Genomics Data 2 (2014) 53–54

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Microarray-based optimization to detect genomic deletion mutations

Eric J. Belfield [a], Carly Brown [a], Xiangchao Gan [a,b,c], Caifu Jiang [a], Dilair Baban [b], Aziz Mithani [a,d], Richard Mott [b], Jiannis Ragoussis [b,e], Nicholas P. Harberd [a,*]

[a] Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK
[b] Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK
[c] Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg, Cologne 50829, Germany
[d] Department of Biology, LUMS School of Science and Engineering, Sector UDHA, Lahore 54792, Pakistan
[e] McGill University and Genome Quebec Innovation Centre, 740 DR Penfield Ave., Montreal H3A 0G1, Canada

ARTICLE INFO

Article history:
Received 10 April 2014
Accepted 11 April 2014
Available online 24 April 2014

Keywords:
Microarray
Mutation
Deletion
Genome
CGH

ABSTRACT

We performed array comparative genome hybridization (aCGH) analyses of five *Arabidopsis thaliana* mutants with genomic deletions ranging in size from 4 bp to >5 kb. We used the Roche NimbleGen *Arabidopsis* CGH 3 × 720 K whole genome custom tiling array to optimize deletion detection. Details of the microarray design and hybridization data have been deposited at the NCBI GEO repository with accession number GSE55327.

Specifications	
Organism/cell line/tissue	*Arabidopsis thaliana*
Sex	Not applicable
Sequencer or array type	NimbleGen CGH 3 × 720 K whole genome tiling array
Data format	Raw and processed microarray hybridization data
Experimental factors	Genomic DNA hybridizations of deletion mutants versus controls
Experimental features	The design of a microarray to determine the probe resolutions required to reliably detect genomic deletions of various sizes
Consent	Not applicable
Sample source location	Not applicable

Direct link to deposited data

Deposited data can be found here: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55327.

Experimental design, materials and methods

Microarrays are a valuable tool for genomic studies. We used a customized version of the Roche NimbleGen CGH 3 × 720 K microarray to perform optimization experiments in identifying *Arabidopsis thaliana* genomic deletions [1]. The standard version of the microarray features 2.1 million 50–75-mer isothermal (target Tm 76 °C) probes that represent the complete *A. thaliana* genome (120 Mb) at 49 bp spacing. The probes are synthesized using maskless array technology and placed randomly on the array. We designed an additional ~15,000 custom probes that were added to the array design at spacings every 2, 6, 10, 12, 15, 17, 20, 22, 25, 27, 30, 32, 35, 37, 40, 42, 45, 47 and 49 bp. These probes were used to determine the optimal density of probes to detect deletions that were present in five *A. thaliana* lines that ranged in size from 4 bp to over 5 kb [1], see Table 1.

Genomic DNA from the five mutant and corresponding control plant lines were sent to the NimbleGen custom microarray services facility, labelled with Cy3 or Cy5 (respectively) using a NimbleGen Dual-Color DNA Labelling Kit and co-hybridized to the arrays (see NCBI GEO supplementary file: NimbleGen Arrays User's Guide_9.1). The arrays were scanned using a NimbleGen MS 200 Microarray Scanner and signal intensity data at 532 nm (Cy3 data) and 635 nm (Cy5 data) were extracted from the scanned images using Roche NimbleGen NimbleScan software (see: http://www.nimblegen.com/downloads/support/NimbleScan_v2p6_UsersGuide.pdf). The raw data generated has been deposited in the NCBI GEO database. Table 2 shows the raw intensity file names associated with each mutant and control hybridisation experiment, and the corresponding NCBI GEO sample names. The table also shows the names of the processed data files associated with each of the mutant versus control hybridization experiments. The processed

[*] Corresponding author.
E-mail address: nicholas.harberd@plants.ox.ac.uk (N.P. Harberd).

http://dx.doi.org/10.1016/j.gdata.2014.04.005
2213-5960/

Table 1
A table showing the size of deletions in five *Arabidopsis* mutants used to optimize microarray-based deletion detection.

A. thaliana mutant	Deletion size (bp)
ga1-3	5051
FN1148	523
E124	104
E99	28
E207	4

data files contain the $\log_2$ intensity values for each array feature reported for both the 532 nm channel and the 635 nm channel, as well as the ratio of the two channels.

The conversion of raw data values to processed data values involves two steps. The first step (performed by NimbleGen) uses a locally weighted polynomial regression (LOESS, see: http://www.obgyn.cam.ac.uk/genearray/loess-normalisation.htm) and is applied practically to all two-color arrays to adjust the signal intensity of each feature based on the X and Y coordinates of the probes on the array. Specifically, this involves spatial correction, when variation that occurs across the array (with respect to both length and width) are resolved. The second step involves normalization using qspline normalization [2] to compensate for signal-dependent differences between the cyanine (Cy3 and Cy5) dyes. Following spatial correction and normalization, $\log_2$-ratios of the mutant versus control sample for each probe are generated.

In addition to the deposited raw and processed data files, three other files are included. The first is a design file (NCD file: GPL18327_090325_Athal_EB_CGH_HX1.ndf) containing the complete information necessary to synthesis the design array. The second is a positions file (POS file: GPL18327_090325_Athal_EB_CGH_HX1.pos) used for applications like CGH, expression tiling, or ChIP-chip. This POS file contains important information such as the genomic positions of each probe, probe selection criteria, probe sequence and probe container IDs for each probe spacing of 2, 6, 10, … bp etc. (as above). The third document is a general feature format file (GFF file: GPL18327_090325_Athal_EB_CGH_HX1_probe_locations). The details of the NimbleGen microarray data are supplied in GFF format for viewing in GFF viewers. The GFF files supplied are tab-delimited, with the following format: <seqname> <source> <feature> <start> <end> <score> <strand> <frame> [attributes] [comments]. For further details see the NCBI GEO supplementary file: GPL18327_NimbleGen_data_formats.pdf.

Table 2
A table listing the raw and processed data files deposited to NCBI GEO and accession names.

NCBI accession number	Title	Raw intensity files	Processed intensity files
GSM1334296	*ga1-3* v control	GSM1334296_ga1-3.txt GSM1334296_ga1-3_control.txt	GSM1334296_ga1_3_matrix.txt
GSM1334297	FN1148 v control	GSM1334297_FN1148.txt GSM1334297_FN1148_control.txt	GSM1334297_FN1148_matrix.txt
GSM1334298	E124 v control	GSM1334298_E124.txt GSM1334298_E124_control.txt	GSM1334298_E124_matrix.txt
GSM1334299	E99 v control	GSM1334299_E99.txt GSM1334299_E99_control.txt	GSM1334299_E99_matrix.txt
GSM1334300	E207 v control	GSM1334300_E207.txt GSM1334300_E207_control.txt	GSM1334300_E207_matrix.txt

Competing interests

The authors declare that there are no competing interests in the work published.

Acknowledgements

This work was supported by a grant from the Biotechnological Biological Science Research Council (BB/F020759/1).

References

[1] E.J. Belfield, C. Brown, X. Gan, C. Jiang, D. Baban, A. Mithani, R. Mott, J. Ragoussis, N.P. Harberd, Microarray-based ultra-high resolution discovery of genomic deletion mutations. BMC Genomics 15 (2014) 224, http://dx.doi.org/10.1186/1471-2164-15-224.

[2] C. Workman, L.J. Jensen, H. Jarmer, R. Berka, L. Gautier, H.B. Nielser, H.-H. Saxild, C. Nielsen, S. Brunak, S. Knudsen, A new non-linear normalization method for reducing variability in DNA microarray experiments. Genome Biol. 3 (0048) (2002) 1–0048, http://dx.doi.org/10.1186/gb-2002-3-9-research0048 (16).