

SCIENTIFIC REPORTS



OPEN

In vivo and *in vitro* ageing results in accumulation of *de novo* copy number variations in bulls

Tamas Revay¹, Olutobi Oluwole¹, Tom Kroetsch² & W. Allan King¹

We have identified *de novo* copy number variations (CNVs) generated in bulls as they age. Blood samples from eight bulls were collected and SNP arrayed in a prospective design over 30 months allowing us to differentiate *de novo* CNVs from constant CNVs that are present throughout the sampling period. Quite remarkably, the total number of CNVs doubled over the 30-month period, as we observed an almost equal number of *de novo* and constant CNVs (107 and 111, respectively, i.e. 49% and 51%). Twice as many *de novo* CNVs emerged during the second half of the sampling schedule as in the first part. It suggests a dynamic generation of *de novo* CNVs in the bovine genome that becomes more frequent as the age of the animal progresses. In a second experiment *de novo* CNVs were detected through *in vitro* ageing of bovine fibroblasts by sampling passage #5, #15 and #25. *De novo* CNVs also became more frequent, but the proportion of them was only ~25% of the total number of CNVs (21 out of 85). Temporal generation of *de novo* CNVs resulted in increasing genome coverage. Genes and quantitative trait loci overlapping *de novo* CNVs were further investigated for ageing related functions.

Copy number variants (CNVs) are deleted or duplicated segments of the genome that have been identified as a prominent source of inter-individual genetic variation. A current CNV map estimates that 4.8–9.5% of the human genome can be affected by these events without apparent phenotypic consequences¹. CNVs however alter the balance of DNA content and have been associated with disease conditions, such as diabetes, autism and schizophrenia². Recently the CNV landscape of many other species including domestic animals has also been investigated³. The potential to affect gene expression or regulation defines CNVs as candidate selection markers for modern breeding schemes; e.g. milk yield in cattle and fertility in pigs have been associated with various CNVs^{4,5}.

The mechanisms of DNA replication and repair are not only essential for the divisions producing the trillions of cells of a whole organism, but are also major factors in generating CNVs and genetic diversity⁶. The resulting intra-individual somatic variability has been observed at the level of CNVs among differentiated tissues of humans and cattle^{7,8}. The progressive and systematic changes of cellular structures are hallmarks of ageing, thus the temporal generation of CNVs is a logical subject of longevity or ageing related studies. On the other hand, the dynamics of *de novo* somatic CNV formation is an understudied aspect of the biology of ageing. Kuningas *et al.*⁹ have observed two common CNV regions and the burden of large common CNVs were associated with higher mortality among the participants of the Rotterdam and Farmingham human ageing studies. Similarly, the increasing total length of genome-wide CNVs were associated with mortality in another population study¹⁰. Accumulation of CNVs in human blood and colon cells over time was also described^{11,12}. Although cattle CNVs have been investigated in a number of studies, most are oriented to reveal variability in populations and potential associations with production traits¹³ and temporal changes of the CNVs landscape has never been described.

The aim of our study was to investigate the dynamics of CNV formation *in vivo* and *in vitro* during prospective sampling of individual bulls and fibroblast cultures. Our aim was not to create a snapshot from the genomes of very mature individuals, as in many genetic studies of ageing, instead to carry out an experiment in bulls where the sequential sampling provides follow up information on the same individual genomes. These *in vivo* and *in vitro* models of the changing CNV landscape provide novel information of the extent of *de novo* CNVs in bulls as well as their potential connection to ageing related genomic changes.

¹Department of Biomedical Sciences, University of Guelph, Guelph, ON, Canada. ²Semex, 5653 HWY6 North, Guelph, ON, Canada. Correspondence and requests for materials should be addressed to W.A.K. (email: waking@uoguelph.ca)

Results

CNV analysis. The experiments aimed to detect and analyze CNVs from data collected on the Bovine SNP50k chip were performed on two sets of samples. First, the *in vivo* 'AGE' array consisted of testing DNA from blood samples of eight bulls collected at three consecutive time points (at 14 ± 3 month of age, then 18 months and 30 months later, Supplementary Figure S1). Moreover, three bovine fibroblast cell lines were established and DNA samples extracted from three consecutive passages (P5, P15, P25) tested on the *in vitro* 'FIBRO' array.

Several quality control steps were performed for both arrays prior to CNV scoring and analysis. Analysing samples by either the derivative log ratio data or the genomic waves characteristic yielded no outlier samples. Although the 24 DNA samples from the eight bulls (AGE) were optimally processed together since they completely filled one array (24 positions), we investigated potential batch effects by principal component analysis (Supplementary Figure S2). This revealed no clustering of the raw data according to the sampling time and minimal tendency of grouping based on the genetic origin (animals). All nine fibroblast samples were arrayed together on the FIBRO array as well. The PCA did not detect clusters by cell passages, but similarity of samples from the same cell lines were observed.

The Univariate-CNAM algorithm of the SVS software detected CNVs by identifying segments of logR ratio values that are significantly different from the neighbouring values. This procedure searches along the genome in individual samples, generating a list of segments that were then critically evaluated (using the segment mean histogram and visualization of the genomic loci) to form a final list of 218 CNVs in total from the *in vivo* and 85 CNVs from the *in vitro* samples. That makes a very similar average number of CNVs/sample for the two dataset (27 CNVs/AGE samples and 28 CNVs/FIBRO samples). The majority of CNVs were deletions, as only eight and seven gains were detected *in vivo* and *in vitro*, respectively, as also observed in earlier studies using the same platform¹⁴. The mean and median length of a CNV was slightly, but not significantly, longer for AGE than for FIBRO (mean: 61 kb vs. 47 kb, median: 31.9 kb vs. 31.7 kb, respectively). The individual CNVs were named and numbered according to the dataset, as CNV-A1 to CNV-A218 for AGE and CNV-F1 to CNV-F85 for FIBRO (Tables S2, S3 lists all CNVs).

The next analysis step was to group and compare CNV calls among the three consecutive samples from the same bulls or cell lines (Table S1). CNVs detected at all three time points represent events that exist in the genome since the beginning of the sampling scheme e.g. a deletion that was present in the genome of one particular bull at the age of 12 months and detected again at later times in the 2nd and 3rd sample collected from the same animal. Similarly, there are CNVs that could be found in all passage samples of a particular fibroblast line. These events are referred to as "constant", reflecting the fact that they did not change over the sampling period (Table S1). We found 111 constant CNVs *in vivo* and 64 *in vitro*. On the other hand, we looked for CNV events that were not present in the first sample, but arose later in the sampling schedule. If a CNV was detected in the 2nd sample (18 months later or P15) a scoring criterion was its presence in the 3rd time point as well. There were 36 *in vivo* and five *in vitro* CNVs detected in the 2nd and 3rd set of samples but not in the 1st set. Moreover, there were 71 *in vivo* and 16 *in vitro* CNVs of the latter class, those detected only in the 3rd sample (30 month later or P25). Due to the fact that these last two categories of CNVs were not present at the beginning of the experiments, but presumably generated *de novo* during the time the sampling was scheduled we refer to them as "*de novo*". There were 107 *in vivo* and 21 *in vitro de novo* CNVs. The ratios of total *de novo* vs. total constant CNVs for both *in vivo* and *in vitro* samples were significantly different ($p < 0.0001$, binomial test), as were the ratios of average *de novo* vs. average constant CNVs per sample (Fig. 1a,b). The genome coverage was calculated for each sample and reflected an increasing length when *de novo* CNVs were added at the 2nd and 3rd time points to the portion covered originally by the constant CNVs (Fig. 1c,d).

No constant or *de novo* CNV was detected on BTA26 (for AGE) or on BTA13, BTA18, BTA20 for FIBRO (Supplementary Figure S4). The largest number of CNVs were detected on BTA2, BTA4, BTA8 and BTA12 for AGE and on BTA2, BTA3, BTA11 for FIBRO, while eight other chromosomes had only one CNV. Individual counts of either constant or *de novo* (2nd & 3rd or 3rd) CNVs varied among the bulls or cell cultures (Supplementary Figure S5). Constant CNVs were in excess in most cases, especially for the *in vitro* samples. However, the number of merged *de novo* events resulted in inverting this difference in six out of eight *in vivo* samples, while it did not change the surplus of constant CNVs in all *in vitro* samples (Supplementary Figure S5c,d). There were on average 14 constant and 13 *de novo* CNVs per *in vivo* sample (median = 14 for both CNV classes), and 21 constant and 7 *de novo* CNVs per cell line (median = 22 and 7, respectively).

A quantitative PCR (qPCR) validation of 28 *de novo* CNVs (20 *in vivo* and 8 *in vitro*) were performed and the results were in 80% agreement with the array-based status predictions (Table S4).

CNV gene content. Positions of all identified CNVs were used to search for overlapping bovine RefSeq genes. 35 AGE CNVs and 26 FIBRO CNVs encompassed a total of 63 genes with 25 unique to *de novo* CNVs (Fig. 2). The highest number of genes (16) were associated with *de novo* AGE CNVs and nine to *de novo* FIBRO CNVs, as listed in Table 1. We found three molecular functions, including "unfolded protein binding", "GTPase regulator activity", "nucleoside-triphosphatase regulator activity" enriched significantly among the genes overlapping FIBRO *de novo* CNVs. The observed common biological functions for AGE *de novo* gene set were "protein complex subunit organization", "macromolecular complex assembly" or "regulation of cellular ketone metabolic process", although they were not significantly enriched (Table 2).

CNVs overlapping QTL regions. The genomic positions of the identified CNVs were compared to that of mapped bovine QTLs. Approximately one third of the CNVs overlapped with QTL regions (Supplementary Figure S6). The average number of QTLs falling into a CNV region varied between 8 and 17, the most observed in *de novo* AGE CNVs (Supplementary Figure S6). Each CNV locus was annotated by the six major QTL classes (Health, Meat and carcass, Milk, Production, Reproduction, Exterior traits) mapped to the position of the CNV.

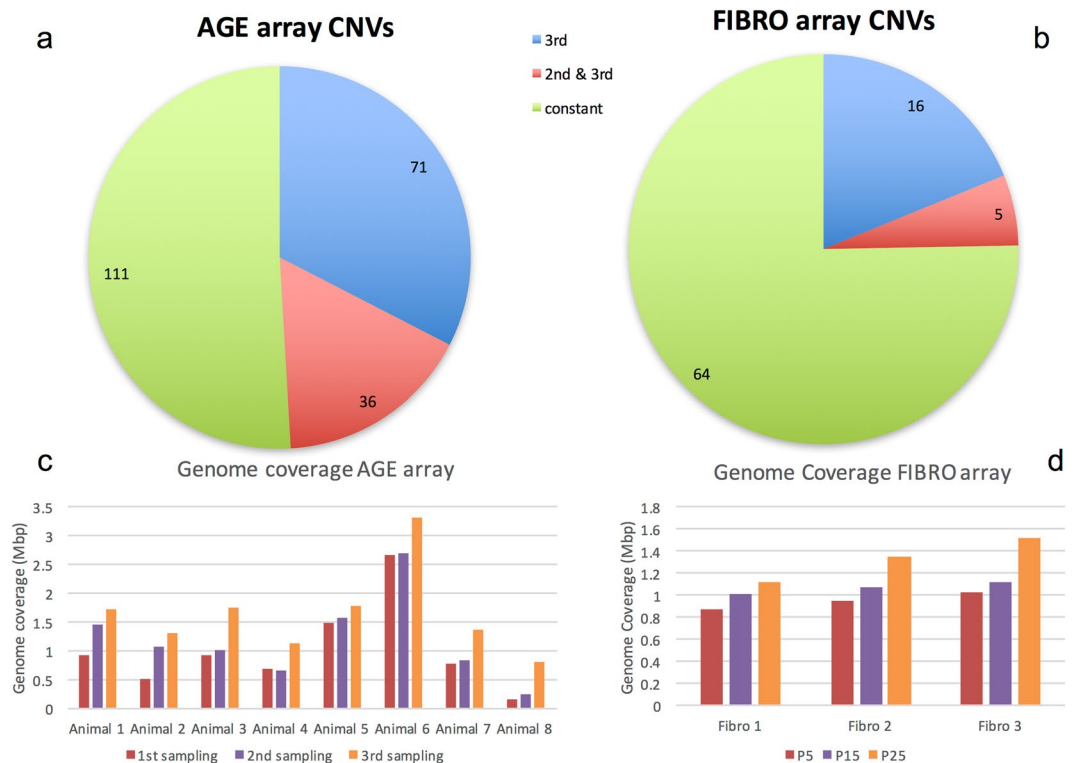


Figure 1. Pie charts presenting the total number of CNVs identified as “constant”, “2nd & 3rd” time point or “3rd” sampling. These latter two classes were scored as “*de novo*”. The ratio of *de novo* vs. constant CNVs was significantly different between AGE (a) and FIBRO (b) ($p < 0.0001$). Genome coverage of AGE array samples (c) and FIBRO array samples (d) at the three time points. Cumulative increase of the portion of genome involved in CNVs are visible as *de novo* CNVs added to the constant CNVs over time.

The most frequent QTL classes were Meat and carcass, Reproduction and Health traits, e.g. out of 34 *de novo* AGE CNVs that encompassed any class of QTLs, 18 overlapped with Meat QTLs. Significant enrichment of Health ($p = 0.036$, binomial test) and Reproduction QTLs ($p = 0.009$, binomial test) were observed in *de novo* vs. constant *in vivo* CNVs, while no significant enrichment of Production trait QTLs was found (Fig. 3). No statistical enrichment was found in any QTL classes between *de novo* and constant *in vitro* CNVs (Fig. 4).

Discussion

In this study, we have identified *de novo* CNVs generated in bulls as they age. The samples were collected in a prospective design allowing us to differentiate *de novo* generated CNVs from constant CNVs that were present throughout the sampling period. This is the first time that individual bulls were sequentially sampled - three samples collected over 30 months - in order to detect *de novo* CNVs that were further characterized for genomic signatures of ageing. Quite remarkably, the total number of *in vivo* CNVs doubled over the 30-month period, as we observed an almost equal number of *de novo* and constant CNVs (107 and 111, respectively, i.e. 49% and 51%, Fig. 1a). Interestingly, the median size of constant and *de novo* CNVs were not different from each other (32 kb and 36 kb). Thus, the total genome coverage (calculated by multiplying the median CNV length by the number of CNVs) was also not different between constant and *de novo* CNVs (3.5 Mbp vs. 3.9 Mbp, respectively). Investigating the 2nd and 3rd sampling time points, it was noted that twice as many *de novo* CNVs emerged during the second half of the sampling schedule as in the first part (a total of 71 and 36, respectively, i.e. on average 9 and 5 per animal, Fig. 1a). It suggests a dynamic generation of *de novo* CNVs in the bovine genome that becomes more frequent as the age of the animal progresses. This results in the temporal cumulative increase in genome coverage in all samples (Fig. 1c), as the *de novo* CNVs were generated and added to the constant CNV fraction. This is in agreement with the observation of a burden of CNVs in human population studies that compare young vs elder subjects^{9,10}, however our study design was able to reveal the activity of *de novo* CNV formation, as an explanation for this observation. Hsieh *et al.*¹¹ have also found accumulation of structural chromosomal changes in human colon crypts.

Consistently, the *in vitro* FIBRO array also showed an increasing number of *de novo* CNVs over the tested consecutive passages of bovine fibroblast cell lines (a total of five at the 2nd sampling vs. 16 at the 3rd, Fig. 1b). However, the proportion of *de novo* CNVs detected in fibroblasts was only ~25% of the total number of CNVs (21 *de novo* vs. 64 constant CNVs, Fig. 1b). This is a highly significant difference, as compared to the 49% *de novo* CNVs *in vivo* ($p < 0.0001$, binomial test). This could be due to the different characteristics of the two experimental models, that are *in vivo* blood-forming cells over 30 months vs. *in vitro* fibroblasts cultures over 3–5 months of passages. CNVs are generated by various mechanisms (recombination, DNA repair) during the cell cycle⁶ and the

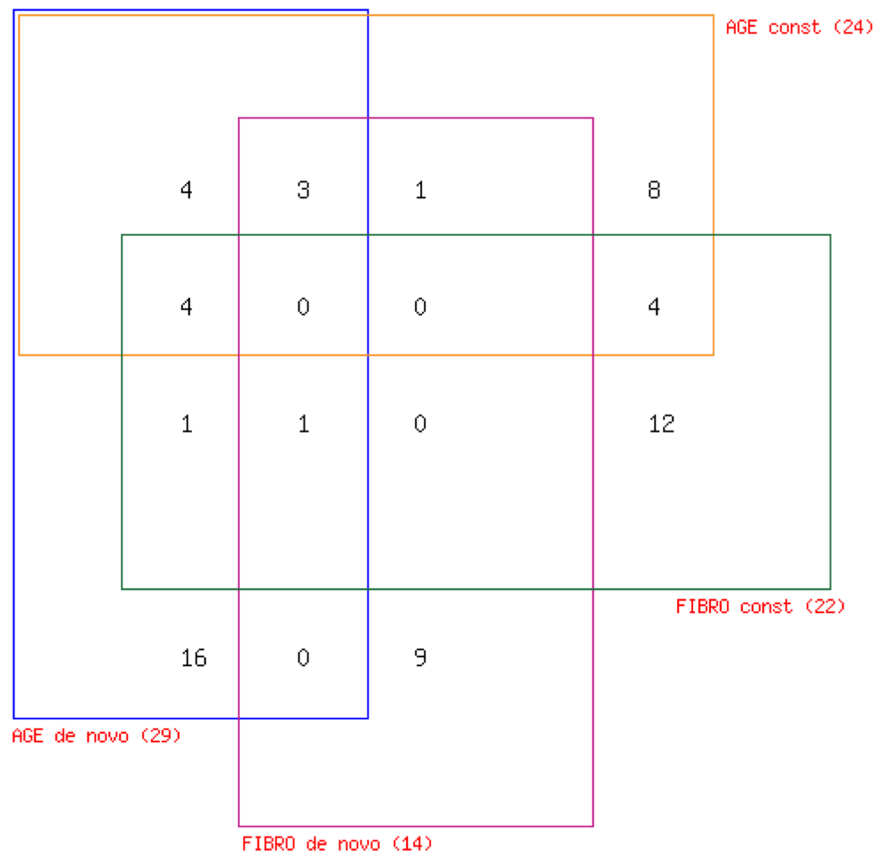


Figure 2. Venn diagram representing genes shared by or unique to AGE and FIBRO CNVs.

AGE <i>de novo</i>	FIBRO <i>de novo</i>
SLC25A24	CELF4
BOLA-DRB3	DAXX
ACN9	ELMOD1
ARHGAP26	MIR2376
B3GALNT2	PFDN6
C2CD5	RGL2
CERS6	SLC25A21
DOCK10	TAPBP
DZIP3	WDR46
EIF2AK3	
FANCC	
KCND3	
KIAA1524	
MATN2	
SLAMF8	
SLC39A11	

Table 1. List of genes overlapping genomic positions with *de novo* AGE or FIBRO array CNVs.

number of cell divisions that occurred during the two experiments could have been different. We can estimate ~50 divisions *in vitro* between P5–P25 with an average of 2.5 population doublings per passage, compared to a minimum of 120 divisions *in vivo* considering less than a week renewal time for lymphocytes^{15,16}. The difference of the rate of *de novo* CNV generation remains even after being corrected by the estimated number of divisions ($107/120 = 0.9$ *in vivo* versus $21/50 = 0.4$ *in vitro*). However, additional factors, such as the impact of the individual genetics, the unknown effects of culture conditions or *in vivo* stress, etc. cannot be excluded. The median size of constant and *de novo* CNVs were not significantly different from each other (33 kb and 30 kb, respectively), however the smaller number of FIBRO samples added up to lower total genome coverage than the AGE samples

Array	Molecular function	GO ID	Statistics	adjusted P-value
FIBRO	unfolded protein binding	0051082	C = 127; O = 2; E = 0.06; R = 34.38; rawP = 0.0014	adjP = 0.0112
	GTPase regulator activity	0030695	C = 446; O = 2; E = 0.20; R = 9.79; rawP = 0.0162	adjP = 0.0453
	nucleoside-triphosphatase regulator activity	0060589	C = 458; O = 2; E = 0.21; R = 9.53; rawP = 0.0170	adjP = 0.0453
AGE	regulation of cellular ketone metabolic process	0010565	C = 168; O = 2; E = 0.11; R = 17.42; rawP = 0.0056	adjP = 0.2743
	macromolecular complex assembly	0065003	C = 995; O = 3; E = 0.68; R = 4.41; rawP = 0.0262	adjP = 0.3065
	protein complex subunit organization	0071822	C = 894; O = 3; E = 0.61; R = 4.91; rawP = 0.0197	adjP = 0.2743

Table 2. Functional enrichment analysis of genes overlapping *de novo* CNVs. Where C = number of reference genes in the category, O = observed number of genes in the gene set from the category, E = expected number in the category, R = Ratio of enrichment, rawP = p value from hypergeometric test, adjusted p-value = p value adjusted by the multiple test adjustment.

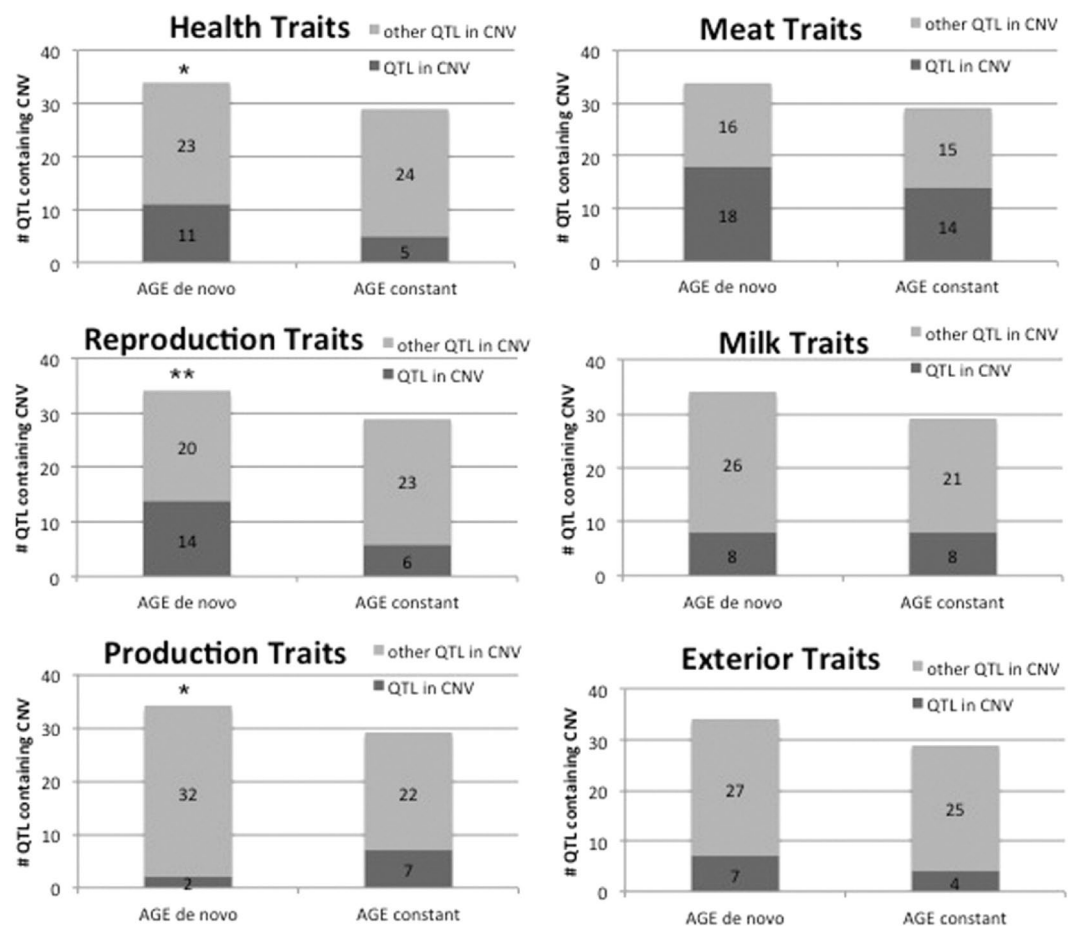


Figure 3. Number of major QTL classes assigned to the *in vivo* AGE CNVs. * $p < 0.05$, ** $p < 0.01$.

(2.2 Mbp constant and 630 kbp *de novo*). Nonetheless, we consider these two experiments as complementary to each other providing information on both *in vivo* and *in vitro* dynamics of *de novo* CNV formation in the bovine genome.

CNVs in the cattle genome have been detected in a number of studies designed to explore population diversity or to identify associations with production traits¹³. Although none of these studies have differentiated between *de novo* and constant CNVs, the general properties of CNVs analysed using the SNP array technology are comparable to the ones identified here. The median length of AGE and FIBRO CNVs was very close to each other (~32 kb), that is on the shorter end of the 21 kb–5.6 Mb size range published for cattle^{14,17,18}. This variability of CNV calls was mostly caused by the difference of the algorithms used to segment the raw data, as compared by Xu *et al.*¹⁴,

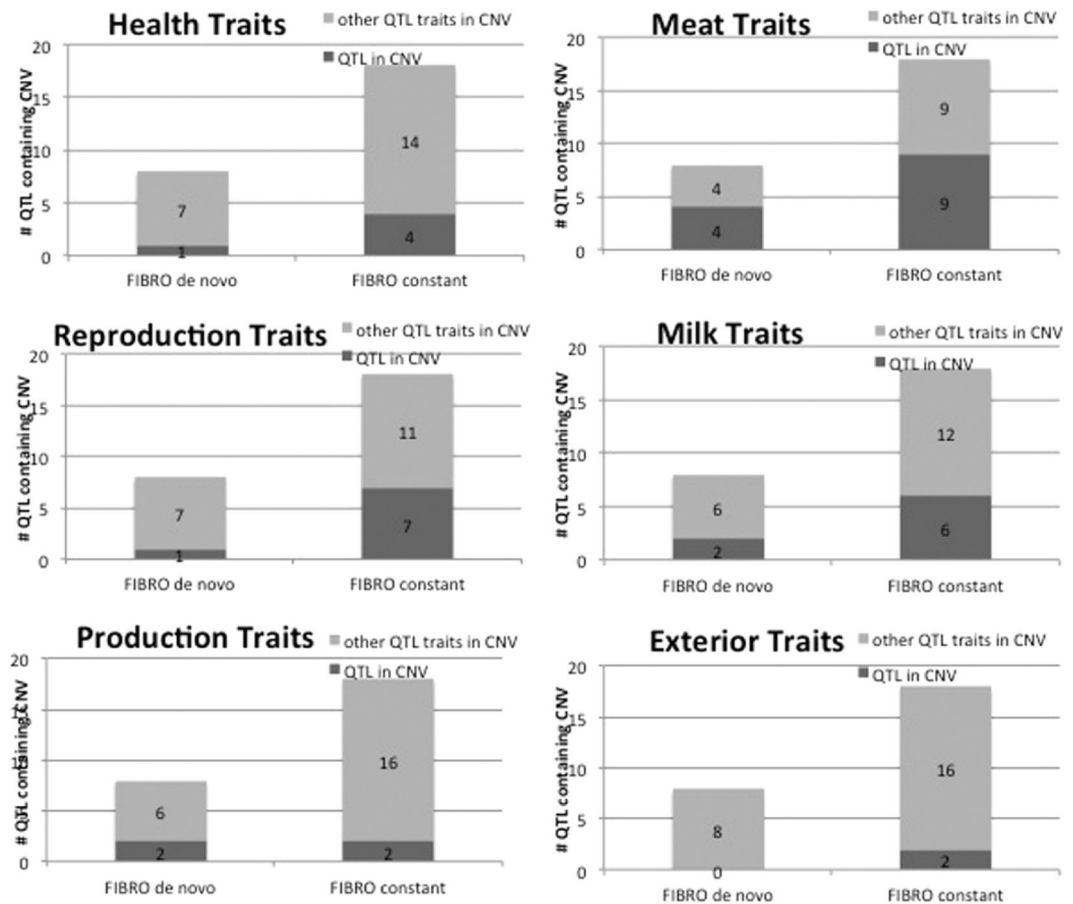


Figure 4. Number of major QTL classes assigned to the *in vitro* FIBRO CNVs.

but the local variability of probe density could also play a role. We have chosen the SVS software based on our previous experience⁴ and the rich available options for quality control and visualization.

Functional characterization of *de novo* CNVs was attempted by locating genes overlapping *de novo* CNV loci. We found that nine genes map to the region covered by the 21 *in vitro* CNVs, while only 16 genes are encompassed by the 107 *in vivo* CNVs. This could be interpreted as 3x higher chance for a *de novo* CNV in fibroblast to be generated in coding region - thus potentially affect gene function - than it is *in vivo* in blood cells, although the smaller *in vitro* sample size could have affected this comparison.

The functional characteristic of genes overlapping *de novo* CNVs (both AGE and FIBRO) were further investigated using the various tools embedded in the Webgestalt software package as well as the comprehensive collection of ageing related changes of the GenAge database. Gene ontology analysis revealed functions connected to proteostasis, that is one of the hallmarks of not only fibroblast ageing, but also the general ageing related alterations of cellular functions^{19,20}. Furthermore, ARHGAP26 - a gene overlapping an AGE CNV -, connects to “GTPase regulator activity” and the GenAge member ARHGAP1. This latter gene is implicated in the regulation of cytoskeleton, affecting cellular adhesion and motility that are linked to premature ageing²¹ and age-related diseases²². Several SLC genes were found in *de novo* CNVs (SLC25A21, SLC25A24, SLC39A11) that are involved in body-fat composition and distribution in human and mice and found to be associated with ageing in a large Korean cohort study²³. Interestingly another member of this gene family (SLC6A14) was also affected by a CNV in humans²⁴. Another gene overlapping with an AGE CNV is ceramide synthase 6 (CERS6), which is related to mitochondrial dysfunction by acting as a pro-apoptotic factor²⁵. Another pro-apoptotic protein DAXX which overlapped with a FIBRO CNV, has been identified in *Drosophila* to affect longevity²⁶, and also acts as a target of the REST transcription factor, that protects against Alzheimer’s and normal ageing²⁷.

Another advantage of our bovine model for characterization of *de novo* CNVs is the availability of a detailed map of genomic regions associated with variation of quantitative traits (QTL map). The analysis of overlapping QTL classes provided an additional layer of functional characterization to the CNVs. Surprisingly, health and reproduction traits are over-represented in *de novo* AGE CNVs. In contrast, milk traits - that give ¼ of all 42000 cattle QTLs and production traits QTLs, such as feed intake or daily gain, are under-represented in *de novo* CNVs. On the other hand, previous research into bovine CNVs found associations with milk production or residual feed intake in diverse selected populations of cattle^{5,18}.

The current study identified *de novo* CNVs in ageing bulls and bovine fibroblast cultures. The highly dynamic generation of these CNV events as well as their potential links to ageing-related genomic functions were observed

and thus warrant further studies of larger populations and the application of higher-resolution arrays or genome sequencing, which will improve the sensitivity of CNV detection.

Methods

Animal samples for AGE array. Peripheral blood samples were collected from eight Holstein bulls housed and managed under the same conditions at the same Canadian farm. Sampling was done according to the Canadian Council on Animal Care (CCAC, http://www.ccac.ca/en/_standards/policies) and the University of Guelph's Animal Care Committee (ACC, <http://www.uoguelph.ca/research/services-divisions/animal-care-services>) guidelines by licensed farm technicians or CFIA (Canadian Food Inspection Agency) veterinarians. The animals were not selected for research purposes, blood samples left over from regular health check events were used for DNA extraction. The animals were available for three consecutive sampling over the course of almost three years. The bulls were approximately one year old at the start of the sampling. The collection time points (14 ± 3 month, 18 months later: 32 ± 3 m, 12 months later: 44 ± 3 m) are shown in Figure S6.

Genomic DNA was extracted from the blood samples using a traditional Proteinase K digestion, salting out and phenol/chloroform method²⁸ and concentrations were determined by a NanoDrop photometer. The three samples from the eight animals (total of 24) were submitted to a genomic service provider (DNA LandMarks, QC, Canada) where they were genotyped on the Illumina Bovine SNP50k array. The 24 samples were handled on the same day, processed on one array slide according to the laboratory SOP. This dataset was referred throughout the manuscript as the AGE or *in vivo* dataset.

Fibroblast cultures. Primary fibroblast cell lines were established by using standard culture techniques²⁹ from ear notch samples of three unrelated Holstein bulls slaughtered at a local abattoir. These animals were not related to the ones genotyped on the AGE array. Briefly, the samples were digested in 0.5% collagenase type 1 for 6 h at 38 °C then cultured in DMEM with 10% FBS, 1% PenStrep (Invitrogen Canada Inc. Burlington, ON, Canada) and 2 mM L-glutamine (Sigma-Aldrich Canada Ltd, Oakville, ON, Canada). The primary fibroblasts were sequentially cultured (seeded at 3×10^4 cells/25 cm² flask and harvested at $\sim 1.5 \times 10^5$ cells/flask, ~ 2.5 population doubling/passage) and sampled at passages 5, 15 and 25 (P5, P15, P25) for DNA extraction.

DNA samples were genotyped on the Illumina Bovine SNP50k array at the same facility (DNA LandMarks), as the AGE array according to the same protocol. The nine samples (3×3) were processed on one array slide on the same day, although it was a different day than for the AGE array. This dataset was named FIBRO or *in vitro* dataset.

CNV analysis. The raw data, that is the log R ratio, represents a normalized signal intensity value for each probe on the array, thus it can be related to the amount of DNA hybridized to over 54,000 loci along the bovine genome³⁰. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus³¹, and are accessible through GEO Series accession number GSE95612 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE95612>).

The log R ratio values were calculated using the Illumina GenomeStudio software, before transferring the data to the SNP and Variation Suite (SVS version 8.4.3., GoldenHelix Inc.) for CNV analysis. First, several quality control steps were taken to identify low quality samples. Noise in the log R ratio values generally comes from the sample preparation or genotyping procedures and could be identified by testing for outliers in the median derivative log ratio values. Genomic waves that represent the variation of log R values according to the DNA composition targeted by the probes were detected by Diskin *et al.*³². The X and Y chromosomal data were excluded from the data. Potential batch effects were tested by principal component analysis (PCA) of logR values. SNP annotations were according to the representative genome Bos_taurus_UMD_3.1.1 (2014).

CNVs were identified in SVS by the Univariate CNAM algorithm that scans log R values of each sample separately and identifies DNA segments that are significantly different from their neighbors. The segmentation conditions were set to 0.005 maximum pairwise segment p value, ≥ 2 markers/segment with outlier removal function and the segment means were filtered to be < -0.35 for losses or > 0.35 for gains. The resulting list of CNVs (named as CNV-A for AGE and CNV-F for FIBRO array) was classified according to their presence in consecutive samples from the same animal or cell culture. CNVs present in all three-samples corresponding to the three time points of sampling (blood samples collected at different ages of the same bull or passages P5, P15 and P25 of the same cell line) were named as "constant". Alternatively, CNVs detected in the 2nd and 3rd or only at the 3rd sampling time were classified as "*de novo*". The ratio of *de novo* vs constant CNVs were compared between AGE and FIBRO with the binomial test, while the CNV length were compared using the Kruskal-Wallis test in Prism 6 (GraphPad). CNVs that could not be scored either as "constant" nor "*de novo*" were inconsistent to our model and not considered further. These types of CNVs (listed in Table S6) were the ones detected only at the 1st or the 2nd time point or the 1st and 2nd time points and probably were caused by changes of the abundance of cells carrying the variant or its mosaic status reaching below detection limit. No CNVs were identified at 1st and 3rd time points, showing consistency of detection.

qPCR tests were designed to validate 28 different 3rd time point *de novo* loci including all samples. The tests were run in triplicate across all three time points and a control sample. This control sample should function as a reference, stably representing diploid copy number for all of the tested CNV loci and was created as a mixture of all 1st time point samples. qPCRs were run in a CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad) using a thermal profile of 98 °C, 2 min; $45 \times (98 \text{ °C}, 10 \text{ sec}; 59 \text{ °C}, 10 \text{ sec})$, followed by the registration of a melting curve between 68 °C to 95 °C in 0.5 °C/sec increments. The 10 μ l reaction was composed of 1 \times SsoFast EvaGreen Supermix (Bio-Rad), 3 mM primers and 20 ng genomic DNA. Relative quantity was calculated in the Bio-Rad CFX Manager software against the control sample and two normalizer genes SRY and BTF3, both widely used for qPCR validation of bovine CNVs. Then the ratio of 2nd and 3rd time point data to the 1st was calculated to

reveal copy number decrease or increase. The PCR efficiency was calculated using the actual amplification curves using the LinReg PCR software³³. Primers were designed using the NCBI Primer-BLAST tool, except for BTF3³⁴ (Table S5).

Functional annotation of CNVs. Genomic locations of all bovine QTLs were downloaded from the Animal Genome Database (release 28³⁵). Enrichment of a particular trait category in *de novo* vs. constant CNVs was detected by the binomial test.

The RefSeq genes were downloaded by using the UCSC Table browser³⁶. The resulting list of annotated genes for each array classes (AGE, FIBRO or *de novo*, constant) were compared by generating a four-way Venn diagram using the web-tool developed by Chris Seidel and available from <http://www.pangloss.com/seidel/Protocols/venn4.cgi>.

Both *de novo* gene lists (AGE and FIBRO) were further analyzed against the GenAge database (<http://genomics.senescence.info/genes/index.html>³⁷) and for functional enrichment in Gene Ontology (GO) terms, using the WEB-based Gene Set Analysis Toolkit (WebGestalt)³⁸. The bovine gene names were converted to the corresponding human ones and the human genome was used as reference set for the default statistical test (Benjamini-Hochberg, adjusted p-value < 0.01).

References

- Zarrei, M., MacDonald, J. R., Merico, D. & Scherer, S. W. A copy number variation map of the human genome. *Nat Rev Genet* **16**, 172–183, doi:10.1038/nrg3871 (2015).
- Stankiewicz, P. & Lupski, J. R. Structural variation in the human genome and its role in disease. *Annu Rev Med* **61**, 437–455, doi:10.1146/annurev-med-100708-204735 (2010).
- Clop, A., Vidal, O. & Amills, M. Copy number variation in the genomes of domestic animals. *Anim Genet* **43**, 503–17, doi:10.1111/j.1365-2052.2012.02317.x (2012).
- Revay, T., Quach, A. T., Maignel, L., Sullivan, B. & King, W. A. Copy number variations in high and low fertility breeding boars. *BMC Genomics* **16**, 280, doi:10.1186/s12864-015-1473-9 (2015).
- Xu, L. *et al.* Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. *BMC Genomics* **15**, 683, doi:10.1186/1471-2164-15-683 (2014).
- Hastings, P. J., Lupski, J. R., Rosenberg, S. M. & Ira, G. Mechanisms of change in gene copy number. *Nat Rev Genet* **10**, 551–564, doi:10.1038/nrg2593 (2009).
- Piotrowski, A. *et al.* Somatic mosaicism for copy number variation in differentiated human tissues. *Hum Mutat* **29**, 1118–1124, doi:10.1002/humu.v29:9 (2008).
- Oluwole, O. A., Revay, T., Mahboubi, K., Favetta, L. A. & King, W. A. Somatic mosaicism in bulls estimated from genome-wide CNV array and TSPY gene copy numbers. *Cytogenet Genome Res* (2016).
- Kuningas, M. *et al.* Large common deletions associate with mortality at old age. *Hum Mol Genet* **20**, 4290–4296, doi:10.1093/hmg/ddr340 (2011).
- Nygaard, M. *et al.* Copy number variation associates with mortality in long-lived individuals: a genome-wide assessment. *Aging Cell* **15**, 49–55, doi:10.1111/accel.2016.15.issue-1 (2016).
- Hsieh, J. C. F., Van Den Berg, D., Kang, H., Hsieh, C. L. & Lieber, M. R. Large chromosome deletions, duplications, and gene conversion events accumulate with age in normal human colon crypts. *Aging Cell* **12**, 269–279, doi:10.1111/accel.2013.12.issue-2 (2013).
- Forsberg, L. A. *et al.* Age-Related Somatic Structural Changes in the Nuclear Genome of Human Blood Cells. *Am J Hum Genet* **90**, 217–228, doi:10.1016/j.ajhg.2011.12.009 (2012).
- Bickhart, D. M. & Liu, G. E. The challenges and importance of structural variation detection in livestock. *Front Genet* **5**, 37, doi:10.3389/fgene.2014.00037 (2014).
- Xu, L., Hou, Y., Bickhart, D., Song, J. & Liu, G. Comparative Analysis of CNV Calling Algorithms: Literature Survey and a Case Study Using Bovine High-Density SNP Data. *Microarrays* **2**, 171–185, doi:10.3390/microarrays2030171 (2013).
- Rocha, B., Freitas, A. A. & Coutinho, A. Population dynamics of T lymphocytes. *Renewal rate and expansion in the peripheral lymphoid organs. J Immunol* **131**, 2158–2164 (1983).
- Fulcher, D. & Basten, A. B cell life span: A review. *Immunol Cell Biol* **75**, 446–455, doi:10.1038/icb.1997.69 (1997).
- Ciconardi, F. *et al.* Massive screening of copy number population-scale variation in Bos taurus genome. *BMC Genomics* **14**, 124, doi:10.1186/1471-2164-14-124 (2013).
- Hou, Y. *et al.* Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes in Angus cattle. *Funct Integr Genomics* **12**, 81–92, doi:10.1007/s10142-011-0252-1 (2012).
- Tigges, J. *et al.* The hallmarks of fibroblast ageing. *Mech Ageing Dev* **138**, 26–44, doi:10.1016/j.mad.2014.03.004 (2014).
- Waldera-Lupa, D. M. *et al.* Proteome-wide analysis reveals an age-associated cellular phenotype of in situ aged human fibroblasts. *Aging (Albany NY)* **6**, 856–78, doi:10.18632/aging.v6i10 (2014).
- Wang, L., Yang, L., Debidia, M., Witte, D. & Zheng, Y. Cdc42 GTPase-activating protein deficiency promotes genomic instability and premature aging-like phenotypes. *Proc Natl Acad Sci USA* **104**, 1248–1253, doi:10.1073/pnas.0609149104 (2007).
- Wolfson, M., Budovsky, A., Tacutu, R. & Fraifeld, V. The signaling hubs at the crossroad of longevity and age-related disease networks. *Int J Biochem Cell Biol* **41**, 516–520, doi:10.1016/j.biocel.2008.08.026 (2009).
- Urano, T., Shiraki, M., Sasaki, N., Ouchi, Y. & Inoue, S. SLC25A24 as a novel susceptibility gene for low fat mass in humans and mice. *J Clin Endocrinol Metab* **100**, E655–E663, doi:10.1210/jc.2014-2829, jc20142829 (2015).
- Lee, B.-Y., Shin, D. H., Cho, S., Seo, K.-S. & Kim, H. Genome-wide analysis of copy number variations reveals that aging processes influence body fat distribution in Korea Associated Resource (KARE) cohorts. *Hum Genet* **131**, 1795–804, doi:10.1007/s00439-012-1203-1 (2012).
- Schüll, S. *et al.* Cytochrome c oxidase deficiency accelerates mitochondrial apoptosis by activating ceramide synthase 6. *Cell Death Dis* **6**, e1691, doi:10.1038/cddis.2015.62 (2015).
- Bodai, L. *et al.* Daxx-like Protein of Drosophila Interacts with Dmp53 and Affects Longevity and Arkl mRNA Level. *J Biol Chem* **282**, 36386–36393, doi:10.1074/jbc.M705547200 (2007).
- Lu, T. *et al.* REST and stress resistance in ageing and Alzheimer's disease. *Nature* **507**, 448–54, doi:10.1038/nature13163 (2014).
- Miller, S. A., Dykes, D. D. & Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**, 1215–1215, doi:10.1093/nar/16.3.1215 (1988).
- Mastromonaco, G. F., Favetta, L., Smith, L., Filion, F. & King, W. A. The Influence of Nuclear Content on Developmental Competence of Gaur x Cattle Hybrid In Vitro Fertilized and Somatic Cell Nuclear Transfer Embryos. *Biol Reprod* **76**, 514–523, doi:10.1095/biolreprod.106.058040 (2007).
- Peiffer, D. A. *et al.* High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res* **16**, 1136–48, doi:10.1101/gr.5402306 (2006).

31. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207–210, doi:[10.1093/nar/30.1.207](https://doi.org/10.1093/nar/30.1.207) (2002).
32. Diskin, S. J. *et al.* Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res* **36**, e126–e126, doi:[10.1093/nar/gkn556](https://doi.org/10.1093/nar/gkn556) (2008).
33. Ramakers, C., Ruijter, J. M., Deprez, R. H. L. & Moorman, A. F. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters* **339**, 62–66, doi:[10.1016/S0304-3940\(02\)01423-4](https://doi.org/10.1016/S0304-3940(02)01423-4) (2003).
34. Bae, J. S. *et al.* Identification of copy number variations and common deletion polymorphisms in cattle. *BMC Genomics* **11**, 232, doi:[10.1186/1471-2164-11-232](https://doi.org/10.1186/1471-2164-11-232) (2010).
35. Hu, Z.-L., Park, Ca., Wu, X.-L. & Reecy, J. M. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res* **41**, D871–9, doi:[10.1093/nar/gks1150](https://doi.org/10.1093/nar/gks1150) (2013).
36. Karolchik, D. *et al.* The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res* **42** (2014).
37. Tacutu, R. *et al.* Human Ageing Genomic Resources: Integrated databases and tools for the biology and genetics of ageing. *Nucleic Acids Res* **41** (2013).
38. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res* **41**, W77–83, doi:[10.1093/nar/gkt439](https://doi.org/10.1093/nar/gkt439) (2013).

Acknowledgements

Funding was provided by the Natural Sciences and Engineering Research Council of Canada, the Canada Research Chairs program and the Ontario Veterinary College, University of Guelph.

Author Contributions

O.O. collected samples, cultured fibroblast, extracted DNA and revised the manuscript; T.R. performed array analysis, validation and wrote the manuscript; T.K. managed the sample collection; W.A.K. conceived the study and revised the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-01793-2](https://doi.org/10.1038/s41598-017-01793-2)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017