[Letter to the Editor] Replication timing influences DNA copy number determination by array-CGH

Here we update our preliminary observations on array-comparative genome hybridization (array-CGH) analyses of rapidly dividing cells. In our array-CGH studies, we observed wave patterns in copy number determination on rapidly dividing cell systems, with population doubling times less than one day and high relative S phase fractions (1). DNA obtained from highly proliferative murine leukemic blasts and induced pluripotent stem (iPS) cells generated reproducible oscillation of data points around the baseline. Interestingly, after applying starvation conditions or direct G1 phase arrest during in vitro cultivation, genomic profiles smoothed to reliable copy number states, avoiding false positives and false negatives, respectively. Therefore, we strongly recommend G1 phase arrest prior to DNA

extraction when highly proliferative cell systems are interrogated by array-CGH.

By comparing genomic copy number data from murine iPS cells with replication timing profiles of iPS cells (2), we observed remarkable similarity in data patterns for every murine chromosome (Figure 1B and C; chromosome 14 is shown as an example). In contrast, genomic profiles generated from the parental fibroblasts appeared as a very flat profile (Figure 1A). We extracted log ratio values from the replication domain platform and correlated these to log ratio data from our array-CGH analysis using Pearson's coefficient. Comparing array-CGH data from fibroblast cells to replication timing data of iPS cells, no correlation could be made ($r = 0.007$), whereas array-CGH data from iPS cells indicated a correlation with respective

replication timing data ($r = 0.489$). We also aligned array-CGH data with replication timing data derived from embryonic stem (ES) cells ($r = 0.491$), which matched the best correlation, while raw data from mammary epithelium showed a weaker correlation ($r = 0.235$). Hence, tissue origin could be interpreted by replication timing patterns.

The population doubling time of reprogrammed iPS cells is approximately 18 h, indicating a high proliferation capacity, whereas cell division of murine fibroblast cultures takes about 25 h (3,4). We conclude from our observations that relatively high fractions of S/G2 phase populations result in DNA content between the 2n and 4n state. This unequal DNA quantity for particular loci, which itself may coincide with replication timing, gene expression, and

Figure 1. Array-CGH data profile of murine iPS cells correlates with replication timing. A) Array-CGH data of murine fibroblasts (example: chromosome 14); $4 \times 180k$ microarrays were performed according to the manufacturer's (Agilent) recommendations; data analysis was done using Genomic Workbench (7.0.4.0). B) Array-CGH data of murine iPS cells derived from fibroblasts shown in 1A (again chromosome 14 is shown exemplarily; gain in 14q proves clonogenic origin). C) Replication timing of mouse chromosome 14 (<http://replicationdomain.org/>); chip ID: iPS8915902&9321702CGH.

chromatin state, most likely leads to the observed genomic waves. These observations are independent of the array-CGH platform since we use the Agilent system as standard for diagnostic and scientific purposes, whereas others were able to detect quite similar wave patterns when applying Nimblegen microarrays(5).

During our DNA extraction protocol, we perform an RNase treatment step, since RNA contamination is believed to have an impact on array-CGH data quality (personal communication). Nevertheless, wave patterns were observed even though DNA quality was good, as demonstrated by photometry and agarose gel electrophoresis; derivative log ratio spread (DLRS) values were excellent. In their recent publication, van Heesch and colleagues suggest a stringent proteinase K digestion prior to DNA extraction to better degrade DNA/protein complexes and to reach equimolar DNA coverage along all chromosomal regions (5). GC content was also suggested to influence wave patterns. However, correction can be made by applying suitable algorithms (6). Appropriately, the observation that replication timing appears to be dependent on GC content, with GC-rich regions replicating earlier than AT-rich regions, was reported previously (7).

Replication timing could also directly influence next generation sequencing (NGS) approaches. Along with array-CGH, NGS is a frequently used diagnostic tool to detect genetic aberrations in a high-throughput manner and will gain importance for verifying genomic integrity in scientific investigations. Due to its high sequence coverage and short reads, NGS should resolve distinct genomic loci at a higher level compared with array-CGH. Hence, NGS might bear an even greater risk of false positive or false negative numerical aberration detection in cases of rapidly dividing in vitro cultures, such as iPS cells, or when DNA from highly regenerative tissues is examined. As in the case of array-CGH, this may not be explained exclusively as a technical inaccuracy and might be directly illustrated by replication timing profiles. Furthermore, it would be interesting to correlate replication dynamics of particular loci or whole genomes with global gene expression analyses, applying not only microarray approaches but also RNA deep sequencing, as transcriptome turnover could likely be influenced by replication activity.

Accordingly, array-CGH could be used as an additional tool for the investigation of replication dynamics indicating chromo-

somal hot-spots of breakage or fusion events or in studies of altered replication due to disease. Moreover, array-CGH might be applied for basic research in the replication biology field, in addition to its use for replication timing analysis. Even though it does not reach the sensitivity of sequencing approaches, as Gilbert has pointed out (8), array-CGH could be used for the analysis of consensus binding sites of ORCs (origin recognition complexes) without replication fork arrest or chromatin immunoprecipitation (ChIP), the common techniques for studying replication origins.

As in our previous work, we suggest monitoring the cell division behavior of in vitro cell systems and, in the case of rapidly replicating cells, the induction of G1 phase synchronized populations. Additionally, it would be of great benefit for the array-CGH community to have information not only on donor sex and race for respective reference material provided by companies but also tissue of origin or even proliferation dynamics such as population doubling time or relative cell cycle phases. An appropriate notation in public array-CGH databases would also help.

In conclusion, we have shown a direct biological impact of cell cycle dynamics on DNA copy number detection by array-CGH. The proportion of actively replicating cells in highly proliferating cell cultures or tissues is responsible for the non-equimolar DNA content. This can be directly visualized in array-CGH data analysis as copy number waves, frequently treated by the array-CGH community as a technical error, and shows a remarkable similarity to replication timing profiles of the respective organisms and tissues of origin.

Author contributions

G.M. and D.S. wrote the paper; G.M. and M.T. performed the experiments as well as data and statistical analyses; G.M., M.T. and D.S. took part in critical discussions.

Acknowledgments

The authors would like to thank Sebastian Kuehle and Soeren Turan for providing the fibroblast and iPS cell DNA, Sarah Tauscher for her essential assistance regarding statistical analysis, and Gillian Teicke for her great support in editing the article. G.M. was funded by a grant from the Hannover Biomedical Research School (HBRS), PhD Program Molecular Medicine, Hannover Medical School, Hannover, Germany.

Competing interests

The authors declare no competing interests.

References

1. Manukjan, G., M. Tauscher, T. Ripperger, A. Schwarzer, B. Schlegelberger, and D. Steinemann. 2012. Induced G1 phase arrest of fast-dividing cells improves the quality of genomic profiles generated by array-CGH. *Biotechniques* 53:245-248.
2. Weddington, N., A. Stuy, I. Hiratani, T. Ryba, T. Yokochi, and D.M. Gilbert. 2008. Replication Domain: a visualization tool and comparative database for genome-wide replication timing data. *BMC Bioinformatics* 10:9:530.
3. Takahashi, K. and S. Yamanaka. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126:663-676.
4. Hindley, C. and A. Philpott. 2013. The cell cycle and pluripotency. *Biochem. J.* 451:135-143.e
5. Van Heesch, S., M. Mokry, V. Boskova, W. Junker, R. Mehon, P. Toonen, E. de Bruijn, J.D. Shull, et al. 2013. Systematic biases in DNA copy number originate from isolation procedures. *Genome Biol.* 24;14:R33.
6. Benjamini, Y. and T.P. Speed. 2012. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* 40:e72.
7. Hiratani, I., T. Ryba, M. Itoh, T. Yokochi, M. Schwaiger, C.W. Chang, Y. Lyo, T.M. Townes, et al. 2008. Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* 7;6:e245.
8. Gilbert, D.M. 2010. Evaluating genome-scale approaches to eukaryotic DNA replication. *Nat. Rev. Genet.* 11:673-684.

Received 02 August 2013; accepted 16 September 2013.

Georgi Manukjan, Marcel Tauscher, and Doris Steinemann

Institute of Cell and Molecular Pathology, Hannover Medical School, Hannover, Germany

BioTechniques 55:231-232 (November 2013)
doi 10.2144/000114097

Keywords: replication timing; array-CGH; genomic waves; iPS cells; cell cycle arrest

Address correspondence to Doris Steinemann, Institute of Cell and Molecular Pathology, Hannover Medical School, Hannover, Germany. E-mail: steinemann.doris@mh-hannover.de

To purchase reprints of this article, contact: biotechniques@fosterprinting.com