot spots and genome instability in humans. *Nature genetics*. 40(9):1124–1129.
- Mzoughi S, Tan YX, Low D, Guccione E. 2016. The role of PRDMs in cancer. One family, two sides. *Current opinion in genetics & development*. 36:83–91.
- Nam KT, Lee H-J, Smith JJ, et al. 2010. Loss of Rab25 promotes the development of intestinal neoplasia in mice and is associated with human colorectal adenocarcinomas. *The Journal of clinical investigation*. 120(3):840–849.
- Narasimhan VM, Hunt KA, Mason D, et al. 2016. Health and population effects of rare gene knockouts in adult humans with related parents. *Science (New York, N.Y.)*. 352(6284):474–477.
- Nishikawa N, Toyota M, Suzuki H, et al. 2007. Gene amplification and overexpression of PRDM14 in breast cancers. *Cancer research*. 67(20):9649–9657.
- Nishitsuji H, Abe M, Sawada R, Takaku H. 2012. ZBRK1 represses HIV-1 LTR-mediated transcription. *FEBS letters*. 586(20):3562–3568.
- O'Carroll D, Scherthan H, Peters AH, et al. 2000. Isolation and characterization of Suv39h2, a second histone H3 methyltransferase gene that displays testis-specific expression. *Molecular and cellular biology*. 20(24):9423–9433.
- Oliver PL, Goodstadt L, Bayes JJ, et al. 2009. Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. *PLoS genetics*. 5(12):e1000753.
- Oliver TR, Middlebrooks C, Harden A, et al. 2016. Variation in the Zinc Finger of PRDM9 is Associated with the Absence of Recombination along Nondisjoined Chromosomes 21 of Maternal Origin. *Journal of Down Syndrome & chromosome abnormalities*. 2(2).
- O'Neill RJ, O'Neill MJ, Graves JA. 1998. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature*. 393(6680):68–72.

- Ostertag EM, Goodier JL, Zhang Y, Kazazian HH. 2003. SVA elements are nonautonomous retrotransposons that cause disease in humans. *American journal of human genetics*. 73(6):1444–1451.
- Ozsolak F, Song JS, Liu XS, Fisher DE. 2007. High-throughput mapping of the chromatin structure of human promoters. *Nature biotechnology*. 25(2):244–248.
- Paigen K, Petkov PM. 2018. PRDM9 and Its Role in Genetic Recombination. *Trends in genetics: TIG*. 34(4):291–300.
- Paigen K, Szatkiewicz JP, Sawyer K, Leahy N, Parvanov ED, Ng SHS, Graber JH, Broman KW, Petkov PM. 2008. The recombinational anatomy of a mouse chromosome. *PLoS genetics*. 4(7):e1000119.
- Pan J, Sasaki M, Kniewel R, et al. 2011. A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell*. 144(5):719–731.
- Parvanov ED, Petkov PM, Paigen K. 2010. Prdm9 controls activation of mammalian recombination hotspots. *Science (New York, N.Y.)*. 327(5967):835.
- Parvanov ED, Tian H, Billings T, Saxl RL, Spruce C, Aithal R, Krejci L, Paigen K, Petkov PM. 2017. PRDM9 interactions with other proteins provide a link between recombination hotspots and the chromosomal axis in meiosis. *Molecular biology of the cell*. 28(3):488–499.
- Patel A, Horton JR, Wilson GG, Zhang X, Cheng X. 2016. Structural basis for human PRDM9 action at recombination hot spots. *Genes & development*. 30(3):257–265.
- Patel A, Zhang X, Blumenthal RM, Cheng X. 2017. Structural basis of human PR/SET domain 9 (PRDM9) allele C-specific recognition of its cognate DNA sequence. *The Journal of biological chemistry*. 292(39):15994–16002.
- Pavletich NP, Pabo CO. 1991. Zinc finger-DNA recognition. Crystal structure of a Zif268-DNA complex at 2.1 Å. *Science (New York, N.Y.)*. 252(5007):809–817.
- Peña PV, Hom RA, Hung T, et al. 2008. Histone H3K4me3 binding is required for the DNA repair and apoptotic activities of ING1 tumor suppressor. *Journal of molecular biology*. 380(2):303–312.
- Peng H, Begg GE, Harper SL, Friedman JR, Speicher DW, Rauscher FJ. 2000. Biochemical analysis of the Kruppel-associated box (KRAB) transcriptional repression domain. *The Journal of biological chemistry*. 275(24):18000–18010.
- Pentao L, Wise CA, Chinault AC, Patel PI, Lupski JR. 1992. Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. *Nature genetics*. 2(4):292–300.
- Perelman P, Johnson WE, Roos C, et al. 2011. A molecular phylogeny of living primates. *PLoS genetics*. 7(3):e1001342.
- Persikov AV, Osada R, Singh M. 2009. Predicting DNA recognition by Cys2His2 zinc finger proteins. *Bioinformatics (Oxford, England)*. 25(1):22–29.
- Persikov AV, Singh M. 2014. De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic acids research*. 42(1):97–108.
- Petes TD. 2001. Meiotic recombination hot spots and cold spots. *Nature reviews. Genetics*. 2(5):360–369.
- Petes TD, Merker JD. 2002. Context dependence of meiotic recombination hotspots in yeast. The relationship between recombination activity of a reporter construct and base composition. *Genetics*. 162(4):2049–2052.
- Pineda-Krch M, Redfield RJ. 2005. Persistence and loss of meiotic recombination hotspots. *Genetics*. 169(4):2319–2333.
- Pittman DL, Cobb J, Schimenti KJ, Wilson LA, Cooper DM, Brignull E, Handel MA, Schimenti JC. 1998. Meiotic Prophase Arrest with Failure of Chromosome Synapsis in Mice Deficient for Dmc1, a Germline-Specific RecA Homolog. *Molecular cell*. 1(5):697–705.
- Popa A, Samollow P, Gautier C, Mouchiroud D. 2012. The sex-specific impact of meiotic recombination on nucleotide composition. *Genome biology and evolution*. 4(3):412–422.
- Potocki L, Chen KS, Park SS, et al. 2000. Molecular mechanism for duplication 17p11.2- the homologous recombination reciprocal of the Smith-Magenis microdeletion. *Nature genetics*. 24(1):84–87.
- Potocki L, Bi W, Treadwell-Deering D, et al. 2007. Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype. *American journal of human genetics*. 80(4):633–649.

- Powers NR, Parvanov ED, Baker CL, Walker M, Petkov PM, Paigen K. 2016. The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 and H3K4 at Recombination Hotspots In Vivo. *PLoS genetics*. 12(6):e1006146.
- Pratto F, Brick K, Khil P, Smagulova F, Petukhova GV, Camerini-Otero RD. 2014. DNA recombination. Recombination initiation maps of individual human genomes. *Science (New York, N.Y.)*. 346(6211):1256442.
- Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, Ballinger DG, Przeworski M, Frazer KA, Pääbo S. 2005. Fine-scale recombination patterns differ between chimpanzees and humans. *Nature genetics*. 37(4):429–434.
- Quenneville S, Turelli P, Bojkowska K, Raclot C, Offner S, Kapopoulou A, Trono D. 2012. The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. *Cell reports*. 2(4):766–773.
- Radman-Livaja M, Rando OJ. 2010. Nucleosome positioning. How is it established, and why does it matter? *Developmental biology*. 339(2):258–266.
- Reiter LT, Murakami T, Koeuth T, Pentao L, Muzny DM, Gibbs RA, Lupski JR. 1996. A recombination hotspot responsible for two inherited peripheral neuropathies is located near a mariner transposon-like element. *Nature genetics*. 12(3):288–297.
- Robert T, Nore A, Brun C, Maffre C, Crimi B, Bourbon H-M, Massy B de. 2016. The TopoVIB-Like protein family is required for meiotic DNA double-strand break formation. *Science (New York, N.Y.)*. 351(6276):943–949.
- Rodgers K, McVey M. 2016. Error-Prone Repair of DNA Double-Strand Breaks. *Journal of cellular physiology*. 231(1):15–24.
- Rowe N. 1996. The pictorial guide to the living primates, East Hampton, NY: Pogonias Press.
- Samonte RV, Eichler EE. 2002. Segmental duplications and the evolution of the primate genome. *Nature reviews. Genetics*. 3(1):65–72.
- Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NCT, Schreiber SL, Mellor J, Kouzarides T. 2002. Active genes are tri-methylated at K4 of histone H3. *Nature*. 419(6905):407–411.
- Satkoski Trask JA, Garnica WT, Smith DG, Houghton P, Lerche N, Kanthaswamy S. 2013. Single-nucleotide polymorphisms reveal patterns of allele sharing across the species boundary between rhesus (*Macaca mulatta*) and cynomolgus (*M. fascicularis*) macaques. *Am. J. Primatol.* 75(2):135–144.
- Schneider R, Bannister AJ, Myers FA, Thorne AW, Crane-Robinson C, Kouzarides T. 2004. Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nature cell biology*. 6(1):73–77.
- Schwartz JJ, Roach DJ, Thomas JH, Shendure J. 2014. Primate evolution of the recombination regulator PRDM9. *Nature communications*. 5:4370.
- Seale P, Bjork B, Yang W, et al. 2008. PRDM16 controls a brown fat/skeletal muscle switch. *Nature*. 454(7207):961–967.
- Seale P, Conroe HM, Estall J, Kajimura S, Frontini A, Ishibashi J, Cohen P, Cinti S, Spiegelman BM. 2011. Prdm16 determines the thermogenic program of subcutaneous white adipose tissue in mice. *The Journal of clinical investigation*. 121(1):96–105.
- Seale P, Kajimura S, Yang W, Chin S, Rohas LM, Uldry M, Tavernier G, Langin D, Spiegelman BM. 2007. Transcriptional control of brown fat determination by PRDM16. *Cell metabolism*. 6(1):38–54.
- Ségurel L, Leffler EM, Przeworski M. 2011. The case of the fickle fingers. How the PRDM9 zinc finger protein specifies meiotic recombination hotspots in humans. *PLoS biology*. 9(12):e1001211.
- Sen SK, Han K, Wang J, et al. 2006. Human genomic deletions mediated by recombination between Alu elements. *American journal of human genetics*. 79(1):41–53.
- Shaw CJ, Bi W, Lupski JR. 2002. Genetic proof of unequal meiotic crossovers in reciprocal deletion and duplication of 17p11.2. *American journal of human genetics*. 71(5):1072–1081.
- Shu XO, Perentesis JP, Wen W, Buckley JD, Boyle E, Ross JA, Robison LL. 2004. Parental exposure to medications and hydrocarbons and ras mutations in children with acute lymphoblastic leukemia. A report from the Children's Oncology Group. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 13(7):1230–1235.

- Shu X-s, Geng H, Li L, et al. 2011. The epigenetic modifier PRDM5 functions as a tumor suppressor through modulating WNT/ β -catenin signaling and is frequently silenced in multiple tumors. *PLoS one*. 6(11):e27346.
- Singhal S, Leffler EM, Sannareddy K, et al. 2015. Stable recombination hotspots in birds. *Science (New York, N.Y.)*. 350(6263):928–932.
- Smagulova F, Brick K, Pu Y, Camerini-Otero RD, Petukhova GV. 2016. The evolutionary turnover of recombination hot spots contributes to speciation in mice. *Genes & development*. 30(3):266–280.
- Sommermeier V, Béneut C, Chaplais E, Serrentino ME, Borde V. 2013. Spp1, a member of the Set1 Complex, promotes meiotic DSB formation in promoters by tethering histone H3K4 methylation sites to chromosome axes. *Molecular cell*. 49(1):43–54.
- Stankiewicz P, Lupski JR. 2002. Genome architecture, rearrangements and genomic disorders. *Trends in genetics : TIG*. 18(2):74–82.
- Stanzione M, Baumann M, Papanikos F, et al. 2016. Meiotic DNA break formation requires the unsynapsed chromosome axis-binding protein IHO1 (CCDC36) in mice. *Nature cell biology*. 18(11):1208–1220.
- Stearns SC, Hoekstra RF. 2003. Evolution. An introduction, Oxford: Univ. Press.
- Steiner CC, Ryder OA. 2013. Characterization of Prdm9 in equids and sterility in mules. *PLoS one*. 8(4):e61746.
- Stevison LS, Kohn MH. 2009. Divergence population genetic analysis of hybridization between rhesus and cynomolgus macaques. *Molecular ecology*. 18(11):2457–2475.
- Stevison LS, Woerner AE, Kidd JM, Kelley JL, Veeramah KR, McManus KF, Bustamante CD, Hammer MF, Wall JD. 2016. The Time Scale of Recombination Rate Evolution in Great Apes. *Molecular biology and evolution*. 33(4):928–945.
- Stöver BC, Müller KF. 2010. TreeGraph 2. Combining and visualizing evidence from different phylogenetic analyses. *BMC bioinformatics*. 11:7.
- Strachan T, Read AP. 2005. Molekulare Humangenetik, Heidelberg: Elsevier Spektrum Akad. Verl.
- Strathern JN, Shafer BK, McGill CB. 1995. DNA synthesis errors associated with double-strand-break repair. *Genetics*. 140(3):965–972.
- Sun F, Fujiwara Y, Reinholdt LG, Hu J, Saxl RL, Baker CL, Petkov PM, Paigen K, Handel MA. 2015. Nuclear localization of PRDM9 and its role in meiotic chromatin modifications and homologous synapsis. *Chromosoma*. 124(3):397–415.
- Takano-Shimizu T. 2001. Local changes in GC/AT substitution biases and in crossover frequencies on Drosophila chromosomes. *Molecular biology and evolution*. 18(4):606–619.
- Tan X, Xu X, Elkenani M, Smorag L, Zechner U, Nolte J, Engel W, Pantakani DVK. 2013. Zfp819, a novel KRAB-zinc finger protein, interacts with KAP1 and functions in genomic integrity maintenance of mouse embryonic stem cells. *Stem cell research*. 11(3):1045–1059.
- Tan YS, Kim M, Kingsbury TJ, Civin CI, Cheng W-C. 2014. Regulation of RAB5C is important for the growth inhibitory effects of MiR-509 in human precursor-B acute lymphoblastic leukemia. *PLoS one*. 9(11):e111777.
- Tang W, Liang P. 2019. Comparative genomics analysis reveals high levels of differential DNA transposition among primates. bioRxiv 520387; doi: <https://doi.org/10.1101/520387>.
- The 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature*. 526(7571):68–74.
- The International HapMap Consortium. 2005. A haplotype map of the human genome. *Nature*. 437(7063):1299–1320.
- Thibault-Sennett S, Yu Q, Smagulova F, Cloutier J, Brick K, Camerini-Otero RD, Petukhova GV. 2018. Interrogating the Functions of PRDM9 Domains in Meiosis. *Genetics*. 209(2):475–487.
- Thomas JH, Emerson RO, Shendure J. 2009. Extraordinary molecular evolution in the PRDM9 fertility gene. *PLoS one*. 4(12):e8505.
- Thomas JH, Schneider S. 2011. Coevolution of retroelements and tandem zinc finger genes. *Genome research*. 21(11):1800–1812.
- Thomas NS, Ennis S, Sharp AJ, Durkie M, Hassold TJ, Collins AR, Jacobs PA. 2001. Maternal sex chromosome non-disjunction. Evidence for X chromosome-specific risk factors. *Human molecular genetics*. 10(3):243–250.
- Thomson JP, Skene PJ, Selfridge J, et al. 2010. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*. 464(7291):1082–1086.

- Tian H, Billings T, Petkov PM. 2018. CXXC1 is not essential for normal DNA double-strand break formation and meiotic recombination in mouse. *PLoS genetics*. 14(10):e1007657.
- Tosi AJ, Morales JC, Melnick DJ. 2002. Y-Chromosome and Mitochondrial Markers in *Macaca fascicularis* Indicate Introgression with Indochinese *M. mulatta* and a Biogeographic Barrier in the Isthmus of Kra. *Int J Primatol*. 23(1):161–178.
- Tosi AJ, Morales JC, Melnick DJ. 2003. Paternal, maternal, and biparental molecular markers provide unique windows onto the evolutionary history of macaque monkeys. *Evolution; international journal of organic evolution*. 57(6):1419–1435.
- van Esch H, Hollanders K, Badisco L, et al. 2005. Deletion of VCX-A due to NAHR plays a major role in the occurrence of mental retardation in patients with X-linked ichthyosis. *Human molecular genetics*. 14(13):1795–1803.
- van Schaik CP, van Noordwijk MA. 1985. Interannual variability in fruit abundance and the reproductive seasonality in Sumatran Long-tailed macaques (*Macaca fascicularis*). *Journal of Zoology*. 206(4):533–549.
- Vervoort M, Meulemeester D, Béhague J, Kerner P. 2016. Evolution of Prdm Genes in Animals. Insights from Comparative Genomics. *Molecular biology and evolution*. 33(3):679–696.
- Wacholder AC, Pollock DD. 2019. PRDM9 and an Epidemic of Gene Conversion and Non-Homologous Recombination among Alu Elements in Ancestral Gorillas. bioRxiv 241356; doi: <https://doi.org/10.1101/241356>.
- Watanabe T, Cheng E-c, Zhong M, Lin H. 2015. Retrotransposons and pseudogenes regulate mRNAs and lncRNAs via the piRNA pathway in the germline. *Genome research*. 25(3):368–380.
- Watanabe Y, Kim HS, Castoro RJ, et al. 2009. Sensitive and specific detection of early gastric cancer with DNA methylation analysis of gastric washes. *Gastroenterology*. 136(7):2149–2158.
- Watanabe Y, Toyota M, Kondo Y, et al. 2007. PRDM5 identified as a target of epigenetic silencing in colorectal and gastric cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 13(16):4786–4794.
- Weinbauer GF, Niehoff M, Niehaus M, Srivastav S, Fuchs A, van Esch E, Cline JM. 2008. Physiology and Endocrinology of the Ovarian Cycle in Macaques. *Toxicologic pathology*. 36(7S):7S-23S.
- Williams EJ, Hurst LD. 2000. The proteins of linked genes evolve at similar rates. *Nature*. 407(6806):900–903.
- Winckler W, Myers SR, Richter DJ, et al. 2005. Comparison of fine-scale recombination rates in humans and chimpanzees. *Science (New York, N.Y.)*. 308(5718):107–111.
- Wolf D, Goff SP. 2009. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature*. 458(7242):1201–1204.
- Wolf G, Greenberg D, Macfarlan TS. 2015. Spotting the enemy within. Targeted silencing of foreign DNA in mammalian genomes by the Krüppel-associated box zinc finger protein family. *Mobile DNA*. 6:17.
- Woodward EL, Olsson ML, Johansson B, Paulsson K. 2014. Allelic variants of PRDM9 associated with high hyperdiploid childhood acute lymphoblastic leukaemia. *British journal of haematology*. 166(6):947–949.
- Wu H, Mathioudakis N, Diagouraga B, Dong A, Dombrowski L, Baudat F, Cusack S, Massy B de, Kadlec J. 2013. Molecular basis for the regulation of the H3K4 methyltransferase activity of PRDM9. *Cell reports*. 5(1):13–20.
- Wu Y, Ferguson JE, Wang H, et al. 2008. PRDM6 is enriched in vascular precursors during development and inhibits endothelial cell proliferation, survival, and differentiation. *Journal of molecular and cellular cardiology*. 44(1):47–58.
- Yamada S, Kim S, Tischfield SE, Jasin M, Lange J, Keeney S. 2017. Genomic and chromatin features shaping meiotic double-strand break formation and repair in mice. *Cell cycle (Georgetown, Tex.)*. 16(20):1870–1884.
- Yamamoto T, Isomura M, Xu Y, et al. 2006. PTPN11, RAS and FLT3 mutations in childhood acute lymphoblastic leukemia. *Leukemia research*. 30(9):1085–1089.
- Yamamoto Y, Watanabe T, Hoki Y, et al. 2013. Targeted gene silencing in mouse germ cells by insertion of a homologous DNA into a piRNA generating locus. *Genome research*. 23(2):292–299.

- Yan G, Zhang G, Fang X, et al. 2011. Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nature biotechnology*. 29(11):1019–1023.
- Yoshida K, Kondoh G, Matsuda Y, Habu T, Nishimune Y, Morita T. 1998. The Mouse RecA -like Gene Dmc1 Is Required for Homologous Chromosome Synapsis during Meiosis. *Molecular cell*. 1(5):707–718.
- Yu A, Zhao C, Fan Y, et al. 2001. Comparison of human genetic and sequence-based physical maps. *Nature*. 409(6822):951–953.
- Zerbino DR, Achuthan P, Akanni W, et al. 2018. Ensembl 2018. *Nucleic acids research*. 46(D1):D754–D761.
- Zhang F, Potocki L, Sampson JB, et al. 2010. Identification of uncommon recurrent Potocki-Lupski syndrome-associated duplications and the distribution of rearrangement types and mechanisms in PTL5. *American journal of human genetics*. 86(3):462–470.
- Zhang Y, Liu T, Meyer CA, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome biology*. 9(9):R137.
- Zhou T, Guo Y, Zhou Z, Guo X, Sha J. 2015. An update of the macaque testis proteome. *Data in brief*. 5:95–98.
- Zhou T, Hu Z, Zhou Z, Guo X, Sha J. 2013. Genome-wide analysis of human hotspot intersected genes highlights the roles of meiotic recombination in evolution and disease. *BMC genomics*. 14:67.

7. Internetquellen

<http://crapome.org/>; 24.09.2019

<http://tree.bio.ed.ac.uk/software/figtree/>, Rambaut A. (2009); 24.09.2019

<http://quantpsy.org>, Preacher, K.J. (2001); 24.09.2019

<https://www.abcam.com/protocols/native-chromatin-immunoprecipitation-protocol/>; 24.09.2019

<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, Andrews S. (2010); 24.09.2019

https://www.ebi.ac.uk/Tools/psa/emboss_needle/; 24.09.2019

<https://www.encodeproject.org/data-standards/terms/>; 24.09.2019

<https://www.ncbi.nlm.nih.gov/books/NBK131777/>, Camacho C. (2013); 24.09.2019

<https://www.proteinatlas.org/ENSG00000126856-PRDM7/tissue>; 24.09.2019

<https://www.socscistatistics.com/>; 24.09.2019

<https://www.uniprot.org/uploadlists/>; 24.09.2019

8. Lebenslauf