Evolutionary Distance of Amino Acid Sequence Orthologs across Macaque Subspecies: Identifying Candidate Genes for SIV Resistance in Chinese Rhesus Macaques

Cody T. Ross[1,2]*, Morteza Roodgar[2,3,4], David Glenn Smith[1,2,3]

1 Department of Anthropology, University of California, Davis. Davis, United States of America, 2 Molecular Anthropology Laboratory, University of California, Davis. Davis, United States of America, 3 California National Primate Research Center, University of California, Davis. Davis, United States of America, 4 Graduate Group of Comparative Pathology, University of California, Davis. Davis, United States of America

* ctross@ucdavis.edu

Abstract

We use the Reciprocal Smallest Distance (RSD) algorithm to identify amino acid sequence orthologs in the Chinese and Indian rhesus macaque draft sequences and estimate the evolutionary distance between such orthologs. We then use GOanna to map gene function annotations and human gene identifiers to the rhesus macaque amino acid sequences. We conclude methodologically by cross-tabulating a list of amino acid orthologs with large divergence scores with a list of genes known to be involved in SIV or HIV pathogenesis. We find that many of the amino acid sequences with large evolutionary divergence scores, as calculated by the RSD algorithm, have been shown to be related to HIV pathogenesis in previous laboratory studies. Four of the strongest candidate genes for SIV_{mac} resistance in Chinese rhesus macaques identified in this study are *CDK*9, *CXCL*12, *TRIM*21, and *TRIM*32. Additionally, *ANKRD*30*A*, *CTSZ*, *GORASP*2, *GTF2H*1, *IL*13*RA*1, *MUC*16, *NMDAR*1, *Notch*1, *NT5M*, *PDCD*5, *RAD*50, and *TM*9*SF*2 were identified as possible candidates, among others. We failed to find many laboratory experiments contrasting the effects of Indian and Chinese orthologs at these sites on SIV_{mac} pathogenesis, but future comparative studies might hold fertile ground for research into the biological mechanisms underlying innate resistance to SIV_{mac} in Chinese rhesus macaques.

Introduction

Recent work has produced complete genome sequence data for both the Indian [1] and Chinese rhesus macaque [2]. The existence of these draft sequences plays a very crucial role in permitting comparative genomic study of the differentiation of Chinese and Indian rhesus macaque amino acid sequences.

Amino acid sequence divergence between Chinese and Indian rhesus macaques has the potential to influence phenotype, especially response to immune system challenge, as genes related to immune response are some of the most rapidly evolving across species [3–6]. Additionally, evolution of the regulatory regions of genes is also known to play a very significant role in immunity [7, 8]. Thus, heterogeneity in phenotypic response to pathogens is a function not only of what protein is expressed by a gene, but also a function of the timing [9, 10], tissue type/biological location [11], quantity [12], and interaction [12] of genes being expressed.

It has been shown that Chinese rhesus macaques generally have increased elite controller status and more frequent long-term non-progression to simian-AIDs relative to Indian rhesus macaques [13, 14]. Understanding the genetic sources of heterogeneity in immune response to SIV_{mac} between Chinese and Indian rhesus macaques may thus provide important insights into SIV_{mac} and HIV/AIDS biology more generally.

In this study, we focus on identifying genes related to HIV or SIV_{mac} whose amino acid sequences show high levels of divergence across Chinese and Indian rhesus macaques, as these genes may be strong candidates for the difference in immunity to SIV_{mac} across these subspecies. We do not investigate divergence in non-coding regulatory regions, although we believe such divergence to play a significant role in heterogeneity in immune response across subspecies. Including data from non-coding regions in our analysis would create significant methodological and computational difficulties at this point in time, but we believe that our methods will be amenable to such investigations as computational methods increase in power and efficiency, and genomic data increases in quantity and quality.

Previous comparative genomic investigations for evidence of selection [2] and high throughput single-nucleotide polymorphism (SNP) sequencing and linkage disequilibrium analysis [13] in Chinese and Indian rhesus macaques have found patterns consistent with positive selection on genes such as *TCTE*1, *ZNF*337, and *TRIM*5 [2], and *MAST*2, *ELAVL*4, and *HIVEP*3 [13]. The functional linkages between the gene candidates presented in these papers and resistance to SIV_{mac} in Chinese rhesus macaques still remain ambiguous, however, as further laboratory work is needed to investigate the relevance of Chinese versus Indian orthologs on SIV_{mac} pathogenesis.

In this analysis, we attempt to identify additional candidate genes for SIV_{mac} resistance and test if we can re-identify previously described candidate genes using a new methodology. Our methods are similar at first pass to those used by Yan et al [2], in that we attempt to identify orthologous amino acid sequences between Chinese and Indian rhesus macaques using complete draft sequence data and use synteny mapping to evaluate the performance of the ortholog classifications. Our methods differ in that we do not infer orthology on the basis of synteny, but instead use a Python implementation of the Reciprocal Smallest Distance (RSD) algorithm [15] to evaluate orthology and then use synteny mapping to check the performance of the RSD algorithm.

The RSD algorithm functions to detect putative orthologs using sequence alignment in a way similar to the reciprocal best hit algorithm (RBH) [16]. The RSD algorithm, however, works around a shortcoming of the RBH algorithm that occurs when a forward blast yields a paralog best hit, but a reciprocal blast recovers an ortholog; in such cases, the RBH alogorithm excludes both pairs, but the RSD algorithm can often recover the true ortholog [15]. The RSD algorithm accomplishes this by conducting a forward blast of an amino acid sequence, *i*, from a draft sequence, *I*, onto a secondary draft sequence, *J*, to obtain a list of hits, *H*. Instead of simply checking for reciprocal best hits, the RSD algorithm moves on to compare each hit, *h*, meeting threshold criteria against the query sequence, *i*. A maximum likelihood estimate, *e*, of the evolutionary distance (substitutions per site) separating *h* and *i* is calculated, given an empirical

amino acid substitution rate matrix [15]. Note that these numbers might appear surprisingly large if the focal amino acid sequences are of different lengths. For instance, the inferred orthologous amino acid sequences ENSMMUP00000038625 and ENSP00000407071 (GO identifier RAD50) are of very different lengths, increasing the associated e value to 2.2, even though the percent identify matrix from Clustal 2.1 shows a small but plausible sequence identity of 39.47 percent. Of all considered sequences in H, only the sequence with the smallest evolutionary distance is retained; this sequence, j, is then used for a reciprocal blast against the first draft sequence, I [15]. Hits from this blast are treated analogously, and an orthologous pair is considered to be found if and only if sequences i and j are the sequences with reciprocal smallest evolutionary distances [15].

An attractive feature of the RSD algorithm is that it provides estimates of evolutionary distance [17] between each pair of othologs. Given that the Indian and Chinese rhesus macaque sub-populations are estimated to have separated only about 160,000 years ago [18], a large majority of the genome is expected to be relatively constant across sub-populations. Thus, large differences in amino acid sequence between orthologous pairs in Chinese and Indian rhesus macaques are notable. Given that bacteria or viruses with structural, functional, or morphological similarities to SIV_{mac} stand to place selective pressure on genes which might confer resistance or even elite controller status to such infections, and Chinese and Indian rhesus macaques show different normative phenotypes in relation to SIV_{mac} pathogenesis, orthologs with negligible ($\sim e < 0.05$) evolutionary distance (about 15,448 of the 17,064 amino acid sequences considered in this study) across Chinese and Indian rhesus macaques are unlikely to be responsible for differences in resistance to SIV_{mac}, conditional on the assumption that coding region variation is the key driver of this phenotypic difference. Orthologs with non-negligible evolutionary distance across Chinese and Indian rhesus macaques are candidates for genes that might be diverging due to a disjunctive array of demographic and selection pressures, of which immune system challenge is only a single example.

To narrow down which of these amino acid sequences might contribute to SIV_{mac} resistance in Chinese rhesus macaques, we used GOanna [19] to associate gene symbols (human homologs) and gene function annotations to each rhesus macaque amino acid sequence, and then cross-tabulated the portion of our candidate list with the highest divergence/evolutionary distance scores (492 genes in total with divergence scores $e > 0.15$) with a list of 140 genes known to be involved in HIV-1 pathogenesis from a thorough literature review [20], a list of 169 candidate genes for HIV-1 susceptibility derived from a large-scale genome-wide common-disease common-variant analysis of HIV-1 in Africa [21], and four genome-wide siRNA analyses [22–25].

Additionally, we conducted a literature review of the 25 orthologs with the largest divergence/evolutionary distance scores from our RSD analysis in order to investigate if any of these genes had been associated with HIV or SIV pathogenesis by laboratory studies not cited in the previous literature reviews or association studies [20].

We complete our analysis by: 1) modeling ortholog divergence scores as a function of chromosome and chromosomal location (on the Indian rhesus macaque reference frame) to investigate if there are any distinct chromosomal locations with high concentrations of divergent orthologs, and 2) investigating divergence scores as a function of gene ontology classification categories.

While many disjunctive ecological differences may have led to divergent amino acid sequence evolution between Chinese and Indian rhesus macaques, researchers are especially interested in understanding how divergent evolution in each sub-population has led one sub-population (Chinese rhesus macaques) to become resistant to the simian analog of HIV [13, 14]. While careful laboratory studies are required to uncover the role of specific genes (and

their interactions) on SIV_{mac} and HIV resistance, wide-scale comparative genomic scans may provide useful insights and substantially reduce the list of candidate genes to be investigated with more rigorous laboratory methods; uncovering the functional and biological significance of genetic variation is made easier by integrating the partial strengths of genomic, evolutionary, and biochemical data and methods [26].

Results

Establishing Orthology between Chinese and Indian Rhesus Macaque Reference Sequences

Supplementary S1 Table includes the full list of Chinese and Indian amino acid sequence orthologs, paired with the location information of each ortholog on each draft sequence, the estimated divergence score from the RSD algorithm, and the associated gene symbols from GOanna.

Fig 1 plots the synteny maps implied by the ortholog classification of the RSD algorithm, dropping a small number of amino acid sequences that the RSD algorithm believes to have undergone inter-chromosomal transposition. These maps show that the vast majority of orthologous pairs are contained in expansive blocks of synteny. We observe only a small number of apparent intra-chromosomal transpositions per chromosome, the majority of which are
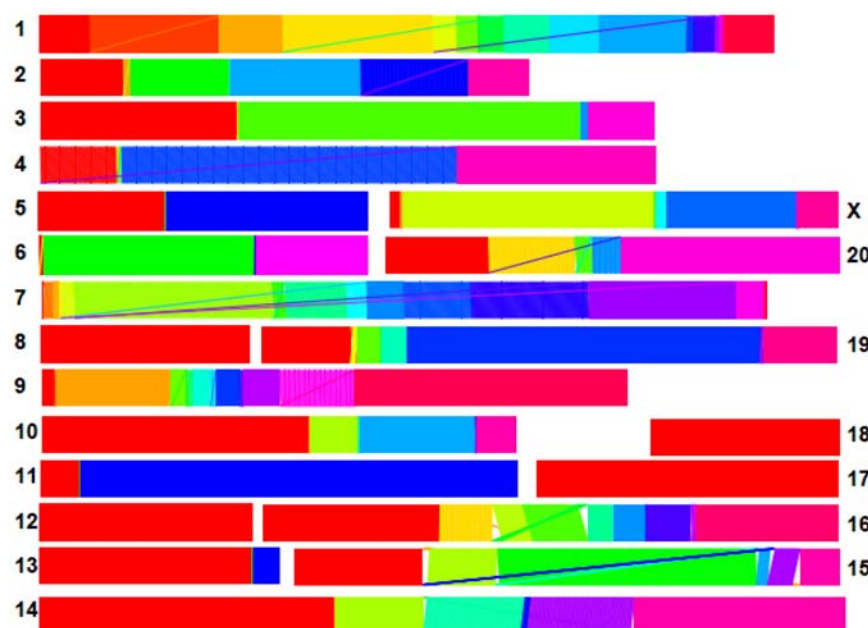

Fig 1. Synteny maps between Chinese and Indian rhesus macaque reference sequences. The large-scale concordance serves to verify that the RSD algorithm is functioning to accurately map orthologous sequences. Each colored line maps an amino acid sequence in the Indian rhesus macaque draft sequence (top) to its corresponding ortholog on the Chinese rhesus macaque draft sequence (bottom). Diagonal lines illustrate transpositions; thin diagonal lines are indicative of single amino acid transposition, and thicker diagonal lines are indicative of block transpositions. The differing colors on each chromosome serve to illustrate the blocks in which synteny is fully conserved. Most chromosomes show almost fully conserved synteny, but chromosomes 15 and 16 show fairly large-scale rearrangement.

doi:10.1371/journal.pone.0123624.g001

inversions of neighboring amino acid sequences. On chromosomes 15, 16, and 19 we observed small intra-chromosomal block transpositions.

Identifying the Orthologs with Highest Divergence between Chinese and Indian Rhesus Macaques

We cross-tabulated our candidate list of 492 genes with divergence values of $e > 0.15$ with several lists of genes (Supplementary S2 Table) suspected to be involved in HIV or SIV pathogenesis from thorough literature review [20], association studies [21], or genome-wide siRNA analyses [22–25].

Using this methodology we identified five genes (*CDK*9, *CXCL*12, *LPL*, *TRIM*32, and *TRIM*21) from the set of genes described by [20], five genes (*CXCL*12, *CYP26B*1, *EFR3A*, *NME*6, and *PIGX*) from the set of genes provided by [21], zero genes from the set of gene provided by [22], four genes (*BCAR*1, *DSP*, *ICAM*4, and *SLC35B*1) from the set of gene provided by [23], six genes (*DAPK*2, *PARVA*, *CTSZ*, *GTF2H*1, *LYPLA*1, and *GORASP*2) from the set of gene provided by [24], and three genes (*ANKRD*30A, *LPL*, and *TM9SF*2) from the set of genes provided by [25]. These twenty genes are listed in Table 1 with their divergence scores and identifiers on the Indian and Chinese rhesus macaque reference frames.

Additionally, we conducted our own literature review of the 20 genes with the largest divergence scores from our RSD analysis; Table 2 presents our results. We find that at least 7 of the

Table 1. Table 1 presents the 20 genes that have been linked to HIV pathogenesis in previous studies [20–25] and also have highly divergent orthologs across Chinese and Indian rhesus macaques. The label "GOname" indicates the gene name assigned by GOanna. The labels "ProteinID.IR" and "ProteinID.CR" indicate the name of the amino acid sequence in the Indian and Chinese rhesus macaque draft sequence files, respectively. The label "*e*" is the estimate of evolutionary distance produced by the RSD alogirthm via implention of PAML [17]. Finally, the labels "Chr.IR" and "Chr.CR" indicate the chromosome on which the amino acid sequences occur in the Indian and Chinese rhesus macaque draft sequence files, respectively.

GOname	ProteinID.IR	ProteinID.CR	*e*	Chr.IR	Chr.CR	HIV-Related	Reference
TRIM32	ENSMMUP00000017213	ENSP00000345464	2.4021	15	4	✓	[20]
ICAM4	ENSMMUP00000012237	ENSP00000342114	0.7103	19	19	✓	[22, 23]
BCAR1	ENSMMUP00000032314	ENSP00000391669	0.51	20	20	✓	[23]
TRIM21	ENSP00000339355	ENSP00000381338	0.4936	14	14	✓	[20]
CXCL12	ENSMMUP00000035953	ENSP00000379140	0.4456	9	9	✓	[20, 21]
DAPK2	ENSMMUP00000023902	ENSP00000408277	0.4175	7	7	✓	[24]
CTSZ	ENSMMUP00000020183	ENSP00000217131	0.3892	10	10	✓	[24]
PARVA	ENSMMUP00000027577	ENSP00000334008	0.2632	14	14	✓	[24]
GORASP2	ENSMMUP00000020692	ENSMMUP00000002335	0.2488	16	16	✓	[24]
EFR3A	ENSMMUP00000016487	ENSMMUP00000016487	0.2401	8	8	✓	[21]
PIGX	ENSMMUP00000005029	ENSP00000296333	0.2396	2	2	✓	[21]
DSP	ENSMMUP00000038586	ENSP00000400669	0.2286	11	11	✓	[23]
TM9SF2	ENSMMUP00000025694	ENSP00000365567	0.196	17	17	✓	[25]
GTF2H1	ENSMMUP00000041182	ENSMMUP00000023608	0.1926	18	14	✓	[24]
NME6	ENSMMUP00000004110	ENSP0000416658	0.1853	2	2	✓	[21]
LPL	ENSMMUP00000006264	ENSP00000309757	0.184	8	8	✓	[20, 25]
ANKRD30A	ENSMMUP00000005876	ENSP00000386398	0.1816	12	12	✓	[25]
CYP26B1	ENSMMUP00000015927	ENSP00000001146	0.1762	13	13	✓	[21]
LYPLA1	ENSMMUP00000017285	ENSP00000397807	0.1729	8	5	✓	[24]
SLC35B1	ENSMMUP00000034903	ENSP00000409548	0.1599	16	16	✓	[23]
CDK9	ENSMMUP00000017333	ENSP00000362362	0.1504	15	15	✓	[20]

doi:10.1371/journal.pone.0123624.t001

Table 2. Table 2 presents the 20 orthologs with the highest levels of divergence according to the RSD algorithm. Of these 20 genes, 7 have been linked to HIV-1 pathogenesis in some way by previous research. The label "GOname" indicates the gene name assigned by GOanna. The labels "ProteinID. IR" and "ProteinID.CR" indicate the name of the amino acid sequence in the Indian and Chinese rhesus macaque draft sequence files, respectively. The label "e" is the estimate of evolutionary distance produced by the RSD alogirthm via implentation of PAML [17]. Finally, the labels "Chr.IR" and "Chr.CR" indicate the chromosome on which the amino acid sequences occur in the Indian and Chinese rhesus macaque draft sequence files, respectively.

GOname	ProteinID.IR	ProteinID.CR	e	Chr.IR	Chr.CR	HIV Related	Reference
TRIM32	ENSMMUP00000017213	ENSP00000345464	2.4021	15	4	✓	[33, 38, 41, 88, 89]
RAD50	ENSMMUP00000038625	ENSP00000407071	2.2032	2	2	✓	[65]
KLHL42	ENSP00000321544	ENSP00000379034	1.4254	8	10	-	-
NMDAR1	ENSMMUP00000009665	ENSP00000381630	1.3902	8	8	✓	[66, 90, 91]
Adcy10	ENSMMUP00000027891	ENSP00000410352	1.2541	4	4	-	-
MUC16	ENSMMUP00000020314	ENSP00000381008	1.1982	19	19	✓	[69, 70]
BCMO1	ENSMMUP00000000974	ENSP00000258168	1.1973	20	20	-	-
IL13RA1	ENSMMUP00000015103	ENSP00000360730	1.1112	X	X	✓	[74]
NT5M	ENSMMUP00000035155	ENSP00000390695	1.0756	16	16	✓	[75]
Prss27	ENSMMUP00000020977	ENSP00000161006	1.0391	20	20	-	-
YY2	ENSMMUP00000029555	ENSP00000368798	1.0059	X	X	-	-
Sec22b	ENSMMUP00000015200	ENSP00000362504	0.9357	4	4	?	[92, 93]
FAM110A	ENSP00000363386	ENSP00000363386	0.9143	1	1	?	[94]
Notch1	ENSMMUP00000036150	ENSP00000398674	0.901	13	13	✓	[76, 95]
RBM38	ENSMMUP00000038172	ENSMMUP00000014941	0.887	10	4	-	-
GRIN3A	ENSMMUP00000021930	ENSP00000234389	0.8687	15	19	-	-
CCDC176	ENSMMUP00000016656	ENSP00000377577	0.817	7	7	-	-
FBXO45	ENSMMUP00000038210	ENSP00000310332	0.8117	2	2	-	-
FRS2	ENSMMUP00000039208	ENSP00000381083	0.8032	11	11	-	-
PDCD5	ENSMMUP00000020187	ENSP00000388543	0.7491	19	19	✓	[78]

doi:10.1371/journal.pone.0123624.t002

20 genes with the highest divergence between the Chinese and Indian orthologs have been linked to HIV or SIV pathogenesis in some way by previous studies.

Identifying Evolutionary Divergence between Chinese and Indian Rhesus Macaque Orthologs as a Function of Chromosomal Region

Figs 2–4 present estimates of evolutionary distance between the Chinese and Indian amino acid sequences as a function of chromosomal location in megabases (Mb). Amino acid sequence divergence levels appear to show weak signs of concentration into regions on a chromosome, although the extent to which this pattern holds may be confounded by the clustering of amino acid sequences themselves into specific regions on each chromosome. To better estimate chromosomal position effects, we use a multi-level, zero-inflated gamma regression model, with random effects on chromosomal position generated using a Gaussian process. Across chromosomes, we see very little evidence for strong chromosomal position effects, as there is very little difference in the estimated probability of a non-zero evolutionary distance or the mean value of evolutionary distance as a function of chromosomal position. These results are plotted in Figs 5–8.

We did observe significant heterogeneity in mean evolutionary distance across chromosomes, with some (e.g. 5, 6, and 17) showing reduced amino acid sequence divergence and others (e.g. 8, 13, 19, and 20) showing increased evolutionary divergence. These results are plotted in Figs 9 and 10.
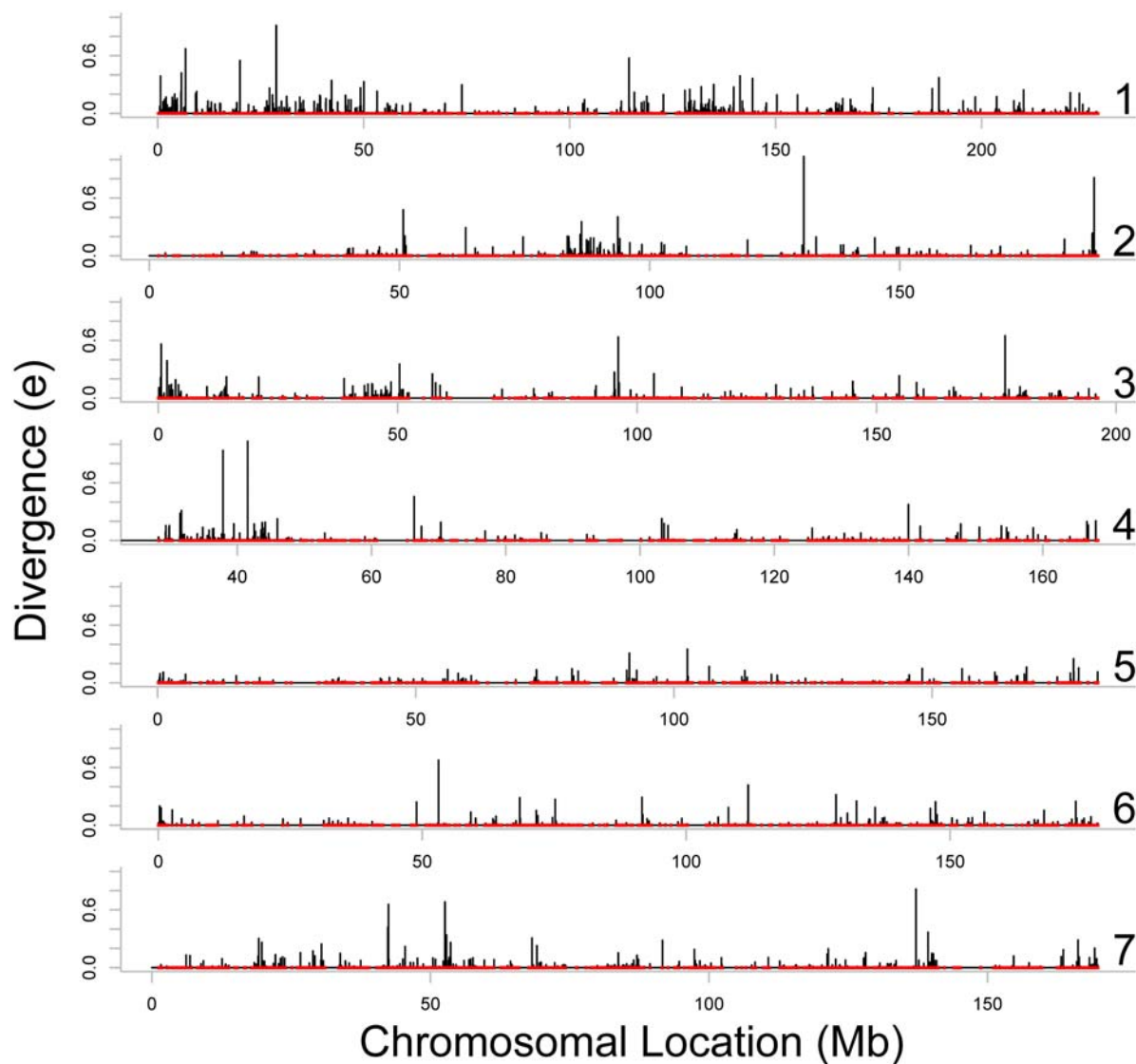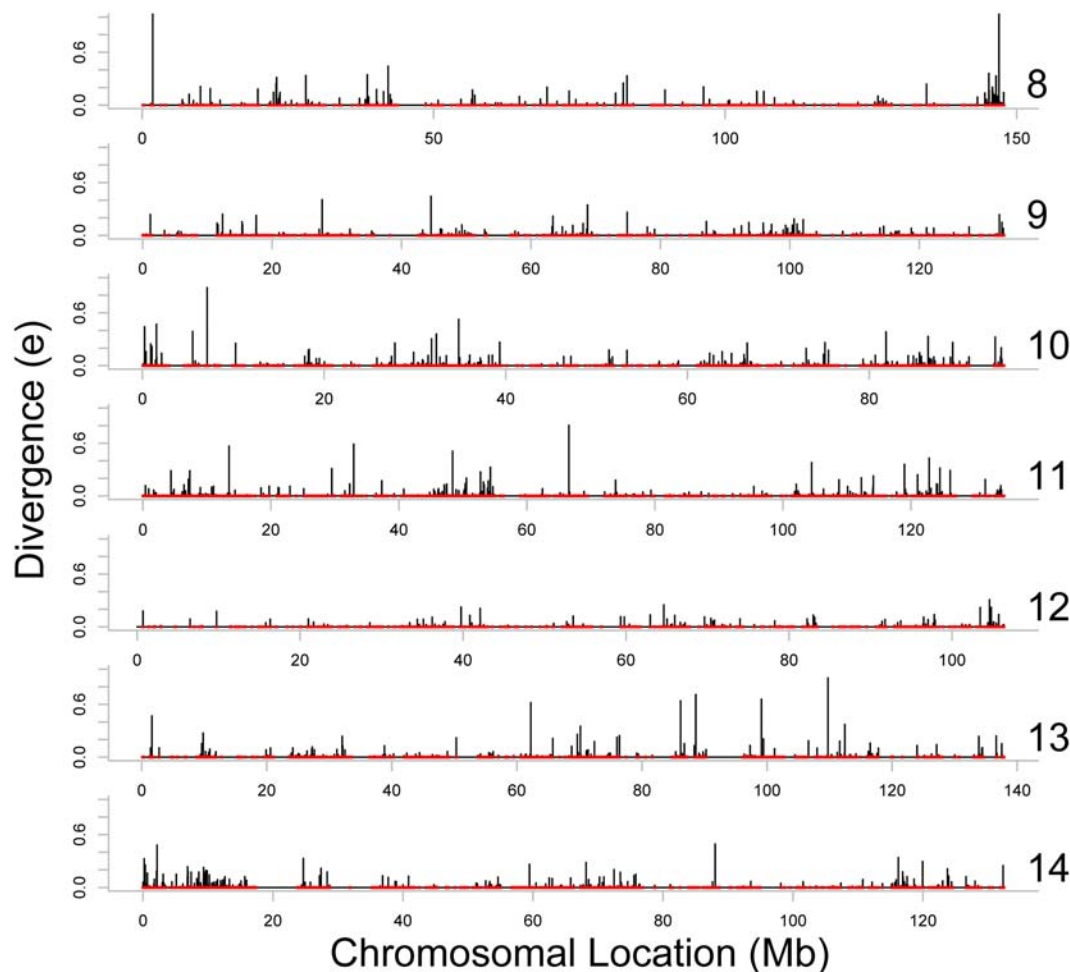
Fig 2. Evolutionary distance between Indian and Chinese rhesus macaque orthologs, localized on the Indian rhesus macaque draft sequence (Chromosomes 1 through 7). The red points illustrate the locations of the orthologs (including the orthologs with $e = 0$) on the horizontal axis, and the black lines represent the divergence scores (e) of these orthologs on the vertical axis. The area of elevated evolutionary distance on the beginning portion of Chromosome 1 corresponds to an area with significant differences in linkage disequilibrium between Indian-derived and Chinese rhesus macaques [13].

doi:10.1371/journal.pone.0123624.g002

Identifying the Gene Function Classes with Highest Evolutionary Divergence between Chinese and Indian Rhesus Macaques

To investigate the functional roles of genes associated with increased ortholog divergence scores, we used GOanna to link gene annotations with ortholog pairs. Table 3 presents the 20 gene function classes (of the 176 containing at least 100 genes) with highest median divergence scores across the rhesus macaque genome. Supplementary S3 Table contains the complete annotation list paired with ortholog divergence scores. Genes involved in viral processes and immunity show elevated levels of evolutionary divergence.
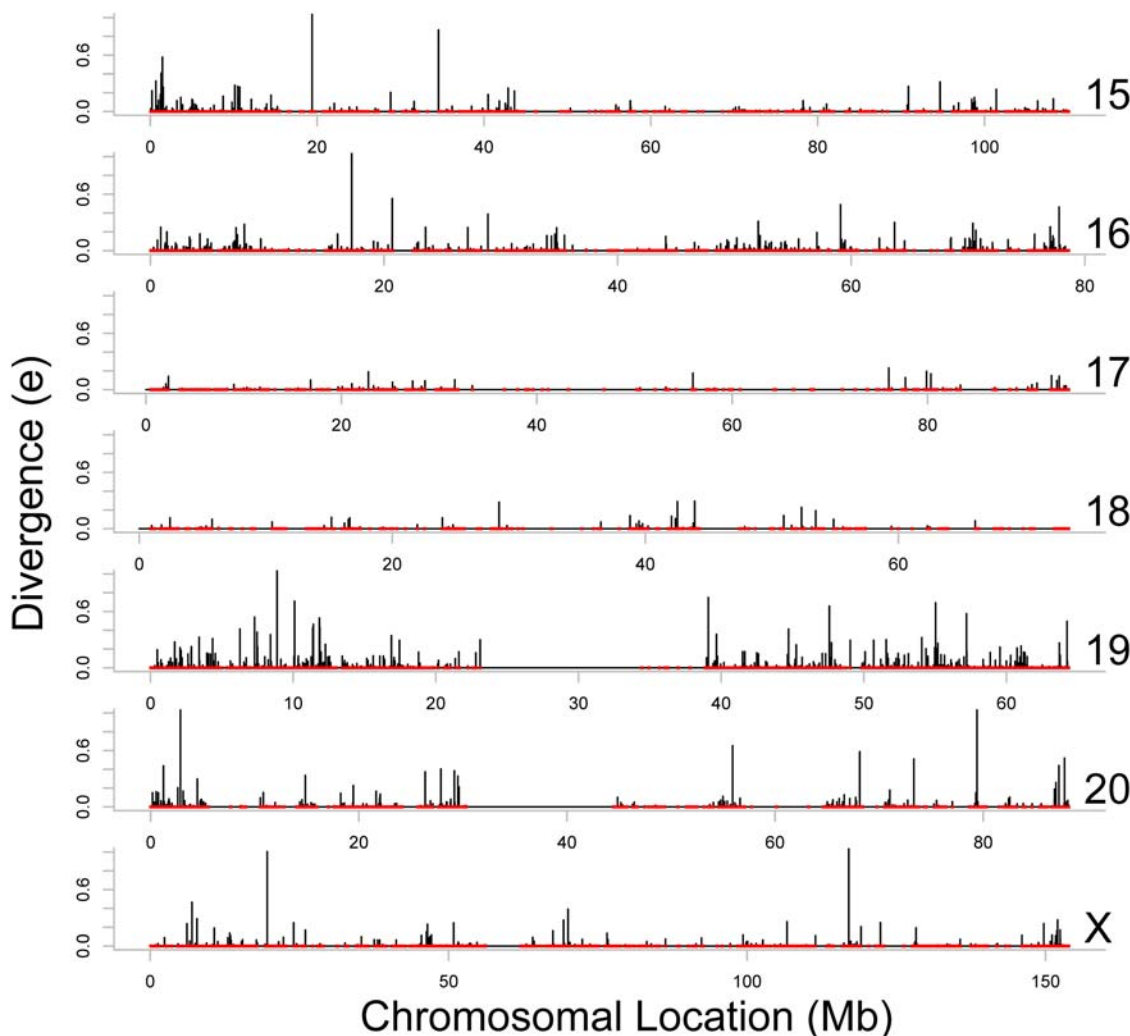
Fig 3. Evolutionary distance between Indian and Chinese rhesus macaque orthologs, localized on the Indian rhesus macaque draft sequence (Chromosomes 8 through 14). The red points illustrate the locations of the orthologs (including the orthologs with *e* = 0) on the horizontal axis, and the black lines represent the divergence scores (*e*) of these orthologs on the vertical axis. The end region of chromosome 8 shows an area of high divergence.

Discussion

RSD Performance

The RSD algorithm appears to function with high accuracy in identifying orthologs, as synteny maps show large blocks of gene order conservation, sometimes spanning the entire chromosome (e.g. on chromosomes 8, 12, 17, and 18). There is, however, evidence of broken synteny on some chromosomes due to single amino acid sequence transpositions (e.g. on chromosomes 1, 4, 7, 9 and 20), as well as block transpositions (e.g. on chromosomes 14, 15, 16, and 19). Additionally, RSD believed some amino acid sequences to have orthologs on different chromosomes.

Some researchers [2] opt to exclude proposed orthologs if the pair break blocks of synteny. We have not excluded any potential orthologs from our results based on synteny breaking, but we have marked (in Supplementary S1 Table) which potential orthologs break synteny or occur on different chromosomes (387 of 17,064 proposed orthologs showed such behavior). It

Fig 4. Evolutionary distance between Indian and Chinese rhesus macaque orthologs, localized on the Indian rhesus macaque draft sequence (Chromosomes 15 through 20, and X). The red points illustrate the locations of the orthologs (including the orthologs with $e = 0$) on the horizontal axis, and the black lines represent the divergence scores (e) of these orthologs on the vertical axis. Chromosome 19 shows consistently high evolutionary divergence.

doi:10.1371/journal.pone.0123624.g004

is possible that potential orthologs which break synteny are either 1) mis-classified as orthologs by RSD, or 2) poorly localized in the draft sequences. If (1) is the case, then our estimates of evolutionary distance are invalid; if (2) is the case, then our estimates of evolutionary distance are valid, but our localizations are invalid.

We, however, find it possible that many of the transpositions discovered by RSD are true transpositions. RSD uses global information and minimum distance criteria to map orthologs, so misclassifications are unlikely conditional on the global information being complete, amino acid sequences being large, and divergence times being small. Transpositions (both intra- and inter-chromosomal) and small inversions are known to occur during the evolutionary process, and are frequently seen in inter-species mappings [27]. It is extremely unlikely to see transposition of blocks in which synteny is conserved if the sequences in the transposed blocks are not true orthologs. In such cases, however, the appearance of transpositions or inversions may be due to errors in localization of orthologs on one or both of the reference sequences.
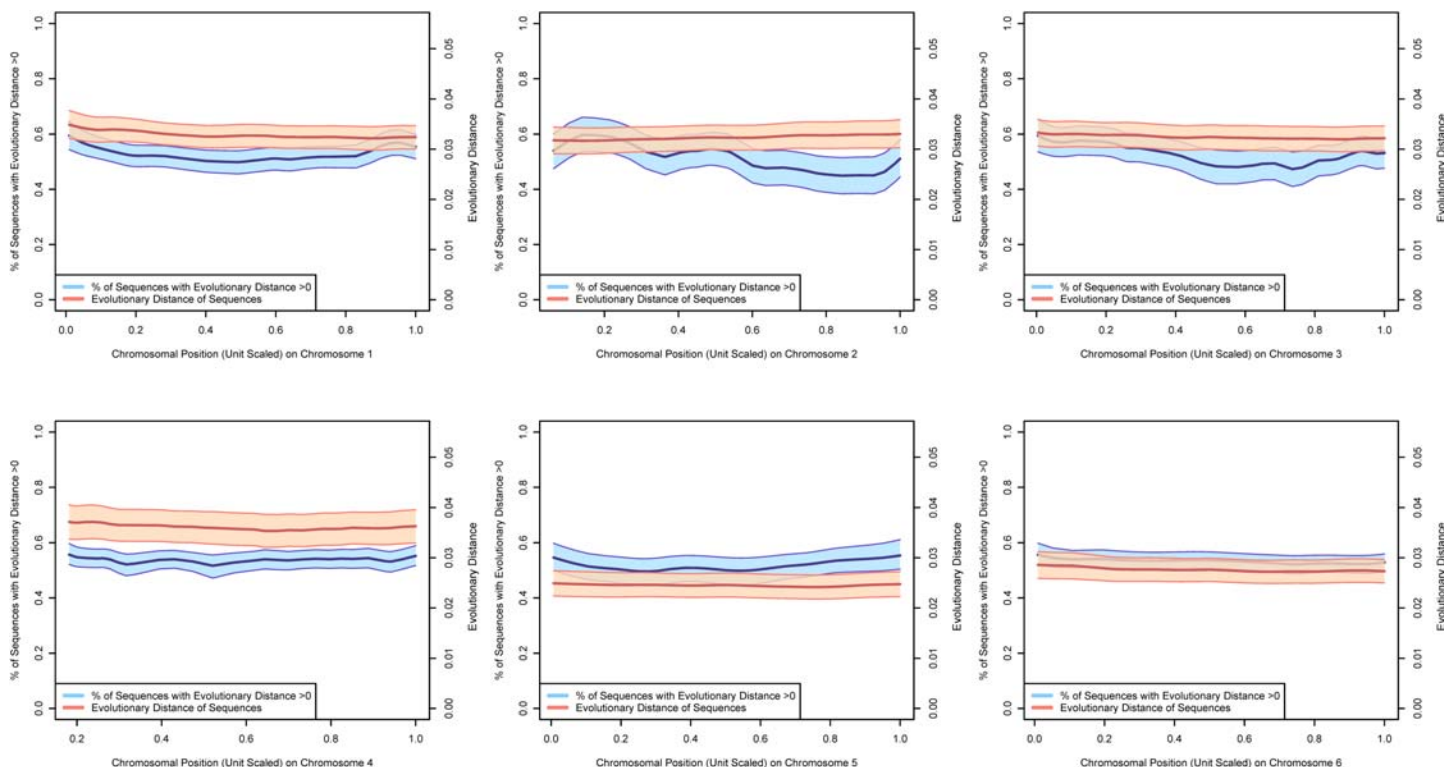
Fig 5. The 70% posterior credibility intervals from the zero-inflated gamma regression model for chromosomes 1–6. The blue confidence band plots the mean probability that an amino acid sequence is divergent as a function of chromosomal position on the left axis and the orange confidence band plots the mean evolutionary distance of the subset of diverging amino acid sequences as a function of chromosomal position on the right axis. Across all chromosomes, we note little difference in divergence of amino acid sequences as a function of chromosomal position, but some differences in mean evolutionary divergence across chromosomes.

Transpositions may be indicative of important inter-gene interactions due to chromosomal position effects [28], epigenetic effects [29], or simply lowered odds of recombination separating alleles that have non-additive or synergistic effects (see Maynard-Smith's argument for the evolutionary origin of chromosomes [30, 31]). By presenting the data and results without exclusion, we leave it to the reader to interpret the findings, while acknowledging where possible confounding methodological factors may inhibit strong inference.

Candidate Genes

We provide a brief and by no means exhaustive review of laboratory studies that address the functional relationships between the candidate genes for SIV_{mac} resistance uncovered in this study and SIV_{mac} or HIV pathogenesis.

Tripartite Motif Proteins 21 and 32 (*TRIM*21 and *TRIM*32). *TRIM*21 and *TRIM*32 are among the strongest candidate genes for SIV_{mac} resistance in Chinese rhesus macaques identified in this study, as they are some of the most divergent orthologs in the entire rhesus macaque genome and are known to play an important role in HIV-1 pathogenesis. These genes are members of a family of tripartite motif (TRIM) proteins that are known to be involved in a variety of cellular functions, including differentiation, apoptosis, and immunity [32]; recently, a number of TRIM proteins have been found to display anti-retroviral activities, and have been
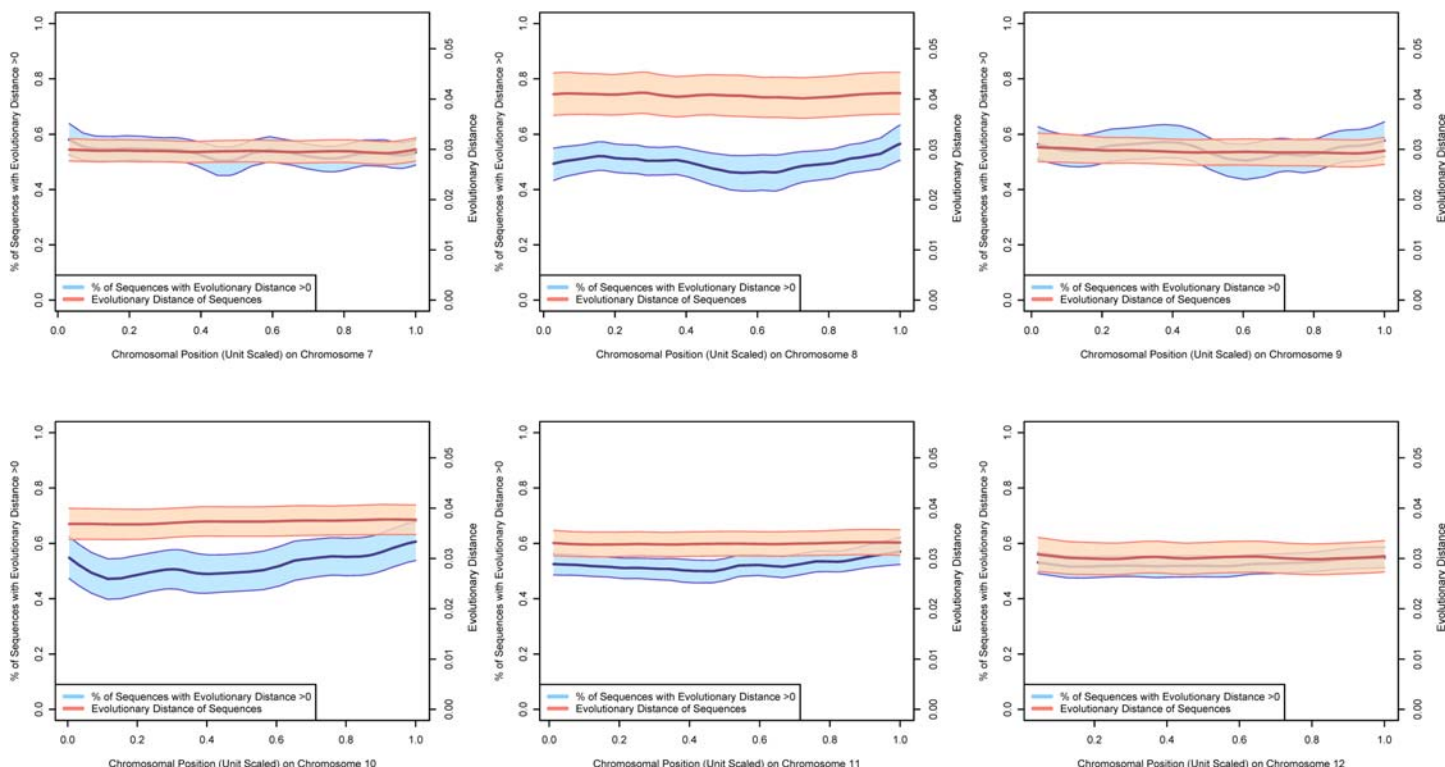
Fig 6. The 70% posterior credibility intervals from the zero-inflated gamma regression model for chromosomes 7–12. The blue confidence band plots the mean probability that an amino acid sequence is divergent as a function of chromosomal position on the left axis and the orange confidence band plots the mean evolutionary distance of the subset of diverging amino acid sequences as a function of chromosomal position on the right axis. Across all chromosomes, we note little difference in divergence of amino acid sequences as a function of chromosomal position, but some differences in mean evolutionary divergence across chromosomes.

doi:10.1371/journal.pone.0123624.g006

associated with innate immunity [32–34]. In our analysis, however, the amino acid sequences corresponding to *TRIM*32, were localized on different chromosomes, and thus have elevated odds of being an ortholog misclassification or being poorly localized in one of the draft sequences.

The exact role of *TRIM*32 in HIV-1 pathogensis is still unclear, as some studies have found that it functions to repress viral gene expression [33], affect trafficking of viral protein [34, 35], or attenuate transcription [36, 37], while others have found that it affected HIV release from cells, but not viral gene expression [38].

*TRIM*32 has been shown to modulate interferon production and cellular antiviral responses by targeting *MITA/STING* [39]. In addition, *TRIM*32 plays a key role in *TNF-α* induced apoptosis [40].

A human study found significant difference in *TRIM*32 expression between natural elite controllers and anti-retroviral theory (ART) suppressed individuals, with *TRIM*32 expression being elevated in the elite controllers [41]. The comparison between elite controllers and an ART suppressed 'control' group is justified by the authors because HIV-1 replication drives expression of many restriction factors (such as TRIM32) and a comparison of gene expression in two 'aviremic' yet infected groups (human elite controllers and ART suppressed individuals) minimizes the confounding differences in HIV-1 antigen levels that would be introduced by uninfected controls, or untreated individuals [41]. This study indicates that heightened
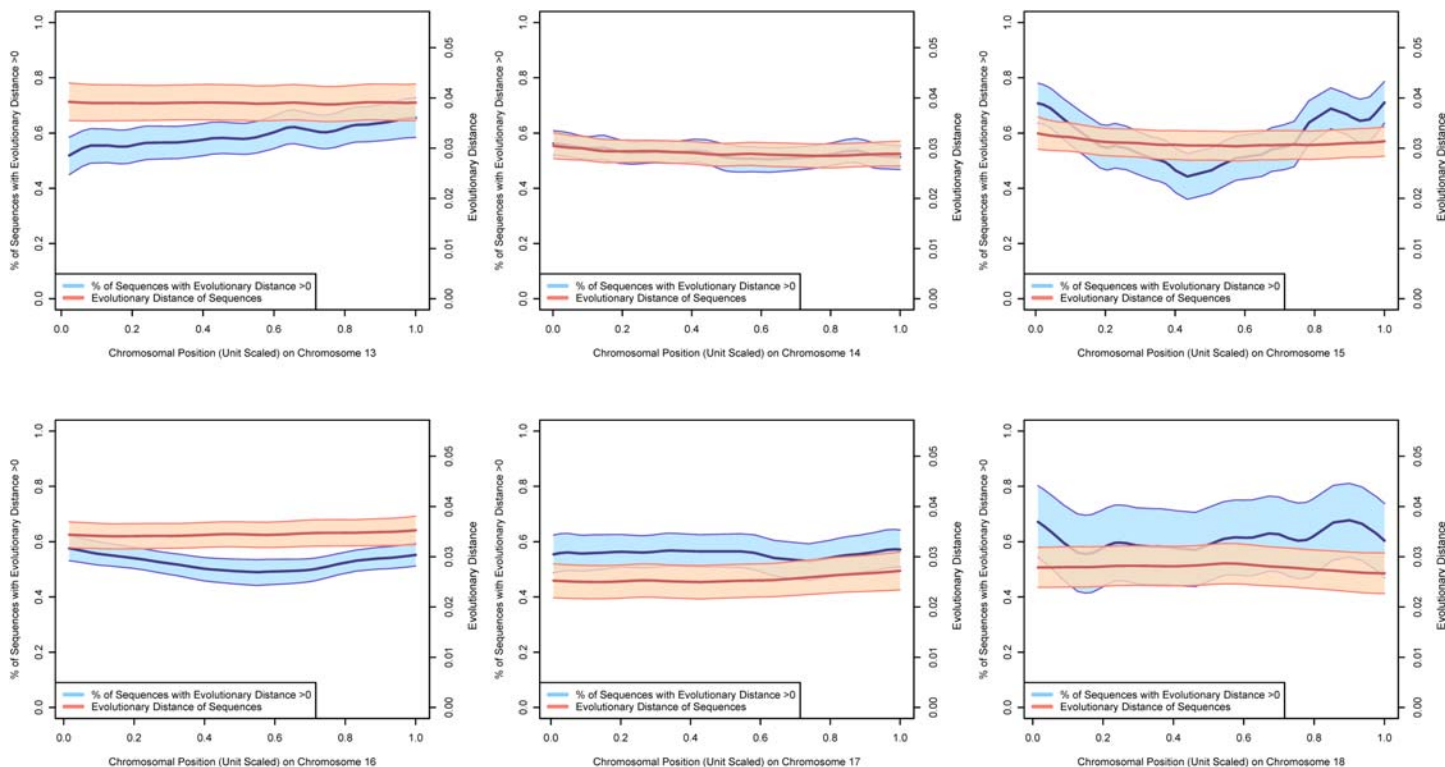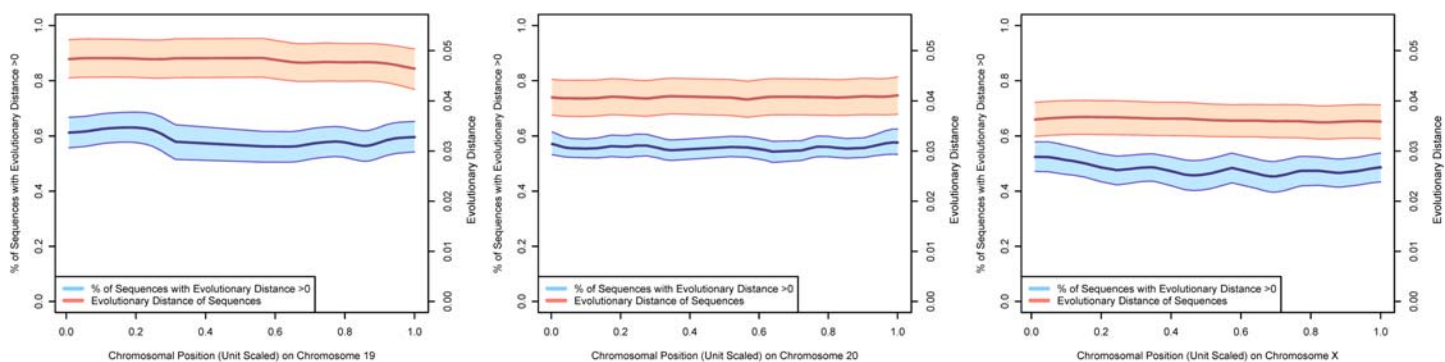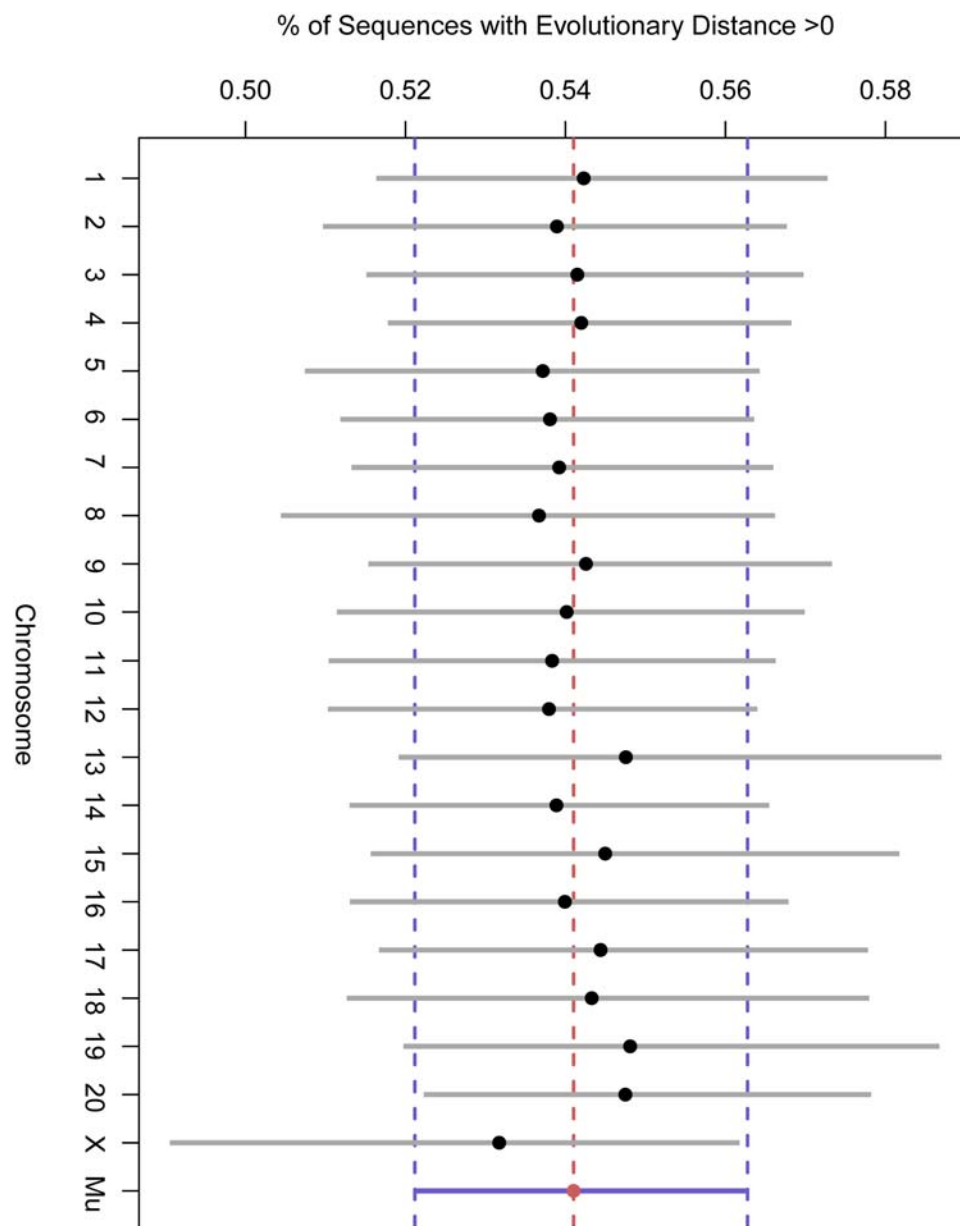
Fig 7. The 70% posterior credibility intervals from the zero-inflated gamma regression model for chromosomes 13–18. The blue confidence band plots the mean probability that an amino acid sequence is divergent as a function of chromosomal position on the left axis and the orange confidence band plots the mean evolutionary distance of the subset of diverging amino acid sequences as a function of chromosomal position on the right axis. Across all chromosomes, we note little difference in divergence of amino acid sequences as a function of chromosomal position, but some differences in mean evolutionary divergence across chromosomes.

doi:10.1371/journal.pone.0123624.g007

Fig 8. The 70% posterior credibility intervals from the zero-inflated gamma regression model for chromosomes 19–20 and X. The blue confidence band plots the mean probability that an amino acid sequence is divergent as a function of chromosomal position on the left axis and the orange confidence band plots the mean evolutionary distance of the subset of diverging amino acid sequences as a function of chromosomal position on the right axis. Across all chromosomes, we note little difference in divergence of amino acid sequences as a function of chromosomal position, but some differences in mean evolutionary divergence across chromosomes.

doi:10.1371/journal.pone.0123624.g008

Fig 9. The 90% posterior credibility intervals of the chromosomal random effects from the Bernoulli sub-model of the zero-inflated gamma regression. The value 'Mu' is the partially-pooled mean estimate across chromosomes. We see that there are no significant differences in the mean probabilty of non-zero divergence in amino acid sequences across chromosomes.
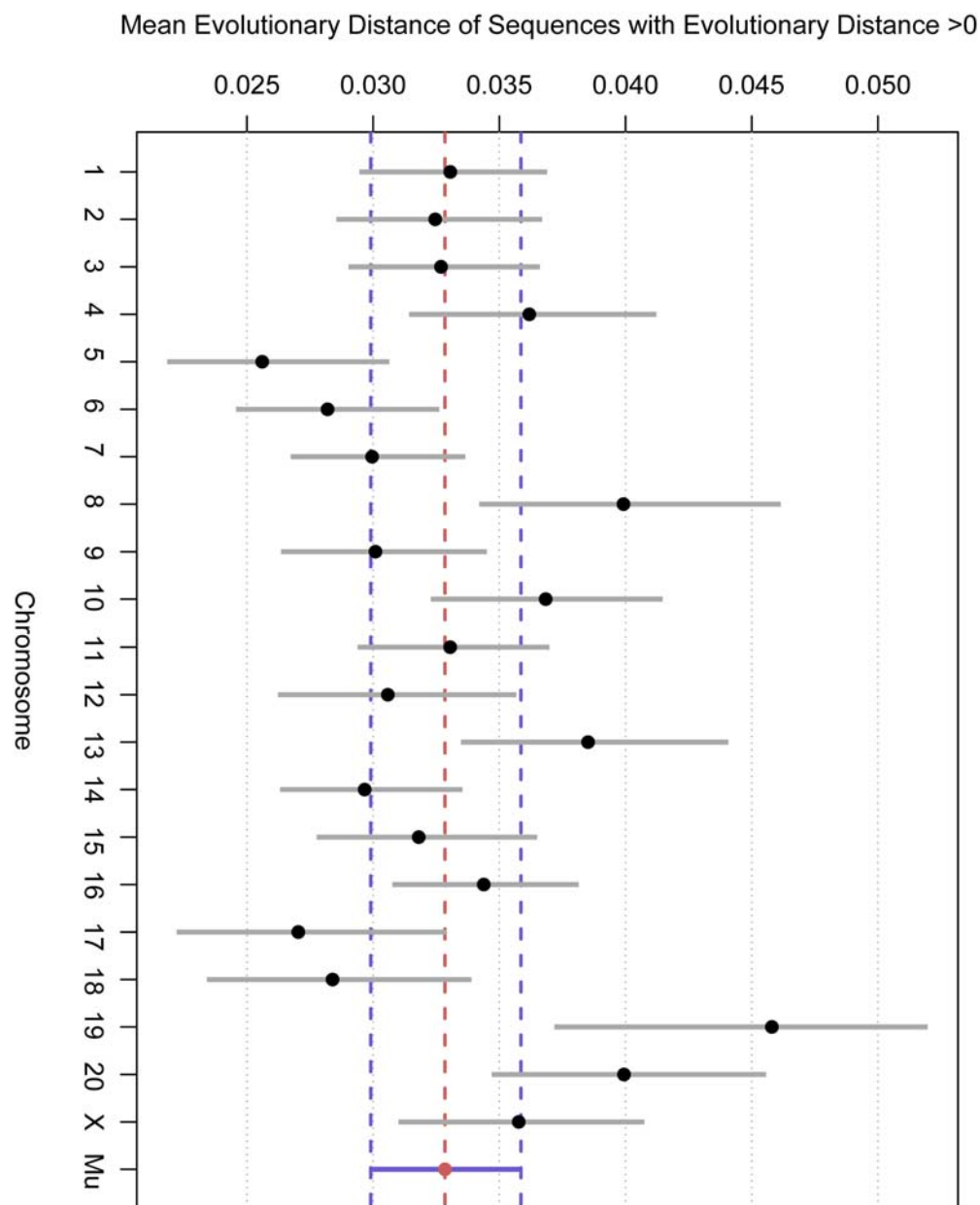
doi:10.1371/journal.pone.0123624.g009

Fig 10. The 90% posterior credibility intervals of the chromosomal random effects from the gamma sub-model of the zero-inflated gamma regression. The value 'Mu' is the partially-pooled mean estimate across chromosomes. We see that there is fairly significant heterogeneity in mean evolutionary divergence of amino acid sequences across chromosomes, with chromosomes 5, 6, and 17 showing reduced evolutionary divergence, and chromosomes 8, 13, 19, and 20 showing increased evolutionary divergence.

doi:10.1371/journal.pone.0123624.g010

Table 3. Table 3 presents the 20 gene ontology (GO) categories (of the 176 GO categories containing at least 100 genes) with the highest median levels of divergence according to the RSD algorithm. The label "Zeros" refers to the number of genes in the GO category with a divergence score of zero. The label "N" refers to the number of genes in the GO category. GO categories related to viral processes and immunity are well represented among the GO categories that are diverging at higher rates relative to most other categories. We note that 3 of the 20 (15%) most divergent GO categories are related to viral processes or immunity, while only 2 of the other 156 (1%) of GO categories are related to viral processes or immunity. Table 3 is a subset of Supplementary S3 Table.

GO Category	Mean	Median	Max	SD	Zeros	N
translational termination	0.012	0.006	0.268	0.027	90	214
viral transcription	0.012	0.006	0.268	0.027	90	214
translational elongation	0.012	0.006	0.268	0.026	93	218
viral life cycle	0.012	0.006	0.268	0.027	95	220
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	0.012	0.005	0.268	0.025	106	241
translational initiation	0.011	0.005	0.268	0.025	103	236
SRP-dependent cotranslational protein targeting to membrane	0.013	0.005	0.29	0.032	107	238
cytosolic large ribosomal subunit	0.011	0.005	0.177	0.023	63	133
protein complex binding	0.024	0.004	0.407	0.054	37	111
viral process	0.013	0.004	0.268	0.029	181	405
anatomical structure morphogenesis	0.017	0.003	0.264	0.043	44	108
detection of chemical stimulus involved in sensory perception of smell	0.008	0.003	0.091	0.013	53	152
olfactory receptor activity	0.008	0.003	0.091	0.013	53	152
sensory perception of smell	0.008	0.003	0.091	0.014	67	176
actin cytoskeleton	0.021	0.003	0.405	0.054	53	139
endoplasmic reticulum lumen	0.025	0.003	0.481	0.061	35	108
oxidation-reduction process	0.033	0.003	1.197	0.128	41	120
cytoskeleton organization	0.012	0.003	0.192	0.028	37	108
cell surface	0.023	0.003	0.337	0.053	93	268
external side of plasma membrane	0.023	0.003	0.446	0.058	47	129

doi:10.1371/journal.pone.0123624.t003

expression levels of *TRIM*32 might decrease viral replication, perhaps through interactions with HIV-1 Tat [34, 37].

A chimeric protein composed of rhesus macaque-derived *TRIM*5α and human-derived *TRIM*21 has been shown to potently inhibit HIV-1 infection [42, 43]. Further, *TRIM*21*CypA*, a *TRIM*21-cyclophilin-A fusion protein, provides highly potent protection against HIV-1 without disturbing normal TRIM activity [44]. Some authors have even suggested that *TRIM*21 itself may be directly involved in anti-HIV activity [45].

More generally, a comparison of human and mouse *TRIM* proteins found that 21% of all mouse, but only 11% of all human *TRIM* proteins interfere with HIV-1 release [38]. We hope to see such comparative methods expanded to include comparison of Chinese and Indian rhesus macaque derived *TRIM* proteins, as *TRIM*21 and *TRIM*32 proteins derived from Chinese rhesus macaques may prove to have especially potent anti-HIV properties relative to those of Indian rhesus macaques.

CXC Chemokine Ligand 12 (*CXCL12*). *CXCL*12 is a natural ligand for the HIV-1 coreceptor CXC Chemokine Receptor 4 (*CXCR*4) [46, 47]. By blocking CXCR4, a key coreceptor for HIV-1 entry, CXCL12 functions as an endogenous inhibitor of some HIV-1 strains [48].

It has been shown that *CXCL*12 is expressed at mucosal surfaces and can strongly reduce the transmission and propagation of X4-HIV at such sites [49], which helps to explain the differential pathogenesis of R5- and X4-SHIV in rhesus macaque models [50]. A naturally occurring splice variant of *CXCL*12 has been shown to be a potent HIV-1 entry inhibitor [51].

It was thought that a SNP (rs1801157) on *CXCL*12 might modulate plasma levels, but this effect was not found [46, 52]. Further, this SNP on *CXCL*12 has been thought to augment risk for HIV-1; an effect has been found in some studies, although the effect size is normally small [46]. A recent meta analysis, however, failed to find any significant association between this SNP and HIV-1 susceptibility [53]. Analysis of genetic variation in *CXCL*12 across Chinese and Indian rhesus macaques, may prove useful in determining if specific variants of *CXCL*12 confer increased resistance to SIV$_{mac}$, which might in turn help clarify the role that *CXCL*12 variants might play in HIV-1 pathogenesis in humans.

Cyclin-dependent kinase 9 (*CDK*9). *CDK*9 is a cyclin-dependent kinase that has been linked to HIV-1 pathogenesis in numerous studies—the CDK9/cyclin T1 enzyme, especially, has been shown to play an important role in HIV-1 transcription [54–56]. Due to the central nature of CDK9/cyclin T1 in HIV-1 transcription, several studies have investigated *CDK*9 as a potential drug target [57, 58].

CDK9 is also one of the main host genes that interacts with two HIV genes, *nef* and *tat* [59–62]. Moreover, CDK9 directly affects HIV gene expression and replication [63, 64].

Little is known about the role of *CDK*9 orthologs in SIV$_{mac}$ pathogenesis *in vivo* in rhesus macaques, but mutant variants of *CDK*9 lacking kinase activity have been shown to augment the basal transcriptional activity of the HIV-1 long terminal repeat [55].

DNA repair protein (*RAD*50). HIV-1 replication depends on integration of virally produced DNA into the host-cell's genome via integrase-dependent linkage [65]. It is argued that this linkage leaves an intermediate state that requires post-integration repair; HIV-1 is thought to exploit host double-strand DNA break (DSB) repair pathways in order to complete such repairs [65]. *RAD*50, along with Nijmegen Breakage Syndrome-1 protein (*NBS*1) and Meiotic Recombination 11 Homologue (*MRE*11), is thought to play a role in this process [65]. Little is known about the relationship between *RAD*50 variants and HIV-1 pathogenesis.

N-methyl-D-aspartate 1 (*NMDAR*1). *NMDAR*1 activation is thought to play a role in HIV-1 pathogenesis by modulating HIV-1 gp120-induced blood-brain barrier leakage, in that *NMDAR*1 antagonists have been shown to protect the blood-brain barrier from gp120-induced damage [66]. *NMDAR*1 has been shown to be significantly down regulated in HIV-1 infected astrocytes as opposed to control astrocytes [67].

Little is known about the relationship between *NMDAR*1 variants and HIV-1 pathogenesis; however, a dissertation on differential expression found that the NR1 subunit was significantly downregulated in SIV-infected Indian rhesus macaques, whereas the SIV-infected Chinese rhesus macaques showed no such signs of NR1 downregulation [68].

Mucin 16 (*MUC*16). *MUC*16 is a large membrane-associated mucin [69, 70]. Mucins are present at the surface of epithelial cells in the male and female genital tracts and are thought to play an important role in immune defense by trapping and eliminating pathogens before they can cause infection [70]. Some mucins have been found to be potent in neutralizing HIV-1 (e.g. *MUC*5B, *MUC*6, and *MUC*7) [71, 72], but the role of *MUC*16 is yet to be investigated.

Other Genes of Interest. *ANKRD*30*A* is a recently-discovered host factor for HIV-1 infection [25]; however, little is known about the role of *ANKRD*30*A* in HIV biology [73].

*IL*13*RA*1 was shown to vary in expression between HIV-1 infected and non-infected controls [74].

Expression of *NT*5*M*, a 5'-nucleotidase gene, was shown to vary (downregulation in macrophages) in response to Tenofovir, a pre-exposure prophylaxis used to prevent HIV-1 infection. Tenofovir functions to cause chain termination when incorporated into viral cDNA, and can thus inhibit viral replication [75]. Although Tenofovir has a direct effect on HIV-1 pathogenesis, immunomodulatory effects of Tenofovir, such as increased expression and secretion of

cytokines and modulated expression of 5'-nucleotidases like *NT5M* and *NT5E*, may also be responsible for its anti-viral activity [75].

The Tat protein of HIV-1 has been shown to interact with *Notch*1 *in vitro* [76]. Further studies have demonstrated linkages between HIV-1 and HIV-associated nephropathy (and progressive renal failure), which is thought to arise through activation of the Notch signaling pathway [77].

*PDCD*5 is a signal gene for apoptosis and was found to be upregulated in HIV-infected individuals relative to controls [78].

CTSZ was one of 221 differentially expressed genes up-regulated in the peripheral blood mononuclear cells of HCV/HIV co-infected individuals versus non-infected controls [79].

*GORASP*2 was found to regulate HIV-1 infection, as it was one of 37 genes that decreased MAGI cell *β*-galactosidase reporter activity at least 2-fold for at least three out of four siRNAs in a study of HIV-responsive phosphoproteins [80].

*GTF*2*H*1 was found to be strongly upregulated after monocyte-derived dendritic cells were experimentally infected with HIV-1 [81].

Several SNPs in the region of *TM9SF*2 showed signs of association with human AIDS progression *in vivo* [82].

LPL, lipoprotein lipase, was associated with HIV in a genome-wide siRNA screen [25]; however, the role that *LPL* plays in HIV pathogenesis, if any, is not well understood. This being said, it has been shown that some HIV protease inhibitors bind to a site on HIV-1 protease that shares approximately 60% homology to a region within *LPL* [83]. For this reason, anti-retroviral therapy has been implicated in peripheral fat wasting (lipodystrophy), central adiposity, hyperlipidaemia, and insulin resistance in human subjects [83].

Knockdown of *ICAM*4 and *BCAR*1 by siRNA enhances the early stages of HIV-1 replication in HIV-infected CD4 cells [22].

No laboratory studies were found that discuss the role of *CYP26B*1, *EFR3A*, *NME*6, *PIGX*, *BCAR*1, *DSP*, *SLC35B*1, *DAPK*2, *PARVA*, or *LYPLA*1 in HIV or SIV pathogenesis; evidence linking these genes to HIV pathogenesis come primarily from genome-wide siRNA analyses [22–25].

Shortcomings of this Study

Notably, 3 of the 20 orthologs in Table 1, and 4 of the 20 orthologs in Table 2, occurred in different chromosomes. As the evolutionary distance between proposed orthologs increases, it becomes more likely that they may actually be misclassified, especially if the pair breaks synteny. We cannot be sure if these transpositions are real, or if they simply result from classification errors by the RSD algorithm.

Furthermore, this study contains a significant shortcoming in that it uses a form of comparative genomic analysis that is more aptly applied to inter-species comparisons than to intra-species, inter-subspecies comparisons. This is because the use of single genome sequences to estimate the evolutionary divergence of orthologs is more defensible when amino acid sequences are more variable between groups than within groups. This assumption is not necessarily met in our study, because there is a strong potential for many of the amino acid sequences included in this analysis to be as variable within groups as between groups.

There are sure to be numerous amino acid sequences in our data for which our results reduce simply to a comparison between *individual animals*, rather than a true comparison between the normative genomic characteristics of *sub-subspecies*. While this weakness is inherent in our study design, it does not necessarily invalidate our results. Numerous genetic comparisons of Chinese and Indians rhesus macaques have shown that they differ genetically in

normative ways at many loci. For instance, of 661 genic SNPs distributed almost uniformly across the genome, 457 were found strictly in one sub-population [84]. Furthermore, discriminant analysis of principal components conducted using 2,808 evenly distributed, (mostly) non-coding SNPs found that the genomic profiles of Chinese rhesus macaques clustered tightly together, with almost no overlap with the distribution of the genomic profiles of Indian rhesus macaques [85]. Further research, however, is needed to investigate amino acid sequence variation in candidate loci across rhesus macaque subspecies using multiple animals from each subspecies.

Conclusions

We used estimates of the evolutionary distance of orthologous amino acid sequences in Chinese and Indian rhesus macaques to create a candidate list of orthlogs that might be diverging due to selection pressures or demographic processes. To the extent that normative phenotypic differences in immune response to SIV_{mac} are due to difference in amino acid sequences across rhesus macaque sub-populations, SIV_{mac} resistance in Chinese, but not Indian rhesus macaques should be associated with divergent orthologs. We cross-tabulated a list of divergent orthologs with lists of genes known to be involved in HIV or SIV pathogenesis. We identified 20 candidate genes that occurred on both gene lists and found that several other highly divergent orthologs had been previously linked to HIV or SIV as well.

There appear to have been few or no prior studies of differences in coding variants or expression of these genes between rhesus macaque sub-populations, nor studies investigating the effect of differing rhesus macaque variants of these genes on SIV or HIV pathogenesis. Such studies in the future may help to shed light on the molecular mechanisms underlying differences in normative SIV_{mac} resistance between these macaque subspecies.

Comparative genomic work can narrow the set of candidate genes to be investigated with more rigorous laboratory methods and case control studies; careful evaluation of genetic differences in these candidate loci across Indian and Chinese rhesus macaque sub-populations, and evaluation of the effect of these differences on SIV_{mac} pathogenesis, is needed in order to identify the structural and functional differences in genes underpinning SIV_{mac} resistance in Chinese, but not Indian rhesus macaques.

Materials and Methods

Data Sources

Amino acid sequence data for the Chinese rhesus macaque was accessed from the Comprehensive Library for Modern Biotechnology, BGI <http://climb.genomics.cn/10.5524/100002>, and amino acid sequence data for the Indian rhesus macaque was accessed from the Monkey Database, BGI <http://macaque.genomics.org.cn/page/species/index.jsp> in February, 2014.

Software, Programming Environment, and Analytical Methods

A Python implementation of the RSD 1.1.6 algorithm [15] was used in data analysis. Local installation of the script was conducted following instructions on GitHub <https://github.com/todddeluca/reciprocal_smallest_distance>. RSD 1.1.6 depends on Python 2.7, NCBI BLAST 2.2.26, PAML 4.4, and Kalign 2.04, so these programs were installed locally as well (February, 2014).

The GOanna [19] program was used to associate amino acid sequence data with human gene IDs and gene ontology annotations.

All other data analysis and visualization was conducted in the the R programming environment [86]; the *seqinr* package was used to read the FASTA sequence files.

Statistical modeling was conducted using the Stan version 2.5 C++ library for Hamiltonian Markov Chain Monte Carlo simulation, using the *RStan* package in R [87].

Statistical Methods

We use a multi-level zero-inflated gamma regression model to estimate differences in the evolutionary distance of orthologs as a function of chromosome and chromosomal position. Random effects on chromosome control for heterogeneity in evolutionary distance across chromosomes, and chromosome-specific vectors of random effects generated using a Gaussian process control for heterogeneity in evolutionary rate as a function of location on each chromosome. More formally, on each chromosome, c, we model the evolutionary distance of the nth amino acid sequence, $Y_{[c,n]}$, using the probability function:

$$Y_{[c,n]} \sim \begin{cases} \text{Gamma}(A_{[c,n]}, B_{[c]}) & \text{if } Y_{[c,n]} > 0, \text{ and} \\ \text{Bernoulli}(\theta_{[c,n]}) & \text{if } Y_{[c,n]} = 0 \end{cases} \tag{1}$$

We note that the mean of a gamma distribution is defined as $\mu = \frac{A}{B}$. As such, $A_{[c,n]}$ is defined from a model of the mean of a gamma distribution as:

$$A_{[c,n]} = \exp(\alpha_{[c]} + \gamma_{[c]} \delta_{[c,Z(n)]}) B_{[c]} \tag{2}$$

and $\theta_{[c,n]}$ is defined through a similar link function:

$$\theta_{[c,n]} = \text{logistic}(\beta_{[c]} + v_{[c]} \phi_{[c,Z(n)]}) \tag{3}$$

The parameters $\alpha_{[c]}$ and $\beta_{[c]} \in \mathbb{R}$ are chromosome specific intercepts. The parameters $\gamma_{[c]}$ and $v_{[c]} \in \mathbb{R}^+$ scale the variance of the random effects, $\delta_{[c,Z(n)]}$ and $\phi_{[c,Z(n)]}$. The vectors $\delta_{[c,1,\ldots,Z]}$ and $\phi_{[c,1,\ldots,Z]}$ (where $Z = 40$ is the number of equally-sized zones modeled per chromosome) are estimated using mean zero Gaussian processes:

$$\delta_{[c,1,\ldots,Z]} \sim \text{Multivariate Normal}((0,\ldots,0)', \Gamma) \tag{4}$$

$$\phi_{[c,1,\ldots,Z]} \sim \text{Multivariate Normal}((0,\ldots,0)', \Pi) \tag{5}$$

where the correlation matrices Γ and Π are formed as functions of the unit-normalized distance, D, between chromosomal zones z_1 and z_2 squared:

$$\Gamma_{[z_1, z_2]} = \begin{cases} \rho_{[c]} \exp(-\zeta_{[c]} D^2_{[z_1, z_2]}) & \text{if } z_1 \neq z_2, \text{ and} \\ 1 & \text{if } z_1 = z_2 \end{cases} \tag{6}$$

$$\Pi_{[z_1, z_2]} = \begin{cases} \kappa_{[c]} \exp(-\xi_{[c]} D^2_{[z_1, z_2]}) & \text{if } z_1 \neq z_2, \text{ and} \\ 1 & \text{if } z_1 = z_2 \end{cases} \tag{7}$$

where the parameters ρ and $\kappa \in (0, 1)$ represent the maximum correlation, and the parameters ζ and $\xi \in \mathbb{R}^+$ modulate the decay of correlation with distance between chromosomal zones. A more finely resolved Gaussian process would use a random effect on each observation, but given that the computational requirements of a Gaussian process model scale at about $O(n^3)$, it is unfeasible to model each data point as arising from a Gaussian process. Instead, we divide each chromosome into $Z = 40$ ordered and equally dense zones, and let the Gaussian process

generate a random effect that is shared among all amino acid sequences in that zone. We denote the zone of the nth amino acid sequence on a chromosome using the function notation $Z(n)$, to return the zone, z, of the nth amino acid sequence.

To complete the Bayesian model, we use multi-level partial pooling to share information across chromosomes:

$$\alpha_{[c]} \sim \text{Normal}(\mu_\alpha, \sigma_\alpha) \tag{8}$$

$$\beta_{[c]} \sim \text{Normal}(\mu_\beta, \sigma_\beta) \tag{9}$$

We specify weakly regularizing priors on μ_α and μ_β:

$$\mu_\alpha \sim \text{Normal}(0, 10) \tag{10}$$

$$\mu_\beta \sim \text{Normal}(0, 10) \tag{11}$$

and vague priors on σ_α and σ_β:

$$\sigma_\alpha \sim \text{Cauchy}(0, 3) T[0, \infty] \tag{12}$$

$$\sigma_\beta \sim \text{Cauchy}(0, 3) T[0, \infty] \tag{13}$$

where the truncation operator, $T[0, \infty]$, indicates truncated support on the interval $(0, \infty)$. Each element of the vector of scale parameters, B, for the gamma regression model is given a weakly informative prior:

$$B_{[c]} \sim \text{Normal}(0, 10) T[0, \infty] \tag{14}$$

Each element of γ and ν, the variance parameters, get vague priors:

$$\gamma_{[c]} \sim \text{Cauchy}(0, 3) T[0, \infty] \tag{15}$$

$$\nu_{[c]} \sim \text{Cauchy}(0, 3) T[0, \infty] \tag{16}$$

Because the elements of ρ and κ are constrained to the unit interval, they get weakly informative Beta priors:

$$\rho_{[c]} \sim \text{Beta}(2, 2) \tag{17}$$

$$\kappa_{[c]} \sim \text{Beta}(2, 2) \tag{18}$$

Finally, the elements of the decay parameter vectors, ζ and ξ, get weakly informative priors:

$$\zeta_{[c]} \sim \text{Normal}(2, 3) T[0, \infty] \tag{19}$$

$$\xi_{[c]} \sim \text{Normal}(2, 3) T[0, \infty] \tag{20}$$

Supporting Information

S1 Table. This file (Excel format) contains the full set of orthologs and evolutionary distances from our analysis. We have marked the genes that break synteny.
(XLSX)

S2 Table. This file (Excel format) contains the full set of previously described HIV-related genes utilized to construct results presented in Table 1.
(XLSX)

S3 Table. This file (Excel format) contains the full set of gene function classifications for which GOanna could amend annotations and for which the RSD algorithm found orthologs.
(XLSX)

S4 Table. This file (Excel format) contains the summary of the posterior parameter estimates from MCMC simulation using the Stan 2.5 C++ library.
(XLSX)

S1 Code. This file (R script) contains the code used to construct our statistical model.
(R)

Acknowledgments

Author Contributions

Conceived and designed the experiments: CTR. Performed the experiments: CTR. Analyzed the data: CTR. Contributed reagents/materials/analysis tools: DGS. Wrote the paper: CTR MR DGS.

References

1. Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, et al. (2007) Evolutionary and biomedical insights from the rhesus macaque genome. Science 316: 222–234. PMID: 17431167

2. Yan G, Zhang G, Fang X, Zhang Y, Li C, et al. (2011) Genome sequencing and comparison of two non-human primate animal models, the cynomolgus and Chinese rhesus macaques. Nature Biotechnology 29: 1019–1023. doi: 10.1038/nbt.1992 PMID: 22002653

3. Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, et al. (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biology 3: e170. doi: 10.1371/journal.pbio.0030170 PMID: 15869325

4. Obbard DJ, Jiggins FM, Halligan DL, Little TJ (2006) Natural selection drives extremely rapid evolution in antiviral rnai genes. Current Biology 16: 580–585. doi: 10.1016/j.cub.2006.01.065 PMID: 16546082

5. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, et al. (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419: 832–837. doi: 10.1038/nature01140 PMID: 12397357

6. Barreiro LB, Quintana-Murci L (2010) From evolutionary genetics to human immunology: how selection shapes host defence genes. Nature Reviews Genetics 11: 17–30. doi: 10.1038/nrg2698 PMID: 19953080

7. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, et al. (2012) Systematic localization of common disease-associated variation in regulatory dna. Science 337: 1190–1195. doi: 10.1126/science.1222794 PMID: 22955828

8. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M (2012) Linking disease associations with regulatory information in the human genome. Genome Research 22: 1748–1759. doi: 10.1101/gr.136127.111 PMID: 22955986

9. Ishida A, Mutoh T, Ueyama T, Bando H, Masubuchi S, et al. (2005) Light activates the adrenal gland: Timing of gene expression and glucocorticoid release. Cell Metabolism 2: 297–307. doi: 10.1016/j. cmet.2005.09.009 PMID: 16271530

10. Hastings MH, Reddy AB, Maywood ES (2003) A clockwork web: Circadian timing in brain and periphery, in health and disease. Nature Reviews Neuroscience 4: 649–661. doi: 10.1038/nrn1177 PMID: 12894240

11. Lage K, Hansen NT, Karlberg EO, Eklund AC, Roque FS, et al. (2008) A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. Proceedings of the National Academy of Sciences 105: 20870–20875. doi: 10.1073/pnas.0810772105

12. Gilad Y, Rifkin SA, Pritchard JK (2008) Revealing the architecture of gene regulation: the promise of eqtl studies. Trends in Genetics 24: 408–415. doi: 10.1016/j.tig.2008.06.001 PMID: 18597885

13. Satkoski Trask J, Garnica W, Malhi R, Kanthaswamy S, Smith D (2011) High-throughput single-nucleotide polymorphism discovery and the search for candidate genes for long-term sivmac non-progression in Chinese rhesus macaques (macaca mulatta). Journal of Medical Primatology 40: 224–232. doi: 10.1111/j.1600-0684.2011.00486.x PMID: 21781130

14. Trichel A, Rajakumar P, Murphey-Corb M (2002) Species-specific variation in siv disease progression between Chinese and indian subspecies of rhesus macaque. Journal of Medical Primatology 31: 171–178. doi: 10.1034/j.1600-0684.2002.02003.x PMID: 12390539

15. Wall D, Fraser H, Hirsh A (2003) Detecting putative orthologs. Bioinformatics 19: 1710–1711. doi: 10.1093/bioinformatics/btg213 PMID: 15593400

16. Hirsh AE, Fraser HB (2001) Protein dispensability and rate of evolution. Nature 411: 1046–1049. doi: 10.1038/35082561 PMID: 11429604

17. Zharkikh A (1994) Estimation of evolutionary distances between nucleotide sequences. Journal of Molecular Evolution 39: 315–329. doi: 10.1007/BF00160155 PMID: 7932793

18. Hernandez RD, Hubisz MJ, Wheeler DA, Smith DG, Ferguson B, et al. (2007) Demographic histories and patterns of linkage disequilibrium in Chinese and indian rhesus macaques. Science 316: 240–243. doi: 10.1126/science.1140462 PMID: 17431170

19. McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, et al. (2006) Agbase: a functional genomics resource for agriculture. BMC Genomics 7: 229. doi: 10.1186/1471-2164-7-229 PMID: 16961921

20. Ortiz M, Guex N, Patin E, Martin O, Xenarios I, et al. (2009) Evolutionary trajectories of primate genes involved in hiv pathogenesis. Molecular Biology and Evolution 26: 2865–2875. doi: 10.1093/molbev/msp197 PMID: 19726537

21. Petrovski S, Fellay J, Shianna KV, Carpenetti N, Kumwenda J, et al. (2011) Common human genetic variants and hiv-1 susceptibility: a genome-wide survey in a homogeneous african population. AIDS 25: 513–518. doi: 10.1097/QAD.0b013e328343817b PMID: 21160409

22. Liu L, Oliveira NM, Cheney KM, Pade C, Dreja H, et al. (2011) A whole genome screen for hiv restriction factors. R etrovirology 8: 1–15.

23. König R, Zhou Y, Elleder D, Diamond TL, Bonamy G, et al. (2008) Global analysis of host-pathogen interactions that regulate early-stage hiv-1 replication. Cell 135: 49–60. doi: 10.1016/j.cell.2008.07.032 PMID: 18854154

24. Zhou H, Xu M, Huang Q, Gates AT, Zhang XD, et al. (2008) Genome-scale rnai screen for host factors required for hiv replication. Cell Host & Microbe 4: 495–504. doi: 10.1016/j.chom.2008.10.004

25. Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, et al. (2008) Identification of host proteins required for hiv infection through a functional genomic screen. Science 319: 921–926. doi: 10.1126/science.1152725 PMID: 18187620

26. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, et al. (2014) Defining functional dna elements in the human genome. Proceedings of the National Academy of Sciences 111: 6131–6138. doi: 10.1073/pnas.1318948111

27. DeLuca TF, Cui J, Jung JY, Gabriel KCS, Wall DP (2012) Roundup 2.0: enabling comparative genomics for over 1800 genomes. Bioinformatics 28: 715–716. doi: 10.1093/bioinformatics/bts006 PMID: 22247275

28. Yáñez RJ, Porter AC (2002) A chromosomal position effect on gene targeting in human cells. Nucleic Acids Research 30: 4892–4901. doi: 10.1093/nar/gkf614 PMID: 12433992

29. Festenstein R, Kioussis D (2000) Locus control regions and epigenetic chromatin modifiers. Current Opinion in Genetics & Development 10: 199–203. doi: 10.1016/S0959-437X(00)00060-5

30. Smith JM, Száthmáry E (1993) The origin of chromosomes i. selection for linkage. Journal of Theoretical Biology 164: 437–446. doi: 10.1006/jtbi.1993.1165 PMID: 8264246

31. Szathmáry E, Smith JM (1993) The evolution of chromosomes ii. molecular mechanisms. Journal of Theoretical Biology 164: 447–454. doi: 10.1006/jtbi.1993.1166 PMID: 7505372

32. Carthagena L, Bergamaschi A, Luna JM, David A, Uchil PD, et al. (2009) Human trim gene expression in response to interferons. PloS One 4: e4894. doi: 10.1371/journal.pone.0004894 PMID: 19290053

33. Ozato K, Shin DM, Chang TH, Morse HC (2008) Trim family proteins and their emerging roles in innate immunity. Nature Reviews Immunology 8: 849–860. doi: 10.1038/nri2413 PMID: 18836477

34. Nisole S, Stoye JP, Saïb A (2005) Trim family proteins: retroviral restriction and antiviral defence. Nature Reviews Microbiology 3: 799–808. doi: 10.1038/nrmicro1248 PMID: 16175175

35. Barr SD, Smiley JR, Bushman FD (2008) The interferon response inhibits hiv particle production by induction of trim22. PLoS Pathogens 4: e1000007. doi: 10.1371/journal.ppat.1000007 PMID: 18389079

36. Tissot C, Mechti N (1995) Molecular cloning of a new interferon-induced factor that represses human immunodeficiency virus type 1 long terminal repeat expression. Journal of Biological Chemistry 270: 14891–14898. doi: 10.1074/jbc.270.25.14891 PMID: 7797467

37. Fridell RA, Harding LS, Bogerd HP, Cullen BR (1995) Identification of a novel human zinc finger protein that specifically interacts with the activation domain of lentiviral tat proteins. Virology 209: 347–357. doi: 10.1006/viro.1995.1266 PMID: 7778269

38. Uchil PD, Quinlan BD, Chan WT, Luna JM, Mothes W (2008) Trim e3 ligases interfere with early and late stages of the retroviral life cycle. PLoS Pathogens 4: e16. doi: 10.1371/journal.ppat.0040016 PMID: 18248090

39. Zhang J, Hu MM, Wang YY, Shu HB (2012) Trim32 protein modulates type i interferon induction and cellular antiviral response by targeting mita/sting protein for k63-linked ubiquitination. Journal of Biological Chemistry 287: 28646–28655. doi: 10.1074/jbc.M112.362608 PMID: 22745133

40. Ryu YS, Lee Y, Lee KW, Hwang CY, Maeng JS, et al. (2011) Trim32 protein sensitizes cells to tumor necrosis factor (tnfα)-induced apoptosis via its ring domain-dependent e3 ligase activity against x-linked inhibitor of apoptosis (xiap). Journal of Biological Chemistry 286: 25729–25738. doi: 10.1074/jbc.M111.241893 PMID: 21628460

41. Abdel-Mohsen M, Raposo RAS, Deng X, Li M, Liegler T, et al. (2013) Expression profile of host restriction factors in hiv-1 elite controllers. Retrovirology 10: 106. doi: 10.1186/1742-4690-10-106 PMID: 24131498

42. Kar AK, Diaz-Griffero F, Li Y, Li X, Sodroski J (2008) Biochemical and biophysical characterization of a chimeric trim21-trim5α protein. Journal of Virology 82: 11669–11681. doi: 10.1128/JVI.01559-08 PMID: 18799572

43. Li X, Li Y, Stremlau M, Yuan W, Song B, et al. (2006) Functional replacement of the ring, b-box 2. and coiled-coil domains of tripartite motif 5α (trim5α) by heterologous trim domains. Journal of Virology 80: 6198–6206. doi: 10.1128/JVI.00283-06 PMID: 16775307

44. Chan E, Schaller T, Eddaoudi A, Zhan H, Tan CP, et al. (2012) Lentiviral gene therapy against human immunodeficiency virus type 1. using a novel human trim21-cyclophilin a restriction factor. Human Gene Therapy 23: 1176–1185. doi: 10.1089/hum.2012.083 PMID: 22909012

45. Diaz-Griffero F (2011) Caging the beast: Trim5α binding to the hiv-1 core. Viruses 3: 423–428. doi: 10.3390/v3050423 PMID: 21994740

46. Petersen DC, Glashoff RH, Shrestha S, Bergeron J, Laten A, et al. (2005) Risk for hiv-1 infection associated with a common cxcl12 (sdf1) polymorphism and cxcr4 variation in an african population. Journal of Acquired Immune Deficiency Syndromes 40: 521–526. doi: 10.1097/01.qai.0000186360.42834.28 PMID: 16284526

47. Feng Y, Broder CC, Kennedy PE, Berger EA (1996) Hiv-1 entry cofactor: functional cdna cloning of a seven-transmembrane, g protein-coupled receptor. Science 272: 872–877. doi: 10.1126/science.272.5263.872 PMID: 8629022

48. Bleul C, Farzan M, Choe H, Parolin C, Clark-Lewis I, et al. (1996) The lymphocyte chemoattractant sdf-1 is a ligand for lestr/fusin and blocks hiv-1 entry. Nature 382: 829–833. doi: 10.1038/382829a0 PMID: 8752280

49. Agace W, Amara A, Roberts A, Pablos J, Thelen S, et al. (2000) Constitutive expression of stromal derived factor-1 by mucosal epithelia and its role in hiv transmission and propagation. Current Biology 10: 325–328. doi: 10.1016/S0960-9822(00)00380-8 PMID: 10744978

50. Harouse JM, Gettie A, Tan RCH, Blanchard J, Cheng-Mayer C (1999) Distinct pathogenic sequela in rhesus macaques infected with ccr5 or cxcr4 utilizing shivs. Science 284: 816–819. doi: 10.1126/science.284.5415.816 PMID: 10221916

51. Altenburg JD, Broxmeyer HE, Jin Q, Cooper S, Basil S, et al. (2007) A naturally occurring splice variant of cxcl12/stromal cell-derived factor 1 is a potent hiv-1 inhibitor with weak chemotaxis and cell survival activities. Journal of Virology: 8140–8148. doi: 10.1128/JVI.00268-07 PMID: 17507482

52. Shalekoff S, Schramm DB, Lassaunière R, Picton AC, Tiemessen CT (2013) Differences are evident within the cxcr4-cxcl12 axis between ethnically divergent south african populations. Cytokine 61: 792–800. doi: 10.1016/j.cyto.2013.01.003 PMID: 23395387

53. Liu S, Zhu H (2011) Association between polymorphism of sdf1 (cxcl12) gene and hiv-1 susceptibility: a meta-analysis. Current HIV Research 9: 112–119. doi: 10.2174/157016211795569113 PMID: 21361865

54. Wei P, Garber ME, Fang SM, Fischer WH, Jones KA (1998) A novel cdk9-associated c-type cyclin interacts directly with hiv-1 tat and mediates its high-affinity, loop-specific binding to tar rna. Cell 92: 451–462. doi: 10.1016/S0092-8674(00)80939-3 PMID: 9491887

55. Amini S, Clavo A, Nadraga Y, Giordano A, Khalili K, et al. (2002) Interplay between cdk9 and nf-kappab factors determines the level of hiv-1 gene transcription in astrocytic cells. Oncogene 21: 5797–5803. doi: 10.1038/sj.onc.1205754 PMID: 12173051

56. Ammosova T, Yedavalli VR, Jerebtsova M, Van Eynde A, Beullens M, et al. (2011) Expression of pp1 inhibitor inhibits cdk9 and hiv-1 transcription. The FASEB Journal 25: 756.

57. Sancineto L, Iraci N, Massari S, Attanasio V, Corazza G, et al. (2013) Computer-aided design, synthesis and validation of 2-phenylquinazolinone fragments as cdk9 inhibitors with anti-hiv-1 tat-mediated transcription activity. ChemMedChem 8: 1941–1953. doi: 10.1002/cmdc.201300287 PMID: 24150998

58. Van Duyne R, Guendel I, Jaworski E, Sampey G, Klase Z, et al. (2013) Effect of mimetic cdk9 inhibitors on hiv-1-activated transcription. Journal of Molecular Biology 425: 812–829. doi: 10.1016/j.jmb.2012.12.005 PMID: 23247501

59. Kumar M, Mitra D (2005) Heat shock protein 40 is necessary for human immunodeficiency virus-1 nef-mediated enhancement of viral gene expression and replication. Journal of Biological Chemistry 280: 40041–40050. doi: 10.1074/jbc.M508904200 PMID: 16179353

60. Narayanan A, Sampey G, Van Duyne R, Guendel I, Kehn-Hall K, et al. (2012) Use of atp analogs to inhibit hiv-1 transcription. Virology 432: 219–231. doi: 10.1016/j.virol.2012.06.007 PMID: 22771113

61. Mameli G, Deshmane SL, Ghafouri M, Cui J, Simbiri K, et al. (2007) C/ebpβ regulates human immunodeficiency virus 1 gene expression through its association with cdk9. Journal of General Virology 88: 631–640. doi: 10.1099/vir.0.82487-0 PMID: 17251582

62. Arhel NJ, Kirchhoff F (2009) Implications of nef: Host cell interactions in viral persistence and progression to aids. In: HIV Interactions with Host Cell Proteins. Springer, pp. 147–175.

63. Khan SZ, Mitra D (2011) Cyclin k inhibits hiv-1 gene expression and replication by interfering with cyclin-dependent kinase 9 (cdk9)-cyclin t1 interaction in nef-dependent manner. Journal of Biological Chemistry 286: 22943–22954. doi: 10.1074/jbc.M110.201194 PMID: 21555514

64. Mbonye UR, Gokulrangan G, Datt M, Dobrowolski C, Cooper M, et al. (2013) Phosphorylation of cdk9 at ser175 enhances hiv transcription and is a marker of activated p-tefb in cd4+ t lymphocytes. PLoS Pathogens 9: e1003338. doi: 10.1371/journal.ppat.1003338 PMID: 23658523

65. Smith JA, Wang FX, Zhang H, Wu KJ, Williams KJ, et al. (2008) Evidence that the nijmegen breakage syndrome protein, an early sensor of double-strand dna breaks (dsb), is involved in hiv-1 post-integration repair by recruiting the ataxia telangiectasia-mutated kinase in a process similar to, but distinct from, cellular dsb repair. Virology Journal 5: 11. doi: 10.1186/1743-422X-5-11 PMID: 18211700

66. Louboutin JP, Strayer DS (2012) Blood-brain barrier abnormalities caused by hiv-1 gp120: Mechanistic and therapeutic implications. The Scientific World Journal 2012: 482575–482575. doi: 10.1100/2012/482575

67. Atluri VSR, Kanthikeel SP, Reddy PV, Yndart A, Nair MP (2013) Human synaptic plasticity gene expression profile and dendritic spine density changes in hiv-infected human cns cells: role in hiv-associated neurocognitive disorders (hand). PloS One 8: e61399. doi: 10.1371/journal.pone.0061399 PMID: 23620748

68. Meisner F (2010) Die Rolle von Dopamin in der Pathogenese der HIV-assoziierten Demenz. Ph.D. thesis.

69. Shukair SA, Allen SA, Cianci GC, Stieh DJ, Anderson MR, et al. (2013) Human cervicovaginal mucus contains an activity that hinders hiv-1 movement. Mucosal Immunology 6: 427–434. doi: 10.1038/mi.2012.87 PMID: 22990624

70. Pudney J, Anderson D (2011) Innate and acquired immunity in the human penile urethra. Journal of Reproductive Immunology 88: 219–227. doi: 10.1016/j.jri.2011.01.006 PMID: 21353311

71. Stax MJ, van Montfort T, Sprenger RR, Melchers M, Sanders RW, et al. (2009) Mucin 6 in seminal plasma binds dc-sign and potently blocks dendritic cell mediated transfer of hiv-1 to cd4+ t-lymphocytes. Virology 391: 203–211. doi: 10.1016/j.virol.2009.06.011 PMID: 19682628

72. Habte HH, Mall AS, De Beer C, Lotz ZE, Kahn D, et al. (2006) The role of crude human saliva and purified salivary muc5b and muc7 mucins in the inhibition of human immunodeficiency virus type 1 in an inhibition assay. Virololgy Journal 3: 99. doi: 10.1186/1743-422X-3-99

73. Meyerson NR, Rowley PA, Swan CH, Le DT, Wilkerson GK, et al. (2014) Positive selection of primate genes that promote hiv-1 replication. Virology 454: 291–298. doi: 10.1016/j.virol.2014.02.029 PMID: 24725956

74. Zhao J, Yi L, Lu J, Yang ZR, Chen Y, et al. (2012) Transcriptomic assay of cd8+ t cells in treatment-naïve hiv, hcv-mono-infected and hiv/hcv-co-infected Chinese. PloS One 7: e45200. doi: 10.1371/journal.pone.0045200 PMID: 23028845

75. Biswas N, Rodriguez-Garcia M, Crist SG, Shen Z, Bodwell JE, et al. (2013) Effect of tenofovir on nucleotidases and cytokines in hiv-1 target cells. PloS One 8: e78814. doi: 10.1371/journal.pone.0078814 PMID: 24205323

76. Shoham N, Cohen L, Yaniv A, Gazit A (2003) The tat protein of the human immunodeficiency virus type 1 (hiv-1) interacts with the egf-like repeats of the notch proteins and the egf precursor. Virus Research 98: 57–61. doi: 10.1016/j.virusres.2003.08.016 PMID: 14609630

77. Sharma M, Callen S, Zhang D, Singhal PC, Heuvel GBV, et al. (2010) Activation of notch signaling pathway in hiv-associated nephropathy. AIDS 24: 2161–2170. doi: 10.1097/QAD.0b013e32833dbc31 PMID: 20706108

78. Motomura K, Toyoda N, Oishi K, Sato H, Nagai S, et al. (2004) Identification of a host gene subset related to disease prognosis of hiv-1 infected individuals. International Immunopharmacology 4: 1829–1836. doi: 10.1016/j.intimp.2004.07.031 PMID: 15531298

79. Gupta P, Liu B, Wu J, Soriano V, Vispo E, et al. (2014) Genome-wide mrna and mirna analysis of peripheral blood mononuclear cells (pbmc) reveals different mirnas regulating hiv/hcv co-infection. Virology 450: 336–349. doi: 10.1016/j.virol.2013.12.026 PMID: 24503097

80. Wojcechowskyj JA, Didigu CA, Lee JY, Parrish NF, Sinha R, et al. (2013) Quantitative phospho-proteomics reveals extensive cellular reprogramming during hiv-1 entry. Cell Host & Microbe 13: 613–623. doi: 10.1016/j.chom.2013.04.011

81. Solis M, Wilkinson P, Romieu R, Hernandez E, Wainberg MA, et al. (2006) Gene expression profiling of the host response to hiv-1 b, c, or a/e infection in monocyte-derived dendritic cells. Virology 352: 86–99. doi: 10.1016/j.virol.2006.04.010 PMID: 16730773

82. Chinn LW, Tang M, Kessing BD, Lautenberger JA, Troyer JL, et al. (2010) Genetic associations of variants in genes encoding hiv-dependency factors required for hiv-1 infection. Journal of Infectious Diseases 202: 1836–1845. doi: 10.1086/657322 PMID: 21083371

83. Carr A, Samaras K, Chisholm DJ, Cooper DA (1998) Pathogenesis of hiv-1-protease inhibitor-associated peripheral lipodystrophy, hyperlipidaemia, and insulin resistance. The Lancet 351: 1881–1883. doi: 10.1016/S0140-6736(98)03391-1

84. Ferguson B, Street SL, Wright H, Pearson C, Jia Y, et al. (2007) Single nucleotide polymorphisms (snps) distinguish indian-origin and chinese-origin rhesus macaques (macaca mulatta). BMC Genomics 8: 43. doi: 10.1186/1471-2164-8-43 PMID: 17286860

85. Kanthaswamy S, Trask JS, Ross CT, Kou A, Houghton P, et al. (2012) A large-scale snp-based genomic admixture analysis of the captive rhesus macaque colony at the California national primate research center. American Journal of Primatology 74: 747–757. doi: 10.1002/ajp.22025 PMID: 24436199

86. R Core Team (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

87. Stan Development Team (2014). Stan: A C++ library for probability and sampling, version 2.5.0. URL http://mc-stan.org/.

88. Kawai T, Akira S (2011) Regulation of innate immune signalling pathways by the tripartite motif (trim) family proteins. EMBO Molecular Medicine 3: 513–527. doi: 10.1002/emmm.201100160 PMID: 21826793

89. Kajaste-Rudnitski A, Pultrone C, Marzetta F, Ghezzi S, Coradin T, et al. (2010) Restriction factors of retroviral replication: the example of tripartite motif (trim) protein 5α and 22. Amino Acids 39: 1–9. doi: 10.1007/s00726-009-0393-x PMID: 19943174

90. Xin KQ, Hamajima K, Hattori S, Cao XR, Kawamoto S, et al. (1999) Evidence of hiv type 1 glycoprotein 120 binding to recombinant n-methyl-d-aspartate receptor subunits expressed in a baculovirus system. AIDS Research and Human Retroviruses 15: 1461–1467. doi: 10.1089/088922299309973 PMID: 10555109

91. Eugenin E, D'aversa T, Lopez L, Calderon T, Berman J (2003) Mcp-1 (ccl2) protects human neurons and astrocytes from nmda or hiv-tat-induced apoptosis. Journal of Neurochemistry 85: 1299–1311. doi: 10.1046/j.1471-4159.2003.01775.x PMID: 12753088

92. Mazzolini J, Herit F, Bouchet J, Benmerah A, Benichou S, et al. (2010) Inhibition of phagocytosis in hiv-1-infected macrophages relies on nef-dependent alteration of focal delivery of recycling compartments. Blood 115: 4226–4236. doi: 10.1182/blood-2009-12-259473 PMID: 20299515

93. Naydenov NG, Harris G, Morales V, Ivanov AL et al. (2012) Loss of a membrane trafficking protein αsnap induces non-canonical autophagy in human epithelia. Cell Cycle 11: 4613–4625. doi: 10.4161/cc.22885 PMID: 23187805

94. Yilmaz S, Boffito M, Collot-Teixeira S, De Lorenzo F, Waters L, et al. (2010) Investigation of low-dose ritonavir on human peripheral blood mononuclear cells using gene expression whole genome microarrays. Genomics 96: 57–65. doi: 10.1016/j.ygeno.2010.03.011 PMID: 20353815

95. Khwaja SS, Liu H, Tong C, Jin F, Pear WS, et al. (2010) Hiv-1 rev-binding protein accelerates cellular uptake of iron to drive notch-induced t cell leukemogenesis in mice. The Journal of Clinical Investigation 120: 2537–2548. doi: 10.1172/JCI41277 PMID: 20516639