

RESEARCH ARTICLE

Open Access

Identification of genome-wide copy number variations among diverse pig breeds by array CGH

Yan Li^{1†}, Shuqi Mei^{2†}, Xuying Zhang¹, Xianwen Peng², Gang Liu¹, Hu Tao¹, Huayu Wu², Siwen Jiang¹, Yuanzhu Xiong¹ and Fenge Li^{1*}

Abstract

Background: Recent studies have shown that copy number variation (CNV) in mammalian genomes contributes to phenotypic diversity, including health and disease status. In domestic pigs, CNV has been catalogued by several reports, but the extent of CNV and the phenotypic effects are far from clear. The goal of this study was to identify CNV regions (CNVRs) in pigs based on array comparative genome hybridization (aCGH).

Results: Here a custom-made tiling oligo-nucleotide array was used with a median probe spacing of 2506 bp for screening 12 pigs including 3 Chinese native pigs (one Chinese Erhualian, one Tongcheng and one Yangxin pig), 5 European pigs (one Large White, one Pietrain, one White Duroc and two Landrace pigs), 2 synthetic pigs (Chinese new line DIV pigs) and 2 crossbred pigs (Landrace × DIV pigs) with a Duroc pig as the reference. Two hundred and fifty-nine CNVRs across chromosomes 1–18 and X were identified, with an average size of 65.07 kb and a median size of 98.74 kb, covering 16.85 Mb or 0.74% of the whole genome. Concerning copy number status, 93 (35.91%) CNVRs were called as gains, 140 (54.05%) were called as losses and the remaining 26 (10.04%) were called as both gains and losses. Of all detected CNVRs, 171 (66.02%) and 34 (13.13%) CNVRs directly overlapped with *Sus scrofa* duplicated sequences and pig QTLs, respectively. The CNVRs encompassed 372 full length Ensembl transcripts. Two CNVRs identified by aCGH were validated using real-time quantitative PCR (qPCR).

Conclusions: Using 720 K array CGH (aCGH) we described a map of porcine CNVs which facilitated the identification of structural variations for important phenotypes and the assessment of the genetic diversity of pigs.

Background

Genetic and archaeological findings suggest that pig domestication began about 9000–10000 years before present (YBP) at multiple sites across Eurasia, followed by their subsequent spread at a worldwide scale [1]. Historically, Europe and China are two major areas of pig breeding [2]. Over the past centuries, pigs have shown marked differences between these two areas, even if many European pig breeds carry far Eastern haplotypes at high frequencies because of an ancient introgression with Chinese swine [1]. The Chinese pigs differ signifi-

cantly from European pig breeds such as the Large White for many traits including fatness and ear traits [3-5]. Genetic variation within the gene pool which produce the above different phenotypes are selected for or against by evolution. Microsatellites, single nucleotide polymorphisms (SNPs) were the main measures of genetic variations in pigs, producing a USMARC pig SNP map (http://www.marc.usda.gov/genome/swine/marker_list.html) and the PorcineSNP60 Genotyping BeadChip with 62163 SNP probes [6]. Recently, structural variations including insertions, duplications, deletions, inversions and translocations of DNA have been shown to contribute to the major phenotypic variations [7]. Copy number variation (CNV) is described as a segment of DNA >1 kb that is copy number variable when compared with a reference genome [8]. This variation may either be inherited or caused by de novo mutation [9-12]. It has become apparent that CNVs

* Correspondence: lifener@mail.hzau.edu.cn

[†]Equal contributors

¹Key Laboratory of Pig Genetics and Breeding of Ministry of Agriculture & Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, Huazhong Agricultural University, Wuhan 430070, PR China

Full list of author information is available at the end of the article

are genome-wide present in the human genome [8] and the genome of farm animals including cattle [13-16], avian [17-19], sheep [20], goat [21]. About a range from 5% to 16% of the human genome was covered by CNVs [22,23]. CNVs can lead to striking phenotypic consequences as a result of altering gene dosage, disrupting coding sequences, or perturbing long-range gene regulation by position effects [24-26]. These striking phenotypic consequences include some common complex diseases such as autism [11], schizophrenia [12], auto-immune Addison's disease [27].

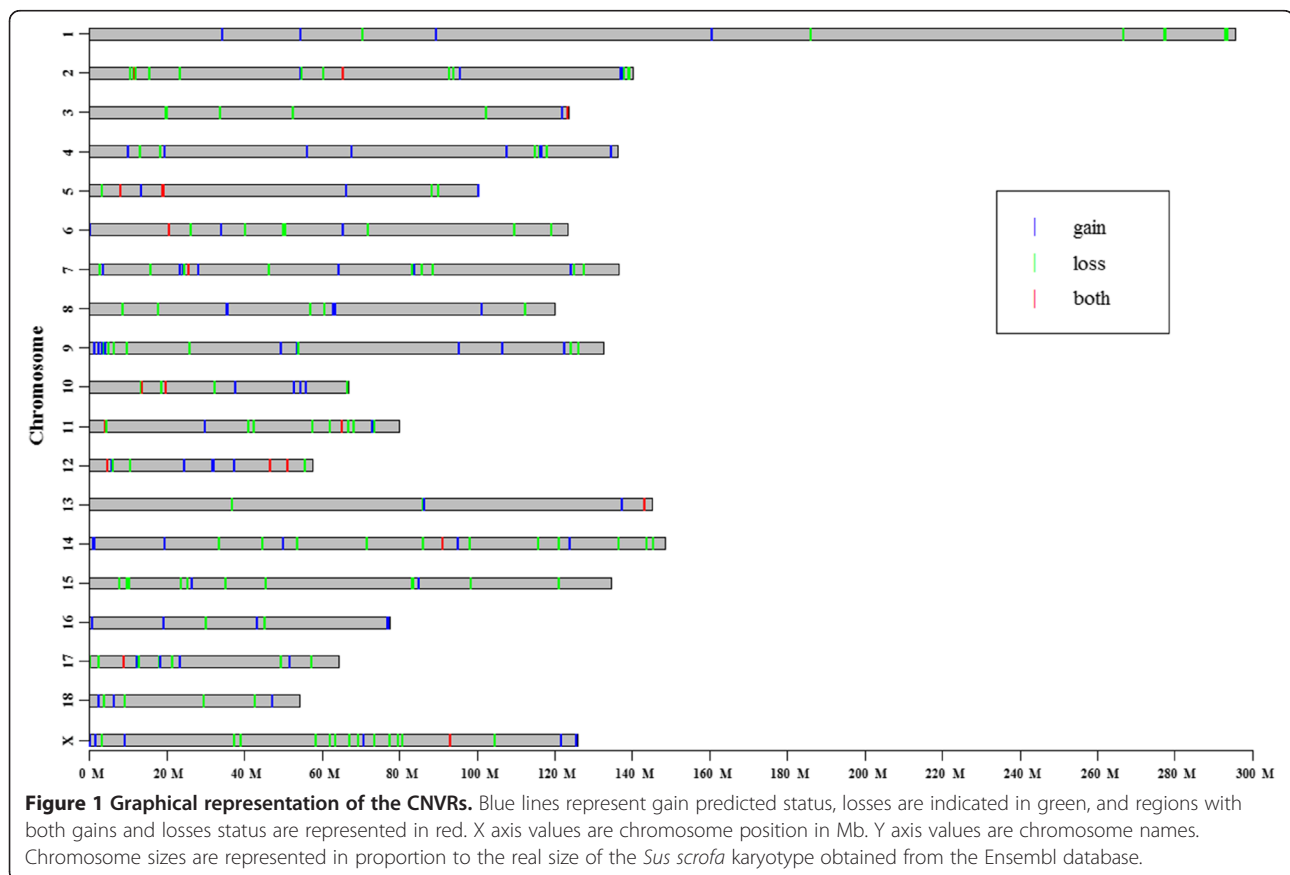
Recently many efforts have been used to detect pig CNVs. By a custom-made tiling oligonucleotide array, 37 CNV regions (CNVRs) across chromosomes 4, 7, 14, and 17 were identified in 12 unrelated Duroc boars [28]. Comparative genome hybridization (CGH) array was also conducted for chromosomes 7 and 8 in 9 different pig populations including Duroc, Large White, Meishan, Pietrain, Hampshire and Wild Boar [29]. By analyzing data from the Porcine SNP60 BeadChip, 49 CNVRs were identified in 55 animals from an Iberian \times Landrace cross (IBMAP) [30] and 382 CNVRs were identified from three purebred populations (Yorkshire, Landrace and Songliao Black) and one Duroc \times Erhualian crossbred population [31]. Up until now, few studies have confirmed the genome-wide presence of CNVs in pigs using array CGH

(aCGH) with high-density probes. Here we reported the use of high-resolution oligonucleotide aCGH to identify the CNV regions in 12 individual pigs from different pig populations. This analysis provided a high-resolution map of copy number variations in the pig genome with a median probe spacing of 2506 bp relative to the latest porcine genome assembly (Sscrofa9.2).

Results and discussion

The overview of CNVR library

Array CGH (NCBI GEO accession no. GPL16165) was carried out using a custom-made array comprising 719,336 oligonucleotide probes covering the whole pig genome assembly with a median probe spacing of 2506 bp (Additional file 1). CNV was assessed by equating the log₂ ratio of signal intensity between the reference (Duroc) and test samples. As we did not perform a self-to-self experiment, a stringent criterion with the mean $|\log_2 \text{ratio}| > 0.5$ was used to reduce the false positive rate of CNV calling according to the studies of Wang et al. [19] and Fadista et al. [28]. Therefore, the segments with at least 5 consecutive probes and a mean $|\log_2 \text{ratio}| > 0.5$ were merged [28,32]. A CNVR was then called if detected in two or more animals. Accordingly, we identified 259 CNVRs (Figure 1, Additional file 2). The CNVRs ranged



in size from 2.30 kb to 1.55 Mb with a mean of 65.07 kb and a median of 98.74 kb, covering 16.85 Mb or 0.74% of the whole genome (Figure 2A, Additional file 2). The largest CNV region, CNVR_85 with 1.55 Mb in size on chromosome 7, showed copy gain in the White Duroc pig, the Pietrain pig, 2 Landrace × DIV pigs and loss in the Yangxin pig and the Large White pig.

Using the custom tiling oligonucleotide aCGH approach, Fadista et al. [28] addressed 37 CNVRs on the *Sus scrofa* chromosomes (SSCs) 4, 7, 14, and 17 of the preliminary assembly of pig genome among 12 Duroc boars. Ramayo-Caldas et al. [30] detected 49 CNVRs using the Porcine SNP60 BeadChip data of 55 animals from an Iberian × Landrace cross. Wang et al. [31] detected 382 CNVRs based on the Porcine SNP60 genotyping data of 474 pigs. Two of the 37 CNVRs (5.41%) detected by Fadista et al. [28], 8 of the 49 CNVRs (16.32%) detected by Ramayo-Caldas et al. [30], 24 of the 382 CNVRs (6.28%) detected by Wang et al. [31] were identical or overlapped with the detected CNVRs in this study (Additional file 2). Totally 39 of the presently detected 259 CNVRs (15.06%) were identical or overlapped with those previously reported pig CNVRs (Additional file 2). The main potential reasons for this less well-overlapping result could be the different genetic backgrounds of pig samples, different platforms and various calling algorithms between the present study and other studies.

Compared with PorcineSNP60 Genotyping BeadChip, the detection power of 720 K aCGH was enhanced by dense marker density, uniform distribution of probes along each chromosome [6,30]. Hence, some small CNVRs can be detected by aCGH technique, as the minimum CNV lengths were 2.30 kb in our present study, and 2.08 kb in the study of Fadista et al. [28], whereas the minimum CNV length detected by SNP chip were 5.03 kb and 44.65 kb, respectively [30,31].

CNVRs chromosome distribution and status

CNVRs were distributed throughout the genome in a non-random manner (Additional file 2), which was coherent with the previous studies on heterogeneous distribution of CNVs in primate genomes [9,14]. Chromosomes 2, 7, 10–12 and 17 had the dense CNVs covering more than 1.00% of genomic sequences (Table 1). A conserved synteny between *Homo sapiens* chromosome 17 (HSA17) and SSC12 had been proposed (<https://www-lgc.toulouse.inra.fr/pig/compare/SSC.htm>). Proportional to its length, HSA17 was especially rich in primate-specific breakpoint regions which would appear to be highly enriched for both segmental duplications (SDs) and CNVs [33,34].

Concerning copy number status, 93 (35.91%) CNVRs were called as gains, 140 (54.05%) were called as losses and the remaining 26 (10.04%) were called as both gains and losses. Previously, it has been suggested that deletions are under stronger purifying selection than duplications [35]. If so, deletions should be both less frequent and shorter than duplications [14]. However, when we compared the length of gains with losses in the CNVRs, loss regions had slightly larger sizes than gain regions with the average length of 57.39 kb and 45.86 kb respectively (T-test not statistically significant at p value > 0.05). The possible reason was that the aCGH approach might favor the identification of deletions [14,15,21,28]. As the samples were collected from 9 different populations, the considerable number of CNVRs status displaying in 'both gains and losses' might be due to the different genetic origins.

Putative population-specific CNVRs and cluster analysis

Some putative population-specific CNVRs were detected. For example, 6 CNVRs including CNVR_132 were purebred Landrace-specific, and CNVR_145 were purebred DIV-specific. CNVR_100 including *KIT* gene contained amplifications specifically in 8 pigs with dominant white

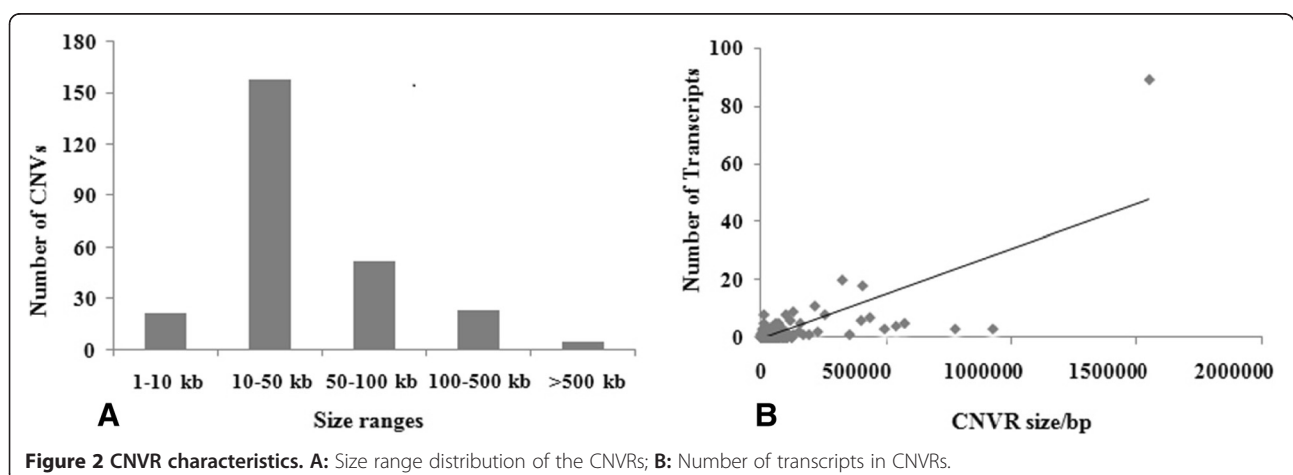


Table 1 Chromosome distribution of CNVRs in pigs

Chr.	No. CNVRs	CNVR Size (bp)	Chr. Size (bp)	Percentage (%)
1	12	1158170	295534705	0.392
2	19	1524080	140138492	1.088
3	9	757763	123604780	0.613
4	13	416439	136259946	0.306
5	13	542338	100521970	0.539
6	11	363611	123310171	0.295
7	19	2572973	136414062	1.886
8	12	1185818	119990671	0.988
9	18	927039	132473591	0.699
10	13	1673511	66741929	2.507
11	12	1220486	79819395	1.529
12	16	885260	57436344	1.541
13	7	198599	145240301	0.137
14	18	665744	148515138	0.448
15	15	407233	134546103	0.303
16	8	325153	77440658	0.419
17	14	869490	64400339	1.350
18	9	360816	54314914	0.664
X	22	798761	125876292	0.634
Total	259	16853284	2262579801	0.745

color and a Pietrain pig with black spots, and CNVR_251 contained gains in pigs without dominant white color such as Yangxin, Erhualian, Tongcheng and Pietrain pigs. However, due to the limited samples used in the present study, the putative population-specific CNVRs need future study. And we also found 3 de novo CNVRs, of which CNVR_IDs 36, 149 were present in 2 Landrace × DIV crossbred pigs but not in their parents, while CNVR_259 were absent in 2 Landrace × DIV crossbred pigs but present in their parents.

Using the cluster tool, average linkage hierarchical clustering based on the CNV profiles of 12 tested pigs was performed. Figure 3 showed the dendrogram of 12 pigs generated by average linkage clustering algorithm of Cluster 3.0 software. Basically, the Chinese native pigs (Erhualian, Yangxin, Tongcheng) clustered together, while the other 9 pigs with European haplotypes belonged to another big cluster. Therefore, CNVs could be used to investigate pig genetic diversity and evolution.

Duplicated sequences colocalize with CNVRs in the pig genome

Although the exact interpretation of mechanisms responsible for generating CNVs is still unclear, previous studies have noted a four- to twenty-fold enrichment of CNVs near SDs in the other mammalian genomes

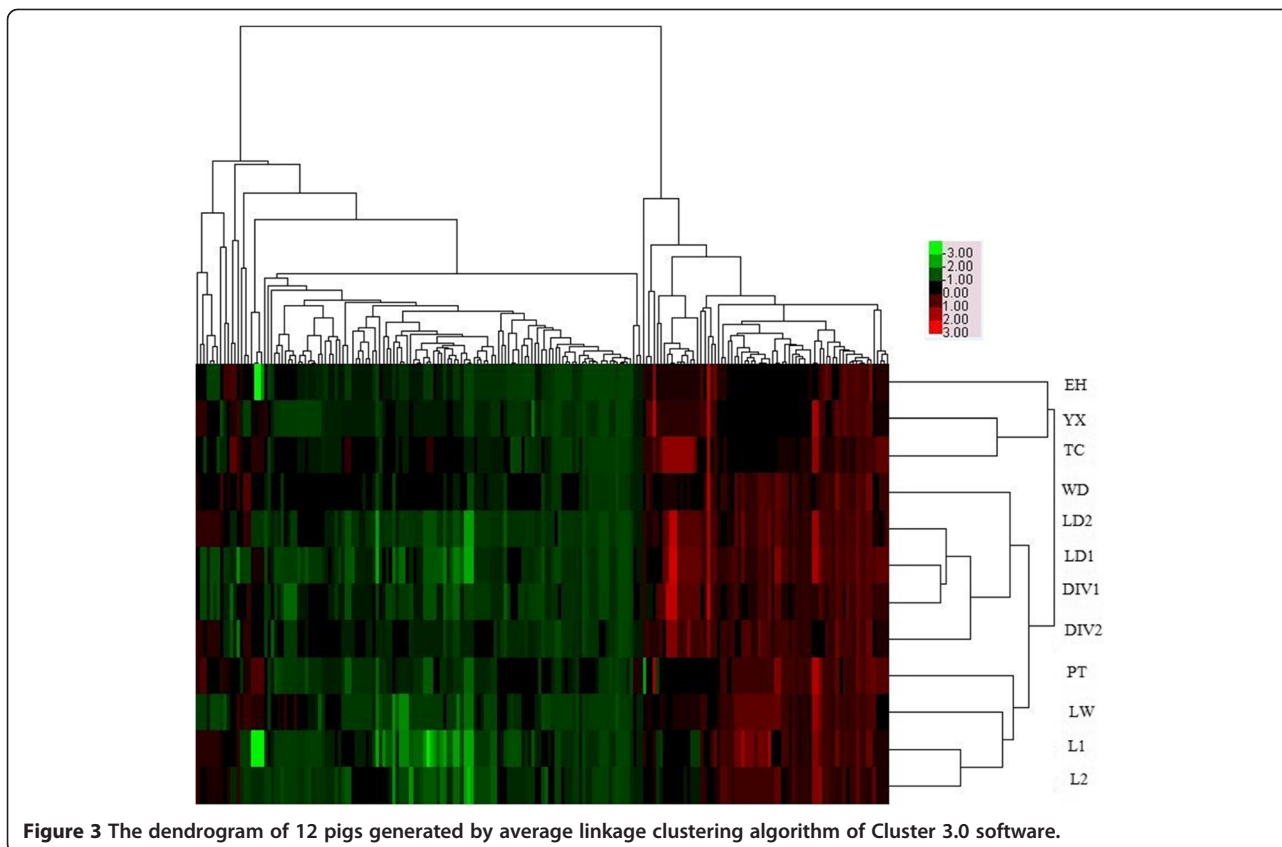


Figure 3 The dendrogram of 12 pigs generated by average linkage clustering algorithm of Cluster 3.0 software.

[22,32,36]. Duplicated sequences are typical segments of DNA which range in size from one to hundreds of kb, share a high level of sequence identity ($\geq 90\%$) and occur at more than one site within the genome [28]. Under the same filter criterion, about 66.02% (171/259) of CNV regions directly overlapping with *Sus scrofa* duplicated sequences were identified through blasting the CNVR sequence against the Ensembl pig genomic sequences. As our present BLAST results did not retain a CNVR overlapping with a duplicated sequence by less than 1000 bp, so the overlaps of CNVs and their targeted duplicated sequences were under reporting. There were 13.5–25.0% CNVRs mapped to duplicated sequences in the previous reports [28,37]. The difference may be related to differences in samples. CNVRs overlapping duplicated sequences were significantly different in average size (87.12 kb versus 22.23 kb, t-test $p < 0.01$) with the CNVRs that did not overlap duplicated sequences, consistent with previous CNV studies reporting a stronger association between duplicated sequences and long CNVRs [9,11].

Gene contents of pig CNV regions

When CNV signals in two or more animals overlapped on a chromosome, they were considered to be high confidence CNVs [19]. Presently, the high confidence CNVRs contained transcripts from 0 to 89. The largest region (CNVR_5) detected in all tested pigs showed an 87.21 kb gain without overlapping any gene or duplicated sequence (Additional file 2). Same as the previous report in chicken [19], our results showed the small CNVs resided in none coding sequences, while larger CNV regions spanned more genes (Figure 2B, Additional file 2). The 259 CNVRs encompassed 372 unique transcripts which corresponded 154 mouse orthologous genes annotated in Ensembl

(Additional file 3). In order to determine the likely biological effects of the 154 mouse orthologous genes, functional annotation analysis was performed with the DAVID tool [38]. Gene Ontology (GO) analysis revealed that CNVR genes belonged to these classes of genes that participated in sensory perception of smell, sensory perception of smell or chemical stimulus, sensory perception, cognition, G-protein coupled receptor protein signaling pathway, olfactory receptor activity and other basic metabolic processes (Table 2). KEGG pathway analyses indicated that 50 genes involved in olfactory transduction ($p < 0.05$) were over-represented in the porcine CNVRs, as previously identified in cattle [15,31,37]. These CNV genes also included ATP-binding cassette, sub-family C (CFTR/MRP), tyrosine-protein kinase Kit (*KIT*) and cytochrome P450 (*CYP450*) as described previously [30,37]. A certain degree of conservation of CNVs across mammals has been observed, which suggests that selective pressure may drive acquisition or retention of specific gene dosage alterations.

To test whether genes unaffected by CNVs exhibited a different selective constraint than the ones affected, we compared the dN/dS ratios for orthologous genes of pigs with those of mouse and human species (Table 3, Additional file 3). Compared with mouse, all pig CNVR genes had dN/dS ratios significantly higher than monomorphic genes by Wilcoxon rank-sum test, which was the same as the previous results [14]. It might indicate a relaxation of purifying selection due to the redundancy fragments generated during the formation process of the variable number of genes [39–42]. However, compared with mouse, the pig CNVR genes with the status of gains had dN/dS ratios lower than monomorphic genes, indicating these genes subjected to stringent purifying selection compared with non-polymorphic genes.

Table 2 Enriched GO terms and KEGG pathway associated with the CNV regions (Modified Fisher Exact P-value ≤ 0.05)

Category	Term	Count	Involved genes/ total genes (%)	P value	Validation
GO_BP	GO:0007608 ~ sensory perception of smell	52	35.14	1.5E-29	Pig [31]
GO_BP	GO:0007606 ~ sensory perception of chemical stimulus	52	35.14	3.3E-28	Pig [31]
GO_BP	GO:0007600 ~ sensory perception	53	35.81	6.8E-26	Pig [31]
GO_BP	GO:0050890 ~ cognition	53	35.81	8.9E-25	Pig [31]
GO_BP	GO:0007186 ~ G-protein coupled receptor protein signaling pathway	56	37.84	1.2E-22	Pig [31]
GO_BP	GO:0050877 ~ neurological system process	53	35.81	3.4E-22	Pig [31]
GO_BP	GO:0007166 ~ cell surface receptor linked signal transduction	60	40.54	6.1E-20	Pig [31]
GO_MF	GO:0004984 ~ olfactory receptor activity	55	37.16	1.6E-28	Pig [31] Cattle [14]
GO_MF	GO:0047961 ~ glycine N-acyltransferase activity	2	1.35	0.025	
GO_CC	GO:0016021 ~ integral to membrane	77	52.03	4.5E-6	Pig [31]
GO_CC	GO:0031224 ~ intrinsic to membrane	78	52.70	9.3E-6	Pig [31]
KEGG_PATHWAY	mmu04740:Olfactory transduction	50	33.78	8.8E-19	Pig [31]

Table 3 Evolutionary rates of pig monomorphic and CNVR genes compared with human and mouse

		CNVR(gain)	CNVR(loss)	CNVR(both)	No CNVR
Human	dN/dS	0.20	0.32	0.27	0.21
	P value	2.41E-03	3.82E-12	1.08E-06	-
Mouse	dN/dS	0.87	0.33	0.22	0.19
	P value	1.72E-07	2.20E-16	2.14E-10	-

dN: nonsynonymous rate; dS : synonymous rate; p-value compares CNVR genes with monomorphic genes by Wilcoxon rank-sum test.

Pig CNVRs overlapped with QTL regions

We queried the animal QTL database that held publicly available QTL data on livestock species. Retrieving all the porcine QTLs (http://www.animalgenome.org/cgi-bin/QTLdb/SS/download?file=gbpSS_10.2) within 2 Mb of our CNVRs resulted that 34 CNVRs overlapped with QTLs for several important traits including average daily gain (ADG) (Additional file 4). However, as the pig QTLs are not fully defined, the contribution of these QTL-overlapping CNVRs to complex traits needs further study.

Validation of CNVRs by real-time quantitative (qPCR)

qPCR was performed to validate 2 CNVRs (CNVR_IDs 100 and 215) detected by the aCGH experiment. Thirteen DNA samples including the reference used in aCGH were used for qPCR analysis. CNVR_100 and CNVR_215 were validated (Additional file 5) with the p threshold values 0.05 as the previous reports [43].

CNVR_100 contained Mast/stem cell growth factor receptor gene, also known as *KIT* gene (ENSSSCT00000009679). In pigs, the dominant white color was associated with a splice mutation leading to the skipping of exon 17 of *KIT* gene [44] and a duplication of a 450 kb fragment encompassing the *KIT* gene [45]. The results of the aCGH array and qPCR analyses revealed that the copy number varied greatly among the different breeds (Figure 4). Coinciding with the previous study [45], 8 pigs with white hair color (one White Duroc pig, one Large White pig, two Landrace × DIV pigs, two Landrace pigs and two DIV pigs) and the Pietrain pig had *KIT* duplication, but 3 Chinese native pigs without pure white color did not have. In addition to the important role in proliferation, survival and migration of melanocytes [45], the *KIT* gene also had effects on follicle and oocyte development [46,47]. Therefore, it was worthy to further investigate the selection impact of white hair color on pig reproduction traits.

Conclusions

In summary, we described a map of porcine CNVs between breeds by a high-resolution array CGH, which was confirmed to be a very valid method to detect porcine genome-wide CNVs. With a stringent CNV calling

criterion, 259 highly reliable CNV regions were reported here among diverse pig breeds. Future studies are required to assess the function of CNVs on pig important phenotypes. Our results facilitated the identification of structural variations for important phenotypes and the assessment of the genetic diversity in pigs.

Methods

Sample preparation

All animal procedures were performed according to protocols approved by the Biological Studies Animal Care and Use Committee of Hubei Province, PR China. Twelve pigs including one White Duroc pig (♀), one Chinese Yangxin pig (♂), one Chinese Erhualian pig (♀), one Chinese Tongcheng pig (♀), one Large White pig (♀), one Pietrain pig (♂), two Landrace pigs (♂), two DIV pigs (♀) and two Landrace × DIV pigs (♀, ♂) were selected to function as test animals. Chinese Erhualian pigs were a strain of Chinese Taihu pig breed. Synthetic Line DIV was a result of cross of Landrace, Large White, Tongcheng or Taihu pigs. An unrelated female Duroc pig was selected as the common reference. The genomic DNA of 13 pig samples was extracted and purified from semen, whole blood or ear notch.

Oligonucleotide aCGH

A 3 × 720 K whole genome tiling aCGH (NCBI GEO accession no. GPL16165) was designed (NimbleGen Systems, <http://www.nimblegen.com>) from the Sscrofa9.2 release (http://www.sanger.ac.uk/Projects/S_scrofa/), which was the new release at the time of the experiment. The probe design fundamentals were described in the NimbleGen technical note (http://www.nimblegen.com/products/lit/probe_design_2008_06_04.pdf). The probes with length of 50–60 bp were integrated into an array design using

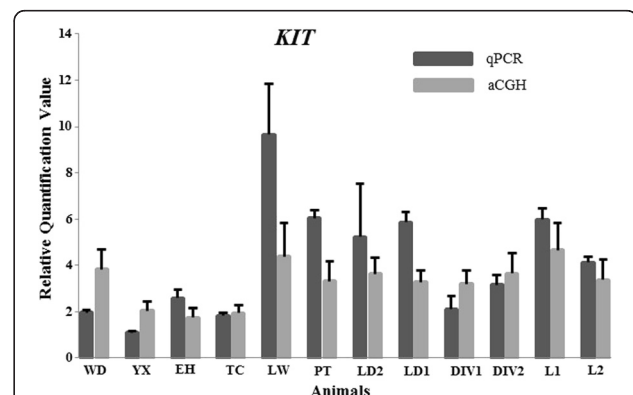


Figure 4 Validation of CNVR_100 (*KIT* gene) detected from the CGH array data using real-time quantitative PCR analysis. The x-axis represents the animals and the y-axis shows the relative quantification value ($2^{-\Delta\Delta C_t}$ values for qPCR; $2^{*(2^{\wedge}\text{Sample signal})}$ values for array CGH).

ArrayScribe™, which resulted in a design with a median probe spacing of 2506 bp. Test DNA and reference DNA samples were independently labeled with either Cy3 or Cy5 dyes. Labeled DNA was co-hybridized to the custom-made NimbleGen CGH array (3 × 720 K). The array format included 3 arrays on single slides containing 719,336 probes. The arrays were scanned using a 5 μm scanner, and NimbleScan software (Roche NimbleGen) 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 Cy5-labeled reference sample were computed. Before normalization and segmentation analysis, spatial correction was applied. Specifically, locally weighted polynomial regression (LOESS) was used to adjust signal intensities based on X, Y feature position [48]. Normalization was then performed using the q-spline method followed by segmentation using the CNV calling algorithm segMNT included in NimbleScan software [11]. CNVRs were called as the segments with at least 5 consecutive probes, a mean |log₂ ratio| of >0.50 and detected in two or more animals [28]. Since the CNV calling pipeline requires at least 5 consecutive probes, our theoretical resolution for CNV detection is 10299 bp (median spacing × 4 + median oligo length × 5). As females had two copies of X-linked genes and males only had one copy, male–female aCGH resulted in an excess of female signals for X-linked genes that can be used to calibrate the threshold values and detection methods [49]. aCGH data have been submitted to the GenBank gene expression omnibus database under the accession number GSE41488. The dendrogram were generated by average linkage clustering algorithm of Cluster 3.0 software [50].

Enrichment analysis

In order to check if the CNVRs overlapped any duplicated sequence, BLAST was used to query the CNVRs sequences against the *Sus scrofa* genome sequence (Sscrofa9.2). Sequences were retained as duplicated sequences if they had ≥ 1 kb and ≥ 90% identity and occurred at more than one site within the genome.

Gene contents in the identified CNVRs were retrieved from the Sscrofa9.2 assembly using the BioMart (<http://www.biomart.org/>) [51]. Gene content of pig CNV regions was assessed using Ensembl transcripts. The DAVID functional annotation tool (<http://david.abcc.ncifcrf.gov/>) was used to perform GO classification and KEGG pathway annotation of CNV mRNAs. Functional annotation terms from the ontologies of "biological processes", "molecular function" and "cellular component" were recorded. Since only a limited number of genes in the pig genome have been annotated, we converted the pig Ensembl transcripts IDs to orthologous mouse and human Ensembl gene IDs

by BioMart, then carried out the GO and pathway analyses, as described previously [31].

All the porcine QTLs data were downloaded from pig QTL database (http://www.animalgenome.org/cgi-bin/QTLdb/SS/download?file=gbpSS_10.2) [52]. The CNVRs were considered to be overlapping pig QTLs if they were within 2 Mb of pig QTLs [14].

Validation of CNVRs by qPCR

Determination of CNVRs by qPCR was performed using the Roche LightCycler® 480 Detection System and obtained the crossing thresholds (Ct) value following the guidelines of the manufacturer. The primers were designed using the Primer Premier 5 software and were available in the Additional file 6. As previously reported [28], the copy number of each CNVR was normalized against the Col10 region, a control region in the genome that did not vary in copy number between the pigs. Triplicate wells of reactions (15 μL) contained 7.5 μL SYBR Green Real-time PCR Master Mix, 1 μL of 10–20 ng/μL gDNA, 0.3 μL 5 μM of each primer and 0.1 μL ROX. The cycling conditions consisted of 1 cycle at 95°C for 10 min, followed by 40 cycles at 94°C for 20 sec, 60°C for 20 sec, and 72°C for 20 sec, with fluorescence acquisition at 74°C in single mode. The specific PCR products were confirmed by the results of melting curve analysis and agarose gel electrophoresis. Analysis of resultant crossing thresholds (Ct) was performed using the -ΔΔCt method [53].

Additional files

Additional file 1: Probe summary of the 720 K custom-made CGH array designed by Roche NimbleGen.

Additional file 2: Description of the CNVRs detected by a whole-genome CGH array. The genomic coordinates were expressed in bp and were relative to the *Sus scrofa* genome sequence assembly (Sscrofa9.2). BLAST was used to query the CNVRs sequences against the *Sus scrofa* genome sequence (Sscrofa9.2). Sequences were retained as duplicated sequences if they had ≥ 1 kb and ≥ 90% identity and occur at more than one site within the genome. WD: White Duroc (♀); YX: Yangxin (♂); EH: Erhualian (♀); TC: Tongcheng (♀); LW: Large White (♀); PT: Pietrain (♂); LD1: Landrace × DIV pig 1 (♂); LD2: Landrace × DIV pig 2 (♀); DIV1: Chinese new pig line DIV 1 (♀); DIV2: Chinese new pig line DIV 2 (♀); L1: Landrace 1 (♂); L2: Landrace 2 (♂).

Additional file 3: Gene contents of CNVRs.

Additional file 4: QTLs overlapped with the CNVRs. All the porcine QTLs within 2 Mb (http://www.animalgenome.org/cgi-bin/QTLdb/SS/download?file=gbpSS_10.2) of our CNVRs were counted.

Additional file 5: The validation of the aCGH results using qPCR method.

Additional file 6: The primers of qPCR to validate the CNVRs detected by aCGH.

Abbreviations

CNV: Copy number variation; CNVR: CNV region; PCR: Polymerase chain reaction; CGH: Comparative genome hybridization; aCGH: Array CGH; qPCR: Real-time quantitative PCR; RQ: Relative quantification value;

QTL: Quantitative trait locus; KIT: Tyrosine-protein kinase Kit; CYP450: Cytochrome P450 gene family; SNP: Single nucleotide polymorphism; HSA: *Homo sapiens* chromosome; SSC: *Sus scrofa* chromosome; GO: Gene ontology; DAVID: The database for annotation, visualization and integrated discovery; KEGG: kyoto encyclopedia of genes and genomes; LOESS: locally weighted polynomial regression; Ct: crossing thresholds; SD: Segmental duplication.

Competing interests

The authors have declared that no financial competing interests exist.

Authors' contributions

YL, SM, FL carried out most of bioinformatics analysis and lab works. XZ, XP, HW, GL, HT participated in the animal samples collection and statistical analysis. FL, SJ, YX participated in the experiment design and coordination. FL conceived the study and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We thank the anonymous reviewers for critical reading and discussions of the manuscript. We are grateful to Prof. Alan Archibald (The Roslin Institute) for the suggestions for this study, and to CapitalBio Corporation for the technical assistance with NimbleGen CGH analysis. The authors also acknowledge the farmers for providing pig samples.

Author details

¹Key Laboratory of Pig Genetics and Breeding of Ministry of Agriculture & Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, Huazhong Agricultural University, Wuhan 430070, PR China. ²Hubei Key Laboratory of Animal Embryo Engineering and Molecular Breeding, Hubei Academy of Agriculture Science, Wuhan 430070, PR China.

Received: 17 August 2012 Accepted: 19 December 2012

Published: 24 December 2012

References

1. Amills M, Clop A, Ramirez O, Perez-Enciso M: **Origin and genetic diversity of pig breeds.** In *Encyclopedia of Life Sciences (ELS)*. Chichester: John Wiley & Sons, Ltd; 2010.
2. Megens HJ, Crooijmans RP, San Cristobal M, Hui X, Li N, Groenen MA: **Biodiversity of pig breeds from China and Europe estimated from pooled DNA samples: differences in microsatellite variation between two areas of domestication.** *Genet Sel Evol* 2008, **40**:103-128.
3. Haley CS, Agaro E, Ellis M: **Genetic components of growth and ultrasonic fat depth traits in Meishan and Large White pigs and their reciprocal crosses.** *Anim Prod* 1992, **54**:105-115.
4. Haley CS, Lee GJ, Ritchie M: **Comparative reproductive-performance in Meishan and Large White pigs and their crosses.** *Anim Sci* 1995, **60**:259-267.
5. Wei WH, de Koning DJ, Penman JC, Finlayson HA, Archibald AL, Haley CS: **QTL modulating ear size and erectness in pigs.** *Anim Genet* 2007, **38**:222-226.
6. Ramos AM, Crooijmans RP, Affara NA, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R, Dehais P, Hansen MS, Hedegaard J, Hu ZL, Kerstens HH, Law AS, Megens HJ, Milan D, Nonneman DJ, Rohrer GA, Rothschild MF, Smith TP, Schnabel RD, Van Tassel CP, Taylor JF, Wiedmann RT, Schook LB, Groenen MA: **Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology.** *PLoS One* 2009, **4**:e6524.
7. Sebat J: **Major changes in our DNA lead to major changes in our thinking.** *Nat Genet* 2007, **39**:S3-S5.
8. Feuk L, Carson AR, Scherer SW: **Structural variation in the human genome.** *Nat Rev Genet* 2006, **7**:85-97.
9. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zhang J, Aramengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME: **Global variation in copy number in the human genome.** *Nature* 2006, **444**:444-454.
10. Greenway SC, Pereira AC, Lin JC, DePalma SR, Israel SJ, Mesquita SM, Ergul E, Conta JH, Korn JM, McCarroll SA, Gorham JM, Gabriel S, Altshuler DM, Quintanilla-Dieck Mde L, Artunduaga MA, Eavey RD, Plenge RM, Shadick NA, Weinblatt ME, De Jager PL, Hafler DA, Breitbart RE, Seidman JG, Seidman CE: **De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot.** *Nat Genet* 2009, **41**:931-935.
11. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, Leotta A, Pai D, Zhang R, Lee YH, Hicks J, Spence SJ, Lee AT, Puura K, Lehtimäki T, Ledbetter D, Gregersen PK, Bregman J, Sutcliffe JS, Jobanputra V, Chung W, Warburton D, King MC, Skuse D, Geschwind DH, Gilliam TC, Ye K, Wigler M: **Strong association of De Novo copy number mutations with Autism.** *Science* 2007, **316**:445-449.
12. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M: **Strong association of de novo copy number mutations with sporadic schizophrenia.** *Nat Genet* 2008, **40**:880-885.
13. Bae JS, Cheong HS, Kim LH, NamGung S, Park TJ, Chun JY, Kim JY, Pasaje CF, Lee JS, Shin HD: **Identification of copy number variations and common deletion polymorphisms in cattle.** *BMC Genomics* 2010, **11**:232.
14. Fadista J, Thomsen B, Holm LE, Bendixen C: **Copy number variation in the bovine genome.** *BMC Genomics* 2010, **11**:284.
15. Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, Mitra A, Alexander LJ, Coutinho LL, Dell'Aquila ME, Gasbarre LC, Lacalandra G, Li RW, Matukumalli LK, Nonneman D, Regitano LC, Smith TP, Song J, Sonstegard TS, Van Tassel CP, Ventura M, Eichler EE, McDanel TG, Keele JW: **Analysis of copy number variations among diverse cattle breeds.** *Genome Res* 2010, **20**:693-703.
16. 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.** *Dev Biol (Basel)* 2008, **132**:231-237.
17. Griffin DK, Robertson LB, Tempest HG, Vignal A, Fillon V, Crooijmans RP, Groenen MA, Deryusheva S, Gaginyskaya E, Carré W, Waddington D, Talbot R, Völker M, Masabanda JS, Burt DW: **Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution.** *BMC Genomics* 2008, **9**:168.
18. Skinner BM, Robertson LB, Tempest HG, Langley EJ, Ioannou D, Fowler KE, Crooijmans RP, Hall AD, Griffin DK, Völker M: **Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis.** *BMC Genomics* 2009, **10**:357.
19. Wang X, Nahashon S, Feaster TK, Bohannon-Stewart A, Adefope N: **An initial map of chromosomal map of chromosomal segmental copy number variations in the chicken.** *BMC Genomics* 2010, **11**:351.
20. Fontanesi L, Beretti F, Martelli PL, Colombo M, Dall'olio S, Occidente M, Portolano B, Casadio R, Matassino D, Russo V: **A first comparative map of copy number variations in the sheep genome.** *Genomics* 2011, **97**:158-165.
21. Fontanesi L, Martelli PL, Beretti F, Riggio V, Dall'Olio S, Colombo M, Casadio R, Russo V, Portolano B: **An initial comparative map of copy number variations in the goat (*Capra hircus*) genome.** *BMC Genomics* 2010, **11**:639.
22. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, Krauss RM, Myers RM, Ridker PM, Chasman DJ, Mefford H, Ying P, Nickerson DA, Eichler EE: **Population analysis of large copy number variants and hotspots of human genetic disease.** *Am J Hum Genet* 2009, **84**:148-161.
23. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemes J, Wyszoker A, Shapero MH, de Bakker PI, Maller JB, Kirby A, Elliott AL, Parkin M, Hubbell E, Webster T, Mei R, Veitch J, Collins PJ, Handsaker R, Lincoln S, Nizzari M, Blume J, Jones KW, Rava R, Daly MJ, Gabriel SB, Altshuler D: **Integrated detection and population-genetic analysis of SNPs and copy number variation.** *Nat Genet* 2008, **40**:1166-1174.
24. 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.
25. Orozco LD, Cokus SJ, Ghazalpour A, Ingram-Drake L, Wang S, van Nas A, Che N, Araujo JA, Pellegrini M, Lusis AJ: **Copy number variation influences gene expression and metabolic traits in mice.** *Hum Mol Genet* 2009, **18**:4118-4129.

26. Butler MW, Hackett NR, Salit J, Strulovici-Barel Y, Omberg L, Mezey J, Crystal RG: **Glutathione S-transferase copy number variation alters lung gene expression.** *Eur Respir J* 2011, **38**:15–28.
27. Brønstad I, Wolff AS, Løvås K, Knappskog PM, Husebye ES: **Genome-wide copy number variation (CNV) in patients with autoimmune Addison's disease.** *BMC Med Genet* 2011, **12**:111.
28. Fadista J, Nygaard M, Holm LE, Thomsen B, Bendixen C: **A snapshot of CNVs in the pig genome.** *PLoS One* 2008, **3**:e3916.
29. Tang H, Li F, Finlayson HA, Smith S, Lu Z, Langford C, Archibald A: **Structural And Copy Number Variation In The Pig Genome.** In *Book Structural And Copy Number Variation In The Pig Genome.* City: Plant & Animal Genomes XVIII Conference. Town; 2010. January 9-13, 2010.
30. Ramayo-Caldas Y, Castelló A, Pena RN, Alves E, Mercadé A, Souza CA, Fernández AI, Perez-Enciso M, Folch JM: **Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip.** *BMC Genomics* 2010, **11**:593.
31. Wang J, Jiang J, Fu W, Jiang L, Ding X, Liu J, Zhang Q: **A genome-wide detection of copy number variations using SNP genotyping arrays in swine.** *BMC Genomics* 2012, **13**:273.
32. Nicholas TJ, Cheng Z, Ventura M, Mealey K, Eichler EE, Akey JM: **The genomic architecture of segmental duplications and associated copy number variants in dogs.** *Genome Res* 2009, **19**:491–499.
33. Armengol L, Pujana MA, Cheung J, Scherer SW, Estivill X: **Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary rearrangements.** *Hum Mol Genet* 2003, **12**:2201–2208.
34. Kemkemer C, Kohn M, Cooper DN, Froenicke L, Högel J, Hameister H, Kehrer-Sawatzki H: **Gene synteny comparisons between different vertebrates provide new insights into breakage and fusion events during mammalian karyotype evolution.** *BMC Evol Biol* 2009, **9**:84.
35. Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, Schwartz S, Albertson DG, Pinkel D, Altshuler DM, Eichler EE: **Linkage disequilibrium and heritability of CNPs within duplicated regions of the human genome.** *Am J Hum Genet* 2006, **79**:275–290.
36. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Seagraves R, Oseroff W, Albertson DG, Pinkel D, Eichler EE: **Segmental duplications and copy-number variation in the human genome.** *Am J Hum Genet* 2005, **77**:78–88.
37. Hou Y, Liu GE, Bickhart DM, Cardone MF, Wang K, Kim ES, Matukumalli LK, Ventura M, Song J, VanRaden PM, Sonstegard TS, Van Tassell CP: **Genomic characteristics of cattle copy number variations.** *BMC Genomics* 2011, **12**:127.
38. Huang DW, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nature Protoc* 2009, **4**:44–57.
39. Kondrashov FA, Kondrashov AS: **Role of selection in fixation of gene duplications.** *J Theor Biol* 2006, **239**:141–151.
40. Nguyen DQ, Webber C, Ponting CP: **Bias of selection on human copy-number variants.** *PLoS Genet* 2006, **2**:e20.
41. Ohno S: *Evolution by gene duplication.* 1st edition. New York Heidelberg Berlin: Springer-Verlag; 1970.
42. 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.
43. Van Belle G, Fisher LD, Heagerty PJ, Lumley T: **Association and prediction: linear models with one predictor variable.** In *Biostatistics: A Methodology For the Health Sciences*, 9. 2nd edition. New Jersey: Wiley; 2004:291–356.
44. Giuffra E, Evans G, Törnsten A, Wales R, Day A, Looft H, Plastow G, Andersson L: **The Belt mutation in pigs is an allele at the Dominant white (I/KIT) locus.** *Mamm Genome* 1999, **10**:1132–1136.
45. Giuffra E, Törnsten A, Marklund S, Bongcam-Rudloff E, Chardon P, Kijas JM, Anderson SI, Archibald AL, Andersson L: **A large duplication associated with dominant white color in pigs originated by homologous recombination between LINE elements flanking KIT.** *Mamm Genome* 2002, **13**:569–577.
46. Wehrle-Haller B: **The role of Kit-ligand in melanocyte development and epidermal homeostasis.** *Pigment Cell Res* 2003, **16**:287–296.
47. Hutt KJ, McLaughlin EA, Holland MK: **Kit ligand and c-Kit have diverse roles during mammalian oogenesis and folliculogenesis.** *Mol Hum Reprod* 2006, **12**:61–69.
48. Smyth GK, Speed T: **Normalization of cDNA microarray data.** *Methods* 2003, **31**:265–273.
49. Zhou J, Lemos B, Dopman EB, Hartl DL: **Copy-number variation: the balance between gene dosage and expression in *Drosophila melanogaster*.** *Genome Biol Evol* 2011, **3**:1014–1024.
50. de Hoon MJ, Imoto S, Nolan J, Miyano S: **Open source clustering software.** *Bioinformatics* 2004, **20**:1453–1454.
51. Guberman JM, Ai J, Arnaiz O, Baran J, Blake A, Baldock R, Chelala C, Croft D, Cros A, Cutts RJ, Di Génova A, Forbes S, Fujisawa T, Gadaleta E, Goodstein DM, Gundem G, Haggarty B, Haider S, Hall M, Harris T, Haw R, Hu S, Hubbard S, Hsu J, Iyer V, Jones P, Katayama T, Kinsella R, Kong L, Lawson D, Liang Y, Lopez-Bigas N, Luo J, Lush M, Mason J, Moreews F, Ndegwa N, Oakley D, Perez-Llamas C, Primig M, Rivkin E, Rosanoff S, Shepherd R, Simon R, Skarnes B, Smedley D, Sperling L, Spooner W, Stevenson P, Stone K, Teague J, Wang J, Wang J, Whitty B, Wong DT, Wong-Erasmus M, Yao L, Youens-Clark K, Yung C, Zhang J, Kasprzyk A: **BioMart Central Portal: an open database network for the biological community.** *Database (Oxford)* 2011, **18**:bar041.
52. Hu ZL, Fritz ER, Reecy JM: **AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond.** *Nucleic Acids Res* 2007, **35**:D604–D609.
53. 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.

doi:10.1186/1471-2164-13-725

Cite this article as: Li et al.: Identification of genome-wide copy number variations among diverse pig breeds by array CGH. *BMC Genomics* 2012 **13**:725.

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

