HETEROLOGOUS HYBRIDIZATION USING THE HUMAN EXOME - A MOLECULAR TOOL TO TARGET AND IDENTIFY MAJOR GENES

Iannucelli N.[1], Cabau C.[2], Sarry J.[1], Bouchez O.[3], Billon Y.[4], Riquet J.[1], Allain D.[1], Demars J.[1]*

[1]GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326, Castanet Tolosan, France
[2]Sigenae, INRAE, F-31326, Castanet Tolosan, France
[3]GenoToulGenomicsPlatform (GeT-PlaGe&Bioinfo), INRAE, F-31326, Castanet Tolosan, France
[4]GenESI, INRAE, doi:10.15454/1.5572415481185847E12, F-17700 Surgères, France
*Corresponding author: julie.demars@inrae.fr

ABSTRACT

Identifying causal mutations responsible of phenotypes in a time and cost-effective manner remains a challenge still for all species. Although the portfolio of next-generation sequencing methodologies offers a broad range of opportunities nowadays, having the right and suitable molecular genetic tools in non-model organisms is often an issue. For the European rabbit (*Oryctolagus cuniculus*), considered as an animal model for experimental research, the OryCun2.0 reference genome was released in 2009 and the Affymetrix AxiomOrcun SNP Array was made available in 2016. This novel beadchip gives the possibility to develop conveniently genome-wide association studies (GWAS) in rabbits without however freeing itself from the subsequent steps of causal mutation characterization. Whole exome sequencing (WES) appears as a direct and suitable strategy to target and sequence variants all at once. Unfortunately, no exome enrichment tool exist for rabbits. Here, we proposed to capture the rabbit exome using the marketed human panel since human and rabbit genomes are closely related on the phylogenetic tree of species. We performed heterologous hybridization for 48 rabbits from 7 different populations and generated 2 billion sequencing reads for the whole dataset. The quality of the heterologous hybridization allowed the capture of the rabbit exome and enabled a dataset of 40,000 confident variants that specifically target both protein-coding and extended exons. This method also presents a unique opportunity to investigate both *(i)* the genetic diversity of different rabbits' breeds and *(ii)* the mendelian determinism of specific phenotypes in rabbits without resorting to costly customized tools.

Key words: human, exome, capture, sequencing, variants.

INTRODUCTION

Genome-wide association studies (GWAS) are widely used to identify loci associated with phenotypic traits. However, GWAS require *(i)* molecular genomic tools that are not available for a broad range of species and *(ii)* important effort for fine-mapping to identify causal mutations. Whole exome sequencing (WES) occured as a direct strategy to target and sequence variants all at once since especially for major genes for which causal mutations are often non-synonymous polymorphisms (Warr et al. 2015).

WES focuses on pre-specified target regions, generally the exonic part of protein-coding regions. While exome enrichment tools are provided by manufacturers for various species such as human and dogs, specific designs are requested for other model organisms. To overcome this issue, heterologous capture have been performed using whole human exome enrichment tools to analyze non-human primate exomes (Vallender 2011). Indeed, the best homology between distinct genomes correspond to protein-coding regions with a moderate to high degree of conservation given the phylogenetic distance.

To determine whether human and rabbit genomes are sufficiently homologous for exome capture, we applied the human exome kit to rabbit DNA samples in order to potentially investigate the mendelian determinism of specific phenotypes in rabbits without resorting to costly customized tools.

MATERIALS AND METHODS

Animals and experimental design

The experimental design includes 48 animals from 7 rabbit populations that were all bred in the experimental INRA farm (UE GenESI, Surgères, France), in accordance with the national regulations for animal care and use of animals in agriculture. The different breeds have been selected for carrying major genes for fur quality or coloration traits. Castor strains include both wt (n=7) and diluted (n=6) animals, Chinchilla breeds include wt (n=7) and diluted (n=7) rabbits as developed in Demars et al., 2018. In addition, Angora strains (Rafat, Allain, and De Rochambeau 2009) include wt (n=6), Angora rabbits as described in Mulsant, de Rochambeau and Thébault, 2004 (n=6) and recent Angora (n=8).

DNA extraction and quantification

Ear punch biopsies (or blood sample) were collected and genomic DNA was extracted from samples with a home-made protocole: proteinase K lysis following by salt-based DNA extraction and ethanol precipitation. Briefly, ear punch biopsies were digested at 56°C for 3h using a 1ml solution including 10mM Tris HCl, 25mM EDTA pH=8, 0,5% SDS and 0.2 mg proteinase K. After overnight incubation at 37°C, 1/3 volume of saturated (6M) NaCl were added and slightly mixed before a centrifugation step (30 min at 4°C and 21000g). Supernatant was mixed with 2,5 volumes of 100% ethanol. DNA was retrieved and resuspended in classic buffer for 1 hour at 60°C before an overnight resuspension at 37°C. Total genomic DNA was quantified using both the Nanodrop 8000 (ND8000LAPTOP, Thermo Fisher Scientific, USA) and the Qubit2.0 (Q32866, Life Technologies, USA).

Capture and Sequencing Analyses

Once quality and quantity of DNA were validated, libraries preparation and exomes enrichment were performed using the Nextera® Rapid Capture Exome kit (Illumina, USA) that includes 214,405 human coding exons (covered by 425,554 probes) for a specific coverage of 37 Mb. The protocole was followed as recommended by the manufacturer. Briefly, tagmentation and PCR amplification allow fragments preparation before the samples pooling (by 6) and the denaturation of double-stranded DNA library. Hybridizations (first at 55°C and second at 58°C) to target regions is performed thanks to biotinylated probes and the use of streptavidin beads allows the enrichment step. A next step of elution is necessary before sequencing. Library quality was assessed using an Advanced Analytical Fragment Analyser (Fragment Analyzer™) and libraries were quantified by QPCR using the Kapa Library Quantification Kit. Sequencing was performed on an Illumina HiSeq3000 with the Illumina Reagent Kits in paired end 2x150nt mode (sequencing service provider using Illumina HiSeq3000 was the GeT-PlaGe core facility, INRA Toulouse). The whole dataset of 48 samples was sequenced on 4 lanes.

Bioinformatic Analysis

The reference genome of rabbit was used in the present study, with a size of 2,734.47 Mb (ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000003625.1_OryCun2.0). Reads were aligned with BWA-MEM version 0.7.12-r1039 (Li 2013) to the OryCun2.0 reference genome. We used picard -tools-2.1.1 (https://broadinstitute.github.io/picard/) to mark and remove duplicates and we applied GATK version 3.7 realignment (-RealignerTargetCreator and -IndelRealigner), base quality score recalibration (-BaseRecalibrator) using a known dataset (Bertolini et al. 2014) and performed SNP and InDel discovery and genotyping (-HaplotypeCaller) across all samples simultaneously using standard hard filtering parameters according to GATK Best Practices recommendations. Annotation of variants and filtering of interesting variants was performed using the SNPEff (Cingolani, Platts, et al. 2012) and SNPSift (Cingolani, Patel, et al. 2012) programs, respectively.

RESULTS AND DISCUSSION

The quality of the heterologous hybridization allows the capture of the rabbit exome

To evaluate the effectiveness of the Nextera Exome Enrichment Kit with rabbit DNA, we firstly analyzed the sequencing depth and coverage. A dataset of 2 billions of sequences was generated in total, with approximately 45 millions of reads contributed from each sample. This corresponds to a mean coverage of 200x considering a number of exons similar between human and rabbit species. Despite the expected 200x sequencing depth, it occurred than a significant number of reads (30% per sample) did not overlap exons or extended exons suggesting *(i)* a moderate effectiveness of the heterologous capture or *(ii)* a poor annotation of the rabbit genome. However, when we considered the mean coverage per transcript, a homogenous representation of the rabbit exome was obtained as shown on Figure 1. Many rabbit transcripts are covered with several probes and show a high mean coverage for all samples. This suggests that the heterologous exon capture worked quite well.

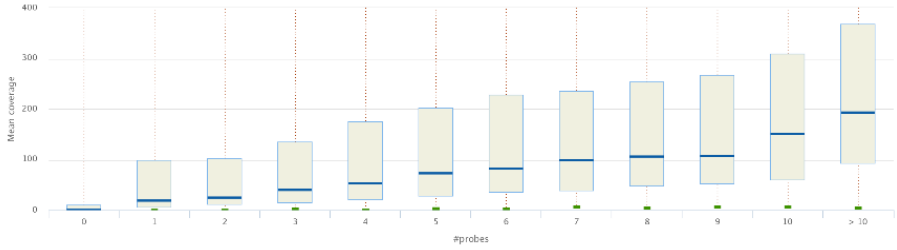

Figure 1: Transcript mean coverage in all samples. The mean coverage of one transcript is the mean of the average coverage across all its exons. The y axis represents the mean coverage and the x axis represents the number of probes.

A Variants discovery targeting rabbit exons allows identification of known mutations

To assess the rabbit exome as a pertinent molecular tool for genetic studies in rabbits, we evaluated different criteria such as the depth coverage (DP), the quality (QUAL) and the heterozygosity for the discovery of genetic variants. In a first row of analysis, we considered a common DP of markers in all samples. The number of identified polymorphisms varied from 25,000 to 57,000 depending on the DP, DP30 is a correct coverage for heterozygous individual detection (Figure 2A). A homogenous density of variants on both autosomes (Figure 2B) and scaffolds (data not shown) was obtained. In addition, previously identified loci associated with different phenotypes were captured and sequenced as shown on Figure 2C with the non-extension mutation (e allele) of the *MC1R* gene corresponding to a 30 bp deletion (Fontanesi et al. 2006). Altogether, these results provide support for future studies in rabbit genetics.

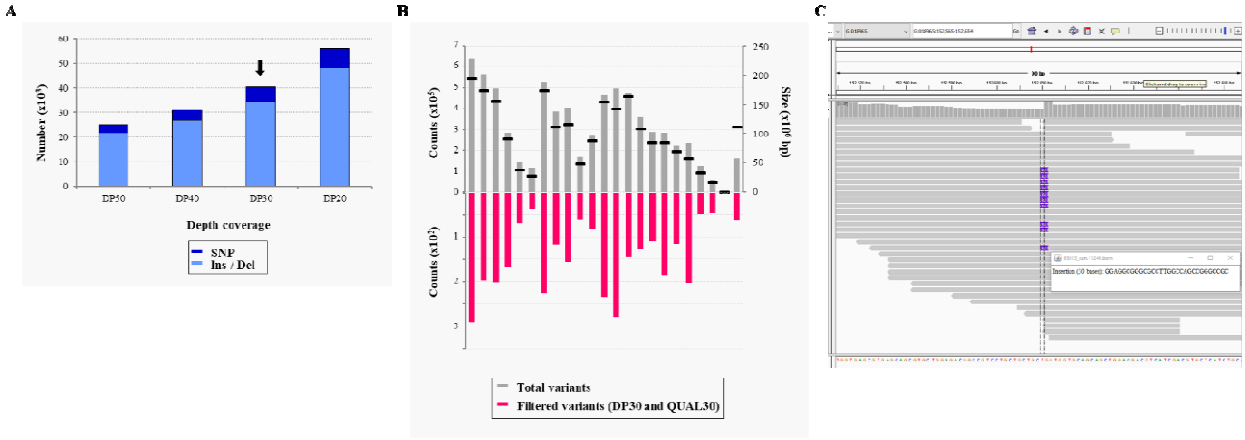

Figure 2: Genetic variants estimation in the whole dataset. A- Genetic variants detection depending on depth coverage. B- Distribution of polymorphisms along the rabbit chromosomes. C- IGV (Integrative Genome Viewer) screenshot of one *MC1R* mutation (e = c304_333del30) (Fontanesi et al. 2006).

CONCLUSIONS

To determine whether the high-throughput variants dataset was a pertinent molecular genetic tool for rabbits, we evaluated the genetic diversity of the 7 rabbit populations *via* a principal component analysis (PCA). The set of 40,000 polymorphisms, based on DP30 and QUAL30 criteria, perfectly discriminated the different rabbit breeds (Figure 3). In addition, the heterologous hybridization using the human exome might be a pertinent alternative to the Affymetrix AxiomOrcun SNP Array for genetic studies to a straightforward identification of causal mutations in coding regions. As example, all known rabbit mutations for coat color, located in *ASIP*, *MC1R*, *TYR*, *TYRP1*, *MLPH* genes, are covered by this tool..

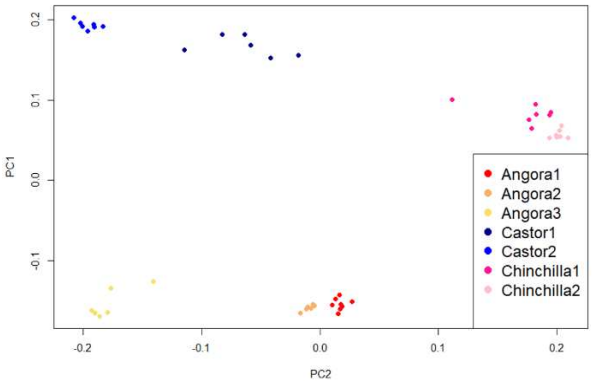

Figure 3: Principal component decomposition of the high-throughput variants dataset.

ACKNOWLEDGEMENTS

We thank G. Auvinet, B. Bayle and L. Weidman, who carefully looked after the animals (GenESI, INRA, 2018. Pig phenotyping and Innovative breeding facility, doi:10.15454/1.5572415481185847E12). The Genetic Animal division from INRAE and the GenPhySE laboratory have funded this study. We thank the Get-PlaGe facility and all those who have contributed to this work for discussion and support.

REFERENCES

Bertolini, F., G. Schiavo, E. Scotti, A. Ribani, P. L. Martelli, R. Casadio, and L. Fontanesi. 2014. “High-Throughput SNP Discovery in the Rabbit (Oryctolagus Cuniculus) Genome by next-Generation Semiconductor-Based Sequencing.” *Animal Genetics*.

Cingolani, Pablo, Viral M. Patel, Melissa Coon, Tung Nguyen, Susan J. Land, Douglas M. Ruden, and Xiangyi Lu. 2012. “Using Drosophila Melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift.” *Frontiers in Genetics* 3:35.

Cingolani, Pablo, Adrian Platts, Le Lily Wang, Melissa Coon, Tung Nguyen, Luan Wang, Susan J. Land, Xiangyi Lu, and Douglas M. Ruden. 2012. “A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff.” *Fly*.

Demars, Julie, Nathalie Iannuccelli, Valerio Joe Utzeri, Gerard Auvinet, Juliette Riquet, Luca Fontanesi, and Daniel Allain. 2018. “New Insights into the Melanophilin (MLPH) Gene Affecting Coat Color Dilution in Rabbits.” *Genes*.

Fontanesi, L., M. Tazzoli, F. Beretti, and V. Russo. 2006. “Mutations in the Melanocortin 1 Receptor (MC1R) Gene Are Associated with Coat Colours in the Domestic Rabbit (Oryctolagus Cuniculus).” *Animal Genetics*.

Li, Heng. 2013. “[Heng Li - Compares BWA to Other Long Read Aligners like CUSHAW2] Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM.” *ArXiv Preprint ArXiv*.

Mulsant, P., H. de Rochambeau, and RG Thébault. 2004. “A Note on Linkage between the Angora and Fgf5 Genes in Rabbits.” *World Rabbit Science*.

Rafat, S. A., D. Allain, and H. De Rochambeau. 2009. “Genetic Description of a Divergent Selection Experiment in Angora Rabbits with Overlapping Generations.” *Journal of Animal Breeding and Genetics*.

Vallender, Eric J. 2011. “Expanding Whole Exome Resequencing into Non-Human Primates.” *Genome Biology*.

Warr, Amanda, Christelle Robert, David Hume, Alan Archibald, Nader Deeb, and Mick Watson. 2015. “Exome Sequencing: Current and Future Perspectives.” *G3: Genes, Genomes, Genetics*.