nature genetics

Distribution and functional impact of DNA copy number variation in the rat

Victor Guryev[1], Kathrin Saar[2], Tatjana Adamovic[3], Mark Verheul[1], Sebastiaan A A C van Heesch[1], Stuart Cook[4,5], Michal Pravenec[6,7], Timothy Aitman[4], Howard Jacob[3], James D Shull[8], Norbert Hubner[2] & Edwin Cuppen[1]

The abundance and dynamics of copy number variants (CNVs) in mammalian genomes poses new challenges in the identification of their impact on natural and disease phenotypes. We used computational and experimental methods to catalog CNVs in rat and found that they share important functional characteristics with those in human. In addition, 113 one-to-one orthologous genes overlap CNVs in both human and rat, 80 of which are implicated in human disease. CNVs are nonrandomly distributed throughout the genome. Chromosome 18 is a cold spot for CNVs as well as evolutionary rearrangements and segmental duplications, suggesting stringent selective mechanisms underlying CNV genesis or maintenance. By exploiting gene expression data available for rat recombinant inbred lines, we established the functional relationship of CNVs underlying 22 expression quantitative trait loci. These characteristics make the rat an excellent model for studying phenotypic effects of structural variation in relation to human complex traits and disease.

A CNV is defined as a DNA segment that is >1 kb in size and is present at a variable copy number in comparison with a reference genome[1]. Present estimates suggest that 5 to 20% of the human genome resides in CNVs ranging from kilobase- to megabase-sized segments[2–4], including complete genes, but also gene-regulatory elements. As a result, CNVs can have a considerable impact on the organization of chromatin and the gene-expression landscape. Recent studies indicate that distribution of CNV regions may be shaped by natural selection[5] and propose their role in facilitating rapid evolutionary innovations[6]. Although most CNVs are inherited as normal mendelian traits and can be tagged by neighboring genetic markers[2], *de novo* CNVs have been observed as well[7,8]. Notably, in mouse, spontaneous copy number variation is a highly nonrandom process, and recurrent events are frequent[7]. Although segmental duplications have the potential to function as gene nurseries that allow for diversification or specialization of the encoded genes[6], one of the most obvious direct effects of gene duplication is on the expression level. Indeed, recent studies in human HapMap trios show that CNVs may be causing up to 20% of detected genetic variation in gene expression[9].

Unfortunately, studies on the functional consequences of structural variation in humans are hampered by various issues, including technical and ethical limitations. Rodent models provide a renewable source of samples for hundreds of inbred strains for which thousands of expression and physiological QTLs have already been defined[10–15]. Furthermore, laboratory mice and rats provide valuable genetic resources such as consomic, congenic and recombinant inbred (RI) strains to address the phenotypic effect of individual SNPs or CNVs in defined genetic backgrounds. However, recent surveys on CNVs in mouse strains have shown that they do not differ much from a random background spectrum and are characterized by distinct functional properties as compared to human[16,17].

Here, we have combined computational approaches with experimental platforms to identify and extensively characterize CNVs in commonly used laboratory rat strains. We identified 643 CNVs regions as well as over a hundred genomic loci in the reference BN genome assembly that require assembly reconsideration. Analysis of the genome-wide transcript expression levels in five different tissues in a widely studied rat RI strain panel identified a set of CNVs that may directly affect 22 expression quantitative trait loci (eQTLs).

RESULTS

Computational prediction of CNVs and assembly artifacts

The availability of whole-genome shotgun (WGS) sequencing data from Brown Norway (BN) and Sprague-Dawley (SD) rat strains

[1]Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences & University Medical Centre Utrecht, Uppsalalaan 8, 3584 CT, Utrecht, The Netherlands. [2]Max-Delbrück Center for Molecular Medicine, Robert-Rössle-Strasse 10, D-13092 Berlin, 13125, Germany. [3]Human and Molecular Genetics Center, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, Wisconsin 53226, USA. [4]Medical Research Council Clinical Sciences Centre, Imperial College London, Hammersmith Hospital Campus, Du Cane Road, London, W12 0NN, UK. [5]National Heart and Lung Institute, Imperial College, Dovehouse Street, London, SW3 6LY, UK. [6]Institute of Physiology, Academy of Sciences of the Czech Republic, 1 Videnska 1083, Prague 14220, Czech Republic. [7]Institute of Biology and Medical Genetics, First Medical Faculty, Charles University, Albertov 4, 12800 Prague 2, Czech Republic. [8]Department of Genetics, Cell Biology and Anatomy, 985805 University of Nebraska Medical Center, Omaha, Nebraska, 68198-5805, USA. Correspondence should be addressed to E.C. (e.cuppen@niob.knaw.nl).

Received 14 November 2007; accepted 17 March 2008; published online 28 April 2008; doi:10.1038/ng.141

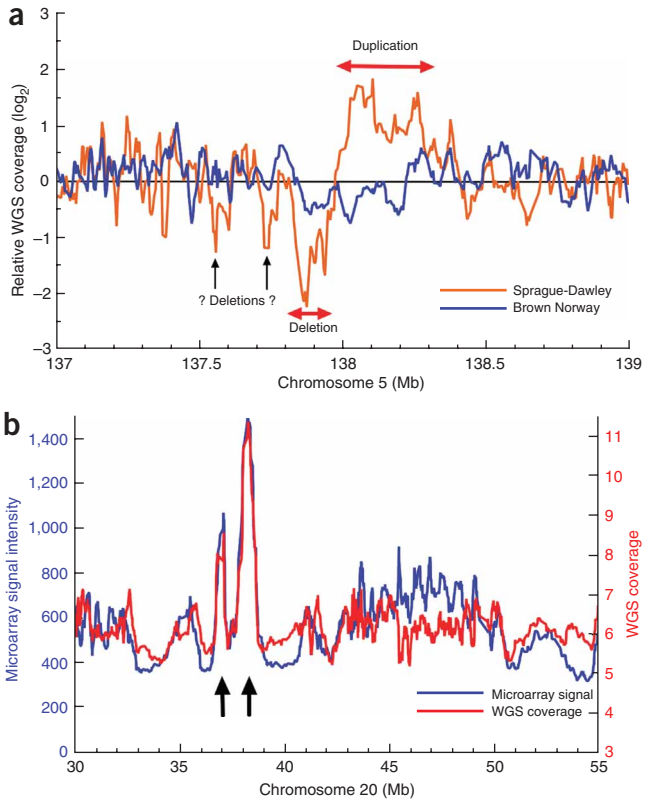

Figure 1 Detection and verification of CNVs and genome assembly artifacts. (**a**) A characteristic example of a genomic region with differential WGS read coverage for BN (blue) and SD (orange) strains. Both a deletion and a duplication in SD are detected in this region (red horizontal arrows). Two potential small deletions, which were not detected computationally using the stringent parameters we used, are indicated with black arrows. (**b**) A very strong correlation between microarray probe intensity and coverage by WGS reads was observed. Average microarray hybridization intensities using BN genomic DNA (blue line) and average WGS sequencing read coverage (red line) were calculated for windows of 100 probes and 10 kb, respectively. The graph displays about half of rat chromosome 20. Two large genomic segments (arrows) between 35 and 40 Mb are clearly collapsed in the current genome assembly, as witnessed by increased WGS coverage and hybridization signals.

(approximately ×4 and ×2 genome coverage, respectively[18]) allows for a computational genome-wide comparison of segmental copy number variation. We mapped WGS reads back to the current genome assembly and analyzed the level of over- or under-representation of WGS reads per genomic bin, similarly as described for human and macaque[19,20] (**Fig. 1**). For approximately 2.5% of WGS reads from BN, we could not find significant homology to the reference genome assembly. Since this genome is based on the BN strain only, the recognized variation includes duplications in BN and SD rats, deletions in SD, and regions with complex patterns of overlapping CNVs. In total, we identified 24 deletions and 22 duplications, encompassing over 15 Mb, between BN and SD (**Supplementary Table 1** online). The false discovery rate estimate for our computational method is below 0.5%. Notably, we identified a large 2.5 Mb telomeric region on chromosome 7 that is completely absent in the SD strain and contains at least 29 olfactory receptors (**Supplementary Fig. 1** online).

The computational approach also allowed for the identification of regions that are not polymorphic between BN and SD but that show significantly increased or decreased WGS coverage, suggesting assembly collapses and overpredictions, respectively. Although the rat genome assembly[18] is generally of very high quality, a low degree of misassembly is inherent to the methods used for DNA shotgun sequencing and genome assembly. Our analysis indicates that there are 37 duplicated regions in the current genome build that should be collapsed into a single locus and 73 regions that should be split into more than one (**Supplementary Table 1**). Twelve and 29 of these regions, respectively, overlap with large gaps in the current genome assembly (>50 kb, **Supplementary Table 2** online). Both types of regions, hereafter named 'genome misassemblies', do contain protein-coding genes (**Supplementary Fig. 2** online).

Identification of CNVs by aCGH

To validate *in silico* predicted CNVs and genome misassemblies and to discover new CNVs, we performed array-based comparative analysis of hybridization signals (aCGH) using various technology platforms. First, we selected ten inbred rat strains selected from divergent branches of the rat evolutionary tree[21] and included one outbred and two RI strains and one wild rat sample for hybridization to Affymetrix Rat Exon Array (RaEx). This platform features 5 million probes per array, which are biased toward gene-coding regions (four probes per exon). We identified 626 CNVs (**Supplementary Table 3** online), with a median size of 5.2 kb, which encompass a total of 22 Mb. The inclusion of two RI strains[22], along with their parental inbred strains, provides the possibility of identifying *de novo* CNVs. Although they were rare, we identified three genomic segments that differed between the parental and RI strains (all in HXB2). Notably, one of the *de novo* duplicated genomic segments (RNO5: 161.8 Mb) was also variable in copy number between other tested rat inbred strains and may thus represent a CNV hot spot[7]. We selected these CNVs and one *de novo* candidate that was called with less stringent parameters for verification by quantitative PCR. All *de novo* CNVs—three gains and one loss of a copy—were experimentally confirmed (**Supplementary Fig. 3** online). As the breeding history of the RI lines is complex because of continuous inbreeding, and as the *de novo* CNV numbers are low, it is not possible to give an estimate on the frequency of these events. Although at present 60 to 80 generations separate the parental strains from the RI lines, our study most likely revealed only a fraction of all CNVs, and therefore the potential phenotypic effects of *de novo* CNV events should not be neglected.

Second, we used the 385k NimbleGen Rat genome tiling path arrays (RN34_WG), which provide less dense but more uniform genome coverage (average probe density of one per ~5 kb in the nonrepetitive part of the genome). We hybridized to these arrays three inbred strains (BN, SS, COP) that were also used on the RaEx platform. We detected 33 large CNVs, with a median size of 256 kb, spanning a total of 15.5 Mb (**Supplementary Table 4** online). There was a good agreement between CNVs called by the different platforms: 19 out of the 31 CNVs (61%) that were called with RN34_WG and had enough probes on the RaEx array to be scored were indeed identified. Conversely, from 12 RaEx CNVs typed as polymorphic among BN, SS and COP strains and that had enough probes on the RN34_WG platform to be scored, 9 (75%) were actually identified. Of the 15 CNVs that were not cross-validated by either of the microarray platforms, 4 were identified in the computational analysis, thereby providing independent support for their validation, and 3 were located in regions of assembly collapse, where CNV calling is less sensitive.

Table 1 Summary of the number and total sizes of CNVs and genome assembly artifacts discovered by three platforms

	CNV or assembly artifact type			
	WGS mapping[a]	RaEx[b]	RN34_WG[c]	Total nonredundant
Gains compared to BN	20 (2.8 Mb)	313 (4.1 Mb)	9 (2.2 Mb)	327 (7.8 Mb)
Losses compared to BN	26 (12.3 Mb)	171 (8.9 Mb)	22 (13.1 Mb)	185 (23.5 Mb)
Gains and losses compared to BN[d]	–	142 (9.1 Mb)	2 (0.2 Mb)	144 (9.3 Mb)
Assembly collapses	73 (26.4 Mb)	–	–	–
Assembly overprediction	37 (7.6 Mb)	–	–	–
Not in assembly, loss in BN	–	31 probe sets	–	N/A
Not in assembly, gain in BN	–	17 probe sets	–	N/A
Total CNVs	46 (15.1 Mb)[e]	626 (22.0 Mb)[f]	33 (15.5 Mb)	643 (36.8 Mb)[g]

[a]WGS mapping included data from the BN and SD strains. [b]Affymetrix exon microarray hybridization experiments included inbred strains BN-Lx/Cub, BN/NHsdMcwi, BS/Ztm, COP/CrCrl, DA/Se, E3/Ztm, F344/NTac, LOU/CZtm, SHR/OlaIpcv and SS/JrHsdMcwi; outbred strains SD/HanBdr and WIST/Crl; and RI strains HXB2 and BXH3. [c]NimbleGen whole genome microarray hybridization experiments included strains BN/NHsdMcwi, COP/CrCrl and SS/JrHsdMcwi. [d]Regions that show both increased and decreased copy numbers as compared with BN when considering multiple strains. [e]Includes 138 nonredundant regions accounting for 38 Mb if assembly collapses and overpredictions. [f]Excluding probe sets missing from the genome, for which gene length cannot be estimated. [g]Several gain and loss regions clustered together, slightly reducing the number of total nonredundant CNVs.

DNA from the SD rat that was used for whole genome shotgun sequencing was not available, and therefore we used another animal from the outbred SD strain. Although this could result in lower confirmation rates owing to different genetic backgrounds, we actually did observe a good overlap between WGS mapping and microarray results. Most of the CNV regions scored on the RN34_WG platform (33) were supported by WGS mapping as CNVs (14) or collapsed regions (6). Similarly, many regions scored by WGS mapping were confirmed by one or both microarray platforms, including 13 deletions and 6 duplications. Notably, the microarray-based approaches discovered CNVs in 21 regions that were not scored as variable between BN and SD rats in the WGS mapping approach but that were marked as regions that are collapsed into a single locus in the current genome build. This supports the idea that recent segmental duplications, which are hard to account for in genome assemblies, may frequently harbor structural variants when multiple strains are compared. In contrast, we observed only a single CNV in a region that is incorrectly annotated as a genomic duplication. All detected CNVs and genome misassemblies may be visualized in a separate track of Ensembl and UCSC genome browsers (**Supplementary Data** online).

We verified a random set of five CNVs that were identified on both RN34_WG and RaEx platforms between BN and SS rats and 15 platform-specific CNVs between BN and SS or between BN and COP using genomic quantitative PCR assays. All regions showed the expected copy number change, with copy number differences varying between 1.4 and 90 (**Supplementary Fig. 3**). We found that the overall cross-validation rate between platforms (different microarray platforms as well as the computational approach) was above 60%, which is comparable to that in previous studies and reflects imperfection of existing methods for CNV discovery, especially for small-sized and multicopy variants. From 71 large CNVs that were identified on either microarray platform and that exceed 100 kb in length, most (56) were confirmed by at least one other method, indicating that there is a considerable proportion of false negatives for any individual platform, which was substantiated by the quantitative PCR verifications. New technologies that provide both higher sensitivity as well as dynamic range—for example, paired-end mapping[23]—should improve CNV detection accuracy.

Missing sequences in the current rat genome assembly

The RaEx microarray was designed for studying expressed sequences in the rat genome and was found to contain almost 300 probe sets that could not be mapped to the current rat genome assembly (RGSC3.4). These sequences are based on expression evidence from EST and cDNA resources, mainly originating from the SD strain, and may represent regions that are absent in BN rats. Using the same parameters as for genome-wide CNV detection, we were able to detect copy number differences in a relatively high proportion of these probe sets (48 out of 292). Thirty-one of the probe sets with CNVs (65%) confirmed segmental losses in the BN strain and are candidates for lineage-specific deletions. We calculated the coverage of the unmapped probe-set sequences in different WGS and genome assembly datasets (**Supplementary Table 5** online) and found that 18 of them did not have appreciable support by either BN or SD reads, and another 14 showed greater sequence overlap when matched against the SD WGS, despite the overall higher coverage for BN reads. The remaining probe sets do probably represent expressed regions in both the SD and BN genome but could correspond to genes that reside in complex genomic regions that are difficult to assemble. Note that our study is likely to provide only an underestimate of the number of genes that are absent in the current genome assembly, as genes with modest or tissue-specific expression are not likely to be included into the RaEx

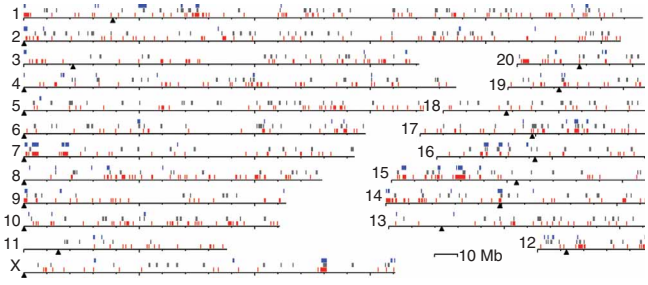

Figure 2 Distribution of CNVs, genome misassemblies and large gaps in the rat. Analysis was done in bins of 400 kb. Genome misassemblies (blue), large gaps (>100 kb, gray) and CNVs (red) are given as separate tracks. The estimated position of centromeres are given by black triangles below each chromosome. A more detailed representation of this information is available as **Supplementary Figure 4**.

Table 2 Comparative properties of CNV regions in human, rat and mouse[a]

	Human		Rat		Mouse	
Property	Observed (expected)	*P* value[b]	Observed (expected)	*P* value[b]	Observed (expected)	*P* value[b]
Genes	837 (623)	↑ 2.4×10^{-3}	850 (567)	↑ $<10^{-3}$ [f]	492 (473)	↑ 0.32
Morbid genes[c]	22 (35)	↓ 8.9×10^{-3}	34 (32)	↑ 0.39	ND	ND
Simple tandem repeats	7.9 (5.3 Mb)	↑ 7.4×10^{-3}	0.9 (1.4 Mb)	↓ $<10^{-3}$ [d,e,f]	1.7 (1.7 Mb)	↑ 0.41
Tandem paralogs	68 (32)	↑ $<10^{-3}$	124 (15)	↑ $<10^{-3}$ [e,f,g]	10 (16)	↓ 5.7×10^{-2}
SignalP[d]	237 (213)	↑ 2.8×10^{-2}	216 (90)	↑ $<10^{-3}$ [e,f,g]	91 (115)	↓ 5.0×10^{-3}
Median Ka/Ks	0.112 (0.094)	↑ 1.7×10^{-2}	0.128 (0.109)	↑ 10^{-2} [f]	0.081 (0.095)	↓ 3.3×10^{-3}
Median Ks	0.653 (0.593)	↑ 1.5×10^{-3}	0.596 (0.599)	↓ 0.42 [e]	0.694 (0.587)	↑ $<10^{-3}$

[a]Table adapted from ref. 17. [b]Arrows indicate the direction of the change of the trait for which the significance level is given. [c]OMIM morbid gene set in rat represents one-to-one orthologs of human OMIM Morbid genes. ND, not determined. [d]Proteins partially or completely encoded within CNVs and predicted by the SignalP algorithm to be secreted. [e]$P < 0.01$ for CNV set predicted by WGS mapping. [f]$P < 0.01$ for CNV set predicted by comparative analysis of hybridization intensities on Affymetrix RaEx microarrays. [g]$P < 0.01$ for CNV set predicted by comparative analysis of hybridization intensities on NimbleGen RN34_WG microarrays.

used here or other microarray designs and are thus also never interrogated in gene expression studies. Yet such genes could be highly relevant for studies in or between other rat strains.

Our observations highlight an important qualitative difference between human and murine genome assemblies. The human genome sequence is composed from DNA of many individuals, and the assembly process may have a significant bias toward high copy number alleles (discussed in ref. 17). In contrast, both the mouse and rat genome assembly were generated from the DNA of a single, inbred animal. Consequently, these assemblies may lack genomic segments that are present in other laboratory strains or wild isolates. This may be especially true for the rat, as the BN strain sequenced is known to be genetically divergent from other commonly used laboratory rat strains[21]. Considering the unprecedented degree of structural variation in mammalian genomes, including that reported here for the rat, the strategy for sequencing and refining genome assemblies may need to be reassessed.

Distribution and properties of CNV regions

In total, a nonredundant set of 643 CNVs covering 36.8 Mb, equaling approximately 1.4% of the genome, was discovered in this study, which included only a limited number of rat strains (**Table 1**). CNVs were found to be distributed nonrandomly over the chromosomes (**Fig. 2**; **Supplementary Fig. 4** online). To begin with, there is a clear tendency for CNVs to occur near telomeres or centromeres. In addition, there are substantial differences in distribution of CNVs among chromosomes (for example, compare 18 and X with 1 and 14) and between chromosome arms (for example, RNO11). Most notably,

Table 3 Expression QTLs overlapping with CNV regions

							Tissues in which eQTL was detected				
Gene[a]	Chr	CNV start	CNV end	CNV type	eQTL type	Overlap	Adrenal	Fat	Heart	Kidney	Muscle
1380768_at	1	2,011,250	2,160,125	BN-Lx>SHR	BN-Lx>SHR	Partial	N/A[b]	N/A[b]	+	N/A[b]	+
1381129_at	2	52,004,759	52,032,085	BN-Lx>SHR	BN-Lx>SHR	Partial	N/A[b]	N/A[b]	–	N/A[b]	+
1391296_at	2	86,046,487	86,072,013	BN-Lx>SHR	BN-Lx>SHR	Partial	N/A[b]	N/A[b]	–	N/A[b]	+
1394465_at	2	159,382,584	159,384,198	BN-Lx>SHR	BN-Lx>SHR	Partial	N/A[b]	N/A[b]	+	N/A[b]	+
Pdcl	3	17,052,940	17,058,940	SHR>BN-Lx	SHR>BN-Lx	Partial	–	–	–	+	+
1378143_at	3	17,052,940	17,058,940	SHR>BN-Lx	SHR>BN-Lx	Partial	N/A[b]	N/A[b]	+	N/A[b]	+
LOC685953 (CD36)	4	13,464,241	13,640,594	BN-Lx>SHR	BN-Lx>SHR	Partial	+	+	+	+	+
1386901_at	4	13,464,241	13,640,594	BN-Lx>SHR	BN-Lx>SHR	Partial	+	+	+	+	+
Serpina3m	6	128,197,974	128,199,438	BN-Lx>SHR	BN-Lx>SHR	Partial	–	+	–	–	–
Prim1	7	1,229,675	1,323,042	BN-Lx>SHR	BN-Lx>SHR	Partial	+	+	+	+	+
1392778_at	7	18,714,971	19,108,582	SHR>BN-Lx	BN-Lx>SHR	Full	+	–	+	+	–
LOC300024 (Ly6b)	7	113,231,593	113,248,199	BN-Lx>SHR	SHR>BN-Lx	Full	–	–	–	+	–
Ly6c	7	113,231,593	113,248,199	BN-Lx>SHR	SHR>BN-Lx	Partial	N/A[b]	N/A[b]	+	N/A[b]	–
Acaa1	8	124,183,674	124,188,529	BN-Lx>SHR	BN-Lx>SHR	Partial	–	–	–	+	–
Sult1c2	9	826,982	832,063	SHR>BN-Lx	BN-Lx>SHR	Partial	–	–	–	+	–
1370943_at	9	826,982	832,063	SHR>BN-Lx	SHR>BN-Lx	Partial	–	–	–	+	–
Trap1	10	11,762,398	11,763,113	BN-Lx>SHR	SHR>BN-Lx	Partial	+	+	+	+	+
Acbd4	10	92,285,668	92,287,244	BN-Lx>SHR	BN-Lx>SHR	Partial	+	+	+	+	+
Mdc1	20	3,039,326	3,042,509	SHR>BN-Lx	BN-Lx>SHR	Full	+	+	+	+	+
RT1-CE10, 1388202_at	20	3,039,326	3,042,509	SHR>BN-Lx	BN-Lx>SHR	Partial	+	+	+	+	+
RT1-CE10, 1388203_x_at	20	3,039,326	3,042,509	SHR>BN-Lx	BN-Lx>SHR	Partial	+	+	+	+	+
RT1-CE1, H2-T23	20	3,039,326	3,042,509	SHR>BN-Lx	BN-Lx>SHR	Partial	+	+	+	+	+

[a]The probe set name was given when no gene was annotated in Rat Ensembl 47 release or multiple probes per gene were called as eQTLs. [b]N/A, not available. Expression of these genes was only assayed in heart and skeletal muscle using Affymetrix RAE230B microarrays. Chr, chromosome.

RNO18 did not reveal any CNVs when assayed on the RN34_WG microarray platform, none of the 138 structural variants obtained by WGS mapping were assigned to it, and the RaEx microarray platform revealed only a few small CNVs covering only 94 kb, whereas up to 689 kb would be expected by random distribution of CNVs. Notably, RNO18 is the only chromosome that did not undergo any major rearrangements during rodent evolution (it has full conservation of synteny with mouse MMU18 and to a large extent with human chromosome 18) and does not show any intrachromosomal segmental duplications[24]. This chromosome is characterized by the lowest protein-coding gene density of all chromosomes, although it has a normal degree of nucleotide conservation. The depletion of CNVs on chromosome 18 is conserved in the mouse[25], where only 17 CNV candidates were identified on MMU18, as compared to 163 and 126 for the similarly sized chromosomes MMU17 and MMU19, respectively.

Most comprehensive CNV studies have focused on human, but recently several extensive studies have been performed to characterize copy number variation in laboratory mice[16,17,26,27]. A comparison with these published results revealed 36 CNV regions containing 55 genes possibly highlighting regions of inherent instability in the rodent genome (**Supplementary Table 6** online). Previously, the characteristics of mouse CNV regions have been found to differ from those observed for human[17]. However, a similar analysis on the characteristics of the rat CNVs revealed much more similarity with human CNVs (**Table 2**). To begin with, rat CNVs, like human CNVs, overlap more genes than would be expected based upon a random distribution. This trend was seen for CNVs obtained with all three methods and was highly significant ($P < 0.001$) for the merged CNV set. Furthermore, tandem paralogous genes, genes coding for secreted proteins, and genes with increased evolutionary rates (increased ratio of the rate of nonsynonymous to the rate of synonymous substitutions, K_a/K_s) were over-represented, as for human. In contrast, we did not see over-representation of simple tandem repeats, but this may be because, in the current rat genome assembly, many simple sequences are not properly assembled. In line with this, we did observe twice as many sequence gaps in rat CNV regions occupying 7.3 Mb, whereas the expected size is 3.4 Mb. We also found none of the 481

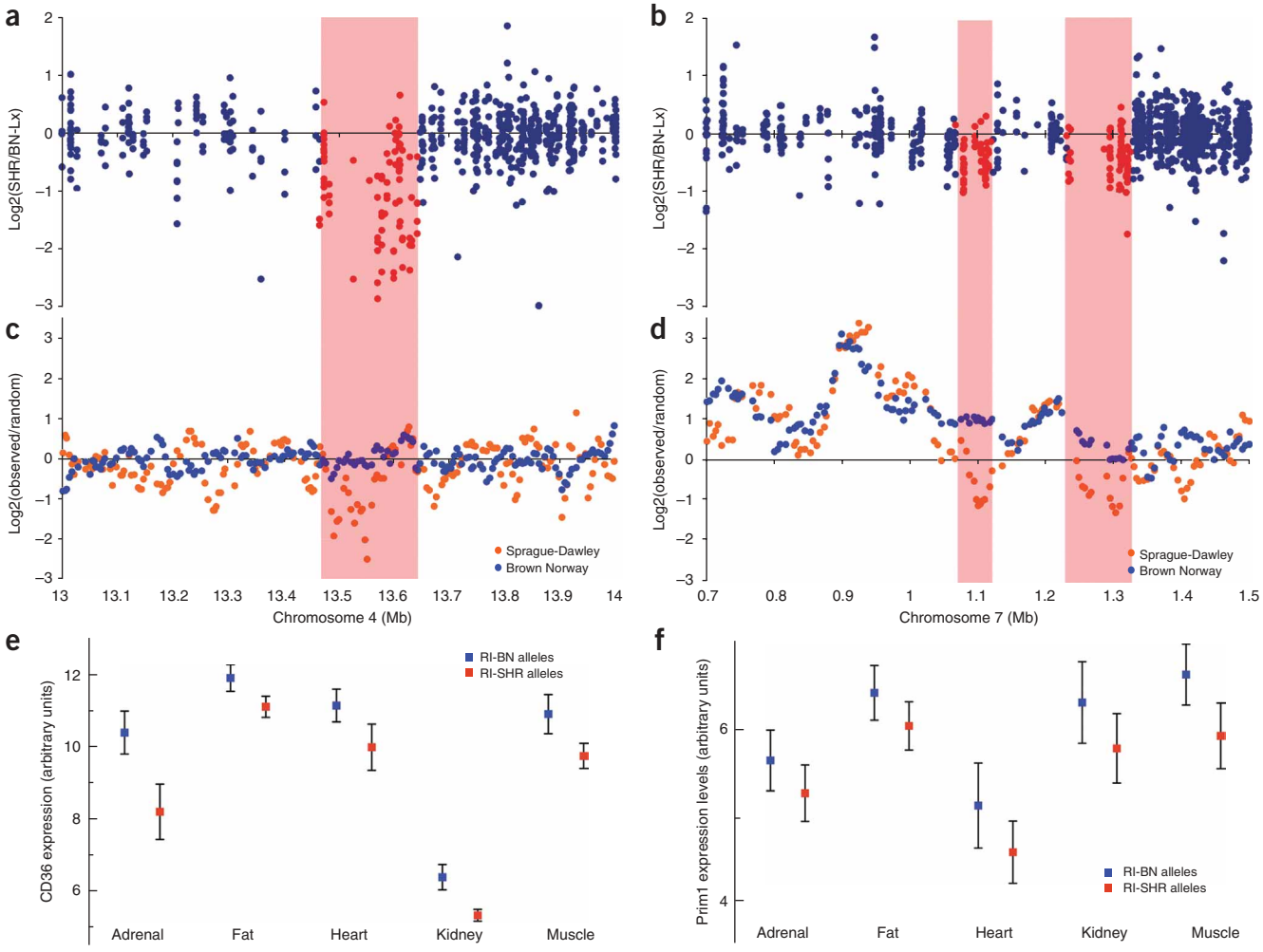

Figure 3 Gene expression levels correlate with copy number status in the RI strains. (**a**–**d**) Comparative analysis of hybridization intensities on Affymetrix RaEx revealed genomic segments varying in copy number in the RI founder strains BN-Lx and SHR on chromosomes 4 (**a**) and 7 (**b**, two regions), encompassing the *CD36* and *Prim1* gene, respectively. (**c**,**d**) Both regions were found to vary in copy number between BN and SD using WGS mapping as well. (**e**,**f**) CNV status of *CD36* (**e**) and *Prim1* (**f**) in the individual RI strains correlates with gene expression as detected by whole-genome expression analysis in all five tissues analyzed. RI strains were grouped by the ancestral haplotype for the *CD36* (**e**) and *Prim1* (**f**) locus and the average expression level of the locus was determined for the BN-Lx and SHR haplotypes.

ultraconserved elements[28] in the CNV regions, whereas five of them would be expected by chance (permutation $P = 0.02$).

Systematic analysis of the genes that overlap with CNV regions using funcassociate[29] revealed that genes with the following Gene Ontology functional classes were overrepresented: (i) response to stimulus ($P = 4.3 \times 10^{-9}$) (including at least 45 annotated olfactory receptor genes), (ii) multicellular organismal process ($P = 3 \times 10^{-6}$), (iii) membrane part ($P = 4 \times 10^{-6}$) and (iv) antigen processing and presentation ($P = 1.1 \times 10^{-5}$). This functional bias is highly reminiscent of the bias reported for human genes that reside in CNV regions[2,3,23,30,31].

The effect of rat CNVs on gene expression

To study the effects of structural variation on gene expression, we defined genetic segments where the BN-Lx and SHR strains have the same (515 Mb) and different haplotypes (1,110 Mb), per rat HapMap data[21], and where we detected structural variation (136 CNVs, 6.1 Mb). We used genome-wide transcript expression data from adrenal, fat, heart, kidney[12] and muscle tissues to define a set of 2,777 genes that were differentially expressed in the two strains (>1.5 fold change) in at least one of the tissues investigated. In genomic regions with the same or different haplotypes, 20 and 22% of the genes, respectively, were found to be differentially expressed (median fold change of 1.75 and 1.78), whereas in CNV regions, up to 44% of the genes were differentially expressed (median fold change of 1.89). Notably, a specific effect on genes that are located in the relative vicinity (<0.5 Mb) of CNVs was still detectable (data not shown), suggesting long-range dosage effects of regulatory elements that reside in the CNV region.

Next, we used the rat RI panel HXB/BXH to assess the robustness of CNV effects in mixed genetic backgrounds. This RI panel consists of 32 inbred strains that are homozygote mosaics of genomic segments derived from the two founder strains BN-Lx and SHR (reviewed in ref. 32). This panel has been used for extensive genome-wide expression profiling in five different tissues (adrenal gland, kidney, heart, fat tissue and skeletal muscle) and the identification of thousands of eQTLs[12,33]. Based on dense, genome-wide SNP genotyping of the RI strains[21], the copy number status of every genomic locus in each strain could be deduced. Although this is only justified when CNVs are tandemly duplicated, linkage analysis of SNP and CNV variation in the HXB2 and BXH3 RI strains confirmed that CNV alleles were predominantly inherited along with SNP alleles (**Supplementary Table 7** online), indicating that the vast majority of CNVs were tandem or at least local.

Sixty-five expression array probe sets overlapped with the 136 CNVs that were called as polymorphic in the direct comparison of the BN-Lx and SHR strains. Notably, 22 out of the 65 transcripts that were assayed in these regions were called as eQTLs (**Table 3**). Not all transcripts were included for all five tissues (owing to the use of different microarray design versions), and not all genes are expressed in the five tissues analyzed, but these numbers were significantly higher than would be expected by chance ($P = 0.0148$). One of the regions included *LOC685953 (CD36)* (**Fig. 3a**), which is a duplicated gene that has previously been identified to cause a defect in insulin action and fatty acid metabolism in the SHR strain[34]. Regions harboring newly identified differentially expressed genes were also identified (**Fig. 3b**), with 14 of them showing differential expression in two or more tissues and nine of them in all five tissues studied. Notably, ten of the eQTLs showed expression changes that were opposite to the copy number status of the genomic segment. Six of these resided in only two different CNVs. For example, the duplication of a 3-kb region on chromosome 20 in the SHR strain overlapped with four eQTLs that all showed a decreased expression in RI strains in which this segment was inherited from the SHR strain. However, such observations are in line with findings that increased copy number can be both positively[35,36] or negatively[37] correlated with gene expression. Such mechanisms could be explained by, for example, the amplification of a transcriptional repressor or an increased or decreased distance between regulator and gene. In line with this and as already indicated above, CNVs are able to affect their direct genomic environment. In addition, CNVs seem to be major tissue-independent regulators of gene expression, as 4.6% (9 out of 196) of the *cis* eQTLs that we detected in all five tissues directly overlapped with CNVs, whereas only 0.22% would be expected by chance. Furthermore, ~20% of the transcripts located within 0.5 Mb of a CNV were identified as *cis* eQTLs in all five tissues, whereas this number is expected to be only 4.5% by chance.

DISCUSSION

The rat is one of the most popular model organisms for studying complex diseases[14], and our results show that the versatile rat genetic tools and powerful genomics resources that are available offer many benefits for studying the phenotypic effects of structural genomic variation. Here, we have presented a first catalog of rat genomic copy number variation obtained using both computational and experimental methods, with high cross-validation rates. Almost 1.5% of the rat genome was found to reside in regions that are variable. These numbers are likely to be significant underestimates of the real number of structural variants in laboratory rat strains, as we surveyed only a limited number of commonly used rat strains. Furthermore, many genomic regions were not effectively surveyed owing to technical limitations (for example, probe density or non-random distribution) of the technology platforms that were used. Nevertheless, the high probe density of the RaEx platform in specific genetic regions allowed us to detect many small CNVs, resulting in a median CNV size in this study of 5.8 kb. Similarly to that found in recent high-resolution studies in human[23], the size of 65% of the CNVs was below 10 kb.

Besides structural variants, more than 100 genomic regions in the current rat genome assembly were highlighted that may require improvements, including genome assembly collapses that are most likely due to recent segmental duplications. Although most of these regions were not found to vary in the strains studied here, they could be variable when more rat strains are analyzed. Furthermore, such regions could also be sources for *de novo* CNVs. Most importantly, however, these misassembled regions harbor many protein-coding genes (**Supplementary Table 8** online), which are currently incorrectly annotated as having paralogs or as being unique. For example, genes such as *Rhag* and *Mut* that reside in a pseudo-duplicated region on chromosome 16 (**Supplementary Fig. 2**) have always been found as a single copy in all animals with sequenced genomes. Hence, these misannotations may complicate or compromise functional studies, genome-wide experiments and comparative studies.

CNVs are speculated to function as gene nurseries to allow for diversification or specialization of paralogous genes[6]. However, one of the direct effects of a gene duplication including all regulatory elements could be a change in gene expression levels. Indeed, an association analysis of genome-wide transcript expression levels with SNPs and CNVs, in individuals who are part of the International HapMap project, suggested that up to 20% of heritable gene expression variation could be explained by copy number variation[9].

However, feedback mechanisms may partially or completely reduce such effects, complicating experimental detection by, for example, expression microarrays. In the case of humans, robust expression analysis is often limited by insufficient independent cases or inaccessibility to relevant tissue samples, resulting in limited statistical power or limited biological scope, respectively. The use of RI strains, as illustrated here, allows for the repeated sampling of tissues from animals with an (almost) identical genetic background, resulting in noise reduction not only in gene expression studies, but also in, for example, physiological or behavioral characterization studies. In our analysis on the CNV status and gene expression levels in the rat RI panel, we found many CNVs that directly overlapped expressed sequences and 22 different eQTLs that could be explained by differences in gene copy number. In all cases, this relationship was supported by expression data from several different tissues.

Finally, the similar properties of human and rat CNV regions demonstrate the favorable characteristics of the rat as a model organism for studying phenotypic effects of structural variations relevant to the human situation and human complex disease. In contrast, mouse CNVs did not show such overlap; this could be because of differences in breeding history[21] and/or selective pressure. Furthermore, we identified 113 one-to-one orthologous genes that overlapped CNV regions in both rat and human, and 80 of these are listed in the Online Mendelian Inheritance in Man (OMIM) database (**Supplementary Table 9** online). This significant overrepresentation of human-rat orthologous disease genes that are affected by CNVs in both species (chi-squared $P = 0.0028$) highlights the importance of copy number–fragile regions for disease etiology. The recent evidence that copy number polymorphisms in the same gene can predispose to disease in both humans and rats[38] supports this idea. Clearly, higher resolution data on CNVs is needed, along with extensive and robust phenotypic information, to fully understand the biological relevance of structural variation in mammalian genomes.

METHODS

Predicting assembly artifacts and copy number variation from density distribution of whole genome shotgun reads. WGS reads were aligned to RGSC 3.4 rat genome assembly using GMAP software[39]. Only the best hit, with at least 100 bp matching, was considered as mapping position for every shotgun read. Detection of assembly collapses and copy number variation was done by applying Smith-Waterman algorithm[40] using a sliding window approach. Every genomic window contained 10 kb of unmasked sequence, and genomic starts of consecutive windows were 5 kb apart. In total we analyzed three datasets—(i) Celera SD WGS, (ii) Celera BN WGS and (iii) GTC BN WGS—and six control sets: (iv–vi) randomized equivalents of sets i–iii; (vii) monotone set with exactly 150 traces per 10 kb window and (viii and ix) sets corresponding to even distribution of shotgun reads with ×1 and ×2 coverage. The Smith-Waterman algorithm was used to find discrepancies in all pairwise comparisons with threshold value $t_0 = 4$ median absolute deviations and hits involving at least five consecutive sliding windows. When considering all nine datasets, only 1.8% of the discrepancies were found to occur between the control sets, indicating a low false-discovery rate. We further required that every candidate regions be called in at least four pairwise comparisons. We chose the central position of the start and end windows as boundaries of the identified CNVs. We classified every candidate region for collapse/overprediction and copy number change, requiring at least four comparisons supporting each classification. To account for possible WGS ascertainment bias for highly homologous segments (>99% over 1 kb), candidate regions with high homology and that were implicated in both assembly collapses and overpredictions were removed from our dataset.

To estimate false discovery rate, we performed a randomization of the mapping positions of WGS reads for every chromosome. We analyzed 100 randomized sets using the same algorithm and estimated the false discovery rate as the average number of CNVs and assembly artifacts scored per randomized set divided by the number of observed events in the experimental dataset.

Comparative analysis of hybridization intensities on microarrays. *Affymetrix RaEx array hybridization.* The Affymetrix GeneChip Rat Exon 1.0ST array covers known, as well as predicted, transcripts with independent probe sets for every exon. For hybridization on the Affymetrix Rat Exon 1.0 ST array, we used a wild rat from the outskirts of Beers, Noord Brabant, The Netherlands (wild1/Hubr), the inbred rat strains BN-Lx/Cub, BN/NHsdMcwi, SHR/OlaIpcv, COP/Crl, DA/Se, F344/NTac, SS/JrHsdMcwi, LOU/CZtm, BS/Ztm and E3/Ztm, and the outbred strains SD/HanBdr (kindly provided by Michael Bader, Max-Delbrück Center) and WIST/Crl. RI strains HXB2 and BXH3 belong to the Prague RI panel[41]. DNA was made available through the rat HapMap STAR consortium (see URLs). We used 250 ng of genomic DNA from each strain for hybridization to the Affymetrix microarray according to the StyI GeneChip protocol. A total of 9 µg of labeled DNA was subjected to hybridization on the rat exon 1.0 ST array. Washing steps were performed according to the manufacturer's protocol for rat exon 1.0 ST array.

NimbleGen RN34_WG array hybridization. The BN/NHsdMcwi and SS/JrHsdMcwi (obtained from the Department of Physiology, Medical College of Wisconsin) and COP/CrCrl (obtained from Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center) inbred rat strains were used for hybridization to the 385k NimbleGen rat genome tiling path arrays, in which the BN sequenced strain served as the reference. DNA was isolated from spleen samples. Every DNA extraction was carried out from 20 to 30 mg of tissue using DNeasy spin columns (Qiagen).

The whole-genome oligonucleotide microarray provided by NimbleGen contained 385,000 isothermal probes, 45–75-mer, spanning the rat genome at a mean probe spacing of 5 kb. The oligonucleotide design, array fabrication, DNA labeling, aCGH experiments, data normalization and calculations of copy number ratio (using $\log_2$ of the ratio of signal from the fluorophores Cy3 and Cy5) were performed at NimbleGen according to recommended and published procedures[42]. The hybridization designs were set up as follows: (i) SS (Cy3) versus BN (Cy5) plus dye swapping and (ii) COP (Cy3) versus BN (Cy5) plus dye swapping.

CNV calling and statistical analysis. Because of the small probe size of 25-mer RaEx probes, many of them had ambiguous mapping to the genome. We excluded all ambiguously mapped probes, leaving 3.7 million probes for our analysis. We used quantile normalization to normalize probe intensities across arrays. The Smith-Waterman algorithm was used to call CNVs with t0 = 1.75 for both RN34_WG and RaEx data. Ten probes were required for the RN34_WG and 13 for the RaEx platform (because of the more pronounced noise of 25-mer probes). For every platform, we required a candidate CNV to be called in at least half of the possible comparisons. The lengthiest span that was covered within at least half of the comparisons was considered to be a candidate region.

RaEx platform detected many more duplications (relative to BN reference) than deletions. By excluding the probes with ambiguous mapping to the rat genome assembly, we partially eliminated duplicated regions of BN strain from the RaEx analysis. However, genome assembly improvements guided by our results could help to design more appropriate assays for more precise quantification of copy number of recent segmental duplications.

For the calculation of overlap between CNVs identified by the different platforms, we included only those regions that had enough probes or sequence windows to be discovered by both platforms. Estimated minimal sizes were approximately 350 bp, 20 kb and 70 kb for RaEx, WGS mapping and RN34_WG platforms, respectively. The completeness of overlaps was affected by uneven distribution of probes between aCGH platforms and the less accurate CNV-boundary approximation inherent to the window-based WGS-mapping approach. Typically, cross-platform overlaps were nearly perfect: 10 of 14 RN34_WG CNVs overlapped over 90% of the corresponding region detected by RaEx. The smallest overlap detected was 47%. Similarly, most (15 of 19) of CNVs called by aCGH experiments overlapped with CNVs predicted by WGS mapping. The procedures for data handling, manipulation and visualization were performed by custom Perl scripts, available on request from the authors.

Preparation of labeled cRNA and hybridization. We used tissues from 29 RI strains (BXH and HXB) that were inbred for over 60 generations. We collected tissues at 6 weeks of age from four males of each RI strain and from five rats from each parental strain, froze them in liquid nitrogen and stored them at –80 °C. Total RNA was extracted using Trizol reagent (Invitrogen) and purified using RNeasy Mini kit (Qiagen) in accordance with the manufacturer's protocol. Double-stranded cDNA was synthesized from total RNA using the One-Cycle cDNA Synthesis Kit (Affymetrix) and cRNA biotinylated from cDNA using the IVT Labeling Kit (Affymetrix). We hybridized the fragmented cRNA samples to Affymetrix RAE 230A (fat, kidney, adrenal) and RAE 230 2.0 (heart, skeletal muscle) expression arrays in accordance with Affymetrix protocols.

Analysis of expression data. We computed gene expression summary values for Affymetrix GeneChip data using the robust multichip average (RMA) algorithm[43], which uses background adjustment, quantile normalization and summarization. Statistical comparisons of expression data in the parental strains were done using ANOVA F statistic after outlier elimination using a Nalimov test with $P < 0.001$. The false discovery rate was calculated as described[12] and a threshold of 5% used to define significant differences in RNA levels.

Analysis of eQTLs. *Cis* eQTLs in HXB/BXH RI strains were detected using QTL Reaper software (see URLs). We used SNP data generated by STAR consortium and required SNP alleles to be within 10 Mb of the corresponding RAE230A probe. To determine the significance of observed *cis* eQTLs, we used a permutation test with 10,000 iterations. Only eQTLs with P values $<10^{-3}$ were considered for further comparative analysis.

URLs. Rat HapMap STAR consortium, http://www.snp-star.eu. QTL Reaper software http://sourceforge.net/projects/qtlreaper. Ensembl database, http://www.ensembl.org.

Accession numbers. ArrayExpress: Microarray data have been deposited with accession codes for the rat RI strains E-MIMR-222 (heart), E-AFMX-7 (fat, kidney), E-TABM-457 (adrenal gland) and E-TABM-458 (skeletal muscle) and for the aCGH arrays E-MEXP-1514 (Affymetrix RaEx) and E-MEXP-1517 (NimbleGen RN34_WG).

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank S. Kurz for technical assistance. This work was supported by the award "Exploiting natural and induced genetic variation in the laboratory rat" to E.C. from the European Heads of Research Councils and European Science Foundation EURYI (European Young Investigator) Award scheme; the EURATools integrated project funded by the Sixth Framework Programme of the European Union; grants from the Ministry of Education of the Czech Republic and support from the Howard Hughes Medical Institute to M.P.; support from the British Heart Foundation and the UK Department of Health to S.C.; and US National Institutes of Health grant CA77876 to J.D.S.

AUTHOR CONTRIBUTIONS

V.G. generated data, performed bioinformatic analyses and wrote the manuscript. K.S., T. Adamovic, M.V., S.A.A.C.v.H. and S.C. performed microarray and quantitative PCR experiments. M.P., T. Aitman., S.C., H.J., J.D.S. and N.H. contributed material, data and discussion. E.C. supervised the project and wrote the manuscript.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Feuk, L., Carson, A.R. & Scherer, S.W. Structural variation in the human genome. *Nat. Rev. Genet.* **7**, 85–97 (2006).
2. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).
3. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
4. Sharp, A.J. *et al.* Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* **77**, 78–88 (2005).
5. Cooper, G.M., Nickerson, D.A. & Eichler, E.E. Mutational and selective effects on copy-number variants in the human genome. *Nat. Genet.* **39**, S22–S29 (2007).
6. Dumas, L. *et al.* Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Res.* **17**, 1266–1277 (2007).
7. Egan, C.M., Sridhar, S., Wigler, M. & Hall, I.M. Recurrent DNA copy number variation in the laboratory mouse. *Nat. Genet.* **39**, 1384–1389 (2007).
8. Sebat, J. *et al.* Strong association of de novo copy number mutations with autism. *Science* **316**, 445–449 (2007).
9. Stranger, B.E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007).
10. Drake, T.A., Schadt, E.E. & Lusis, A.J. Integrating genetic and gene expression data: application to cardiovascular and metabolic traits in mice. *Mamm. Genome* **17**, 466–479 (2006).
11. Flint, J., Valdar, W., Shifman, S. & Mott, R. Strategies for mapping and cloning quantitative trait genes in rodents. *Nat. Rev. Genet.* **6**, 271–286 (2005).
12. Petretto, E. *et al.* Heritability and tissue specificity of expression quantitative trait loci. *PLoS Genet.* **2**, e172 (2006).
13. Jacob, H.J. & Kwitek, A.E. Rat genetics: attaching physiology and pharmacology to the genome. *Nat. Rev. Genet.* **3**, 33–42 (2002).
14. Kwitek, A.E. *et al.* BN phenome: detailed characterization of the cardiovascular, renal, and pulmonary systems of the sequenced rat. *Physiol. Genomics* **25**, 303–313 (2006).
15. Malek, R.L. *et al.* Physiogenomic resources for rat models of heart, lung and blood disorders. *Nat. Genet.* **38**, 234–239 (2006).
16. Graubert, T.A. *et al.* A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genet.* **3**, e3 (2007).
17. Nguyen, D.Q., Webber, C. & Ponting, C.P. Bias of selection on human copy-number variants. *PLoS Genet.* **2**, e20 (2006).
18. Gibbs, R.A. *et al.* Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521 (2004).
19. Bailey, J.A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
20. Gibbs, R.A. *et al.* Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007).
21. The STAR Consortium. SNP and haplotype mapping for genetic analysis in the rat. *Nat. Genet.* advance online publication, doi: 10.1038/ng.124 (28 April 2008).
22. Pravenec, M. & Kren, V. Genetic analysis of complex cardiovascular traits in the spontaneously hypertensive rat. *Exp. Physiol.* **90**, 273–276 (2005).
23. Korbel, J.O. *et al.* Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**, 420–426 (2007).
24. Tuzun, E., Bailey, J.A. & Eichler, E.E. Recent segmental duplications in the working draft assembly of the brown Norway rat. *Genome Res.* **14**, 493–506 (2004).
25. Cutler, G., Marshall, L.A., Chin, N., Baribault, H. & Kassner, P.D. Significant gene content variation characterizes the genomes of inbred mouse strains. *Genome Res.* **17**, 1743–1754 (2007).
26. Li, J. *et al.* Genomic segmental polymorphisms in inbred mouse strains. *Nat. Genet.* **36**, 952–954 (2004).
27. Snijders, A.M. *et al.* Mapping segmental and sequence variations among laboratory mice using BAC array CGH. *Genome Res.* **15**, 302–311 (2005).
28. Bejerano, G. *et al.* Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004).
29. Berriz, G.F., King, O.D., Bryant, B., Sander, C. & Roth, F.P. Characterizing gene sets with FuncAssociate. *Bioinformatics* **19**, 2502–2504 (2003).
30. Iafrate, A.J. *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* **36**, 949–951 (2004).
31. Tuzun, E. *et al.* Fine-scale structural variation of the human genome. *Nat. Genet.* **37**, 727–732 (2005).
32. Pravenec, M. & Kren, V. Genetic analysis of complex cardiovascular traits in the spontaneously hypertensive rat. *Exp. Physiol.* **90**, 273–276 (2005).
33. Hubner, N. Expressing physiology. *Nat. Genet.* **38**, 140–141 (2006).
34. Glazier, A.M., Scott, J. & Aitman, T.J. Molecular basis of the Cd36 chromosomal deletion underlying SHR defects in insulin action and fatty acid metabolism. *Mamm. Genome* **13**, 108–113 (2002).
35. McCarroll, S.A. & Altshuler, D.M. Copy-number variation and association studies of human disease. *Nat. Genet.* **39**, S37–S42 (2007).
36. Somerville, M.J. *et al.* Severe expressive-language delay related to duplication of the Williams-Beuren locus. *N. Engl. J. Med.* **353**, 1694–1701 (2005).
37. Lee, J.A. *et al.* Spastic paraplegia type 2 associated with axonal neuropathy and apparent PLP1 position effect. *Ann. Neurol.* **59**, 398–403 (2006).
38. Aitman, T.J. *et al.* Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature* **439**, 851–855 (2006).
39. Wu, T.D. & Watanabe, C.K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
40. Smith, T.F. & Waterman, M.S. Identification of common molecular subsequences. *J. Mol. Biol.* **147**, 195–197 (1981).
41. Pravenec, M., Klir, P., Kren, V., Zicha, J. & Kunes, J. An analysis of spontaneous hypertension in spontaneously hypertensive rats by means of new recombinant inbred strains. *J. Hypertens.* **7**, 217–221 (1989).
42. Selzer, R.R. *et al.* Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosom. Cancer* **44**, 305–319 (2005).
43. Irizarry, R.A., Ooi, S.L., Wu, Z. & Boeke, J.D. Use of mixture models in a microarray-based screening procedure for detecting differentially represented yeast mutants. *Stat. Appl. Genet. Mol. Biol.* **2**, 1 (2003).