A Snapshot of CNVs in the Pig Genome
João Fadista, Marianne Nygaard, Lars-Erik Holm, Bo Thomsen, Christian Bendixen*

Group of Molecular Genetics and Systems Biology, Department of Genetics and Biotechnology, Faculty of Agricultural Sciences, Aarhus University, Tjele, Denmark

Abstract
Recent studies of mammalian genomes have uncovered the extent of copy number variation (CNV) that contributes to phenotypic diversity, including health and disease status. Here we report a first account of CNVs in the pig genome covering part of the chromosomes 4, 7, 14, and 17 already sequenced and assembled. A custom tiling oligonucleotide array was used with a median probe spacing of 409 bp for screening 12 unrelated Duroc boars that are founders of a large family material. After a strict CNV calling pipeline, 37 copy number variable regions (CNVRs) across all four chromosomes were identified, with five CNVRs overlapping segmental duplications, three overlapping pig unigenes and one overlapping a RefSeq pig mRNA. This CNV snapshot analysis is the first of its kind in the porcine genome and constitutes the basis for a better understanding of porcine phenotypes and genotypes with the prospect of identifying important economic traits.

Introduction
The pig (Sus scrofa) is a cetartiodactyl mammal from a different clade than rodents and primates and last shared a common ancestor with humans approx. 83 million years ago [1]. The porcine genome has an estimated size of 2.7 Gb, consisting of 18 autosomes and the X and Y sex chromosomes [2]. Genomic comparisons between the pig and human have unravelled more structural resemblance than, for example, mouse and human [2–4]. Pig is also a more trustworthy animal model for human disease since its physiological and anatomical resemblance is far greater than any other laboratory species. Consequently, the pig has been used progressively as a model within the human health research in e.g. obesity, cardiovascular disease, arthritis, diabetes, hypertension, cancer, organ transplantation, and Alzheimer’s disease [5–8]. Further to the biomedical relevance, the pig is of great agricultural importance as the main source of animal protein world-wide (Porcine sequencing white paper).

Recently it has been reported that structural variation, like copy number variants (CNVs), is genome-wide present not only in humans [9–17] but also in chimpanzees [18–19], mice [20–23], nematodes [24], fruit fly [25], and cow [26]. CNVs represent segments of DNA larger than 1 kb present at a variable copy number in comparison with a reference genome [27] and they can be responsible for altered gene expression [28] leading to striking phenotypic variance including disease associated traits [29–30]. Despite numerous studies, no assessment of the extent and impact of CNVs in the pig genome has been made until now. According to a pig family material comprising 14 boars, 700 sows and about 12,000 offsprings this paper presents a preliminary analysis of CNVs detected in the genomes of twelve of the boar founders compared to one unrelated Hampshire boar, using high density tiling-path oligonucleotide array comparative genomic hybridization technology [array CGH] [31]. The designed arrays encompass part of the chromosomes 4, 7, 14, and 17 from the August 2007 preliminary assembly release with a median probe spacing of 409 bp. After a stringent pipeline, the analysis led to the identification of 37 copy number variable regions. Chosen CNVs were further confirmed by RT-PCR [32].

As the first of its kind in pig, this study examines the extent and pattern of CNVs in the pig genome, important for future studies associating phenotype to genome architecture.

Results
Study design
Array CGH was carried out using an array comprising 384,979 oligonucleotide probes covering the preliminary pig genome assembly for part of the chromosomes 4, 7, 14, and 17 with a median probe spacing of 409 bp. Copy number variation was assessed by equating the log2ratio of signal intensity between the reference and test samples. Given the relative type of these comparative data, it was not possible to unequivocally ascertain the real status of the CNVs not RT-PCR validated, and hence whether they were deletions or duplications in the reference or in the test samples. Therefore, the status of the copy number variations reported here is in relation to the reference sample.

Since our criteria of CNV detection (Methods and Figure 1) only permit to call a CNV if it is detected in at least two animals, we will be referring to copy number variable regions or CNVRs (merging of overlapping CNVs in two or more animals) instead of CNVs. The possibility that true CNVs exist within the loci discovered only in a single animal is acknowledged, since they may comprise sporadic cases, but in order to minimize the false positive rate, we focused only on CNVs found in two or more animals. Another strong reason to discard CNVs found in only one animal relates to the fact that some CNVs may be somatic and not germline [10–61]. Therefore, in this study, as previously [10], a CNV was considered to be “germline” if it was detected in at least two animals.
Figure 1. Methodological pipeline for assessing copy number variation in this study. See Methods section for detailed description.

doi:10.1371/journal.pone.0003916.g001
Pattern and frequency of CNVRs

Following the methodological copy number pipeline, 37 CNVRs were identified by array CGH across all the four chromosomes queried (Figure 2) with the proportion of any given chromosome amenable to CNVR varying from 0.03% to 0.31%. In summary, 19 (51.4%) CNVRs were called in two animals and the remaining (48.6%) called as losses (Table 1).

Previously, it has been suggested that deletions are under stronger purifying selection than duplications [33]. If so, deletions should be both less frequent and shorter than duplications. When comparing the length and number of gains versus losses in the CNVRs, practically the same number was detected (see above) while the total size of the gains was about 16 kb larger than the size of the losses (although according to the Wilcoxon rank sum test not statistically significant at p value $\leq 0.01$).

Copy number regions were discarded if (1) they overlapped more than one contig and (2) contained gaps due to the high error rate of this preliminary assembly. This was done because contigs can be misassigned and gaps may contain future contigs that can result in the disruption of the CNVRs called (see Figure 3 for an example of a CNVR called). The CNVRs ranged in size from 1.74 kb to 61.92 kb with a mean of 9.32 kb and a median of 6.89 kb, covering 429.269 kb (0.18%) of the 237.76 Mb of sequence addressed (Table 2).

When querying the part of the genome covered by CNVRs for the greatest divergence in genome size between two animals among our set, it was found that animal B had the biggest net gain, spanning 57.56 kb over five CNVRs, while animal E had the biggest net loss with $-106.6$ kb over eleven CNVRs. Comparison of these genomes disclosed a difference of 164.1 kb in size between these two animals.

There are seven CNVRs that are apparently aberrant in the genomes of at least half of the test boars. However, these CNVRs are most likely to be aberrant regions in the reference boar since this is of a different breed (i.e. Hampshire), and probably more structural genomic variation is present between breeds than within the same breed, since all the test boars are of the Duroc breed. This hypothesis remains to be tested.

Previous analyses have reported enrichment of CNVs near segmental duplications (sequences $\geq$1 kb in size with sequence identity $\geq$90% [34]) in humans [9–12,16–17], mice [20–23], and chimpanzees [19]. Segmental duplications have been mapped in the genomes of human [35], chimpanzee [36], and mouse [37], but the incomplete and highly error-prone preliminary assembly of the pig genome prevents us from drawing such a map. Consequently, only the available numbers can be focused on in this study. Among the regions found to contain CNVs, five (13.5%) overlapped segmental duplications (Table 1 and Methods).

Despite the discovery of CNVRs overlapping segmental duplications, it is important to note that our array probe design is biased against the detection of CNVRs that coincide with sites of segmental duplication because it only allowed probes that had a unique match in the genome (Methods).

Functional analysis

In order to assess the gene content within the CNVRs reported, a sequence similarity $\geq$98% search between the pig Unigene database and the CNVRs was made and three pig Unigenes were retrieved (Table 1). Since a CNV can also affect gene expression at long distances [28] an additional search was performed for pig Unigenes that showed $\geq$98% sequence similarity with the contigs where the CNVRs are. Contigs with CNVRs range in length from 5.577 kb from 245.924 kb. Further there were nine new Unigenes in eight of the contigs (CNVR.1 - Ssc.28459; CNVR.11 - Ssc.25025; CNVR.12 - Ssc.26197; CNVR.13 - Ssc.26126; CNVR.14 - Ssc.42797; CNVR.25 - Ssc.8364; CNVR.31 - Ssc.38482, Ssc.14020; CNVR.33 - Ssc.63374).

Searching the Refseq mRNA database for vertebrate mammals, a gene was identified (95% sequence id) as part of the contig containing the CNVR 33. This Refseq mRNA corresponds to the ADRA2 gene, encoding the alpha2A-adrenergic receptor - a transmembrane receptor belonging to the rhodopsin family from which genes have been consistently reported to overlap CNV
regions in other mammals [16,19,21–23]. In fact, this gene is actually overlapped by a putative human CNV, as seen in the Database of Genomic Variants [9].

Validation by RT-PCR

Validation of the results was made with RT-PCR [32] on eight genomic regions (Figure 4) selected to represent a range of amplifications and deletions (CNVR IDs 2, 6, 7, 12, 23, 28, 33, and 37). The Data S1 file contains primer sequences, RT-PCR results, and the correlation between the array CGH and RT-PCR.

From these eight regions, four were confirmed (CNVR IDs 2, 7, 28, and 37). For CNVR 2, an additional animal having this CNVR was found by RT-PCR which was not expected from the array data. Regarding the CNVR 7, a loss was found not only in the four animals predicted by our CNV calling pipeline but also in all the other test animals relative to the reference. For CNVR 37, the PCR was negative in the reference animal, while the test animals gave well-shaped sigmoidal curves, suggesting a loss in the reference animal (data not shown). Three regions were not confirmed (CNVR IDs 12, 33, and 23) and one gave ambiguous results (CNVR ID 6).

Table 1. Distribution, length(bp), status and frequency of the 37 CNVRs detected by array CGH.

| CNVR ID | Chr | Start  | End   | Length | Status | Animals | Pig Unigenes |
|---------|-----|--------|-------|--------|--------|---------|--------------|
| 1       | 4   | 1 695 691 | 1 703 117 | 7 427  | Gain   | 2       |              |
| 2       | 4   | 7 560 864 | 7 564 463 | 3 600  | Loss   | 3       |              |
| 3       | 4   | 21 108 665 | 21 114 596 | 5 932  | Gain   | 2       |              |
| 4       | 4   | 24 510 700 | 24 517 176 | 6 477  | Gain   | 9       |              |
| 5       | 4   | 34 556 025 | 34 562 828 | 6 804  | Loss   | 3       |              |
| 6       | 4   | 41 748 621 | 41 758 338 | 9 718  | Gain   | 2       |              |
| 7       | 4   | 50 753 651 | 50 761 743 | 8 093  | Loss   | 4       |              |
| 8       | 4   | 78 167 375 | 78 177 556 | 10 182 | Loss   | 8       |              |
| 9       | 7   | 4 502 382 | 4 510 505 | 8 124  | Loss   | 2       | Ssc.18508    |
| 10      | 7   | 6 630 532 | 6 636 941 | 6 410  | Loss   | 2       |              |
| 11      | 7   | 17 630 828 | 17 634 369 | 3 542  | Gain   | 2       |              |
| 12      | 7   | 23 821 562 | 23 838 973 | 17 412 | Loss   | 10      |              |
| 13      | 7   | 27 171 334 | 27 203 171 | 31 838 | Loss   | 4       |              |
| 14      | 7   | 27 660 543 | 27 699 166 | 38 624 | Loss   | 3       |              |
| 15      | 7   | 29 138 825 | 29 161 420 | 22 596 | Loss   | 2       |              |
| 16      | 7   | 38 743 324 | 38 746 353 | 3 030  | Gain   | 4       |              |
| 17      | 7   | 51 627 146 | 51 633 742 | 6 597  | Loss   | 2       |              |
| 18      | 7   | 66 692 031 | 66 753 950 | 61 920 | Gain   | 10      |              |
| 19      | 7   | 83 324 440 | 83 331 849 | 7 410  | Loss   | 2       |              |
| 20      | 7   | 86 241 308 | 86 244 612 | 3 305  | Loss   | 3       |              |
| 21      | 7   | 90 510 071 | 90 513 252 | 3 182  | Gain   | 2       |              |
| 22      | 7   | 95 943 885 | 95 949 065 | 5 181  | Loss   | 2       |              |
| 23      | 7   | 115 040 167 | 115 050 693 | 10 527 | Gain   | 3       |              |
| 24      | 7   | 121 271 511 | 121 276 035 | 4 525  | Loss   | 2       |              |
| 25      | 14  | 41 934 539 | 41 940 418 | 5 880  | Loss   | 2       |              |
| 26      | 14  | 53 639 912 | 53 651 468 | 11 557 | Gain   | 3       |              |
| 27      | 14  | 60 889 473 | 60 895 547 | 6 075  | Gain   | 6       |              |
| 28      | 14  | 71 252 362 | 71 281 840 | 29 479 | Gain   | 6       |              |
| 29      | 14  | 72 268 017 | 72 274 799 | 6 783  | Loss   | 2       |              |
| 30      | 14  | 115 689 469 | 115 733 497 | 44 029 | Gain   | 4       | Ssc.7991, Ssc.52309 |
| 31      | 14  | 119 256 322 | 119 263 779 | 7 458  | Loss   | 2       |              |
| 32      | 14  | 122 029 011 | 122 035 919 | 6 909  | Gain   | 2       |              |
| 33      | 14  | 126 779 081 | 126 784 999 | 5 919  | Loss   | 7       |              |
| 34      | 14  | 136 045 554 | 136 047 297 | 1 744  | Gain   | 2       |              |
| 35      | 14  | 137 517 567 | 137 521 449 | 3 883  | Gain   | 2       |              |
| 36      | 17  | 50 258 487 | 50 260 569 | 2 083  | Gain   | 2       |              |
| 37      | 17  | 50 554 656 | 50 559 679 | 5 024  | Gain   | 3       |              |

CNVRs indicated in bold overlap segmental duplications.
The genomic coordinates are relative to the Sus scrofa May 2008 assembly.

doi:10.1371/journal.pone.0003916.t001
Here, using a custom tiling oligonucleotide array CGH approach, we reported the first CNV survey of the pig genome among twelve unrelated healthy boars which are founders of a vast pig family. It should be stressed that only four chromosomes and not the whole genome were screened here. Both gains and losses of different lengths were discovered on part of chromosomes 4, 7, 14, and 17. With the tiling nature of the array, we were able to identify 37 frequently occurring loci of copy number variation.

Natural large-scale genomic size divergence between animals of the same breed was found to vary by at least 164.1 kb, showing that a substantial portion of the pig genome may vary in copy number. In comparison with CNV studies in the “finished” human and mouse genomes [17,22], our study found an order of magnitude less genomic size divergency. This is not surprising since the pig assembly is currently only in its draft form, covering less sequenced data.

With a detection sensitivity ranging from about 2 kb (median spacing*5probes) to 248.471 kb (length of the biggest contig in the

---

**Table 2.** Summary of the CNVR content (in bp) and sequence covered (including gaps) by the oligo array CGH probes.

| Chr | CNVRs | Median size | Mean size | Size range | CNVR Content | Sequence covered | % CNVR |
|-----|-------|-------------|-----------|------------|--------------|------------------|--------|
| 4   | 8     | 7 116       | 7 279     | 10 182–3 600 | 58 223 | 52983989 | 0.11   |
| 7   | 16    | 7 004       | 14 639    | 6 1920–3 030 | 234 223 | 76062953 | 0.31   |
| 14  | 11    | 6 783       | 11 792    | 44 029–1 744 | 129 716 | 83175859 | 0.156  |
| 17  | 2     | 3 554       | 3 554     | 5 024–2 083 | 7 107  | 25535213 | 0.03   |
| All | 37    | 6 894       | 9 316     | 61 920–1 744 | 429 269 | 237758014 | 0.18   |

doi:10.1371/journal.pone.0003916.t002
Sscrofa 6 assembly), at least 0.18% of the mapped pig chromosomes are tolerant to copy number variation.

Concerning the functional sequence content, twelve pig unigene sequences and one Refseq gene were found to be putatively under influence of the CNVRs. The Refseq gene is related to sensory perception, which is a common large and rapidly evolving gene family found to contain many genes overlapped by CNVs in other mammalian genomes [16–17,19,21–22]. This gene family is possibly conserved by natural selection in mammalian species or, with a different view, could mean a relative relaxation of selective pressure on copy number variants for these genes.

In order to confirm the CNVRs found with the array approach, RT-PCR was carried out on some CNVRs and 50% of the selected CNVRs were validated. Although this validation rate seems poor, it should be noted that RT-PCR is not trivial for a highly error-prone preliminary genome assembly. Many factors could account for this discrepancy as explained very thoroughly elsewhere [63], like: (1) The breakpoint estimation of the copy number variable regions may not be correct leading to a primer design upstream or downstream of the true boundaries of the CNVR; (2) CNVRs have a lower probe density than usual because some regions surrounding the Nimblegen probes have a high repeat content which may disturb the PCR reaction; (3) The animals may have SNPs and small indels in the CNVRs compared to the reference genome assembly, which may compromise the RT-PCR reaction but not the CGH hybridization, or at least not so seriously [62], since the RT-PCR primers are shorter and thus less robust than the CGH probes. The source of the disagreement between RT-PCR and array CGH awaits further research.

Further validation was done using 7k SNPs ascertained in-house by mapping their surrounding sequences to the Sus scrofa 6 assembly for the pigs queried in this study (unpublished data). Here it was tested whether the SNPs found in close proximity of the CNVRs validated by RT-PCR gave some information about the presence of copy number variants in those regions (see Methods). In fact, for three of the CNVRs, the SNP alleles in close proximity were found to cluster in 2 groups: animals with the CNVR had one set of alleles while the others had a different set of alleles. Probably due to the low density distribution of SNPs they were uninformative regarding the status of the other putative copy number variable regions.

Since our analytical pipeline for measuring the pig CNV landscape was developed in order to minimize the detection of somatic CNVs and false positives, and since the pig preliminary assembly contains high amounts of unfinished sequence and incorrectly mapped regions, our results are an obvious underestimate of the total number of CNVs in the sequences covered. As an example, when allowing copy number variants to be called in only one animal, there is an increase in the CNVR estimate from 37 to 165 (unpublished data).

It is also important to state that the sequences within a contig might be incorrectly assembled. Consequently, a CNVR detected at a certain position and in a certain orientation within a contig might have a different position and orientation within this contig. This could affect the performance of the calling algorithms. Future pig genome assemblies will shed light on this matter.

With the hypothesis that hundreds or maybe thousands of CNVs exist in the pig genome, this study is still an early step toward a more complete understanding of copy number variation within the pig species. Consequently, more studies are needed to fully understand the extent and functional roles of CNVs. Therefore, integration of previously gathered QTL and SNP (unpublished data) data for the pig families, the CNV data reported here, and a more comprehensive genome-wide CNV study in our group will certainly provide a framework for genetic association studies that will hopefully unravel the biological relevance of genetic variation and their effect upon important economic traits.
Methods

Oligonucleotide array CGH

A custom 385k tiling-path array CGH was designed (Nimblegen Systems, http://www.nimblegen.com) to cover the preliminary Sus Scrofa assembly for chromosomes 4, 7, 14, and 17, from the August 2007 release (http://www.sanger.ac.uk/Projects/S_scorfa/), which was the newest release at the time of the experiment. The tiling-path array covers the chromosome sequences of the August 2007 release up to the old chromosome endpoint coordinates of the previous release (April 2007).

The probe design fundamentals are described in the Nimblegen technical note http://www.nimblegen.com/products/hit/probe_design_2007_11_13.pdf. Briefly, highly repeated elements in the genome were repeat-masked with a strategy similar to the WindowMasker program [40]. Concerning uniqueness, probes having a unique genome sequence match were selected with SSAHA [41]. An isothermal format (Tm = 76 °C) [42] and probe length constraint between 50 and 75 bp were used for probe synthesis.

The probes were integrated into an array design using ArrayScribe™, which resulted in a design with a median probe spacing of 409 bp. The arrays were manufactured by maskless array synthesis technology and the oligonucleotides were synthesized on the arrays by photolithography [43–44].

Sample preparation

From a pig family-material comprising 14 boar founders, 700 sows, and about 12,000 offspring, 12 of the Duroc boar founders (A, B, C, D, E, G, H, J, K, L, M, and N) were selected to function as test animals. An unrelated Hampshire boar was selected as the common reference. We adhered to our institutional guidelines for the ethical use and treatment of animals in experiments.

Genomic DNA from boar N and the reference animal was isolated from lung/liver tissue by the use of Genomic-tip 100/G, the NCBI’s pig Unigene database release 34 [57], based mainly on the Sus Scrofa 6, May 2008 release assembly.

The NCBI’s pig Unigene database release 34 [57], based mainly on the primary Sus Scrofa assembly, the Refseq vertebrate mammalian mRNA database release 27 [57], and the Sus scrofa version 6, May 2008 release (http://www.sanger.ac.uk/Projects/S_scorfa/) was implemented to run on a DeCypher computer (http://www.timelogic.com).

The Tera-BLAST™N sequence similarity algorithm was used to query the CNVR sequences against the pig Unigene and the Refseq vertebrate mammalian mRNA databases. Hits were retained if they had an E-value ≤ 1e-15 and if their sequence aligned ≥ 95% (from Refseq) and ≥ 98% (from Unigene) with a CNVR.

About 7k ~120 bp sequences around SNPs ascertained in-house (unpublished data) from the animals queried in the CNV study were also queried against the Sus scrofa version 6, May 2008 release with an E-value ≤ 1e-15 and they were retained if they had a perfect hit in the chromosomes 4, 7, 14 and 17.

In order to check if the CNVRs overlapped any segmental duplication, Tera-BLAST™N was used to query the CNVRs sequences against the pig Unigene and the Refseq vertebrate mammalian mRNA databases. Hits were retained if they had an E-value ≤ 1e-15 and if their sequence aligned ≥ 95% (from Refseq) and ≥ 98% (from Unigene) with a CNVR.

The full data set from the oligo array CGH experiments has been submitted to GEO [59] under the accession ID GSE10753.

Quantitative Real Time PCR

Determination of copy number variation by quantitative real time PCR was performed using the Applied Biosystems 7900HT Sequence Detection System and analyzed with the SDS 2.2
software following the guidelines of the manufacturer (Applied Biosystems). The primers and probes (Universal ProbeLibrary Probes, Roche Applied Science) were designed using the ProbeFinder software from Roche Applied Science (https://www.roche-applied-science.com/sis/rtpcr/upl/acenter.jsp?id = 030000) and are available in the supplementary data file. A serial dilution of genomic DNA from the common reference animal was used as template for creating a standard curve for each primer pair. The copy number of each CNVR was normalized against a control region in the genome that does not vary in copy number between the pigs. All PCRs (10 μL) were run in triplicate in 1× TaqMan Universal PCR Master Mix, 100 nM of each primer, 250 nM probe and 10 ng of genomic DNA. PCRs were run as follows: 10 min at 95°C followed by 40 cycles at 95°C for 15 sec and 58°C for 10 sec.

References

1. Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392: 917–920.
2. Hart EA, Caccamo M, Harrow JL, Humphrey SJ, Gilbert JG, et al. (2007) Lessons learned from the initial sequencing of the pig genome: comparative analysis of an 8 Mb region of pig chromosome 17. Genome Biol 8: R168.
3. Thomas JW, Touchman JW, Blakesley RW, Bouchard GG, Beckston-Sternberg SM, et al. (2003) Comparative analyses of multi-species sequences from genome-sequenced animal genotypes. Nature 424: 780–793.
4. Rettenberger G, Klett C, Zeher U, Kunz J, Vogel W, et al. (1995) Visualization of the conservation of synteny between pigs and humans by heterologous chromosomal painting. Genomics 26: 372–378.
5. Tubb et al. (1996) Advances in Swine in Biomedical Research. New York: Plenum Press. 905 p.
6. Lai L, Kolber-Simonds D, Park K-W, Cheong H-T, Sun H, Garcia B, et al. (2005) Acute rejection is associated with antibodies to non-Gal antigens in baboons using Gal-knockout pig kidneys. Nat Med 11: 1295–1298.
7. Madsen LB, Thomsen B, Larsen K, Bendixen C, Holm LE, et al. (2007) A high-resolution map of segmental DNA copy number variation in the mouse genome. Nature 443: 444–454.
8. Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Bailey JA, et al. (2007) Genome-wide characterization and temporal expression profiling of presenilins in the developing porcine brain. BMC Neuroscience 8: 72.
9. Ifr et al. (2004) Detection of large-scale variation in the human genome. Nat Genet 36: 949–951.
10. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, et al. (2004) Large-scale copy number polymorphism in the human genome. Science 305: 525–529.
11. Sharp AJ, Thomsen B, Larsen K, Bentzcin C, Holm LE, et al. (2007) Molecular characterization and temporal expression profiling of presenilins in the developing porcine brain. BMC Neuroscience 8: 72.
12. Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, et al. (2005) Fine-scale structural variation of the human genome. Nat Genet 37: 727–732.
13. Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK (2006) A high resolution survey of deletion polymorphism in the human genome. Nat Genet 38: 75–91.
14. Hinds DA, Kloek AP, Jen M, Chen X, Frazer KA (2006) Common deletions and SNPs are in linkage disequilibrium in the human genome. Nat Genet 38: 82–95.
15. McCarthy SA, Hadnot TN, Perry GH, Sabeti PC, Zody MC, et al. (2006) Common deletion polymorphisms in the human genome. Nat Genet 38: 86–92.
16. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al. (2006) Global variation in copy number in the human genome. Nature 444: 444–454.
17. Cheng Z, Venture M, She X, Khaitovich P, Graves T, et al. (2005) A genome-wide comparison of recent chimpanzee and human segmental duplications. Genome Res 15: 1344–1356.
18. Niu X, Cox AJ, Liul Kelin JC, et al. (2002) Regional deletions and conveyed phenotypes. PLoS Genet 1: 627–633.
19. Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler E (2001) Segmental duplications: organization and impact within the current human genome project assembly. Genome Res 11: 1005–17.
20. Bailey J, Go Z, Clark R, Reinert K, Samonte R, et al. (2002) Recent segmental duplications in the human genome. Science 297: 1003–1007.
21. Nevo E, Frankel H, Poustovit R, Nevo J, Hattab S, et al. (2005) The conservation of synteny between humans and pigs by heterologous chromosomal painting. Genomics 26: 372–378.
22. Haddow ET, Tran DT, Li J, Jiang S, Balmain A, et al. (2005) Genomic segmental polymorphisms in inbred mouse strains. Nat Genet 36: 952–954.
23. Lupski JR, Stankiewicz P (2005) Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. PLoS Genet 1: 627–633.
24. Lupski JR (2005) Genomic rearrangements and sporadic disease. Nat Genet 39: 845–847.
25. Pinkel D, Albertson DG (2005) Comparative Genomic Hybridization. Annual Rev Genomics Hum Genet 6: 331–354.
26. Liu G, Van Tassell CP, Stankiewicz P, et al. (2007) Genomic deletion discovery in Caenorhabditis elegans by array comparative genomic hybridization. Genome Res 17: 537–547.
27. Maydan JS, Hribbote S, Edgley ML, Lau J, Selzer RR, et al. (2007) Efficient high-resolution deletion discovery in Caenorhabditis elegans by array comparative genomic hybridization. Genome Res 17: 537–547.
28. Sirbaugh KE, Forrest MS, Dunning M, Ingle CE, Beazley C, et al. (2007) Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes. Science 315(5813): 848.
29. Li J, Jiang T, Mao JH, Balmain A, Peterson L, et al. (2004) Genomic segmental polymorphisms in inbred mouse strains. Nat Genet 36: 952–954.
30. Lupski JR, Stankiewicz P (2005) Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. PLoS Genet 1: 627–633.
31. Lupski JR (2005) Genomic rearrangements and sporadic disease. Nat Genet 39: 845–847.
32. Islas et al. (2005) A genome-wide survey of deletion polymorphism in the human genome. Nat Genet 38: 75–91.
33. Hinds DA, Kloek AP, Jen M, Chen X, Frazer KA (2006) Common deletions and SNPs are in linkage disequilibrium in the human genome. Nat Genet 38: 82–95.
34. McCarthy SA, Hadnot TN, Perry GH, Sabeti PC, Zody MC, et al. (2006) Common deletion polymorphisms in the human genome. Nat Genet 38: 86–92.
35. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al. (2006) Global variation in copy number in the human genome. Nature 444: 444–454.
36. Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Bailey JA, et al. (2007) A comprehensive analysis of common copy-number variations in the human genome. Science 305: 525–529.
37. Nevan et al. (2005) Molecular characterization and temporal expression profiling of presenilins in the developing porcine brain. BMC Neuroscience 8: 72.
38. Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, et al. (2005) Fine-scale structural variation of the human genome. Nat Genet 37: 727–732.
39. Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK (2006) A high resolution survey of deletion polymorphism in the human genome. Nat Genet 38: 75–91.
40. Hinds DA, Kloek AP, Jen M, Chen X, Frazer KA (2006) Common deletions and SNPs are in linkage disequilibrium in the human genome. Nat Genet 38: 82–95.
41. Liu G, Van Tassell CP, Stankiewicz P, et al. (2007) Genomic deletion discovery in Caenorhabditis elegans by array comparative genomic hybridization. Genome Res 17: 537–547.
42. Sirbaugh KE, Forrest MS, Dunning M, Ingle CE, Beazley C, et al. (2007) Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes. Science 315(5813): 848.
43. Li J, Jiang T, Mao JH, Balmain A, Peterson L, et al. (2004) Genomic segmental polymorphisms in inbred mouse strains. Nat Genet 36: 952–954.
48. Smyth GK, Speed TP (2003) Normalization of cDNA microarray data. Methods 31: 265–273.

49. Smith ML, Marioni JC, Hardcastle TJ, Thorne NP (2006) snupCGH: Segmentation, Normalization and Processing of aCGH Data Users’ Guide. Bioconductor.

50. Olshen AB, Venkatraman ES, Lucito R, Wigler M (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5: 557–572.

51. Hupe P, Vranken N, Thiery JP, Radclyffe F, Barillot E (2004) Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. Bioinformatics 20: 3413–3422.

52. Fridlyand J, Snijders AM, Pinkel D, Albertson DG, Jain AN (2004) Hidden Markov models approach to the analysis of array CGH data. J Multivar Anal 90: 132.

53. Lai WR, Johnson MD, Kucherlapati R, Park PJ (2005) Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. Bioinformatics 21(22): 4084–91.

54. Luedtke R, Hooser C (2004) Hardware and software systems for accelerating common bioinformatics sequence analysis algorithms. Biosilico 2(3): 12–17.

55. Aeschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3): 403–10.

56. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K et al. (2007) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 35: D5–12.

57. Gorodkin J, Carrera S, Hedegaard J, Gilchrist MJ, Panitz F et al. (2007) Porcine transcriptome analysis based on 95 non-normalized cDNA libraries and assembly of 1,021,291 expressed sequence tags. Genome Biol 8: R45.

58. Olshen AB, Venkatraman ES, Lucito R, Wigler M (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5: 557–572.

59. Willenbrock H, Fridlyand J (2005) A comparison study: applying segmentation to array CGH data for downstream analyses. Bioinformatics 21(22): 4084–91.

60. Bruder C, Piotrowski A, Gijsbers A, Andersson R, Erickson S et al. (2008) Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. Am J Hum Genet 82(3): 763–71.

61. Piotrowski A, Bruder C, Andersson R, de Stahl TD, Menzel U et al. (2008) Somatic mosaicism for copy number variation in differentiated human tissues. Human Mutat 29(9): 1118–1124.

62. Rennie C, Noyes HA, Kemp SJ, Hulme H, Brass H et al. (2008) Strong position-dependent effects of sequence mismatches on signal ratios measured using long oligonucleotide microarrays. BMC Genomics 9: 317.

63. Emerson J, Cardoso-Moreira M, Bosevitz JO, Long M et al. (2008) Natural Selection Shapes Genome-Wide Patterns of Copy-Number Polymorphism in Drosophila melanogaster. Science 320(5883): 1629–31.