Genomics Data 6 (2015) 202–207

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Identifying rhesus macaque gene orthologs using heterospecific human CNV probes

Jillian Ng [a], Joseph N. Fass [b], Blythe Durbin-Johnson [b], David Glenn Smith [a,c], Sree Kanthaswamy [c,d,*]

[a] Molecular Anthropology Laboratory, Department of Anthropology, University of California, Davis, CA, USA
[b] Genome Center Bioinformatics Core, University of California, Davis, CA, USA
[c] California National Primate Research Center, University of California, Davis, CA, USA
[d] School of Mathematics and Natural Sciences, Arizona State University (ASU) at the West Campus, Glendale, AZ, USA

ARTICLE INFO

Article history:
Received 24 August 2015
Accepted 12 September 2015
Available online 16 September 2015

Keywords:
Macaca mulatta
Copy number variation (CNV)
Homo sapiens
Heterospecific sequence map

ABSTRACT

We used the Affymetrix® Genome-Wide Human SNP Array 6.0 to identify heterospecific markers and compare copy number and structural genomic variation between humans and rhesus macaques. Over 200,000 human copy number variation (CNV) probes were mapped to a Chinese and an Indian rhesus macaque sample. Observed genomic rearrangements and synteny were in agreement with the results of a previously published genomic comparison between humans and rhesus macaques. Comparisons between each of the two rhesus macaques and humans yielded 206 regions with copy numbers that differed by at least two fold in the Indian rhesus macaque and human, 32 in the Chinese rhesus macaque and human, and 147 in both rhesus macaques. The detailed genomic map and preliminary CNV data are useful for better understanding genetic variation in rhesus macaques, identifying derived changes in human CNVs that may have evolved by selection, and determining the suitability of rhesus macaques as human models for particular biomedical studies.

1. Introduction

The last common ancestor of humans and macaques is estimated to have lived approximately 25 million years ago [1]. The macaque biomedical model has been instrumental in studying and combating infectious, cardiovascular, metabolic, and respiratory diseases, as well as understanding reproduction, development, aging, and neuroscience in humans [2–8]. The most commonly used macaque species in human biomedical research are rhesus macaques (*Macaca mulatta*) with those of Indian origin being most popular [9]. The natural range of rhesus macaques spans from Afghanistan in the west to the western shores of the Yellow Sea in the east [10]. Genetic differences between rhesus macaques from India and China have been characterized and can be used to determine an individual's country of origin [11–14]. In a genome comparison of the two regional populations of rhesus macaques, Yan et al. [15] found that 97% of Chinese rhesus macaque sequence scaffolds were successfully placed on Indian rhesus chromosomes, suggesting high levels of genetic similarity. Alternatively, comparisons of the human and Indian rhesus macaque genomes revealed that these two species share 93.54% sequence identity [9], justifying the use of rhesus macaques as human biological models. However, rhesus macaques carry three times the genetic diversity observed in humans [16], and large scale genomic rearrangements have also been observed between these two species [17], circumstances that may either limit or enhance the suitability of the rhesus macaque model.

Single nucleotide polymorphisms (SNPs) and copy number variations (CNVs) are highly variable in the human population and are often associated with human diseases and other phenotypes [9, 18–22]. Kanthaswamy et al. [17] used the commercially available Affymetrix® Genome-Wide Human SNP Array 6.0 to study the conservation of SNPs between humans and rhesus macaques. Of the 906,600 SNPs featured on the array, more than 85,000 were found to be conserved and were used to map over 9400 human gene orthologs in rhesus macaques. Analysis of the SNPs revealed major regions of genomic rearrangement including chromosome fusions and gene inversions, consistent with previous studies [23–25].

Many CNVs are involved in variable copy numbers of genes that may have different phenotypic effects depending on gene dosage. For example, higher copy numbers of the chemokine *CCL3L1* confer greater resistance to HIV (human immunodeficiency virus) and SIV (simian immunodeficiency virus; the non-human primate equivalent to HIV) in humans and rhesus macaques, respectively [18,19]. Chinese rhesus macaques have, on average, more copies of *CCL3L1* than Indian rhesus macaques, contributing to the former's lower susceptibility to SIV infection [19]. The conservation of *CCL3L1* in rhesus macaques makes them good models for studying HIV, and the difference in copy number of

* Corresponding author at: School of Mathematical and Natural Sciences, Arizona State University, Glendale, AZ, USA
E-mail address: sree.kanthaswamy@asu.edu (S. Kanthaswamy).

http://dx.doi.org/10.1016/j.gdata.2015.09.016

this gene between the regional rhesus macaque populations may make one population more suitable as subjects for a particular study than the other. Though few CNVs have been determined to cause disease in humans, it is likely that these variants may affect susceptibility to some diseases. Moreover, the mapping of human CNVs in rhesus macaques can show differences in the location of CNVs shared between the two species and identify genomic rearrangement.

Since CNVs are highly variable and have been linked to some phenotypes, including some diseases, in humans, analyses similar to those done with the SNPs can be used to identify conserved CNVs. In addition to the SNP probes used in Kanthaswamy et al. [17], the Affymetrix® array also features 946,000 CNV detection probes. The present study is based on the analyses of the rhesus macaque CNV data generated using the Affymetrix® array to identify species-specific copy number differences within the orthologous genomic regions in macaques and humans. Genomic similarities and differences in orthologous CNVs of rhesus macaques and humans may reveal if macaques are ideal models for the study of some human traits while being suboptimal for the study of others. Since biomedical research depends heavily on the suitability of animal models for specific studies, a better understanding of the genomic differences between orthologous CNVs of humans and macaques may provide insight on the utility of macaques as models for studying specific human traits.

2. Results

A total of 230,973 CNV markers on the Affymetrix® Genome-Wide Human SNP Array 6.0 were successfully mapped to both human (hg19) and Indian rhesus macaque (rheMac2) genomes. Seven hundred and nine of these markers mapped to multiple sites within the rheMac2 genome. The chromosomal distribution of the markers across both human and Indian rhesus macaque genomes is summarized in Table 1 and the generalized synteny of Indian rhesus macaque chromosomal segments relative to human chromosomes is shown in Fig. 1a. Though the majority of the human markers were mapped to the rhesus macaque chromosome orthologs that Rogers et al. [23] identified using STRs, suggesting generally conserved synteny across the genomes, a few markers did not (the single colored bands within the larger colored blocks), which could indicate genomic rearrangement. From the 230,973 human markers that mapped to the Indian rhesus macaque, 109,044 mapped to 30,642 unique NCBI RefSeq accession numbers and 14,466 unique annotated genes in human, and 24,666 markers mapped to 3809 unique accession numbers and 3782 genes in Indian rhesus macaque.

Similarly, comparing the human (hg19) and Chinese rhesus macaque (rheMac3) genomes yielded a total of 221,268 markers that mapped to both genomes with 951 CNV markers mapping to multiple locations and the majority of markers mapping to their orthologous rhesus macaque chromosomes. The probe distribution across the Chinese rhesus macaque and human chromosomes is shown in Table 1 and the conserved synteny is illustrated in Fig. 1b. While most human markers mapped to their orthologous rhesus macaque chromosomes similar to that of Indian rhesus macaques, more non-orthologous markers were observed in the Chinese rhesus macaque than the Indian. Of the 221,268 markers mapped to both genomes, 105,248 mapped to 30,282 unique accession numbers and 14,337 unique annotated genes in human, and 23,378 markers mapped to 3731 unique accession numbers and 3704 genes in Chinese rhesus macaque.

The comparison between the two rhesus macaques resulted in 205,532 markers that mapped to the genomes of both rhesus macaques and 1146 makers that mapped to multiple locations. From the 205,532 markers that mapped to both Indian and Chinese rhesus macaques, 22,385 markers mapped to 3709 unique accession numbers and 3682 genes in Indian rhesus macaque, and 22,101 markers mapped to 3675 accession numbers and 3650 genes in Chinese rhesus macaque. All markers with their mapped annotated genes for all comparisons are summarized in Table 2 and listed in detail for the Indian rhesus-human, Chinese rhesus-human, and Chinese–Indian rhesus comparisons in Supplementary Files 1–3, respectively.

Chromosomal rearrangements, such as the fusion of rhesus macaque chromosomes 12 and 13 that are orthologous to human chromosome 2, were in agreement with previous observations in great apes [26], baboons [27], and Japanese macaques [28]. Human chromosomes 5, 12, and X and their orthologous rhesus macaque chromosomes showed no microinversions or rearrangements in either the Indian or the Chinese rhesus macaque while human chromosome 19 also showed no rearrangement in the Chinese rhesus macaque. The chromosomal fission of human chromosome 2 corresponding to the orthologous rhesus macaque chromosomes 12 and 13 mentioned above and the chromosomal fusions of human chromosomes 7 and 21 to rhesus macaque chromosome 3, human 14 and 15 to rhesus macaque 7, and human 20 and 22 to rhesus macaque 10 were observed and consistent with previous studies [17,23,24,26,28,29]. Despite being of the same species, rearrangements, albeit small compared to those observed between human and rhesus macaque, were observed in chromosomes 2, 4, 5, 7, 10, 14, 15, 16, 19, 20, and X in the Indian and Chinese rhesus macaque. Fig. 2 shows the large-scale chromosomal rearrangements and synteny observed in rhesus macaque chromosome 10 in both rhesus macaque reference genomes (Fig. 2a: Indian rheMac2; 2b: Chinese rheMac3) in relation to the human genome (hg19) and between the two rhesus macaques (Fig. 2c) for all mapped and annotated genes. Similar genomic rearrangements and regional inversions were observed in both Indian and Chinese rhesus macaques as compared to human with the exception of a single inversion block at the end of rhesus chromosome 10 and human chromosome 22 (Fig. 2b). Although the Chinese rhesus macaque-human inversion above was not observed when the two rhesus macaques were compared, a gene rearrangement was detected, indicating that structural genomic variation exists within the rhesus macaque species. Chromosomal rearrangement and microinversions for each comparison are shown in Supplementary Files 1–3 for Indian rhesus-human, Chinese rhesus-human, and Indian-Chinese rhesus macaques, respectively.

Copy number comparison between Indian rhesus macaque and human identified 1679 copy number regions with 206 regions that had copy number differences greater than or less than a factor of two ("doubled regions"). When comparing Chinese rhesus macaques to humans, the number of total copy number regions decreased to 305 and doubled regions to 32. Comparisons of the two rhesus macaques yielded the greatest number of total copy number regions (2725), but fewer doubled regions (147) than the comparison between Indian rhesus macaque and human.

The double region comparison for Indian rhesus-human resulted in 885 unique accession numbers and 407 unique genes in human and 96 unique accession numbers and genes in Indian rhesus macaque. The doubled regions for the Chinese rhesus-human comparison mapped to 308 unique RefSeq accession numbers and 154 unique genes in human and 26 unique accession numbers and genes in Chinese rhesus macaque. Lastly, despite having the most copy number regions, the Chinese rhesus-Indian rhesus comparison only had 17 unique accession numbers and genes that mapped to the doubled regions of the Indian rhesus macaque and 13 accession numbers and genes to the Chinese rhesus macaque. Of the genes found in the doubled regions for each comparison pair, 66 genes were found in both Indian rhesus macaque and human, eight genes between Chinese rhesus macaque and human, and 13 in both the Chinese and Indian rhesus macaques. All doubled region information, including accession numbers, genes, and counts in log base two, for all comparisons is summarized in Table 3 and listed in detail in Supplementary File 4.

3. Discussion

Using human markers to investigate the rhesus macaque genomes has resulted in the identification of over 200,000 sites that are

Table 1
Probe distribution across the Indian (upper half) and Chinese rhesus macaque (lower half) in comparison to human chromosomes. Bolded numbers indicate probes in orthologous chromosomes. For example, human chromosome 2 corresponds to rhesus macaque chromosomes 12 and 13.

		Human																						
		chr1	chr2	chr3	chr4	chr5	chr6	chr7	chr8	chr9	chr10	chr11	chr12	chr13	chr14	chr15	chr16	chr17	chr18	chr19	chr20	chr21	chr22	chrX	chrY
Indian Rhesus Macaque	chr1	**31244**	5	9	347	5	16	38	14	11	7	13	1	1	6	5	8	27	0	4	0	6	13	13	7
	chr2	21	7	**26221**	9	13	4	3	5	13	3	2	15	16	2	3	1	6	5	9	3	1	8	6	1
	chr3	40	22	9	4	13	2	**20860**	2	16	7	16	2	2	7	9	13	0	0	11	41	**4487**	2	20	0
	chr4	4	6	5	10	19	**23085**	4	1	10	10	15	28	23	4	4	0	4	2	2	34	1	6	0	1
	chr5	26	26	8	**20488**	2	6	17	3	7	13	6	15	1	28	2	5	6	1	2	1	0	1	8	1
	chr6	18	11	18	3	**23282**	6	10	18	6	36	4	14	5	3	17	6	5	2	3	0	1	0	1	2
	chr7	28	10	23	5	0	7	1	2	3	4	7	14	8	**12051**	**11333**	17	0	0	2	0	10	3	7	6
	chr8	29	7	12	8	5	1	10	**18467**	13	14	2	8	5	6	7	0	1	7	3	0	2	0	9	0
	chr9	10	3	3	0	14	40	1	1	15	**22042**	4	11	3	7	9	2	8	0	0	1	4	0	2	1
	chr10	5	19	3	4	0	26	6	1	2	8	7	3	16	8	1	8	11	2	0	**8052**	7	**4389**	0	0
	chr11	4	9	677	18	12	19	2	3	3	2	0	**18349**	2	10	0	4	0	1	3	3	1	9	17	0
	chr12	7	**15448**	2	1	9	6	8	3	2	12	20	3	0	2	1	0	4	0	2	4	5	4	1	0
	chr13	11	**15898**	4	11	17	16	4	4	2	19	2	5	0	1	10	6	0	3	2	0	4	14	5	0
	chr14	4	15	6	9	5	0	3	8	2	**4**	**18068**	0	4	0	10	2	7	0	0	0	2	7	2	0
	chr15	2	5	4	10	7	26	4	7	**16693**	5	2	0	4	3	9	1	3	5	4	1	1	0	4	0
	chr16	9	0	5	0	3	18	10	1	0	2	4	**4**	0	0	0	1	**9327**	0	2	3	0	1	1	0
	chr17	3	11	3	2	11	0	0	3	1	0	2	1	**10217**	0	24	0	0	1	2	0	0	2	2	1
	chr18	3	0	3	39	3	0	0	1	2	0	17	0	3	0	1	1	0	**9881**	0	0	0	1	2	1
	chr19	35	3	9	7	6	1	7	5	7	2	3	9	5	8	8	4	1	1	**3973**	0	25	0	2	0
	chr20	4	5	2	2	1	2	0	1	1	6	1	1	2	4	6	**9151**	3	0	2	0	20	6	2	0
	chrX	17	25	12	4	2	25	11	3	4	20	15	1	9	13	4	2	4	10	7	3	3	0	**17957**	213
Chinese Rhesus Macaque	chr1	**17837**	10	4	154	6	11	28	4	11	7	5	4	1	7	4	0	8	1	2	3	3	6	6	5
	chr2	8	6	**15178**	10	32	30	4	9	9	3	4	9	3	2	4	1	3	3	4	2	2	5	170	2
	chr3	33	7	76	3	5	15	**11938**	3	9	7	4	4	4	3	5	7	0	32	13	2	**2643**	2	3	0
	chr4	4	34	1	4	43	**13628**	9	13	2	7	0	2	18	1	2	1	6	16	8	1	2	2	5	2
	chr5	8	11	2	**13228**	75	5	12	4	2	9	2	5	1	39	2	6	3	1	2	1	25	1	15	1
	chr6	23	9	69	1	**14123**	5	7	4	9	8	1	2	3	4	5	61	6	2	3	0	0	0	17	2
	chr7	10	7	8	16	10	6	2	3	2	0	8	86	2	**6949**	**6535**	6	1	53	1	1	4	2	21	1
	chr8	9	4	4	5	6	1	12	**11108**	1	6	121	2	3	27	3	0	0	1	1	0	3	1	6	1
	chr9	1	5	2	0	6	9	1	2	2	**11202**	4	6	2	3	3	0	8	0	2	1	2	1	1	1
	chr10	3	4	5	2	0	3	7	0	3	2	11	1	15	1	0	20	1	2	2	**4981**	1	**2339**	116	0
	chr11	2	5	282	3	7	3	3	1	2	4	1	**10341**	3	2	2	1	0	1	1	2	0	2	11	0
	chr12	2	**9023**	2	1	4	0	1	2	3	5	62	4	1	3	2	0	3	6	2	1	2	2	2	0
	chr13	124	**10035**	12	3	6	6	7	3	3	3	38	5	0	1	8	2	0	44	1	1	3	4	2	0
	chr14	97	3	33	5	32	0	4	4	7	5	**9981**	1	2	0	2	1	2	37	1	0	0	5	1	0
	chr15	1	5	6	6	1	1	1	2	**9874**	2	75	2	5	3	2	2	1	1	2	2	1	1	4	0
	chr16	4	1	4	0	3	4	10	2	0	2	2	2	0	0	72	1	**5596**	3	1	1	6	1	1	0
	chr17	1	5	1	1	2	2	1	1	2	0	1	7	**7465**	3	10	0	0	0	1	1	1	1	0	1
	chr18	3	0	2	4	1	1	1	2	2	0	2	0	4	60	1	3	0	**6098**	0	0	0	0	1	0
	chr19	4	4	1	2	0	2	5	3	3	2	3	3	1	1	3	3	4	1	**2423**	1	2	1	3	0
	chr20	4	1	2	4	1	18	0	2	2	0	2	2	0	0	4	**5898**	1	0	1	0	0	1	2	0
	chrX	12	30	13	4	2	46	15	3	2	8	5	1	5	2	4	1	22	16	4	8	3	1	**9968**	55

complementary to orthologous probes. The high number of successful orthologous probes also suggests that rhesus macaques have high levels of sequence identity and genetic similarity to humans. Mapping of the orthologous probes provided more detailed observations of synteny between humans and rhesus macaques and between the two geographically distinct rhesus macaques. The genomic rearrangement observed in both human–rhesus macaque comparisons was in agreement with previous findings [17,23,26–28] and provided finer mapping results that could be added to previous comparative genomic maps of the two species. As the genetic maps reported here were based on annotated gene builds and were comprehensive, the results should be verified for particular regions of interest since available gene annotations have been known to contain errors [30]. Zimin et al. [31] independently reconstructed and revised the rheMac2 genome annotation, which may have corrected some of those errors. However, that study did not address potential issues concerning rheMac3. While it could have been more appropriate to include Zimin et al.'s [31] corrections, at present, that is not possible as their information has neither been incorporated into the NCBI database nor built into the UCSC genome browser.

Despite this, the results from this study have identified regions containing CNVs in rhesus macaques that may not have been previously located. These copy number regions were mapped to many annotated genes, some of which are associated in disease susceptibility such as the phosphate and tensin homolog (*PTEN*), cancer susceptibility candidate 3 (*CASC3*), and B-cell lymphoma 6 (*BCL6*), which are involved in the cell cycle and tumor suppression [32], breast carcinoma [33], and non-Hodgkin lymphoma [34], respectively. A more extensive list of genes and disease associations is given in Table 4. Since only two rhesus macaques were genotyped, the number of copy number regions and "doubled regions" was probably underestimated. A more thorough assessment of CNVs in several rhesus macaque samples from different geographic populations is needed to provide better estimates of the number of CNV regions in this species' genome and the copy number at each CNV. This can be achieved through similar analyses as those done by Ng et al. [35] who used the orthologous SNPs from Kanthaswamy et al.'s [17] study to design a custom SNP array and genotype multiple rhesus macaques so population genetic comparisons could be made.

Although the copy number regions identified here may indicate that a gene has a greater copy number in humans than in rhesus macaques, the gene must first be located in rhesus macaques and the copy number determined to verify the species-specific copy number differences. For example, the histocompatibility antigen gene *HLA-B* exhibited over two times the copy number in humans, but was not located in any copy number regions in either rhesus macaque. This was due to the potential orthologous gene in rhesus macaques being identified as *LOC714964*, a gene that is like *HLA-B* but not yet confirmed as orthologous. In addition to genes that have yet to be annotated in the rhesus macaque genome, other factors may dictate why some genes are located in copy number regions in humans but not in rhesus macaques. These discrepancies may be due to the gene being absent in rhesus macaques or not mapping to a copy number region complimentary to the probes used in rhesus macaques. For studies involving CNVs based on these or other results, sequencing of the copy number variant regions or genes should be done to precisely determine the copy number in the samples and verify the correct annotation of genes.

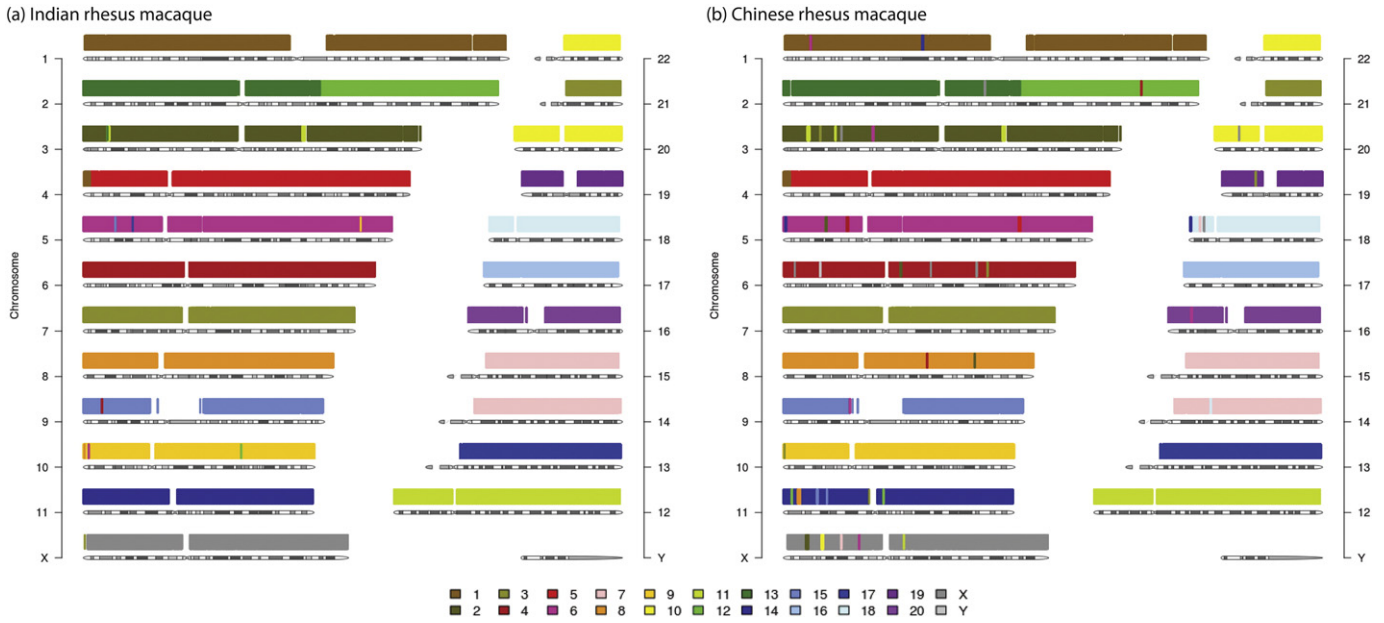

Fig. 1. General conserved synteny of (a) Indian rhesus macaque (rheMac2) and (b) Chinese rhesus macaque (rheMac3) reference genome chromosomal segments relative to human (hg19) chromosomes. The colored bands (above) show orthologous rhesus macaque chromosome segments, with each color corresponding to a different rhesus macaque chromosome, and how they distribute across the human genome represented with G-banded chromosomes (below).

Though rhesus macaques are one of the closest model organisms to humans, the available genetic data repositories are not as extensive or nearly as detailed as other more distantly related model organisms such as mice or yeast. The level of genetic conservation between human and macaques has yet to be fully investigated, but the results from this study will be useful for all biomedical research. The use of orthologous probes provides information regarding the genetic similarities between the model organism and humans and highlights potential markers of interest, especially those known to be associated with human traits and diseases. As with SNPs, CNVs are often associated with traits and diseases, and greater knowledge about these variants in rhesus macaques will be beneficial for studies utilizing the rhesus macaque model.

4. Materials and methods

4.1. CNV probe hybridization

Genomic DNA from one female Chinese and one female Indian rhesus macaque obtained from the California National Primate Research Center and genomic DNA from one human female purchased from Zyagen (San Diego, CA) were hybridized to the Affymetrix® Genome-Wide Human SNP Array 6.0 following manufacturer protocols. The array was processed and all probes assessed and validated following previously published [17] procedures.

Table 2
Summary of all the markers mapped in each comparison and their respective species annotations.

	Comparisons					
	Indian	Human	Chinese	Human	Chinese	Indian
Total markers found in both species	230,973		221,268		205,532	
Repeated markers found in both species	709		951		1146	
Markers mapping to annotations	24,666	109,044	23,378	105,248	22,101	22,385
Number of RefSeq accession numbers associated with the markers	3809	30,642	3731	30,282	3675	3709
Number of genes associated with the markers	3782	14,466	3704	14,337	3650	3682

4.2. Orthologous probe identification and validation

The probe sequences and their corresponding coordinates in the UCSC Genome Browser's (https://genome.ucsc.edu/) [36] human genome (version hg19) were downloaded from the Affymetrix website (http://www.affymetrix.com/) and used to check that probe sequences aligned uniquely and perfectly to the rhesus macaque genomes. All probe sequences were aligned using BWA 0.7.4 [37], SAMtools 0.1.19 [38], and R [39] against both rhesus macaque genome references (UCSC rhesus genome versions rheMac2: Indian rhesus macaque; rheMac3: Chinese rhesus macaque) with all known repeats masked using RepeatMasker [40] to allow for zero mismatches and only perfect alignments. Only alignments with mapping qualities greater than zero were used and separate lists were compiled for CNV probes that aligned to either a single, multiple, or no target region in each rhesus macaque genome. Probes were then confirmed against the UCSC Genome Browser's [36] Primate Chain/Net Comparative Genomics track, specifically the nets [41] for the rhesus macaque to human comparisons (rheMac2-hg19 and rheMac3-hg19), to be in the same conserved blocks.

4.3. Copy number estimation

Copy number estimation was conducted using the algorithm CRMA v2 [42] on human genomic positions (hg19). Smoothed regions with copy number changes in a given sample compared to a reference sample (Chinese rhesus to human, Chinese rhesus to Indian rhesus, and Indian rhesus to human) were estimated using circular binary segmentation [43,44]. Raw and smoothed copy number estimation were both conducted using the R package aroma.affymetrix, version 2.10.1 [45]. The UCSC Genome Browser's [36] LiftOver tool [46] was used to convert the CNV regions from the human genome (hg19) positions to the Indian and Chinese rhesus macaque genome (rheMac2 and rheMac3, respectively) positions for human–rhesus macaque comparisons. For comparisons between the two macaques, Chinese rhesus macaque (rheMac3) positions were converted from Indian rhesus macaque (rheMac2) positions using LiftOver [46] instead of directly from human (hg19) so converted regions could be tracked between the genome builds. Only copy

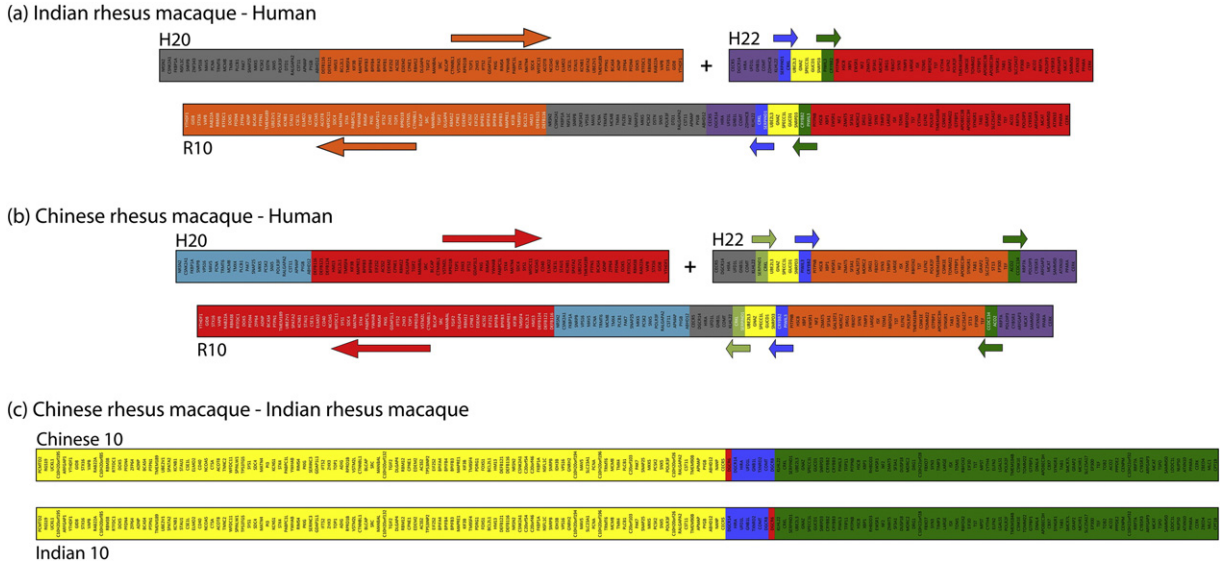

Fig. 2. Observed genetic rearrangement of rhesus macaque (R) chromosome 10 as compared to orthologous human (H) chromosomes 20 and 22 for (a) Indian rhesus macaque and human, (b) Chinese rhesus macaque and human, and (c) between the Indian and Chinese rhesus macaques. The colors indicate different gene blocks and arrows identify blocks with genomic inversions that were observed using the orthologous probes. Since each comparison set was done separately using markers found in both individuals compared, the number of genes identified differed and the color blocks do not represent the same segments between (a–c).

number regions that varied by at least a factor of two were used for gene copy number comparison between the species.

4.4. Probe annotation mapping

All probes were mapped to the human (hg19) and each rhesus macaque (rheMac2 and rheMac3) RefSeq Genes database downloaded from the UCSC Table Browser [36,47] using BEDTools 2.22.0 [48] for gene annotations. Probes identified in annotated genes were ordered according to their chromosomal positions and compared between the orthologous chromosomes of each pair of species. Genes found in each pair of species were used to detect genomic rearrangements and inversions in the orthologous chromosomes. Gene annotations for probes with copy number differences of greater than or equal to two or less than or equal to one-half were determined to identify potential genes with copy number differences between the pair of species.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2015.09.016.

Acknowledgments

This study was supported by the California National Primate Research Center base grant (No. RR000169-48), by an ARRA supplement awarded to S.K. and Nick Lerche, CNPRC (No. RR018144-07), NIH grants RR005090 and RR025871 to D.G.S., and the UC Davis Henry A. Jastro Graduate Research Award to J.N. Animals used in this research were managed in compliance with Institutional Animal Care and Use Committee (IACUC) regulations or in accordance with the National Institutes of Health guidelines or the US Department of Agriculture regulations prescribing the humane care and use of laboratory animals.

Table 3
Summary of the copy number regions and gene annotations found in all comparisons.

	Comparisons					
	Indian	Human	Chinese	Human	Indian	Chinese
Total copy number regions found in both species	1679		305		2725	
"Double regions" where the copy number in one species that was at least twice that of the other	206		32		147	
Number of RefSeq accession numbers associated with copy number regions	96	885	26	308	17	13
Number of genes associated with copy number regions	96	407	26	154	17	13
Number of genes in "doubled regions"	66		8		13	

Table 4
Example of identified genes with their disease associations.

Gene Name	OMIM ID	Disease association
BCL6	109565	B-cell lymphoma
CASC3	606504	Breast cancer
DEK	125264	Acute nonlymphocytic leukemia
EIF2B2	606454	Leukoencephalopathy
EP300	602700	Colorectal cancer
EPHB6	602757	Neuroblastoma
FAM134B	613114	Neuropathy
FGFR1	136350	Osteoglophonic dysplasia
FZD4	604579	Exudative vitreoretinopathy
ICA1	147625	Diabetes
IL18	600953	Coronary artery disease
IL2RG	308380	Severe X-linked combined immunodeficiency
KIT	164920	Piebaldism
LMO4	603129	Breast cancer
MIR10B	611576	Breast cancer
MLH3	604395	Colorectal cancer
ORMDL3	610075	Asthma
PEX3	603164	Peroxisome biogenesis disorder
PRKRA	603424	Dystonia
PTEN	601728	Endometrial carcinoma
PTPN1	176885	Diabetes
PTS	612719	Hyperphenylalaninemia
SLC25A3	600370	Mitochondrial phosphate carrier deficiency
TERC	602322	Dyskeratosis congenita
TMEM165	614726	Congenital disorder of glycosylation
WDR11	606417	Hypogonadotropic hypogonadism

References

[1] S. Kumar, S.B. Hedges, A molecular timescale for vertebrate evolution. Nature 392 (1998) 917–920.
[2] R.L. Coffman, E.M. Hessel, Nonhuman primate models of asthma. J. Exp. Med. 201 (2005) 1875–1879.
[3] M.B. Gardner, P.A. Luciw, Macaque models of human infectious disease. ILAR J. Natl. Res. Counc. Inst. Lab. Anim. Resour. 49 (2008) 220–255.
[4] K.A. Phillips, K.L. Bales, J.P. Capitanio, A. Conley, P.W. Czoty, B.A. 't Hart, W.D. Hopkins, S.-L. Hu, L.A. Miller, M.A. Nader, P.W. Nathanielsz, J. Rogers, C.A. Shively, M.L. Voytko, Why primate models matter. Am. J. Primatol. 76 (2014) 801–827.
[5] H.E. Carlsson, S.J. Schapiro, I. Farah, J. Hau, Use of primates in research: a global overview. Am. J. Primatol. 63 (2004) 225–237.
[6] S.A. Sharpe, H. McShane, M.J. Dennis, R.J. Basaraba, F. Gleeson, G. Hall, A. McIntyre, K. Gooch, S. Clark, N.E.R. Beveridge, E. Nuth, A. White, A. Marriott, S. Dowall, A.V.S. Hill, A. Williams, P.D. Marsh, Establishment of an aerosol challenge model of tuberculosis in rhesus macaques and an evaluation of endpoints for vaccine testing. Clin. Vaccine Immunol. 17 (2010) 1170–1182.
[7] A.A. Bremer, K.L. Stanhope, J.L. Graham, B.P. Cummings, W. Wang, B.R. Saville, P.J. Havel, Fructose-fed rhesus monkeys: a nonhuman primate model of insulin resistance, metabolic syndrome, and type 2 diabetes. Clin. Transl. Sci. 4 (2011) 243–252.
[8] G.S. Roth, J.A. Mattison, M.A. Ottinger, M.E. Chachich, M.A. Lane, D.K. Ingram, Aging in rhesus monkeys: relevance to human health interventions. Science 305 (2004) 1423–1426.
[9] Rhesus Macaque Genome Sequencing Analysis Consortium, R.A. Gibbs, J. Rogers, M.G. Katze, R. Bumgarner, G.M. Weinstock, E.R. Mardis, K.A. Remington, R.L. Strausberg, J.C. Venter, R.K. Wilson, M.A. Batzer, C.D. Bustamante, E.E. Eichler, M.W. Hahn, R.C. Hardison, K.D. Makova, W. Miller, A. Milosavljevic, R.E. Palermo, A. Siepel, J.M. Sikela, T. Attaway, S. Bell, K.E. Bernard, C.J. Buhay, M.N. Chandrabose, M. Dao, C. Davis, K.D. Delehaunty, Y. Ding, H.H. Dinh, S. Dugan-Rocha, L.A. Fulton, R.A. Gabisi, T.T. Garner, J. Godfrey, A.C. Hawes, J. Hernandez, S. Hines, M. Holder, J. Hume, S.N. Jhangiani, V. Joshi, Z.M. Khan, E.F. Kirkness, A. Cree, R.G. Fowler, S. Lee, L.R. Lewis, Z. Li, Y.S. Liu, S.M. Moore, D. Muzny, L.V. Nazareth, D.N. Ngo, G.O. Okwuonu, G. Pai, D. Parker, H.A. Paul, C. Pfannkoch, C.S. Pohl, Y.H. Rogers, S.J. Ruiz, A. Sabo, J. Santibanez, B.W. Schneider, S.M. Smith, E. Sodergren, A.F. Svatek, T.R. Utterback, S. Vattathil, W. Warren, C.S. White, A.T. Chinwalla, Y. Feng, A.L. Halpern, L.W. Hillier, X. Huang, P. Minx, J.O. Nelson, K.H. Pepin, X. Qin, G.G. Sutton, E. Venter, B.P. Walenz, J.W. Wallis, K.C. Worley, S.P. Yang, S.M. Jones, M.A. Marra, M. Rocchi, J.E. Schein, R. Baertsch, L. Clarke, M. Csuros, J. Glasscock, R.A. Harris, P. Havlak, A.R. Jackson, H. Jiang, Y. Liu, D.N. Messina, Y. Shen, H.X. Song, T. Wylie, L. Zhang, E. Birney, K. Han, M.K. Konkel, J. Lee, A.F. Smit, B. Ullmer, H. Wang, J. Xing, R. Burhans, Z. Cheng, J.E. Karro, J. Ma, B. Raney, X. She, M.J. Cox, J.P. Demuth, L.J. Dumas, S.G. Han, J. Hopkins, A. Karimpour-Fard, Y.H. Kim, J.R. Pollack, T. Vinar, C. Addo-Quaye, J. Degenhardt, A. Denby, M.J. Hubisz, A. Indap, C. Kosiol, B.T. Lahn, H.A. Lawson, A. Marklein, R. Nielsen, E.J. Vallender, A.G. Clark, B. Ferguson, R.D. Hernandez, K. Hirani, H. Kehrer-Sawatzki, J. Kolb, S. Patil, L.L. Pu, Y. Ren, D.G. Smith, D.A. Wheeler, I. Schenck, E.V. Ball, R. Chen, D.N. Cooper, B. Giardine, F. Hsu, W.J. Kent, A. Lesk, D.L. Nelson, W.E. O'Brien, K. Prufer, P.D. Stenson, J.C. Wallace, H. Ke, X.M. Liu, P. Wang, A.P. Xiang, F. Yang, G.P. Barber, D. Haussler, D. Karolchik, A.D. Kern, R.M. Kuhn, K.E. Smith, A.S. Zwieg, Evolutionary and biomedical insights from the rhesus macaque genome. Science 316 (2007) 222–234.
[10] J. Fooden, Systematic review of the rhesus macaque, *Macaca mulatta* (Zimmermann 1780). Fieldiana Zool. 96 (2000) 1–180.
[11] S. Kanthaswamy, L. Gill, J. Satkoski, V. Goyal, V. Malladi, A. Kou, K. Basuta, L. Sarkisyan, D. George, D.G. Smith, Development of a Chinese–Indian hybrid (Chindian) rhesus macaque colony at the California National Primate Research Center by introgression. J. Med. Primatol. 38 (2009) 86–96.
[12] B. Ferguson, S.L. Street, H. Wright, C. Pearson, Y. Jia, S.L. Thompson, P. Allibone, C.J. Dubay, E. Spindel, R.B. Norgren Jr., Single nucleotide polymorphisms (SNPs) distinguish Indian-origin and Chinese-origin rhesus macaques (Macaca mulatta). BMC Genomics 8 (2007) 43.
[13] S. Kanthaswamy, J. Satkoski, A. Kou, V. Malladi, D. Glenn Smith, Detecting signatures of inter-regional and inter-specific hybridization among the Chinese rhesus macaque specific pathogen-free (SPF) population using single nucleotide polymorphic (SNP) markers. J. Med. Primatol. 39 (2010) 252–265.
[14] S. Kanthaswamy, D.G. Smith, Effects of geographic origin on captive Macaca mulatta mitochondrial DNA variation. Comp. Med. 54 (2004) 193–201.
[15] G. Yan, G. Zhang, X. Fang, Y. Zhang, C. Li, F. Ling, D.N. Cooper, Q. Li, Y. Li, A.J. van Gool, H. Du, J. Chen, R. Chen, P. Zhang, Z. Huang, J.R. Thompson, Y. Meng, Y. Bai, J. Wang, M. Zhuo, T. Wang, Y. Huang, L. Wei, J. Li, Z. Wang, H. Hu, P. Yang, L. Le, P.D. Stenson, B. Li, X. Liu, E.V. Ball, N. An, Q. Huang, Y. Zhang, W. Fan, X. Zhang, Y. Li, W. Wang, M.G. Katze, B. Su, R. Nielsen, H. Yang, J. Wang, X. Wang, J. Wang, Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. Nat. Biotechnol. 29 (2011) 1019–1023.
[16] Q. Yuan, Z. Zhou, S. Lindell, J. Higley, B. Ferguson, R. Thompson, J. Lopez, S. Suomi, B. Baghal, M. Baker, D. Mash, C. Barr, D. Goldman, The rhesus macaque is three times as diverse but more closely equivalent in damaging coding variation as compared to the human. BMC Genet. 13 (2012) 52.
[17] S. Kanthaswamy, J. Ng, C.T. Ross, J.S. Trask, D.G. Smith, V.S. Buffalo, J.N. Fass, D. Lin, Identifying human–rhesus macaque gene orthologs using heterospecific SNP probes. Genomics 101 (2013) 30–37.
[18] E. Gonzalez, H. Kulkarni, H. Bolivar, A. Mangano, R. Sanchez, G. Catano, R.J. Nibbs, B.I. Freedman, M.P. Quinones, M.J. Bamshad, K.K. Murthy, B.H. Rovin, W. Bradley, R.A. Clark, S.A. Anderson, R.J. O'Connell, B.K. Agan, S.S. Ahuja, R. Bologna, L. Sen, M.J. Dolan, S.K. Ahuja, The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307 (2005) 1434–1440.
[19] J.D. Degenhardt, P. de Candia, A. Chabot, S. Schwartz, L. Henderson, B. Ling, M. Hunter, Z. Jiang, R.E. Palermo, M. Katze, E.E. Eichler, M. Ventura, J. Rogers, P. Marx, Y. Gilad, C.D. Bustamante, Copy number variation of CCL3-like genes affects rate of progression to simian-AIDS in Rhesus Macaques (Macaca mulatta). PLoS Genet. 5 (2009), e1000346.
[20] The International HapMap Consortium, The International HapMap Project. Nature 426 (2003) 789–796.
[21] The International HapMap Consortium, A haplotype map of the human genome. Nature 437 (2005) 1299–1320.
[22] The International HapMap Consortium, A second generation human haplotype map of over 3.1 million SNPs. Nature 449 (2007) 851–861.
[23] J. Rogers, R. Garcia, W. Shelledy, J. Kaplan, A. Arya, Z. Johnson, M. Bergstrom, L. Novakowski, P. Nair, A. Vinson, D. Newman, G. Heckman, J. Cameron, An initial genetic linkage map of the rhesus macaque (Macaca mulatta) genome using human microsatellite loci. Genomics 87 (2006) 30–38.
[24] P.L. Pearson, T.H. Roderick, M.T. Davisson, J.J. Garver, D. Warburton, P.A. Lalley, S.J. O'Brien, Report of the committee on comparative mapping. Cytogenet. Genome Res. 25 (1979) 82–95.
[25] R. Roberto, D. Misceo, P. D'Addabbo, N. Archidiacono, M. Rocchi, Refinement of macaque synteny arrangement with respect to the official rheMac2 macaque sequence assembly. Chromosom. Res. 16 (2008) 977–985.
[26] R.G. Best, D. Diamond, E. Crawford, F.S. Grass, C. Janish, T.L. Lear, D. Soenksen, A.A. Szalay, C.M. Moore, Baboon/human homologies examined by spectral karyotyping (SKY): a visual comparison. Cytogenet. Cell Genet. 82 (1998) 83–87.
[27] J. Yunis, O. Prakash, The origin of man: a chromosomal pictorial legacy. Science 215 (1982) 1525–1530.
[28] J. Wienberg, R. Stanyon, A. Jauch, T. Cremer, Homologies in human and Macasa fuscata chromosomes revealed by in situ suppression hybridization with human chromosome specific DNA libraries. Chromosoma 101 (1992) 265–270.
[29] W. Murphy, R. Stanyon, S. O'Brien, Evolution of mammalian genome organization inferred from comparative gene mapping. Genome Biol. 2 (2001) (reviews0005.0001 - reviews0005.0008).
[30] X. Zhang, J. Goodsell, R. Norgren, Limitations of the rhesus macaque draft genome assembly and annotation. BMC Genomics 13 (2012) 206.
[31] A. Zimin, A. Cornish, M. Maudhoo, R. Gibbs, X. Zhang, S. Pandey, D. Meehan, K. Wipfler, S. Bosinger, Z. Johnson, G. Tharp, G. Marcais, M. Roberts, B. Ferguson, H. Fox, T. Treangen, S. Salzberg, J. Yorke, Norgren Robert, A new rhesus macaque assembly and annotation for next-generation sequencing analyses. Biol. Direct 9 (2014) 20.
[32] D.M. Li, H. Sun, PTEN/MMAC1/TEP1 suppresses the tumorigenicity and induces G1 cell cycle arrest in human glioblastoma cells. Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 15406–15411.
[33] C. Tomasetto, C. Régnier, C. Moog-Lutz, M.G. Mattei, M.P. Chenard, R. Lidereau, P. Basset, M.C. Rio, Identification of four novel human genes amplified and overexpressed in breast carcinoma and localized to the q11–q21.3 region of chromosome 17. Genomics 28 (1995) 367–376.
[34] S.R. Chaganti, W. Chen, N. Parsa, K. Offit, D.C. Louie, R. Dalla-Favera, R.S.K. Chaganti, Involvement of BCL6 in chromosomal aberrations affecting band 3q27 in B-cell non-Hodgkin lymphoma. Genes Chromosom. Cancer 23 (1998) 323–327.
[35] J. Ng, J.S. Trask, D.G. Smith, S. Kanthaswamy, Heterospecific SNP diversity in humans and rhesus macaque (Macaca mulatta). J. Med. Primatol. 44 (2015) 194–201.
[36] W.J. Kent, C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, Haussler, David, The human genome browser at UCSC. Genome Res. 12 (2002) 996–1006.
[37] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25 (2009) 1754–1760.
[38] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map format and SAMtools. Bioinformatics 25 (2009) 2078–2079.
[39] R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012.
[40] A. Smit, R. Hubley, P. Green, RepeatMasker Open-3.0, in. http://www.repeatmasker.org1996–2010.
[41] W.J. Kent, R. Baertsch, A. Hinrichs, W. Miller, D. Haussler, Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 11484–11489.
[42] H. Bengtsson, P. Wirapati, T.P. Speed, A single-array preprocessing method for estimating full-resolution raw copy numbers from all Affymetrix genotyping arrays including GenomeWideSNP 5 & 6. Bioinformatics 25 (2009) 2149–2156.
[43] A.B. Olshen, E.S. Venkatraman, R. Lucito, M. Wigler, Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics (Oxford, England) 5 (2004) 557–572.
[44] E.S. Venkatraman, A.B. Olshen, A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics 23 (2007) 657–663.
[45] H. Bengtsson, K. Simpson, J. Bullard, K. Hansen, aroma.affymetrix: a generic framework in R for analyzing small to very large Affymetrix data sets in bounded memory. Tech Report #745, Department of Statistics, University of California, Berkeley, 2008.
[46] A.S. Hinrichs, D. Karolchik, R. Baertsch, G.P. Barber, G. Bejerano, H. Clawson, M. Diekhans, T.S. Furey, R.A. Harte, F. Hsu, J. Hillman-Jackson, R.M. Kuhn, J.S. Pedersen, A. Pohl, B.J. Raney, K.R. Rosenbloom, A. Siepel, K.E. Smith, C.W. Sugnet, A. Sultan-Qurraie, D.J. Thomas, H. Trumbower, R.J. Weber, M. Weirauch, A.S. Zweig, D. Haussler, W.J. Kent, The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 34 (2006) D590–D598.
[47] D. Karolchik, A.S. Hinrichs, T.S. Furey, K.M. Roskin, C.W. Sugnet, D. Haussler, W.J. Kent, The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 32 (2004) D493–D496.
[48] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26 (2010) 841–842.