Detecting CNVs from Illumina Porcine 60K SNP Beadchip Data

Y. Ramayo[*], A. Castelló[*], E. Alves[†], R.N. Pena, A[+]. Mercadé[*], A.I. Fernández[†] and J.M. Folch[*]

Introduction

Copy numbers variants (CNVs) are DNA segments ranging from kilobases to several megabases in length with a variable number of repeats among individuals (Orozco *et al.* 2009). Recent studies at genome scale have revealed that 5% of the human genome is affected by CNVs (Zhang *et al*., 2009). This type of structural variation can influence gene expression and has been associated with Mendelian and complex genetic disorders and also affects variation in metabolic traits (Orozco *et al.,* 2009). Previous studies in pigs have detected CNVs using the Comparative Genomic Hybridization (CGH) technique (Fadista *et al.* 2008; Tang *et al.* 2010) in arrays designed to cover specific porcine chromosomes.

An alternative method for CNV detection is based on whole genome SNP genotyping arrays (Komura *et al.* 2006; Peiffer *et al.* 2006; Tuefferd *et al.* 2008), but has not yet been tested in swine species. A high density porcine SNP chip has recently been released by Illumina. This tool is a very valuable resource for studies of pig genetic variability and the molecular dissection of complex traits of economic importance. The Illumina's Porcine SNP60 Beadchip contains probes to genotype 62,621 SNPs covering the whole genome (Ramos *et al.* 2009). The average distance between SNPs is 34.6 kb in autosomes and 59.2 kb in chromosome X (on build 7).

The goal of this study was to detect CNVs on autosomal chromosomes in a pedigree from Iberian x Landrace cross using the Porcine SNP60 BeadChip data.

Material and methods

Animal material. We analyzed a total of 55 individuals (13 males and 42 females) belonging to four generations of the IBMAP cross (Perez-Enciso *et al.* 2000; Clop *et al.* 2003). This population was originated by crossing 3 Iberian (Guadyerbas line) boars with 31 Landrace sows. In this study we analyzed the 3 founder Iberian boards, 24 founder Landrace sows, 17 F1, 3 F2, and 8 backcross animals.

Genotyping. The 55 animals were genotyped with the Porcine SNP60 BeadChip (Illumina) using the manufacturer recommendations. Raw data were visualized and analyzed with the GenomeStudio software (Illumina).

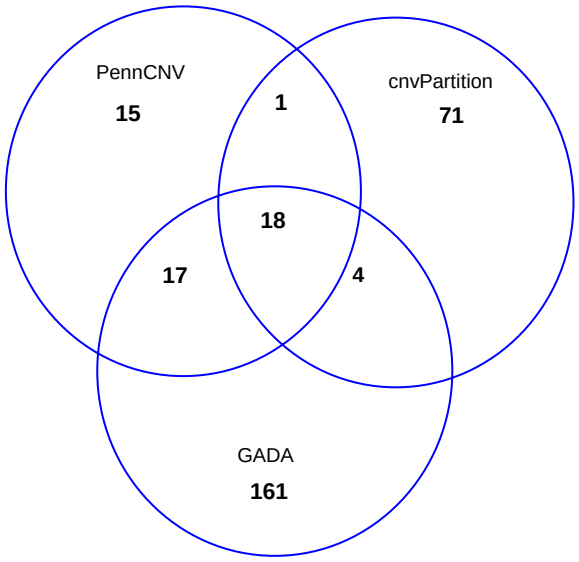
[*]Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain
[†] Departamento de Mejora Animal SGIT-INIA, 28040 Madrid, Spain
[+] Genètica i Millora Animal, IRTA Lleida, 25198 Lleida, Spain

Statistical analyses. We analyzed the Porcine SNP60 BeadChip data using three different softwares as recommended by (Winchester *et al.* 2009) to increase confidence in the analysis and limit the number of false positives. First, we used the Illumina's proprietary software GenomeStudio to check data quality and the cnvPartition v2.4.4 Analysis Plug-in for CNV detection. Then, we exported the signal intensity data of logRratio (LRR) and B allele frequency (BAF) to employ the R package for Genome Alteration Detection Algorithm (GADA) (Pique-Regi *et al.* 2008). This algorithm uses sparse Bayesian learning to predict CNV changes. Next, we used the PennCNV software that was originally developed for Illumina data analysis. This program integrates in a Joint-calling algorithm a Hidden Markov Model (HMM) with family relationships and signal intensities for parent-offspring trios (Wang *et al.* 2007). Finally, we compared the results from the three softwares and annoted the CNVs detected in at least two animals and recalled by at least two softwares.

Results and discussion

The initial number of CNVs called by each software was: cnvPartition (94), GADA (200) and PennCNV (51). Figure 1 summarizes the CNVs identified and compares the results obtained from the three programs.

Figure 1. Overlapping CNVs events from the three programs.



When a more stringent criterion for the selection of CNVs was used, namely, detected in at least two animals, and recalled by at least two programs, a total of 40 CNVs located in 15 of the 18 analyzed chromosomes passed this filter and showed Mendelian inheritance.

The number of CNVs overlapping the PennCNV and GADA programs is higher than between cnvPartition and any of the other programs. The percentage of CNV events confirmed by software was 70.6 % for PennCNV, 19.5 % for GADA and 24.5 % for cnvPartition. Similar results were reported by Winchester *et al.* (2009) comparing different algorithms for CNV detection, suggesting that PennCNV is more accurate in the prediction of CNVs for Illumina platform. This is likely explained by the uniqueness ability of this algorithm to consider family relationships and signal intensities for parent-offspring trios.

Size for the CNVs detected ranged from 49.5 to 1,450 kb, with a median size of 401.4 kb . They contain a total of 55 genes (Table 1): 16 CNVs with one gene and 7 CNVs with more than one. We could not find annotated genes in 17 CNVs and this can be partially explained by the incomplete annotation of the pig Sscrofa9 genome sequence assembly.

Table 1. Description by chromosome of the CNVs events detected.

Chr.	Number of CNV	Mean size (Kb)	Number of genes
1	8	1259.93	31
2	3	337.21	1
3	2	374.56	1
4	4	240.66	1
5	2	187.36	1
7	3	144.97	1
8	2	226.36	1
9	1	49.50	1
10	1	257.71	-
11	2	740.03	2
13	3	215.20	5
14	3	1452.04	3
15	3	230.90	4
16	2	200.997	1
17	1	103.96	2

One of the limitations of the Porcine SNP60 BeadChip for its use in the identification of CNVs is the relatively low density of SNPs in comparison to the human 1 M SNP arrays. Hence, only the largest CNVs are expected to be assessed with the Porcine SNP60 BeadChip. This may explain the differences in the minimum CNV length between our study (49.5 kb) and Fadista *et al.* (2008) (9.3 kb using the CGH technique).

Conclusion

We have assessed the ability of the Porcine SNP60 BeadChip to detect CNVs in an Iberian x Landrace cross. Three algorithms were compared (cnvPartition, GADA, and PennCNV) showing a higher accuracy for the PennCNV program. A total of 40 CNVs were identified with two algorithms and in at least two animals, showing a Mendelian inheritance pattern.

Acknowledgements

This work was funded by MICINN project AGL2008-04818-C03-01/GAN. Y. Ramayo was funded by a FPU PhD grant from Spanish Ministerio de Educación (AP2008-01450). We are indebted to Dr. M. Perez-Enciso for manuscript revision and to Dr. S. Scherrer (Illumina), Dr. R. Pique-Regi, MSc Yang Bin and MSc. A. Esteve for the help during the data analysis.

References

Clop A., Ovilo C., Perez-Enciso M., Cercos A., Tomas A., Fernandez A., Coll A., Folch J.M., Barragan C., Diaz I., Oliver M.A., Varona L., Silio L., Sanchez A. & Noguera J.L. (2003) . *Mamm Genome* **14**, 650-6.

Fadista J., Nygaard M., Holm L.E., Thomsen B. & Bendixen C. (2008). *PLoS One* **3**, e3916.

Komura D., Shen F., Ishikawa S., Fitch K.R., Chen W., Zhang J., Liu G., Ihara S., Nakamura H., Hurles M.E., Lee C., Scherer S.W., Jones K.W., Shapero M.H., Huang J. & Aburatani H. (2006). *Genome Res* **16**, 1575-84.

Orozco L.D., Cokus S.J., Ghazalpour A., Ingram-Drake L., Wang S., van Nas A., Che N., Araujo J.A., Pellegrini M. & Lusis A.J. (2009). *Hum. Mol. Genet.* **18**, 4118-29.

Peiffer D.A., Le J.M., Steemers F.J., Chang W., Jenniges T., Garcia F., Haden K., Li J., Shaw C.A., Belmont J., Cheung S.W., Shen R.M., Barker D.L. & Gunderson K.L. (2006). *Genome Research* **16**, 1136-48.

Perez-Enciso M., Clop A., Noguera J.L., Ovilo C., Coll A., Folch J.M., Babot D., Estany J., Oliver M.A., Diaz I. & Sanchez A. (2000). *J. Anim Sci.* **78**, 2525-31.

Pique-Regi R., Monso-Varona J., Ortega A., Seeger R.C., Triche T.J. & Asgharzadeh S. (2008). *Bioinformatics* **24**, 309-18.

Ramos A.M., Crooijmans R.P.M.A., Affara N.A., Amaral A.J., Archibald A.L., Beever J.E., Bendixen C., Churcher C., Clark R., Dehais P., Hansen M.S., Hedegaard J., Hu Z.-L., Kerstens H.H., Law A.S., Megens H.-J., Milan D., Nonneman D.J., Rohrer G.A., Rothschild M.F., Smith T.P.L., Schnabel R.D., Van Tassell C.P., Taylor J.F., Wiedmann R.T., Schook L.B. & Groenen M.A.M. (2009). *PLoS One* **4**, e6524.

Tang H., Li F., Finlayson H.A., Smith S., Lu Z., Langford C. & Archibald A.L. (2010). Plant & Animal Genomes XVIII Conference. January 9-13, 2010. Town & Country Convention Center, San Diego, CA.

Tuefferd M., Bondt A.D., Wyngaert I.V.D., Talloen W., Verbeke T., Carvalho B., Clevert D.-A., Alifano M., Raghavan N., Amaratunga D., Göhlmann H., Broët P. & Camilleri-Broët S. (2008). *Genes, Chromosomes and Cancer* **47**, 957-64.

Wang K., Li M., Hadley D., Liu R., Glessner J., Grant S.F., Hakonarson H. & Bucan M. (2007). *Genome Res* **17**, 1665-74.

Winchester L., Yau C. & Ragoussis J. (2009).*Briefings in Functional Genomics and Proteomics* **8**, 353-66.