# Detection of CNVs in layer chickens using 42K, 50K and 600K SNP chips

*A. Wolc[1,2], W. Drobik- Czwarno [1,3], J.E. Fulton[2], P. Settar[2], J. Arango[2], D. Lubritz[2], J.C.M. Dekkers[1]*

*[1]Department of Animal Science, Iowa State University, 806 Stange Road, 239E Kildee Hall, Ames, Iowa 50011, USA.*
*awolc@iastate.edu (Corresponding Author)*
*[2]Hy-Line International, 2583 240th Street, Dallas Center, Iowa 50063, USA.*
*[3]Department of Animal Genetics and Breeding, Warsaw University of Life Sciences, Ciszewskiego 8, 02-786 Warsaw, Poland.*

## Summary

The development of SNP chips has enabled rapid genotyping of hundreds of thousands of loci at a relatively low cost. In addition to providing SNP genotype information, copy number variation (CNV) can also be inferred from intensity data generated by the same chips. The aim of this study was to detect and describe CNVs in five lines of layer chickens using different SNP chips.

A total of 18,719 individuals from four pure lines and one commercial cross were genotyped using four different SNP chips (Illumina 42K, Affymetrix 600K, and two customized Affymetrix 50K chips). Analysis software Axiom® CNV Summary Tools and PennCNV were used to identify CNVs from Affymetrix chips and cnvPartition in Genome Studio was used to identify CNV's from the Illumina chip. The CNV regions (CNVR) within lines were defined using the BedTools software, through merging CNVs overlapping by at least 1 bp. CNVRs identified across all panels were selected with BedTools intersect to choose regions with highest confidence. Gene enrichment analysis was performed for genes that overlap identified CNVs to detect overrepresented biological processes and pathways.

The mean number of detected CNVs per individual varied depending on the population and size of the SNP set, from 0.50 with the 50K chip in one of the white layer lines up to 4.87 with the 600K chip in one of the brown layer lines. There were considerable differences in the number of detected CNVs between lines, probably due to breeding history. The length of detected CNVs ranged from 1.16 kb to 3.16 Mb. The mean length of detected CNVs was higher for the 42K and 50K chips than for the 600K chip, which is most likely a result of low detectability of shorter CNVs due to larger distances between markers in comparison to the 600K chip. The low frequency of CNVs detected on the 50K panels could also result from their custom design, which intentionally eliminated poorly clustering SNPs. Most of the detected CNVs had low population frequencies. In total CNVs were merged into 2687 CNVRs and overlapped 495.73 Mb of the genome. Intersecting CNVRs across all lines and panels yielded 4131 CNVRs from which 2139 were observed in at least two individuals.

In conclusion, commonly used SNP chip platforms and analysis using relevant software can be used to identify CNVs in commercially relevant chicken layer lines. The number of detected CNVs and their length depends on the population and density of SNP on the chip.

*Keywords: CNV, SNP chip, layer chicken, association study*

## Introduction

Copy number variations (CNVs), defined as large scale insertions, duplications or deletions of DNA sequence, are an important source of genetic variance. Currently known chicken CNV regions (CNVR) encompass approximately 8.3% of the chicken genome or 9.6% of the ordered assembly (Wang and Bayes, 2014). Differences observed between chromosomes suggest that CNVs are more frequent on micro-chromosomes than on macro-chromosomes (Yi et al., 2014).

Copy number variations can be detected by a number of methods, including array comparative genome hybridization (aCGH), sequencing and SNP arrays. Although SNP arrays are primarily designed for SNP genotyping, they can also be used to detect CNVs due to abnormal hybridization of SNP probes that are located within CNVs. Use of the 60K SNP chip in chickens resulted in a low frequency of detected CNVs (Jia et al., 2013; Rao et al., 2016). In the study by Rao et al. (2016), 1875 CNVs were detected across 554 commercial cross individuals, which were distributed over 383 independent CNVRs and covered 41 Mb (4%) of the genome. Although the highest number of CNVs can be identified using sequence data (Yi et al., 2014; Fan et al., 2013; Yan et al., 2015), it was shown that the 600K SNP array (Kranis et al., 2013) can also yield significant CNV detection (Yi et al., 2015; Gorla et al., 2017).

The aim of this study was to detect and describe CNVs in five lines of layer chickens using different SNP chips. Results were compared across genotyping panels and lines. In order to assess the possible functional impact of the detected CNVs, genes that overlap the CNVR were identified. Gene enrichment analysis was performed to detect overrepresented biological processes and pathways.

## Material and Methods

DNA from 18,719 individuals from 4 pure lines and one commercial cross (Table 1) were provided by Hy-Line International. There was no overlap of individuals genotyped on different chips.

*Table 1. Summary information on the available material*

| Material | Genetics | SNP Chip | N |
|---|---|---|---|
| White commercial cross | Hybrids | 600K Affymetrix chip | 806 |
| White layers pure lines | Line W1 | 600K Affymetrix chip | 253 |
| | | 50K-W Affymetrix chip* | 3350 |
| | Line W2 | 600K Affymetrix chip | 748 |
| | | 50K-W Affymetrix chip* | 3215 |
| | | 50K-B Affymetrix chip* | 2401 |
| Brown layers pure lines | Line B1 | 600K Affymetrix chip | 241 |
| | | 50K-B Affymetrix chip* | 5908 |
| | Line B2 | 42K Illumina Infinium chip | 1797 |
| Summary | N=5 | N=4 | 18719 |

* Two customized 50K chips were used: designed for white layer lines (50K-W) and designed

for brown layer lines (50K-B)

Genotyping and CNV calling was performed within plate due to the potential for large differences in signal intensity between plates. For Affymetrix chips, quality control with a minimum DishQC of 0.82 and a minimum call rate of 97% was used. Axiom™ Analysis Suite and CNV Summary Tools were used for genotyping and for extracting the log R Ratio (LRR) and B allele frequency (BAF). The Penncnv 1.0.3 software integrating Hidden Markov Model, was used to call CNVs based on data from the 600K Affymetrix chip and from the line specific 50K Affymetrix chips. This algorithm incorporates multiple sources of information to call CNVs, including: LRR, BAF at each SNP, the distance between SNPs, and the population frequency of the B allele (PFB). Data from the 42K Illumina panel were processed in Genome Studio using Genotyping module and the cnvPartition CNV Analysis Plugin.

Individual-based CNV calling was performed in the Penncnv software using the -test option. To adjust for genomic waviness (Diskin et al, 2008), the -gcmodel option with the chicken GC content file was used. PFB files were compiled separately for each panel in Penncnv for the individuals listed in Table 1. For filtering, standard deviation of LRR $\leq 0.35$, BAF drift $< 0.01$ and waviness factor $\leq 0.04$ were used. Only CNVs consisting of 3 (50K and 42K), 5 (600K) or more consecutive SNPs were used in the analysis and individuals with more than 30 called CNVs were excluded (N=1013). CNVs were identified only on autosomes (1–28) as Penncnv callings for the sex chromosomes were unreliable and difficult to interpret. CNVRs were defined as CNVs that overlapped across individuals or lines. Genes overlapping CNVRs were identified with the Ensembl BioMart webtool based on the Galgal4 assembly and Ensembl Genes 84 database. Analysis of overrepresented GO terms and pathways was performed using PANTHER Classification System.

## Results and discussion

The mean number of CNVs per individual ranged from 0.50 on the 50K-W panel in line W2 up to 4.87 for the 600K panel in line B1. The highest number of CNVs per individual was identified in the hybrid commercial cross and in line B1, while line W2 had the lowest number of CNVs, which could be caused by higher genetic variability in the commercial cross and brown lines in comparison to pure white lines. Line W2 is characterized by a higher level of inbreeding, which could explain the lower number of CNVs detected with all panels. The length of detected CNVs ranged from 1.16 kb to 3.16 Mb (Table 2). The mean length of detected CNVs was higher for the 42K and 50K chips than for the 600K chip, most likely the result of low detectability of shorter variants due to the greater distance between markers for the smaller chips. The length of CNVs was calculated as the distance from first to the last SNP included in the CNV and therefore could slightly underestimate the true size of CNV.

*Table 2. Summary of identified CNVs called for all lines genotyped on different panels.*

| Population | SNP chip | N | N pass QC | Number of CNVs | Length of CNV, range [kb] | Mean N of CNVs per bird | Mean CNV length [kb] |
|---|---|---|---|---|---|---|---|
| Line W1 | 50K-W | 3350 | 3308 | 2053 | 1.81 – 955.67 | 0.62 | 88.09 |
| | 600K | 253 | 252 | 774 | 1.16 – 271.85 | 3.08 | 25.91 |
| Line W2 | 50K-W | 3215 | 3172 | 1575 | 1.94 – 1294.56 | 0.50 | 76.87 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 50K-B | 2401 | 2253 | 1844 | 1.35 – 1492.99 | 0.82 | 111.30 |
| | 600K | 748 | 714 | 1415 | 1.50 – 428.70 | 1.98 | 37.15 |
| Cross | 600K | 806 | 769 | 2269 | 1.44 – 1116.22 | 2.94 | 31.06 |
| Line B1 | 50K-B | 5908 | 5284 | 6203 | 1.67 – 3160.23 | 1.17 | 215.99 |
| | 600K | 241 | 238 | 1158 | 1.19 – 663.12 | 4.87 | 24.86 |
| Line B2 | 42K | 1797 | 1716 | 2250 | 1.43 – 1658.27 | 1.31 | 90.94 |
| Summary | | 18719 | 17707 | 19541 | 1.16 – 3160.23 | 1.10 | 51.07 |

Merging CNVs across all samples and lines gave 2687 regions with total length equal to 495.73 Mb overlapping 47.12% of the genome. Intersecting CNVRs across all lines and panels gave 4131 CNVRs. The CNVRs that were identified in at least two individuals were selected for further analysis (N=2139). The total length of these CNVRs was equal to 117.26 Mb, which corresponds to 12.7% of the genome. In total, 29.8% of these CNVRs were overlapped by 3510 genes included Ensembl Genes 84 database, of which 2994 genes mapped to Panther biological categories. GO enrichment analysis of these genes revealed several significant terms involved in antigen processing and presentation, cellular defence response, cell adhesion and chromatin organization, which may represent biological processes that are influenced by CNVs.

## Conclusion

Commonly used SNP chip platforms with analysis using relevant software identified the presence of CNVs segregating in commercially relevant chicken layer lines. The number of detected CNVs and their length depends on the population and density of SNP on the chip.

## List of References

Diskin, S.J., Li, M., Hou, C., Yang, S., Glessner, J., Hakonarson, H., Bucan, M., Maris, J.M. & K. Wang, 2008. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. Nucleic Acids Res. 36(19): e126.

Fan, W.L., Ng, C.S., Chen, C.F., Lu, M.Y., Chen, Y.H., Liu, C.J., Wu, S.M., Chen, C.K., Chen, J.J., Mao, C.T., Lai, Y.T., Lo, W.S., Chang, W.H. & W.H. Li, 2013. Genome-Wide Patterns of Genetic Variation in Two Domestic Chickens. Genome Biol. Evol. 5(7):1376-92.

Gorla, E., Cozzi, M.C., Román-Ponce, S.I., Ruiz López, F.J., Vega-Murillo, V.E., Cerolini, S., Bagnato, A. & M.G. Strillacci, 2017. Genomic variability in Mexican chicken population using copy number variants. BMC Genetics 18: 61.

Jia, X., Chen, S., Zhou, H., Li, D., Liu, W. & N. Yang, 2013. Copy number variations identified in the chicken using a 60K SNP BeadChip. Anim. Genet. 44:276-284.

Kranis, A., Gheyas, A.A., Boschiero, C., Turner, F., Yu, L., Smith, S., Talbot, R., Pirani, A., Brew, F., Kaiser, P., Hocking, P.M., Fife, M., Salmon, N., Fulton, J., Strom, T.M., Habrer, G., Weigend, S., Preisinger, R., Gholami, M., Quanbari, S., Simianer, H., Watson, K.A., Woolliams, J.A. & D.W. Burt, 2013. Development of a high density 600K SNP genotyping array for chicken. BMC Genomics 14: 59.

Rao, Y.S., Li, J., Zhang, R., Lin, X.R., Xu, J.G., Xie, L., Xu, Z.Q., Wang, L., Gan, J.K., Xie, X.J., He, J. & X.Q. Zhang, 2016. Copy number variation identification and analysis of the

chicken genome using a 60K SNP BeadChip. Poult. Sci. 95(8):1750-6.
Wang, X. & S. Byers, 2014. Copy Number Variation in Chickens: A Review and Future Prospects. Microarrays 3: 24-38.
Yan, Y., Yang, N., Cheng, H.H., Song, J. & L. Qu, 2015. Genome-wide identification of copy number variations between two chicken lines that differ in genetic resistance to Marek's disease. Genomics 16: 843.
Yi, G., Qu, L., Liu, J., Yan, Y., Xu, G. & N. Yang, 2014. Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing. BMC Genomics. 7:962.
Yi, G., Qu, L., Chen, S., Xu, G. & N. Yang, 2015. Genome-wide copy number profiling using high-density SNP array in chickens. Anim. Genet. 46:148-157.
Zhang, H., Du, Z.Q., Dong, J.Q., Wang, H.X., Sh, H.Y., Wang, N., Wang, S.Z. & H. Li, 2014. Detection of genome-wide copy number variations in two chicken lines divergently selected for abdominal fat content. BMC Genomics 15: 517.