540. Allele specific expression as an indication of ploidy in pig IPECJ2 and chicken SL-29 cell lines

J. de Vos[1*], R.P.M.A. Crooijmans[1], M.F.L. Derks[1], S.L. Kloet[2], M.A.M. Groenen[1] and O. Madsen[1]

[1]Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands; [2]Leids Universitair Medisch Centrum, the Netherlands; jani.devos@wur.nl

Abstract

Cell lines are useful for investigating traits of interest e.g. intestinal absorption, feed efficiency and immunity in farm animals. We earlier investigated an intestinal cell line in pig and a fibroblast cell line in chicken and found chromosomal abnormalities by whole genome sequence (WGS) data analysis. Results from RNA-seq allele-specific expression analysis (ASE) in 4 cell lines showed aneuploidy in some chromosomes. In this paper we show that RNA-seq can be used to detect whole/partial chromosomal abnormalities based on ASE analysis.

Introduction

Pig and chicken are the primary sources for meat production worldwide and are also important for use as biomedical models and to study embryonic development. Cell lines are a valuable tool for gaining insight into genomic architecture and regulatory regions of genomes. Cell lines are derived from specific tissues of a species, and can either continue to divide endlessly or perish after a specific number of divisions (passages) (Verma *et al.*, 2020). Immortalized cell lines (i.e. that can be grown indefinitely) often show aneuploidy (presence of an abnormal number of chromosomes in a cell) or heteroploidy which is most pronounced in cancer cell lines (Verma *et al.*, 2020; Molina *et al.*, 2021). Animal cell lines from pig or chicken have proven useful to obtain insight in e.g. intestinal transport and immune function (Nossol *et al.*, 2015).

The phenotype of an individual is determined by gene expression which is a process where information from genes, encoded within DNA, is translated into proteins through RNA (Hartwell *et al.*, 2015). The phenomenon of unequal expression between alleles in diploid cells caused by either genetic variation or epigenetic regulation (cis-regulated gene expression) is known as ASE. Furthermore, non-haploid chromosomes show imbalances in gene transcriptional activity, implying that chromosomal abnormalities can influence gene expression of different alleles, resulting in ASE. Our aim was to investigate the usability of RNA-seq data to identify whole chromosomal abnormalities, with the pig IPEC-J2 and chicken SL-29 cell lines as model.

Materials & methods

Cell lines. The pig IPECJ67 cell line (IPECJ2 cells grown for 67 passages) and chicken SL-29 cell line (grown for 4 passages) used were obtained from the cell repositories at DSMZ (https://www.dsmz.de/collection/catalogue/details/culture/ACC-701) and ATCC (https://www.lgcstandards-atcc.org/products/all/CRL-1590.aspx?geo_country=nl#generalinformation) respectively. The pig cell line is derived from intestinal epithelial cells while the chicken cell line is derived from embryonic fibroblast cells.

Data analysis. Pig (Sus Scrofa 11.1) and chicken (Gallus gallus GRCg6a) reference genomes, together with ENSEMBL annotations (Sus Scrofa 11.1 release 103 & Gallus gallus GRCg6a release 94) were utilized for all data analyses of our study.

Whole genome sequencing. Whole genome sequences were trimmed using Sickle v1.33 (https://github.com/najoshi/sickle) in paired end mode, followed by alignment (bwa mem v0.7.15 (Li., 2013)) of the

R.F. Veerkamp and Y. de Haas (eds)

DOI: 10.3920/978-90-8686-940-4_540, © J. de Vos *et al.* 2022

trimmed reads, and removal of duplicates (Samblaster v0.1.26 (Faust and Hall, 2014)). Mate coordinates were added using samtools (v1.9, Li *et al.,* 2009). Single nucleotide variants (SNVs) were called using FreeBayes (v1.3.1), thereafter the read support ratio was assessed for heterozygous variants within the VCF file.

RNA sequencing. RNA-seq data (pig IPECJ67 and chicken SL-29) were trimmed for adapters, and minimum length using TrimGalore v0.6.4 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). FastQC v0.11.9 (Andrews *et al.*, 2010) was implemented to evaluate the quality of the data. RSEM (Li and Dewey, 2011) was used for alignment of the trimmed reads, as well as gene quantification (STAR v2.7.3a as aligner) (Dobin *et al.*, 2013). Average gene expression level per chromosome was calculated and plotted using a custom python script with the Seaborn package. Additional raw RNA-seq data from the PRJNA610529 project was downloaded from ENA. This data comprises a pig Jejunum organoid sample (12 weeks (SAMN14300021), a 5 week old pig Jejunum tissue sample (SAMN14300018), cell lines IPECJ87, an IPECJ2 cell line grown for 87 passages (SAMN14300016), and IPECJ91 an IPECJ2 cell line grown for 91 passages (SAMN14299997). We trimmed, aligned, and completed gene quantification of this data following the same procedure as the above for the IPECJ67 cell line. These samples were used to compare the average gene expression levels per chromosome in the tissue, organoid, and cell lines to the IPECJ67 cell line. Initial ASE analysis was completed with variant calls from WGS data (FreeBayes, Garrison and Marth, 2012) and aligned reads from RNA-seq using GATK ASEReader (McKenna *et al.,* 2010). Variant calling from RNA-seq alignments for SL-29, IPECJ67, IPECJ87, and IPECJ91 was completed using FreeBayes and thereafter allele specific expression analysis was completed using GATK ASEReader implementing the bam and VCF files. Results from the ASE analysis was plotted using Seaborn package in python.

Results

Chromosomal abnormalities within the cell-line genomes. In a previous study of the cell lines IPECJ67 and SL-29 we analysed the structure of the genome in multiple aspects using whole genome sequence data (de Vos *et al.* manuscript in preparation). In pig IPECJ67 we observed that chromosome 16 is diploid and chromosome 17 is triploid based on the frequency of the alleles (data not shown). Additionally, aneuploidy and structural variations were observed in many other chromosomes in the IPECJ67 cell line as well as chicken SL-29 (e.g. chromosome 20 is tetraploid).

Allele specific expression. RNA-seq data provides insight into gene expression levels at a chromosome-wide level. We investigated the expression of genes and transcripts in the pig IPECJ67, IPECJ87, IPECJ91 cell lines, jejunum intestinal tissue and jejunum derived organoid sample (data not shown). We found elevated gene expression on chromosome 17 in the cell lines compared to organoid and tissue whereas for the diploid chromosomes 15 and 16 similar gene expression levels across tissue, organoids and cell lines, with slightly higher expression levels in cell lines were observed. Investigating the allele specific expression of the IPECJ67 cell line (Figure 1) shows that the frequency of the allele expression displays a pattern that confirms chromosome 17 as triploid and chromosomes 15 and 16 as diploid.

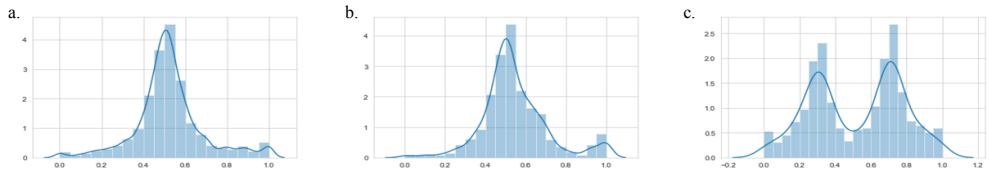

Figure 1. Allele specific expression of chromosomes 15(a), 16(b), 17(c) for the IPECJ67 cell line supports the ploidy level of the three chromosomes – allele expression of ~0.5 for diploids and allele expression of ~0.33 and ~0.7 for triploids, respectively.

Further investigation of the ASE and chromosomal abnormalities of the IPECJ2 derived cell lines cultured for a longer time also showed triploidy for chromosome 17 (Figure 2 c,f) in both cell cultures. Interestingly, chromosome 16 was observed as triploid (Figure 2 b,e) for these cell cultures, which is different from IPECJ67 which was diploid for chromosome 16 (Figure 1b).

Similar patterns of allele distribution for aneuploid chromosomes detected using ASE are observed on other chromosomes in pig IPECJ67 and in chicken SL-29 (data not shown). Lastly, we called the genome variation from the RNA-seq data and repeated the ASE analysis based on the RNA-seq called genome variation. This resulted in the same ASE results as using the variation called from the genome sequencing data (data not shown). Thus, genome sequencing is not required to detect genome aberrations based on ASE.

Discussion

Detecting chromosome abnormalities in e.g. cell lines is important as such abnormalities may influence the results obtained in cell line assays, especially if genes involved in the process studied show copy number variation as this will result in deviating expression (Figure 1c). Chromosome ploidy can be investigated by e.g. cytogenetic and whole genome sequencing methods, but these methods are generally not applied in standard cell-line research. As RNA-seq is (becoming) the general method used for gene expression detection, also in cell-lines, we investigated the potential of implementing RNA-seq as a tool for detecting chromosomal abnormalities through variant calling, ASE analysis and plotting the allele ratios. We observed similar patterns of allele support ratios from WGS variant calling and RNA-seq allelic expression variants. This suggests we do not need WGS data for detection of chromosomal abnormalities. We show that a larger number of passages of the IPECJ2 cell line results in an increased level of aneuploidy. Thus, our approach, using RNA-seq, is a sufficient and cost-effective tool to detect aneuploidy, which we suggest to be regularly applied in cell line experiments.

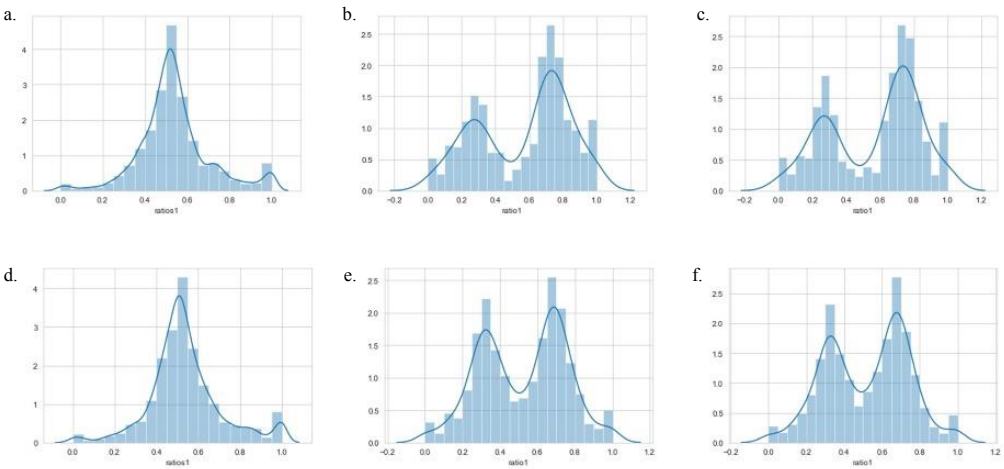

Figure 2. ASE for IPECJ2 derived cell lines cultured for a longer time (passages). IPECJ87 (a) chromosome 15, (b) chromosome 16, (c) chromosome 17. IPECJ91 (d) chromosome 15, (e) chromsome16, (f) chromosome17.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the grant agreement no. 817998. Due to the restriction of the page number for this paper we were unable to include figures for all the results. The results not included can be obtained by contacting the corresponding author(s).

References

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Berlin: ScienceOpen, Inc.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), pp.15-21.

Faust, G. G., & Hall, I. M. (2014). SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics, 30(17), 2503-2505.

Garrison E, Marth G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907.

Hartwell, L., Goldberg, M.L., Fischer, J.A., Hood, L.E., Aquadro, C.F., Bejcek, B. (2015). Genetics: from genes to genomes (pp.). New York: McGraw-Hill.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res, 20:1297-303. https://doi.org/10.1101/gr.107524.110.

Molina, O., Abad, M.A., Solé, F. and Menéndez, P. (2020). Aneuploidy in Cancer: Lessons from Acute Lymphoblastic Leukemia. Trends in Cancer.

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics, 12(1), 1-16.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics, 25(16), pp.2078-2079.

Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN].

Nossol, C., Barta-Böszörményi, A., Kahlert, S., Zuschratter, W., Faber-Zuschratter, H., Reinhardt, N., Ponsuksili, S., Wimmers, K., Diesing, A.K. and Rothkötter, H.J., (2015). Comparing two intestinal porcine epithelial cell lines (IPECs): morphological differentiation, function and metabolism. PloS one, 10(7), p.e0132323.

Verma, A., Verma, M. and Singh, A., (2020). Animal tissue culture principles and applications. In Animal Biotechnology (pp. 269-293). Academic Press.