Fadista *et al. BMC Genomics* 2010, **11**:284
http://www.biomedcentral.com/1471-2164/11/284

BMC Genomics

RESEARCH ARTICLE Open Access

Copy number variation in the bovine genome

João Fadista, Bo Thomsen, Lars-Erik Holm and Christian Bendixen*

Abstract

Background: Copy number variations (CNVs), which represent a significant source of genetic diversity in mammals, have been shown to be associated with phenotypes of clinical relevance and to be causative of disease. Notwithstanding, little is known about the extent to which CNV contributes to genetic variation in cattle.

Results: We designed and used a set of NimbleGen CGH arrays that tile across the assayable portion of the cattle genome with approximately 6.3 million probes, at a median probe spacing of 301 bp. This study reports the highest resolution map of copy number variation in the cattle genome, with 304 CNV regions (CNVRs) being identified among the genomes of 20 bovine samples from 4 dairy and beef breeds. The CNVRs identified covered 0.68% (22 Mb) of the genome, and ranged in size from 1.7 to 2,031 kb (median size 16.7 kb). About 20% of the CNVs co-localized with segmental duplications, while 30% encompass genes, of which the majority is involved in environmental response. About 10% of the human orthologous of these genes are associated with human disease susceptibility and, hence, may have important phenotypic consequences.

Conclusions: Together, this analysis provides a useful resource for assessment of the impact of CNVs regarding variation in bovine health and production traits.

Background

Cattle, part of the Cetartiodactyl order of eutherian mammals [1], is an important source of human nutrition worldwide as well as the most studied ruminant model of metabolism, reproduction, and disease [2]. Following the milestone publication of the cattle genome assembly along with annotation of functional elements and variation [2,3], we are now enabled to search for genomic regions that impact the genetic variation of important phenotypic traits.

Genomic structural variation, including insertions, duplications, deletions, inversions and translocations of DNA, has long been known to be present in animal genomes [4,5] but had predominantly been assumed to be to rare events and often associated with disease. This notion changed in 2004 when two groups of researchers published the first genome-wide maps of copy number variation in seemingly healthy individuals [6,7]. Copy number variant (CNV) is described as a segment of DNA > = 1 kb that is copy number variable when compared with a reference genome [8]. Before these landmark studies, it was thought that SNPs were the major source of genetic variation between individuals [9] but genomic structural genetic variation is now known to cover more base pairs [10-17], and to have a higher per-locus mutation rate than SNPs do [18].

There are indications that CNVs appear throughout the genome not only in humans, but also in other primates [19-21], rodents [22-30], flies [31,32], dogs [33], chickens [34] and cattle [35]. Nevertheless, other than humans and mice [29,36-40], little is known about how CNVs contribute to normal phenotypic variation and disease susceptibility. Up until now, relatively few studies have confirmed the presence of CNVs in cattle [35,41,42], of which only one study focused on genome-wide detection of CNVs [35], but at low resolution using version 3 of bovine genome assembly [2].

Here we report the use of high-resolution oligonucleotide array comparative genomic hybridization (array CGH) to identify 304 CNV regions in 20 animals (14 Holsteins, 2 Red Danish, 3 Simmental and 1 Hereford). With an average probe spacing of 420 bp relative to the latest bovine genome assembly (BT4, 2007) [2], this analysis provides the highest-resolution map of copy number variation in the cattle genome to date.

* Correspondence: christian.bendixen@agrsci.dk
[1] Group of Molecular Genetics and Systems Biology, Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, Aarhus University, Blichers Allé 20, DK-8830 Tjele, Denmark
Full list of author information is available at the end of the article

BioMed Central © 2010 Fadista et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Results

Experiment design

The goal of our study was to characterize levels and patterns of copy number variation among bovine animals. Therefore, to assess the bovine CNV landscape, the genomic DNA of 20 bovine samples from two dairy (14 Holsteins, 2 Red Danish) and two beef breeds (3 Simmental, 1 Hereford) were analyzed. Assessment of copy number variation between samples was done using a set of Nimblegen HD2 CGH arrays that tile across the genome with approximately 6.3 million unique oligo probes with a mean probe spacing of 420 bp, using the latest genome assembly (BT4) [2].

We opted for a dye-swap loop array design, rather than a common reference design, so that each sample was hybridized to two different samples in two different dye orientations. Dye swap is used to compensate for dye bias, while the loop design (known to be more efficient than the reference design [43,44]) is applied to help assign the CNV gain and loss status more accurately for each sample based on the number of samples with the CNV.

Array CGH evaluation

For evaluation of our array CGH platform, four sex-mismatched arrays and one self-self hybridization (all in dye swaps) were used to assess the false positive rate (see Methods). Probes were interpreted as revealing a copy-number difference if the standard error of the log-intensity ratio was beyond an intensity-ratio threshold. The adequacy of this threshold in detecting copy-number differences was confirmed by conducting sex-mismatched hybridizations, comparing the number of X-linked probes beyond the threshold. From the 88.52 Mb length of chromosome X, 3.21 Mb were greater than the threshold, yielding an estimate of 3.62% for the rate of false positives (FP). The false positive rate (FPR) is conservatively overestimated due to: (1) the assumption that there are no CNVs in the chromosome X of sex-mismatched arrays; (2) calling for FP was done at individual arrays rather than if they were detected in both dye swaps; (3) and because the self-self hybridization array yielded much lower FPR at individual array calling (0.0085%), and zero FPR when calling CNVs detected in both dye swap.

Pattern and frequency of CNV regions

Since copy-number changes are relative for array CGH data, unambiguous ascertainment of the ancestral state of a CNV and (subsequent) identification of duplications and deletions is challenging. We have therefore chosen a design where a dye swap is coupled with a loop design, with each animal sample hybridized with two other animal samples, enabling us to distinguish between a deletion and duplication as well as the animal origin of the CNV (Figure 1). Since identical CNVs, when called in different animals, might be assigned different boundaries due to technical and/or biological sources of variability, overlapping CNVs were handled as a whole and named copy number variable regions (CNVRs) [10,45].

After applying a stringent CNV calling pipeline with a theoretical 1.5 kb resolution for CNV detection (Figure 2 and Methods), 304 putative CNVRs were identified, averaging 47 CNVs per animal (Additional file 1, Table S1), with 70% (212) of the CNVRs observed in more than one animal. Although CNVRs detected in more than one animal of different families and/or breeds could be defined as frequent, the relationship between some of the animals precludes such classification (Additional file 2, Figure S1). The relatively poor breakpoint estimation also prevents information regarding whether these CNVRs are identical-by-descent (arisen before the divergence of these cattle breeds), or separate events occurred independently in different breeds (in putative structurally fragile genomic regions).

CNVRs were detected on all chromosomes, but were distributed throughout the genome in a non-random manner (Figure 3) with little correlation between CNVRs occurrence and chromosome length (Additional file 3, Figure S2). This is coherent with previous studies on heterogeneous distribution of CNVs in primates [10,21], but the low number of samples used in this study prevents us from drawing any conclusions regarding putative genomic CNV hotspots.

Out of the total number of CNVRs detected, 49.7% (151) partially overlap gaps in the assembly (BT4), which indicates that the CNVRs have a high probability of being linked with gaps within the reference cattle genome assembly (permutation test, p-value < 0.001). This stresses the need for unraveling of these genomic regions of high structural complexity. The CNVRs detected vary in size from 1.7 kb to 2 Mb with a median size of 16.7 kb, and encompass approximately 23 Mb or 0.68% of the bovine genome (Table 1 and Figure 4). The biggest region showing copy loss is 2.03 Mb on chromosome 13 in animal 16, while the biggest region showing copy number gain was detected in animals 6 and 7 showing a 417 kb amplification without overlapping any gene nor SD (Additional file 1, Table S1).

When comparing the size distribution between the 202 losses and the 102 gains, no significant difference was found (Wilcoxon rank sum test, p-value > 0.05), although we detected significantly more losses than gains (exact binomial test, p-value = 1.01e-08). This bias in detecting more deletions could be due to both biological and technical reasons. One of the main mechanisms responsible for the CNV formation, non-allelic homologous recombination (NAHR), has been shown to generate more deletions than duplications [46]. As also noted by others

Figure 1 Example of an identification procedure for CNV gain or loss status. Y axis represents log2ratios and X axis represents genomic positions along chromosome 7. (A) animals *2 vs. 19*; (B) *19 vs. 6* (gain in 19 or loss in 6); (C) *6 vs. 20* (loss in 6 or gain in 20) and (D) *20 vs. 17*. The only plots that show a CNV are B and C, and since the only animal common to those hybridizations is animal 6, we classify this CNV as a deletion in animal 6.

[10,31,47], a technical bias favoring the detection of deletions may be responsible since our CNV detection pipeline have more power to detect a loss (log2(1/2) = -1) than a gain (log2(3/2) = 0.58). With 69% of the CNVRs described within 50 kb in size (figure 4), it should be noted that a significant proportion of the CNVs are near our effective resolution of 1.5 kb. This indicates that the experimental detection of 304 CNVRs may greatly underestimate the actual number of CNVs in the cattle genome, and that a substantial proportion of CNVs could be smaller than 1.5 kb in size. CNVs >2 Mb in size were not detected, which may be a consequence of the number and size of sequence gaps in the current outline of the cattle genome sequence assembly (75 654 gaps spanning 5.8% of the assembly).

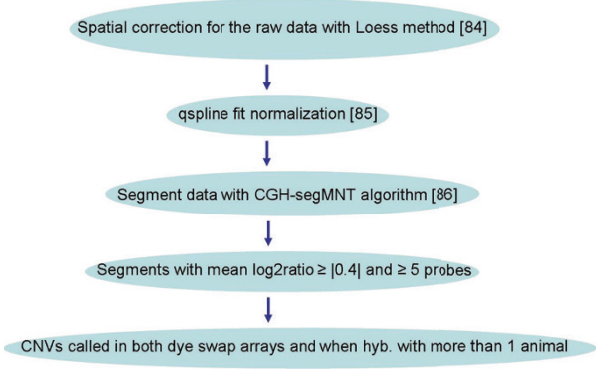

Figure 2 CNV calling pipeline (details in Methods section). Data analysis procedure to discover copy number variations in array CGH data.

When assessing hybridization signals in the unassembled chromosome (ChrUn), it was verified that only the male vs. female hybridizations were detecting CNVs in some regions. Although the number of females in this study is small (n = 2), the findings suggest that these regions may be from the bovine chr Y (Additional file 11, Table S9). It is known that the *Bos taurus* genome assembly was not only composed by a female animal, but also had a BAC library sequenced from a male animal, from which the corresponding Y chromosome scaffolds were unlabeled and placed in the ChrUn [48]. Consequently,

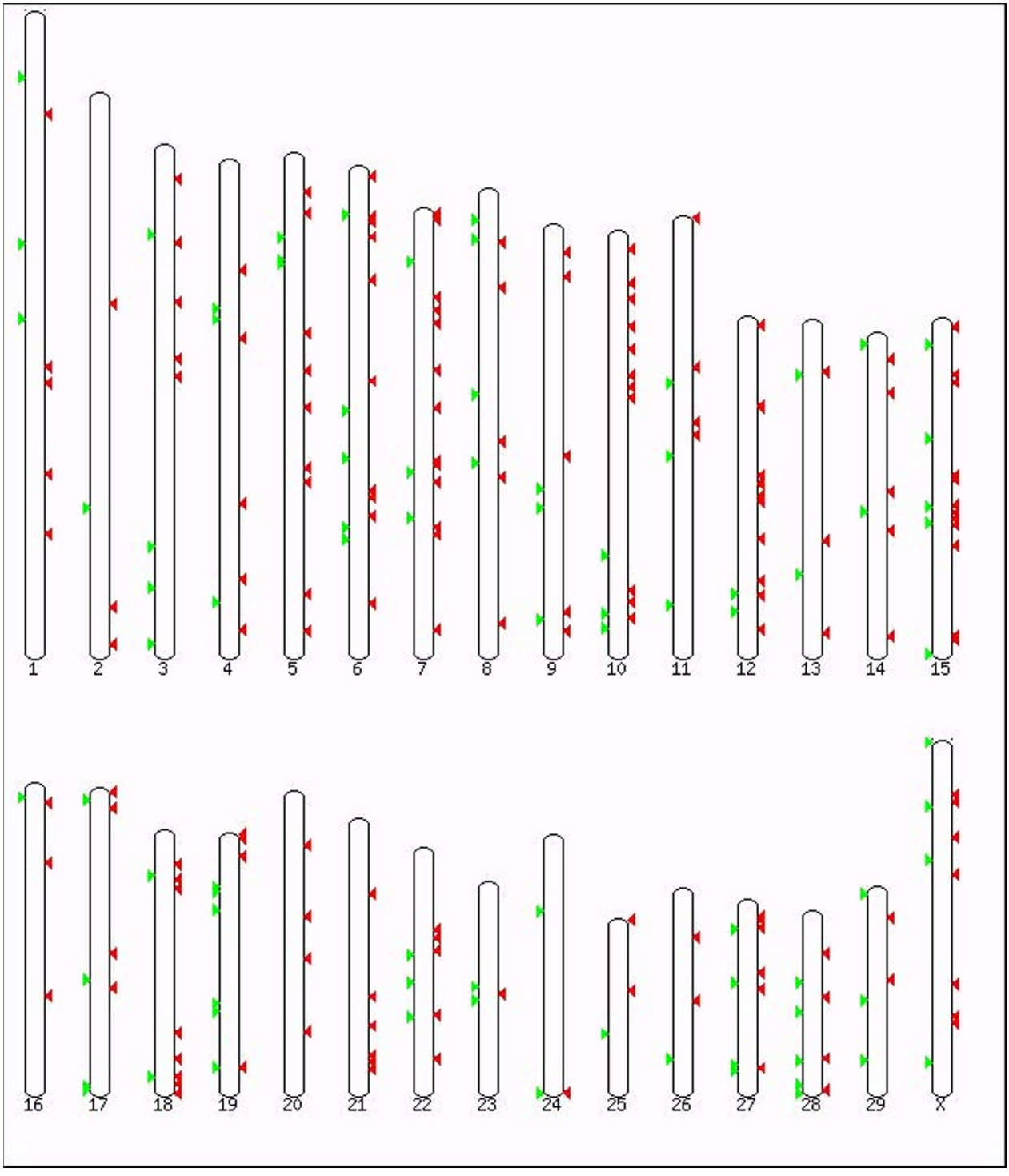

Figure 3 Bovine karyotype, with CNVR distribution. Green arrows correspond to gains and red arrows correspond to losses.

this study highlights regions for future genome assembly improvements.

In accordance with analyses conducted in humans [10,11], we detected that the GC content of the CNVRs (43.6%) are slightly larger than of the whole genome (41.8%), which supports the notion that CNVs arise more often in GC rich regions.

Table 1: Characteristics of the CNV regions, with sizes in base pairs (bp).

Type	CNVRs	Mean size	Median size	Size range	CNVR Content	Sequence covered	% CNVR
Loss	202	77,420	16,543	1,716 - 2,031,343	15,638,770		0.482
Gain	102	62,147	19,580	1,737 - 416,858	6,339,041	3,247,516,410	0.195
All	304	72,295	16,678	1,716 - 2,031,343	21,977,811		0.677

SDs are associated with CNVs

Segmental duplications (SDs), defined as regions of length > = 1 kb with at least 90% sequence identity [49], are important elements in the formation of CNVs via non-allelic homologous recombination (NAHR) throughout the mammalian lineage [39,50,51]. To test whether the non-random association between CNVs and SDs was preserved in our high-resolution data, the overlap of CNVs with segmental duplications was determined. Segmental duplications were overlapped by 20% (61) of the CNVRs, which implies that CNVs are enriched near segmental duplication (permutation test, p-value < 0.001). It should be noted that the enrichment is increased when testing only the CNVRs bigger than 20 kb, with segmen-

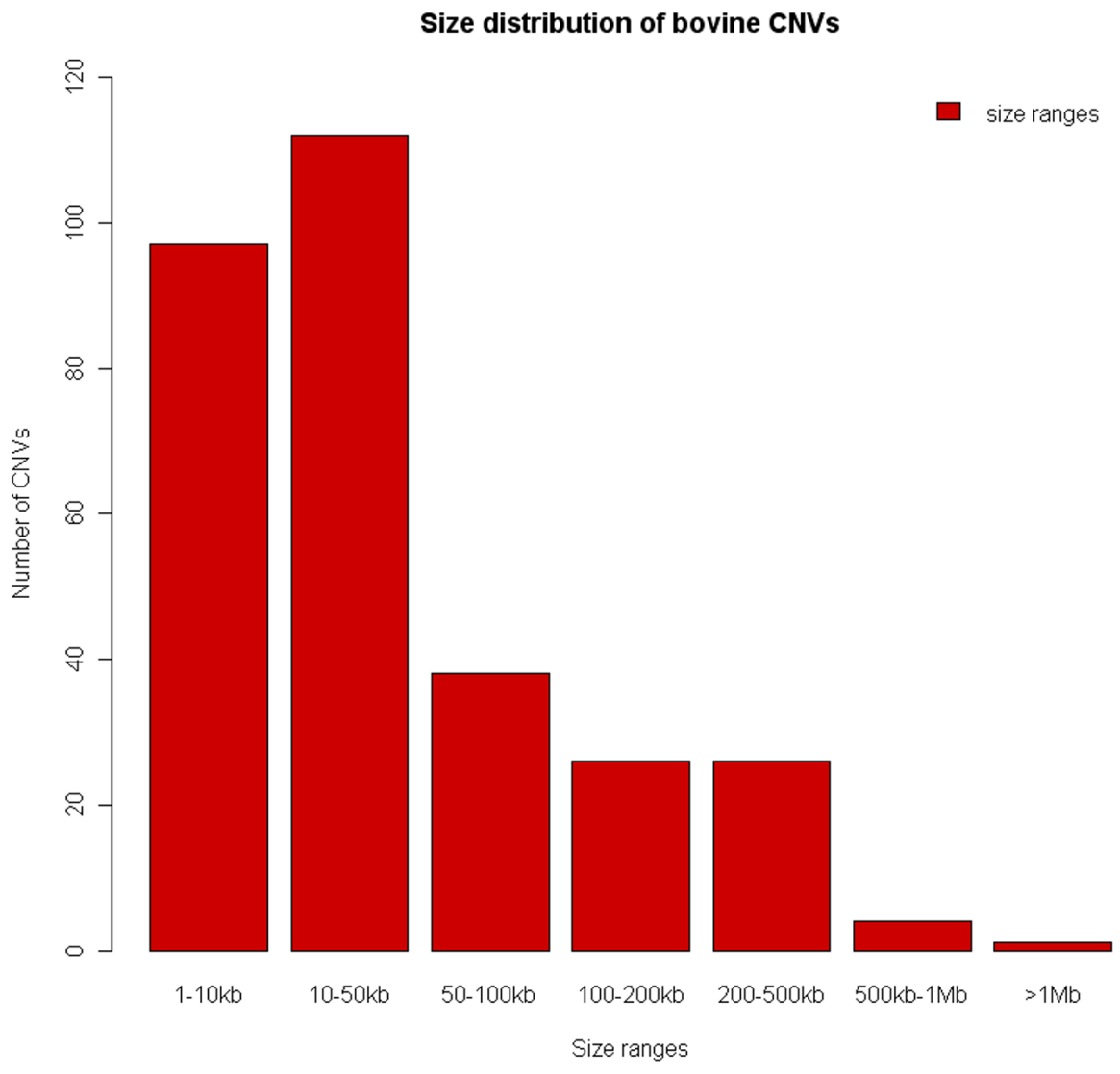

Figure 4 Size range distribution of the CNVRs detected.

tal duplications overlapping 47% of those CNVRs. This is also consistent with previous CNV studies reporting a stronger association between segmental duplications and long CNVRs [10,45].

Functional analysis

Nearly 30% (90) of the CNVRs encompassed 348 full-length genes as annotated in Ensembl [52] (Additional file 4, Table S2), but contrary to segmental duplications, the enrichment of CNVRs in genic areas is not significant. This indicates that the gene content of the CNVRs does not significantly differ from the whole bovine genome. The fact that none of the 481 ultraconserved elements [53], nor the 611 new long conserved noncoding sequences in vertebrates [54], were found in the CNV regions (whereas six of them would be expected by chance, permutation p-value < 0.001), further supports the notion that CNVs are significantly depleted in highly conserved functional elements.

In order to determine the likely biological effects of the 348 copy number variant genes, a gene ontology (GO) analysis was performed with the EasyGO tool [55]. Genes that were not completely included within the CNVRs were excluded from the GO analysis, since the breakpoint definition of CNVRs can be equivocal [56,57]. Table 2 shows that genes involved in environmental response are over-represented in the bovine CNVRs, as also seen in other studies of mammalian genomes [10,21,25,33].

The following step was to test if genes unaffected by CNVs exhibited a different selective constraint than the ones affected. To test this, the dN/dS ratios for orthologous genes between the cow and human species were compared (Table 3 and Additional file 5, Table S3). Knowing that dN is the number of nucleotide differences per non synonymous site and dS the number of nucleotide differences per synonymous site, dN/dS < 1 suggests that amino acid change is selectively constrained (purifying selection), while dN/dS ≥ 1 suggests a relaxation of that same selection (positive selection). It was determined that both deleted and duplicated genes have dN/dS ratios significantly higher than those for non-polymorphic genes. This result, as for the over represented set of 'environmental response' genes, might indicate a relaxation of constrains due to the redundancy expected from the variable number of gene copies [58-61].

When examining the human orthologs for the cattle genes [52] affected by CNV, we studied 167 human ortholog genes of which 84 overlap with the genomic coordinates of previously reported human structural variation, as seen in the Database of Genomic Variants (DGV) [7]. Since it is unlikely that CNVs in the human-cow common ancestor would have been conserved, the overlapping CNVs most probably reflect the existence of orthologous genomic regions of structural instability that are prone to recurrently generate polymorphisms in both species. However, this may also indicate that the CNVs annotated in DGV, being derived using different technological and analytical platforms, have a large variance in CNV resolution which may overestimate CNV sizes [14]. Consequently, even if the annotated CNVs represent true structural variation it is difficult to estimate the actual boundaries of the CNV and subsequently the overlap of DGV CNVs with the CNVs identified here.

CNV affecting genes associated with disease

Querying for copy number variant genes that had an orthologous human gene with OMIM morbid ID reference [62], revealed that 19 of these genes have been associated in human disease (susceptibility to sarcoidoisis and Alzheimer's disease, myopathy, encephalopathy, ataxia, etc - Additional file 6, Table S4). Likewise, when probing orthologous human genes involved in genome-wide association studies (GWAS) [63], 12 genes associated with human disease traits were found (Additional file 7, Table S5). We also queried the Animal QTL database [64] that holds publicly available QTL data on livestock species. Retrieving all the bovine QTLs within 2 Mb of our CNVRs resulted in 110 QTLs, which can hold putative valuable information for some important traits of interest (Additional file 8, Table S6). The database of Online Mendelian Inheritance in Animals (OMIA) [65] was also queried, and 21 cow phenotypes within 2 Mb of CNVRs were retrieved (Additional file 9, Table S7).

Comparison with other mammalian CNV studies

Next, we compared the number of CNVRs detected here with CNVRs from other studies (Table 4). To minimize technical CNV detection biases we: (1) used data only from the same platform (when available), (2) used data from the highest resolution genome-wide survey published on each species queried and (3) required that the study was non biased to any particular genomic region. The main finding of this comparison is that an increased resolution of the array platform increases the number of detected CNVs. This supports our finding that the bulk of CNVs in mammalian genomes are small events, implying that the characterization of the mammalian CNV landscape is far from complete. A comparison with other non-human CNV studies shows that the number of CNVRs/sample does not follow the same trend. This is expected since we have employed a number of related animals as well as the overall genetic variation in cattle is known to be reduced relative to mouse and man. Another contributing factor might be that our stringent CNV calling criteria hampers the detection of putative true CNVs. Concerning the platform used to assess CNVs in humans [17]; with a resolution to find CNVs 3 times bigger than ours, a similar difference when detecting CNVR/sample

Table 2: Enriched GO terms associated with the CNV regions (FDR p-value ≤ 0.01).

Cellular_Component		
GO term	GO name	p-value
GO:0031224	intrinsic to membrane	3.08e-10
GO:0016021	integral to membrane	3.08e-10
GO:0016020	membrane	2.73e-10
GO:0044425	membrane part	2.16e-10
Molecular_Function		
GO term	GO name	p-value
GO:0060089	molecular transducer activity	1.36e-29
GO:0004867	serine-type endopeptidase inhibitor activity	0.00448
GO:0004984	olfactory receptor activity	1.6e-26
GO:0004872	receptor activity	8.73e-31
GO:0004871	signal transducer activity	1.36e-29
GO:0004930	G-protein coupled receptor activity	1.6e-26
GO:0005044	scavenger receptor activity	1.53e-06
GO:0004888	transmembrane receptor activity	9.94e-32
Biological Process		
GO term	GO name	p-value
GO:0007165	signal transduction	7.62e-16
GO:0007166	cell surface receptor linked signal transduction	3.74e-21
GO:0050789	regulation of biological process	3.84e-10
GO:0065007	biological regulation	9.61e-10
GO:0007186	G-protein coupled receptor protein signaling pathway	1.13e-21
GO:0050794	regulation of cellular process	5.01e-10
GO:0006952	defense response	0.0107
GO:0007154	cell communication	1.82e-15

would be expected. This is not the case because CNV counts are known to be inversely proportional to their size, as seen here and elsewhere (DGV [7]).

Validation of the CNVRs

To evaluate the accuracy of the copy number assignments, quantitative real time-PCR was used as described previously [47]. Briefly, two control regions, one site on the X chromosome and one site on an autosome, plus ten potential CNV regions were selected. Six quantitative PCRs immediately confirmed the existence of copy number variation in these regions (Additional file 10, Table S8), whereas primer sets in four regions did not work satisfactorily (see Methods). Primer sets were therefore re-

Table 3: Evolutionary rates for monomorphic and CNV genes.

	Duplication CNV	Deletion CNV	No CNV
dN/dS	0.339	0.392	0.165
Wilcoxon rank-sum test, P value	1.39E-06	< 2.2e-16	-

dN - nonsynonymous rate; dS - synonymous rate; p-value compares CNV genes with monomorphic genes.

Table 4: Comparison between this and other mammalian CNV studies using array CGH.

Species	Samples	CNVRs	CNVR/ sample	Mean size	Platform	Mean spacing	Resolution	Ref
Cow	20	304	15	72	3*2.1 M	0.4	2	This
Cow	5	25	5	128	385 k	6	28.8	[35]
Dog	9	60	7	310	385 k	4.7	23.5	[33]
Mouse	20	1333	66	64	2.1 M	1	5	[45]
Rat	3	33	11	256	385 k	5	25	[30]
Macaque	10	123	12	102	385 k	6.5	32.5	[21]
Human	40	8599	215	20	20*2.1 M	0.06	0.5	[17]

Mean size, Mean spacing and Resolution are given in kilobases (kb).

designed within these four CNV regions, and using these new primers the PCR reactions were performed successfully. Thus, the existence of copy number variation in all ten regions was confirmed by quantitative PCR.

Discussion

The study outlined in this paper yields the highest-resolution analysis of bovine CNVs to date. Using a genome-wide tiling oligo array CGH, the largest number of CNV regions yet reported in cattle (304 CNVRs; average of 47 per genome) have been identified. Almost all (98%) of the CNV regions discovered here are novel relative to previous reports [35,42], thereby vastly expanding our insight of genome structural variation in cattle. With an effective resolution of 1.5 kb in detecting CNVs, resulting in a median CNV size of 16.7 kb, our data shows that at least 0.68% of the cattle genome can vary in copy number in seemingly healthy animals. This is most probably an underestimate of the true genomic fraction that is tolerant to copy number due to the low number of animals sampled and their close relatedness.

As detected previously, not only in cattle [35] but also in other species [10,25,27,33], CNVs are strongly associated with segmental duplications (SDs). This SD relation creates a lack of probe coverage in and around duplicated sequences [66], which significantly hampers the applicability of genome-wide association studies using SNP arrays to tag SDs-driven CNVs.

In addition, our data suggests that smaller CNVs (<50 kb) are much more frequent than larger ones, which is in agreement with other high resolution studies [14,17,45]. If this can be extrapolated for the whole cattle genome, the commercially available Illumina 50 k SNP panel (with an average probe spacing of 54 kb [67]) would not be sufficient to detect the bulk of existing CNVs. Consequently, further characterization of cattle CNVs should be done with similar high-density array CGH or using next-generation sequencing technologies. The latter identifies a more complete size and class ranges of structural variation [68-79].

As previously shown, copy number variants can have an impact on phenotypic variation mainly due to gene-dosage effects [80], and are often associated with disease susceptibility [38-40,81]. In this analysis, CNVRs were found to be enriched for genes with functions related to environmental response, such as immune and sensory functions previously noticed in other species [10,25,27,33,35]. The enrichment aspect is an interesting finding, since variation in immunity related genes have been associated with disease. In particular, genes of the major histocompatibility complex (MHC) http://www.ebi.ac.uk/ipd/mhc/, of which some are included in our dataset, are reported to be responsible for differences in predisposition to diseases like mastitis, dermatophilosis and other tick infections [41]. Concerning the genetics of milk production and lactation, we found none of the 197 unique milk protein genes and the over 6000 mammary-related genes within our CNV regions. This is expected since these genes are known to be highly conserved and evolving more slowly than other genes in the bovine genome [82].

Many genes and QTLs associated with human and cow diseases were found to be copy number variable or located nearby CNV regions. The fact that some bovine CNVs occurred in regions orthologous to human CNVs, reflect most likely recurrent CNV formation, rather than ancestral CNVs maintained in both species. These regions could be hotspots of CNV genesis due to their fragile structural architecture that prompts frequent rearrangements.

Conclusions

In summary, the data presented here extends and establish the fact that a significant part of cattle genome is copy number variable within and between breeds and that our high-resolution array CGH is a valid method to detect bovine CNVs in a genome-wide manner. With a

limited amount of sampled animals and breeds, and a stringent CNV calling criteria, the CNV regions reported here are believed to be highly reliable, but the number might greatly underestimate true number of CNVs in cattle populations. Consequently, future studies are required to assess the functional significance of CNVs and their impact on health and productive efficiency in cattle.

Methods

Sample preparation

The genomic DNA of 20 bovine samples was obtained from 4 dairy and beef breeds (14 Holsteins, 3 Simmental 2 Red danish and 1 Hereford). The pedigree scheme for the related animals is in Additional file 2, Figure S1. DNA was extracted and purified from blood as described elsewhere [47], in order to pass Nimblegen quality control requirements. We adhered to our national and institutional guidelines for the ethical use and treatment of animals in experiments.

Array CGH

DNA fragmentation, labeling, hybridization, washing and array imaging were carried out according to the manufacturer's protocol and done as previously described [83]. Briefly, the genomic DNA samples were fragmented by sonication and labeled with fluorescent dyes Cy3 and Cy5. According to the dye swap loop design (Additional file 12, Table S10), samples were co-hybridized with a MAUI hybridization system (BioMicro Systems) to custom-made cattle CGH 2.1 M (HD2) arrays (Roche NimbleGen, Madison, WI). In order to cover the latest bovine genome assembly (bt4) with high density, the custom CGH arrays were planned in 3 designs. Each design covered a specific set of chromosomes with 2.1 million probes, which yielded 420 bp of average probe spacing (301 bp median probe spacing). The probe design fundamentals are described by the array manufacturer and elsewhere [47].

The arrays were scanned using a 5 μm scanner, and Nimblescan software (Roche Nimblegen, Madison, WI) was used to retrieve fluorescent intensity raw data from the scanned images of the oligonucleotide tiling arrays. For each spot on the array, log2-ratios of the Cy3-labeled test sample versus Cy-5 reference sample were computed. Before normalization and segmentation analysis, spatial correction was applied. Spatial correction reduces some artifacts observed in CGH data from 2.1 M arrays, adjusting position-dependent non-uniformity of signals across the array. Specifically, locally weighted polynomial regression (loess) was used to adjust signal intensities based on X, Y feature position [84]. Normalization was then performed using the q-spline method [85], followed by segmentation using the CNV calling algorithm segMNT [86]. This algorithm is shown to outperform both DNACopy [87], which is one of the most widely used CNV calling algorithm in the literature, and StepGram [88], the algorithm used by Agilent for CGH arrays. The segments with mean log2ratio ≥ |0.4| and at least 5 consecutive probes were retained. From these, a CNV was called if it was detected in both dye swap arrays and detected at least in two different dye swap hybridizations (i.e. in two hybridizations with an animal in common). Since the CNV calling pipeline requires at least 5 consecutive probes before calling a region copy number variant, our theoretical resolution for CNV detection is 1465 bp (median spacing*4 + median oligo length*5).

False positive rate

The false positive rate was calculated based on the 8 sex-mismatched arrays in this study: the length of chromosome × (from all the 8 hybs.) having a log2-ratio with a different signal than it should (given the sex-mismatched hybridization), dividing by the length of chrX multiplied by the number of sex-mismatched arrays (25,694,212/(88,516,663*8) = 3.62%). Since the FPR can be overestimated from sex-mismatched arrays, due to the assumption that no CNV exist in the chrX of sex-mismatched arrays, the FPR was also calculated from the self-self experiment and was determined as the length of sequence that would normally be called a CNV with our CNV calling pipeline. The FPR was determined as being 0.0085%.

Enrichment analysis

Bovine segmental duplication (SD) data was retrieved from [89]. They used two independent approaches to detect segmental duplications: WGAC (whole-genome assembly comparison), which is a BLAST-based analysis of all assembled sequence that detects self alignments (>90% and 1 kb); and WSSD (whole-genome shotgun sequence detection), which is an assembly-independent approach that examines the reference sequence for an increase in WGS read depth-of-coverage. This strategy has been used previously to map SDs in the human [49] and mouse [27] genomes. From their global data we choose to filter out those SDs bigger than 94% identity using WGAC if they were not also confirmed by WSSD. The reason for this relates to the fact that the assembly of highly similar duplicated sequences will often be missed, collapsed or mis-assigned [90,91].

The association of CNVRs with genomic features (SDs, assembly gaps, genes and conserved elements) was tested by randomly permuting the genomic position of each CNVR 10 000 times and determining the sequence content of the resulting region or flanking regions.

Real-Time PCR

Validation with RT-PCR was executed as previously described [47], with the Applied Biosystems 7900HT

Sequence Detection System used for the Taqman assays, and downstream analysis performed with SDS 2.2 software. The full sequence of the CNVR was BLASTN-searched against the bovine genome sequence in order to identify a subregion that was unique and specific to the chromosomal location of the CNVR. PCR primers and probes were designed in this subregion of the CNVR using the ProbeFinder software from Roche Applied Science (Additional file 10, Table S8). Criteria for classifying a "not working" primer involved two parameters: reaction efficiency below 85%, and Pearson correlation of each standard curve below 0.95. Only ten of the original twenty bovine samples were used due to lack of DNA availability. For each target, the relative quantification analysis with a reference female sample was done to calculate estimated copy numbers of each sample.

Data availability

The full data set and designs from the oligo array CGH experiments have been submitted to GEO [92] under the accession ID GSE18174.

Additional material

Additional file 1 Excel file includes Table S1. CNV regions detected and the number of CNVs per sample.

Additional file 2 Powerpoint file includes Figure S1. Pedigree for the related animals. The numbers correspond to the sample ID, described in Table S1.

Additional file 3 Excel file includes Figure S2. Correlation between chromosome length and number of CNVRs.

Additional file 11 Excel file includes Table S9. Sequences in ChrUn that putatively belong to chromosome Y.

Additional file 4 Excel file includes Table S2. CNVRs and ensembl genes within CNVRs.

Additional file 5 Excel file includes Table S3. Cow-human ortholog genes with the respective dN and dS values (data from Ensembl database [52]).

Additional file 6 Text file includes Table S4. Genes within CNV regions for which cow-human orthology exist and have a OMIM morbid ID [62].

Additional file 7 Excel file includes Table S5. Genes within CNV regions for which cow-human orthology exist and have been associated with a phenotype by a GWAS study.

Additional file 8 Excel file includes Table S6. All CNVRs and QTLs within 2 Mb of a CNVR.

Additional file 9 Excel file includes Table S7. All CNVRs and OMIA genes within 2 Mb of a CNVR.

Additional file 10 Excel file includes Table S8. Primers and result plots for the RT-PCR experiments.

Additional file 12 Excel file includes Table S10. Array CGH hybridizations. Extra information on the samples and experiment design is in GEO [92].

Authors' contributions

JF, BT, LEH and CB designed the project. JF performed data analysis and drafted the manuscript. BT planned the RT-PCR validation experiments. CB was the principal investigator of the project. All the authors have contributed to writing this manuscript and have read and approved the contents of the final submitted version.

Acknowledgements

We thank the Bovine Genome Sequencing and Analysis Consortium for providing access to the *Bos Taurus* genome sequence assembly. We thank Søren Svendsen and Hanne Jørgensen for technical assistance. This work was conducted as part of the SABRETRAIN Project, funded by the Marie Curie Host Fellowships for Early Stage Research Training, as part of the 6th Framework Programme of the European Commission. We also acknowledge the support from The Danish Food Industry Agency, Danish Ministry of Food, Agriculture and Fisheries as well as support from Viking Genetics.

Author Details

Group of Molecular Genetics and Systems Biology, Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, Aarhus University, Blichers Allé 20, DK-8830 Tjele, Denmark

Received: 11 December 2009 Accepted: 6 May 2010
Published: 6 May 2010

References

1. Tellam RL, Lemay DG, Van Tassell CP, Lewin HA, Worley KC, Elsik CG: **Unlocking the Bovine Genome.** *BMC Genomics* 2009, **10:**193.
2. Bovine Genome Sequencing and Analysis Consortium: **The genome sequence of taurine cattle: A window to ruminant biology and evolution.** *Science* 2009, **324:**522-528.
3. Bovine HapMap Consortium: **Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds.** *Science* 2009, **324:**528-32.
4. Bridges CB: **The Bar "gene" duplication.** *Science* 1936, **83:**210-11.
5. Lejeune J, Gautier M, Turpin R: **Study of somatic chromosomes from 9 mongoloid children.** *C R Hebd Seances Acad Sci* 1959, **248:**1721-1722.
6. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Månér S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M: **Large-scale copy number polymorphism in the human genome.** *Science* 2004, **305:**525-528.
7. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C: **Detection of large-scale variation in the human genome.** *Nat Genet* 2004, **36:**949-951.
8. Feuk L, Carson AR, Scherer SW: **Structural variation in the human genome.** *Nat Rev Genet* 2006, **7:**85-97.
9. International HapMap Consortium: **A haplotype map of the human genome.** *Nature* 2005, **437:**1299-1320.
10. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, et al.: **Global variation in copy number in the human genome.** *Nature* 2006, **444:**444-454.
11. Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ, Eichler EE, Lam WL: **A comprehensive analysis of common copy-number variations in the human genome.** *Am J Hum Genet* 2007, **80:**91-104.
12. Zogopoulos G, Ha KC, Naqib F, Moore S, Kim H, Montpetit A, Robidoux F, Laflamme P, Cotterchio M, Greenwood C, Scherer SW, Zanke B, Hudson TJ, Bader GD, Gallinger S: **Germ-line dna copy number variation frequencies in a large north american population.** *Hum Genet* 2007, **122:**345-353.
13. de Smith AJ, Tsalenko A, Sampas N, Scheffer A, Yamada NA, Tsang P, Ben-Dor A, Yakhini Z, Ellis RJ, Bruhn L, Laderman S, Froguel P, Blakemore AIF: **Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males: implications for association studies of complex diseases.** *Hum Mol Genet* 2007, **16:**2783-2794.
14. Perry GH, Ben-Dor A, Tsalenko A, Sampas N, Rodriguez-Revenga L, Tran CW, Scheffer A, Steinfeld I, Tsang P, Yamada NA, Park HS, Kim JI, Seo JS, Yakhini Z, Laderman S, Bruhn L, Lee C: **The Fine-Scale and Complex Architecture of Human Copy-Number Variation.** *Am J Hum Genet* 2008, **82:**685-695.
15. Shaikh TH, Gai X, Perin JC, Glessner JT, Xie H, Murphy K, O'Hara R, Casalunovo T, Conlin LK, D'Arcy M, et al.: **High-resolution mapping and analysis of copy number variations in the human genome: A data resource for clinical and research applications.** *Genome Res* 2009, **19:**1682-1690.

16. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, et al.: **The diploid genome sequence of an individual human.** *PLoS Biol* 2007, **5:**e254.
17. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Wellcome Trust Case Control Consortium, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME: **Origins and functional impact of copy number variation in the human genome.** *Nature* 2010, **464:**704-712.
18. Lupski JR: **Genomic rearrangements and sporadic disease.** *Nat Genet* 2007, **39(Supplement):**S43-S47.
19. Perry GH, Tchinda J, McGrath SD, Zhang J, Picker SR, Caceres AM, Iafrate AJ, Tyler-Smith C, Scherer SW, Eichler EE, Stone AC, Lee C: **Hotspots for copy number variation in chimpanzees and humans.** *Proc Natl Acad Sci USA* 2006, **103:**8006-8011.
20. Kehrer-Sawatzki H, Cooper DN: **Structural divergence between the human and chimpanzee genomes.** *Hum Genet* 2007, **120:**759-778.
21. Lee AS, Gutierrez-Arcelus M, Perry GH, Vallender EJ, Johnson WE, Miller GM, Korbel JO, Lee C: **Analysis of copy number variation in the rhesus macaque genome identifies candidate loci for evolutionary and human disease studies.** *Hum Mol Genet* 2008, **17:**1127-1136.
22. Cutler G, Marshall LA, Chin N, Baribault H, Kassner PD: **Significant gene content variation characterizes the genomes of inbred mouse strains.** *Genome Res* 2007, **17:**1743-1754.
23. Li J, Jiang T, Mao JH, Balmain A, Peterson L, Harris C, Rao PH, Havlak P, Gibbs R, Cai WW: **Genomic segmental polymorphisms in inbred mouse strains.** *Nat Genet* 2004, **36:**952-954.
24. Adams DJ, Dermitzakis ET, Cox T, Smith J, Davies R, Banerjee R, Bonfield J, Mullikin JC, Chung YJ, Rogers J, Bradley A: **Complex haplotypes, copy number polymorphisms and coding variation in two recently divergent mouse strains.** *Nat Genet* 2005, **37:**532-536.
25. Graubert TA, Cahan P, Edwin D, Selzer RR, Richmond TA, Eis PS, Shannon WD, Li X, McLeod HL, Cheverud JM, Ley TJ: **A high-resolution map of segmental DNA copy number variation in the mouse genome.** *PLoS Genet* 2007, **3:**e3.
26. Egan CM, Sridhar S, Wigler M, Hall IM: **Recurrent DNA copy number variation in the laboratory mouse.** *Nat Genet* 2007, **39:**1384-1389.
27. She X, Cheng Z, Zollner S, Church DM, Eichler EE: **Mouse segmental duplication and copy number variation.** *Nat Genet* 2008, **40:**909-914.
28. Snijders AM, Nowak NJ, Huey B, Fridlyand J, Law S, Conroy J, Tokuyasu T, Demir K, Chiu R, Mao JH, Jain AN, Jones SJ, Balmain A, Pinkel D, Albertson DG: **Mapping segmental and sequence variations among laboratory mice using BAC array CGH.** *Genome Res* 2005, **15:**302-311.
29. Henrichsen CN, Vinckenbosch N, Zöllner S, Chaignat E, Pradervand S, Schütz F, Ruedi M, Kaessmann H, Reymond A: **Segmental copy number variation shapes tissue transcriptome.** *Nat Genet* 2009, **41:**424-429.
30. Guryev V, Saar K, Adamovic T, Verheul M, van Heesch SA, Cook S, Pravenec M, Aitman T, Jacob H, Shull JD, Hubner N, Cuppen E: **Distribution and functional impact of DNA copy number variation in the rat.** *Nat Genet* 2008, **40:**538-545.
31. Dopman EB, Hartl DL: **A portrait of copy-number polymorphism in Drosophila melanogaster.** *Proc Natl Acad Sci USA* 2007, **104:**19920-19925.
32. Emerson JJ, Cardoso-Moreira M, Borevitz JO, Long M: **Natural selection shapes genome-wide patterns of copy-number polymorphism in Drosophila melanogaster.** *Science* 2008, **320:**1629-1631.
33. Chen WK, Swartz JD, Rush LJ, Alvarez CE: **Mapping dna structural variation in dogs.** *Genome Res* 2009, **19:**500-509.
34. Crooijmans RPMA, Groenen MAM: **Copy Number Variation Analysis in Chicken.** *Proceedings of Plant & Animal Genomes XVII Conference; 10-14 January 2009; San Diego* 2009:W397.
35. Liu GE, Van Tassel CP, Sonstegard TS, Li RW, Alexander LJ, Keele JW, Matukumalli LK, Smith TP, Gasbarre LC: **Detection of germline and somatic copy number variations in cattle.** *Developments in Biologicals* 2008, **132:**231-237.
36. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavaré S, Deloukas P, Hurles ME, Dermitzakis ET: **Relative impact of nucleotide and copy number variation on gene expression phenotypes.** *Science* 2007, **315:**848-853.
37. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter NP, Lee C, Stone AC: **Diet and the evolution of human amylase gene copy number variation.** *Nat Genet* 2007, **39:**1256-1260.
38. Shastry BS: **Copy number variation and susceptibility to human disorders (review).** *Molecular Medicine reports* 2009, **2:**143-147.
39. Lupski JR, Stankiewicz P: **Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes.** *PLoS Genet* 2005, **1:**e49.
40. Reymond A, Henrichsen CN, Harewood L, Merla G: **Side effects of genome structural changes.** *Curr Opin Genet Dev* 2007, **17:**381-386.
41. Ibeagha-Awemu EM, Kgwatalala P, Ibeagha AE, Zhao X: **A critical analysis of disease-associated DNA polymorphisms in the genes of cattle, goat, sheep, and pig.** *Mamm Genome* 2008, **19:**226-245.
42. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TP, Sonstegard TS, Van Tassell CP: **Development and Characterization of a High Density SNP Genotyping Assay for Cattle.** *PLoS ONE* 2009, **4:**e5350.
43. Rosa GJ, de Leon N, Rosa AJ: **Review of microarray experimental design strategies for genetical genomics studies.** *Physiol Genomics* 2006, **28:**15-23.
44. Bueno Filho JS, Gilmour SG, Rosa GJ: **Design of microarray experiments for genetical genomics studies.** *Genetics* 2006, **174:**945-957.
45. Cahan P, Li Y, Izumi M, Graubert TA: **The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells.** *Nature Genet* 2009, **41:**430-437.
46. Turner DJ, Miretti M, Rajan D, Fiegler H, Carter NP, Blayney ML, Beck S, Hurles ME: **Germline rates of de novo meiotic deletions and duplications causing several genomic disorders.** *Nature Genet* 2008, **40:**90-95.
47. Fadista J, Nygaard M, Holm LE, Thomsen B, Bendixen C: **A Snapshot of CNVs in the Pig Genome.** *PLoS ONE* 2008, **3:**e3916.
48. Liu Y, Qin X, Song XZH, Jiang H, Shen Y, Durbin KJ, Lien S, Kent MP, Sodeland M, Ren Y, Zhang L, Sodergren E, Havlak P, Worley KC, Weinstock GM, Gibbs RA: **Bos taurus genome assembly.** *BMC Genomics* 2009, **10:**180.
49. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Evan EE: **Recent segmental duplications in the human genome.** *Science* 2002, **297:**1003-1007.
50. Shaw CJ, Lupski JR: **Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease.** *Hum Mol Genet* 2004, **13:**R57-R64.
51. Hastings PJ, Lupski JR, Rosenberg SM, Ira G: **Mechanisms of change in gene copy number.** *Nat Rev Genet* 2009, **10:**551-564.
52. Hubbard TJB, Aken BL, Ayling S, Ballester B, Beal K, Bragin E, Brent S, Chen Y, Clapham P, Clarke L, et al.: **Ensembl 2009.** *Nucl Acids Res* 2009, **37(Suppl 1):**D690-D697.
53. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D: **Ultraconserved elements in the human genome.** *Science* 2004, **304:**1321-1325.
54. Sakuraba Y, Kimura T, Masuya H, Noguchi H, Sezutsu H, Takahasi KR, Toyoda A, Fukumura R, Murata T, Sakaki Y, Yamamura M, Wakana S, Noda T, Shiroishi T, Gondo Y: **Identification and characterization of new long conserved noncoding sequences in vertebrates.** *Mamm Genome* 2008, **19:**703-712.
55. Xin Z, Zhen S: **EasyGO: Gene Ontology-based annotation and functional enrichment analysis tool for agronomical species.** *BMC Genomics* 2007, **8:**246.
56. Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders AC, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M: **Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome.** *Science* 2007, **318:**420-426.
57. Nozawa M, Kawahara Y, Nei M: **Genomic drift and copy number variation of sensory receptor genes in humans.** *Proc Natl Acad Sci USA* 2007, **104:**20421-20426.
58. Kondrashov FA, Kondrashov AS: **Role of selection in fixation of gene duplications.** *J Theor Biol* 2006, **239:**141-151.
59. Nguyen DQ, Webber C, Ponting CP: **Bias of Selection on Human Copy-Number Variants.** *PLoS Genet* 2006, **2:**e20.
60. Ohno S: *Evolution by gene duplication* New York: Springer-Verlag; 1970.

61. Nguyen DQ, Webber C, Hehir-Kwa J, Pfundt R, Veltman J, Ponting CP: **Reduced purifying selection prevails over positive selection in human copy number variant evolution.** *Genome Res* 2008, **18:**1711-1723.
62. **Online Mendelian Inheritance in Man (OMIM)** [http://www.ncbi.nlm.nih.gov/omim/]
63. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA: **Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.** *Proc Natl Acad Sci USA* 2009, **106:**9362-9367.
64. Hu Z, Reecy JM: **Animal QTLdb: Beyond a Repository - A Public Platform for QTL Comparisons and Integration with Diverse Types of Structural Genomic Information.** *Mamm Genome* 2007, **18:**1-4.
65. **Online Mendelian Inheritance in Animals (OMIA)** [http://omia.angis.org.au/]
66. Cooper GM, Zerr T, Kidd JM, Eichler EE, Nickerson DA: **Systematic assessment of copy number variant detection via genome-wide SNP genotyping.** *Nat Genet* 2008, **40:**1199-1203.
67. Villa-Angulo R, Matukumalli LK, Gill CA, Choi J, Van Tassell CP, Grefenstette JJ: **High-resolution haplotype block structure in the cattle genome.** *BMC Genetics* 2009, **10:**19.
68. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM: **The complete genome of an individual by massively parallel DNA sequencing.** *Nature* 2008, **452:**872-876.
69. Bentley D, Balasubramanian S, Swerdlow H, Smith G, Milton J, Brown C, Hall K, Evers D, Barnes C, Bignell H, et al.: **Accurate whole human genome sequencing using reversible terminator chemistry.** *Nature* 2008, **456:**53-59.
70. Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Guo Y, et al.: **The diploid genome sequence of an Asian individual.** *Nature* 2008, **456:**60-65.
71. Ahn SM, Kim TH, Lee S, Kim D, Ghang H, Kim DS, Kim BC, Kim SY, Kim WY, Kim C, Park D, Lee YS, Kim S, Reja R, Jho S, Kim CG, Cha JY, Kim KH, Lee B, Bhak J, Kim SJ: **The first Korean genome sequence and analysis: Full genome sequencing for a socio-ethnic group.** *Genome Res* 2009, **19:**1622-1629.
72. McKernan KJ, Peckham HE, Costa GL, McLaughlin SF, Fu Y, Tsung EF, Clouser CR, Duncan C, Ichikawa JK, Lee CC, et al.: **Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two base encoding.** *Genome Res* 2009, **19:**1527-1541.
73. Kim JI, Ju YS, Park H, Kim S, Lee S, Yi JH, Mudge J, Miller NA, Hong D, Bell CJ, et al.: **A highly annotated whole-genome sequence of a Korean individual.** *Nature* 2009, **460:**1011-1015.
74. Chiang DY, Getz G, Jaffe DB, O'Kelly MJ, Zhao X, Carter SL, Russ C, Nusbaum C, Meyerson M, Lander ES: **High-resolution mapping of copy-number alterations with massively parallel sequencing.** *Nat Methods* 2008, **6:**99-103.
75. Xie C, Tammi M: **CNV-seq, a new method to detect copy number variation using high-throughput sequencing.** *BMC Bioinformatics* 2009, **10:**80.
76. Yoon S, Xuan Z, Makarov V, Ye K, Sebat J: **Sensitive and accurate detection of copy number variants using read depth of coverage.** *Genome Res* 2009, **19:**1586-1592.
77. Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O, Sahinalp SC, Gibbs RA, Eichler EE: **Personalized copy number and segmental duplication maps using next-generation sequencing.** *Nat Genet* 2009, **41:**1061-1067.
78. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP, Shi X, Fulton RS, Ley TJ, Wilson RK, Ding L, Mardis ER: **Breakdancer: an algorithm for high-resolution mapping of genomic structural variation.** *Nat Methods* 2009, **6:**677-681.
79. Lee S, Hormozdiari F, Alkan C, Brudno M: **Modil: detecting small indels from clone-end sequencing with mixtures of distributions.** *Nat Methods* 2009, **6:**473-474.
80. Korbel JO, Kim PM, Chen X, Urban AE, Weissman S, Snyder M, Gerstein MB: **The current excitement about copy-number variation: how it relates to gene duplications and protein families.** *Curr Opin Struct Biol* 2008, **18:**366-374.
81. Cook H, Scherer SW: **Copy-number variations associated with neuropsychiatric conditions.** *Nature* 2008, **455:**919-923.
82. Lemay DG, Lynn DJ, Martin WF, Neville MC, Casey TM, Rincon G, Kriventseva EV, Barris WC, Hinrichs AS, Molenaar AJ, Pollard KS, Maqbool NJ, Singh K, Murney R, Zdobnov EM, Tellam RL, Medrano JF, German JB, Rijnkels M: **The bovine lactation genome: insights into the evolution of mammalian milk.** *Genome Biol* 2009, **10:**R43.
83. Selzer RR, Richmond TA, Pofahl NJ, Green RD, Eis PS, Nair P, Brothman AR, Stallings RL: **Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH.** *Genes Chrom Cancer* 2005, **44:**305-319.
84. Smyth GK, Speed TP: **Normalization of cDNA microarray data.** *Methods* 2003, **31:**265-273.
85. Workman C, Jensen LJ, Jarmer H, Berka R, Gautier L, Nielser HB, Saxild HH, Nielsen C, Brunak S, Knudsen S: **A new non-linear normalization method for reducing variability in DNA microarray experiments.** *Genome Biol* 2002, **3:**1-16.
86. Molla M: **Novel Uses for Machine Learning and Other Computational Methods for the Design and Interpretation of Genetic Microarrays.** In *PhD thesis* University of Wisconsin-Madison Department of Computer Sciences; 2007.
87. Olshen A, Venkatraman E, Lucito R, Wigler M: **Circular binary segmentation for the analysis of array-based DNA copy number data.** *Biostatistics* 2004, **5:**557-572.
88. Lipson D, Aumann Y, Ben-Dor A, Linial N, Yakhini Z: **Efficient calculation of interval scores for DNA copy number data analysis.** *J Comput Biol* 2006, **13:**215-228.
89. Liu GE, Ventura M, Cellamare A, Chen Li, Cheng Z, Bin Zhu, Li C, Song J, Eichler EE: **Analysis of recent segmental duplications in the bovine genome.** *BMC Genomics* 2009, **10:**571.
90. She X, Jiang Z, Clark RA, Liu G, Cheng Z, Tuzun E, Church DM, Sutton G, Halpern AL, Eichler EE: **Shotgun sequence assembly and recent segmental duplications within the human genome.** *Nature* 2004, **431:**927-930.
91. Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE: **Segmental duplications: organization and impact within the current human genome project assembly.** *Genome Res* 2001, **11:**1005-1017.
92. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muertter RN, Edgar R: **NCBI GEO: archive for high-throughput functional genomic data.** *Nucleic Acids Res* 2009:D885-D890.

doi: 10.1186/1471-2164-11-284
Cite this article as: Fadista *et al.*, Copy number variation in the bovine genome *BMC Genomics* 2010, **11**:284

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central