Genomics Data 14 (2017) 1–4

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in Brief

Genome-wide miRNA, gene and methylation analysis of triple negative breast cancer to identify changes associated with lymph node metastases

Kelly A. Avery-Kiejda[a,b,*], Andrea Mathe[a,b,1], Rodney J. Scott[a,b,c]

[a] Centre for Information Based Medicine, Hunter Medical Research Institute, NSW 2305, Australia
[b] Priority Research Centre for Cancer Research, Innovation and Translation, School of Biomedical Sciences and Pharmacy, Faculty of Health and Medicine, University of Newcastle, NSW 2308, Australia
[c] Pathology North, John Hunter Hospital, Newcastle, NSW 2305, Australia

ABSTRACT

Triple negative breast cancer (TNBC) is a particularly important breast cancer subtype with an aggressive clinical phenotype that is associated with a higher likelihood of metastasis. This subtype is characterized by an absence of the estrogen (ER) and progesterone (PR) receptors, as well as the human epidermal growth factor receptor 2 (HER2/HER neu). The absence of the three receptors significantly reduces targeted treatment options for patients with TNBC and as such, there is an urgent need to identify novel treatment targets. Here, we provide detailed information regarding the design of a multi-platform dataset that describes genome-wide assessment of miRNA (assessed by microarray, GSE38167) and gene expression (assessed by microarray, GSE61723), as well as methylation (assessed by Illumina HM450K BeadChip, GSE78751) in TNBCs, matched normal adjacent tissues and matched lymph node metastases. The use of this multi-platform dataset is likely to uncover novel markers and key pathways involved in progression to lymph node metastasis in TNBC.

Specifications	
Organism/cell line/tissue	Homo sapiens/breast cancer specimens (triple negative subtype)
Sex	Female
Sequencer or array type	• Agilent Human miRNA microarrays (Sanger Release 14.0) • Affymetrix Human Gene 2.0 ST Array (HuGene-2_0-st) • Illumina HumanMethylation450 BeadChip (HumanMethylation450_15017482)
Data format	Raw (.tar) and processed (.txt)
Experimental factors	Matched normal adjacent tissue, tumour tissue, matched lymph node metastases
Experimental features	This study has investigated miRNA, gene and methylation profiles in primary TNBC cases (miRNA (n = 31), gene (n = 33), methylation (n = 23)) and matched lymph node metastases (miRNA (n = 13), gene (n = 15), methylation (n = 12)) compared with matched normal breast tissues (miRNA (n = 23), gene (n = 17), methylation (n = 11, 3 pooled samples and 1 single sample)).
Consent	This study complies with the Helsinki Declaration with ethical approval from the Hunter New England Human Research Ethics Committee (Approval number: 09/05/20/5.02). In accordance with the National Statement on Ethical Conduct in Research Involving Humans, a waiver of consent was granted for this study.
Sample source location	Newcastle, NSW, Australia

1. Direct link to deposited data

miRNA arrays (GSE38167): https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE38167.

Gene expression arrays (GSE61723): https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE61723.

Illumina HM450K BeadChip (GSE78751): https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE78751.

* Corresponding author.
E-mail address: kelly.kiejda@newcastle.edu.au (K.A. Avery-Kiejda).
[1] Present address: Gastrointestinal Research Group, Hunter Medical Research Institute, New Lambton Heights, NSW, Australia.

http://dx.doi.org/10.1016/j.gdata.2017.07.004
Received 11 May 2017; Received in revised form 28 June 2017; Accepted 4 July 2017
Available online 06 July 2017

Table 1
Labelling of patient samples in individual studies of whole genome gene, miRNA and methylation analysis. Cases shaded in grey are lymph node negative and unshaded cases are lymph node positive.

Gene expression analysis (GSE61723)	miRNA analysis (GSE38167)	Methylation analysis (GSE78751)
Normal adjacent tissues (NAT)		
Patient 3 NAT	Patient 26_NAT	
	Patient 1_NAT	
Patient 5 NAT	Patient 20_NAT	
Patient 6 NAT	Patient 34_NAT	
Patient 7 NAT	Patient 5_NAT	
Patient 8 NAT	Patient 27_NAT	
Patient 9 NAT	Patient 11_NAT	Patient 7 NAT
Patient 10 NAT	Patient 13_NAT	
Patient 11 NAT	Patient 18_NAT	
Patient 12 NAT	Patient 16_NAT	
	Patient 15_NAT	
Patient 18 NAT	Patient 25_NAT	
Patient 19 NAT	Patient 17_NAT	
	Patient 19_NAT	
Patient 23 NAT	Patient 6_NAT	
	Patient 9_NAT	
Patient 26 NAT	Patient 14_NAT	
	Patient 7_NAT	
Patient 28 NAT	Patient 33_NAT	
	Patient 10_NAT	
Patient 30 NAT	Patient 12_NAT	
Patient 32 NAT	Patient 21_NAT	
Patient 33 NAT	Patient 3_NAT	
		NAT pool 1: Patient 21, 30, 31
		NAT pool 2: Patient 1, 3, 32
		NAT pool 3: Patient 5, 7, 14, 33
Invasive ductal carcinomas (IDC)		
Patient 1 IDC	Patient 23_IDC	Patient 17 IDC
Patient 2 IDC	Patient 4_IDC	Patient 6 IDC
Patient 3 IDC		Patient 16 IDC
Patient 4 IDC	Patient 1_IDC	
Patient 5 IDC	Patient 20_IDC	Patient 3 IDC
Patient 6 IDC	Patient 34_IDC	Patient 18 IDC
Patient 7 IDC	Patient 5_IDC	
Patient 8 IDC	Patient 27_IDC	Patient 14 IDC
Patient 9 IDC	Patient 11_IDC	Patient 7 IDC
Patient 10 IDC	Patient 13_IDC	Patient 4 IDC
		Patient 10 IDC
Patient 11 IDC	Patient 18_IDC	
		Patient 11 IDC
Patient 12 IDC	Patient 16_IDC	
Patient 13 IDC	Patient 22_IDC	
Patient 14 IDC	Patient 31_IDC	
Patient 15 IDC		
Patient 16 IDC		Patient 9 IDC
Patient 17 IDC	Patient 15_IDC	Patient 23 IDC
Patient 18 IDC		
Patient 19 IDC	Patient 17_IDC	
Patient 20 IDC	Patient 19_IDC	Patient 5 IDC
Patient 21 IDC	Patient 2_IDC	Patient 2 IDC
Patient 22 IDC	Patient 8_IDC	
Patient 23 IDC	Patient 6_IDC	Patient 21 IDC
Patient 24 IDC	Patient 9_IDC	
Patient 25 IDC	Patient 32_IDC	
Patient 26 IDC	Patient 14_IDC	Patient 1 IDC
Patient 27 IDC	Patient 7_IDC	Patient 19 IDC
Patient 28 IDC	Patient 33_IDC	
Patient 29 IDC	Patient 10_IDC	Patient 20 IDC
Patient 30 IDC	Patient 12_IDC	Patient 8 IDC
	Patient 29_IDC	Patient 12 IDC
		Patient 22 IDC
Patient 32 IDC	Patient 21_IDC	
Patient 33 IDC	Patient 3_IDC	
Patient 34 IDC	Patient 24_IDC	Patient 15 IDC
	Patient 28_IDC	Patient 13 IDC

Lymph node (LN) metastases		
Patient 4 LN	Patient 1_LN	Patient 27 LN
Patient 5 LN	Patient 20_LN	Patient 3 LN
Patient 7 LN	Patient 5_LN	Patient 25 LN
Patient 9 LN	Patient 11_LN	Patient 7 LN
Patient 11 LN	Patient 18_LN	Patient 30 LN
Patient 14 LN	Patient 31_LN	Patient 29 LN
Patient 16 LN	Patient 30_LN	
Patient 18 LN		Patient 26 LN
Patient 20 LN	Patient 19_LN	Patient 5 LN
Patient 25 LN	Patient 32_LN	
Patient 30 LN	Patient 12_LN	Patient 8 LN
Patient 31 LN		
Patient 32 LN	Patient 21_LN	Patient 28 LN
Patient 33 LN	Patient 3_LN	Patient 24 LN
Patient 35 LN	Patient 28_LN	Patient 13 LN

2. Experimental design, materials and methods

2.1. Study cohort

A total of 38 grade 3, triple negative, invasive ductal carcinomas (IDCs), 23 matched normal adjacent tissues (NAT), and 15 lymph node metastases (LNmet) were used for this analysis. All samples were formalin-fixed, paraffin-embedded (FFPE) and obtained from the archives of Pathology North, John Hunter Hospital, Newcastle, Australia. This cohort has been described previously [1–3]. Due to a lack of material not all samples were included in all studies and the number of samples in each of the studies is as follows: TNBC cases (miRNA (n = 31), gene (n = 33), methylation (n = 23)), matched lymph node metastases (miRNA (n = 13), gene (n = 15), methylation (n = 12)) and matched normal breast tissues (miRNA (n = 23), gene (n = 17), methylation (n = 11, 3 pooled samples and 1 single sample)). Because the analysis of miRNA, gene and methylation was performed at distinct times, the patient identifiers do not match between the individual studies described in Gene Expression Omnibus and Table 1 shows the corresponding patient identifiers between the studies so that miRNA, gene and methylation data can be linked to an individual patient.

2.2. RNA extraction

Total RNA was extracted using the miRNeasy FFPE kit (Qiagen, Doncaster, VIC, Australia). RNA was quantified using the Quant-it RiboGreen RNA Assay kit (Invitrogen, Mulgrave, VIC, Australia) and purity assessed by $A_{260/A280}$ and $A_{260/230}$ ratios (> 1.8) using the Nanodrop. The RNA integrity of selected samples was analysed using the 2100 Bioanalyser and the RNA 6000 Nano kit (Agilent Technologies, Mulgrave, VIC, Australia).

2.3. DNA extraction

The Gentra Puregene Tissue Kit (Qiagen, Venlo, Limburg, Netherlands) was used to isolate DNA from FFPE tissue following the manufacturers' instruction with minor modifications as previously described [2]. DNA was quantitated using the Qubit dsDNA BR Assay Kit according to the manufacturers' instructions (Life Technologies, Carlsbad, CA, United States of America).

2.4. miRNA profiling

miRNA profiling was performed as previously described [1]. Briefly, 100 ng of total RNA was dephosphorylated and labelled with Cy3 using the miRNA Complete Labelling and Hyb Kit (Agilent Technologies). Labelled RNA was hybridised to Human miRNA microarrays (Sanger Release 14.0) according to the manufacturers' instructions (Agilent Technologies) and scanned on an Agilent High-resolution C scanner. Data from 15,000 probe features representing 904 unique miRNAs was extracted using Agilent Feature Extraction software (v10.7.3.1) and converted to background subtracted signal intensities. The extracted data was imported into Genespring GX (Agilent Technologies) where it was log2 transformed and median normalised. 570 miRNA transcripts had a signal intensity threshold above background in at least one sample. Unpaired t-tests identified miRNAs with significantly altered expression (> 2-fold, $p < 0.05$, FDR < 5%).

2.5. Gene expression profiling

Gene expression profiling was performed as previously described [3]. Briefly, 100 ng total RNA was amplified (Ovation FFPE WTA kit) and biotinylated (Encore Biotin module) according to the manufacturers' instructions (Nugen, San Carlos, California, USA). The samples were hybridised to HuGene 2.0 arrays (Affymetrix, Santa Clara, California, USA) and 17 h later washed and stained. The HumanGene 2.0 arrays (Affymetrix) contain probe features representing 11,000 lncRNAs, 24,000 genes, and 30,000 coding transcripts. The arrays were scanned on a GeneChip Scanner 3000 7G (Affymetrix). The data was imported to Genomic Suite 6.6 (Partek, St Louis, Missouri, USA) and a robust multi-array analysis (RMA) was performed, which included log2 transformation, background correction, quantile normalisation and summarisation of the probe features resulting in a set of expression signal intensities. Unpaired t-tests identified genes that were found to be significantly different between samples (p-value < 0.05; fold change > 1.5 or < −1.5). Correction for multiple testing was performed using Benjamini – Hochberg procedure.

2.6. DNA methylation profiling

DNA methylation profiling was performed as previously described [2]. The Infinium HD FFPE quality control (QC) Assay (Illumina, San Diega, CA, USA) was used to assess the integrity of the DNA used for the analysis according to the manufacturers' instructions. Bisulfite conversion was performed using the EZ-96 DNA Methylation Kit (Zymo Research, Irvine, CA, USA). FFPE Restoration was undertaken following the Infinium HD FFPE Restore Protocol (Illumina). The Infinium HD FFPE Methylation Assay (Illumina) with the hybridisation, washing and staining of the arrays as well as the scanning (iScan) of the HumanMethylation 450 K BeadChip arrays was performed using the manufacturers' instructions. The data from all samples was imported in form of the idat files into Genomic Suite 6.6 (Partek) and Illumina normalisation was performed. ANOVA analysis was performed to detect differentially methylated loci between groups (e.g. IDC versus NAT, LNmet versus NAT, and IDC versus LNmet). Significance was granted if $p < 0.05$ and the estimated difference between groups (Δß) was <

-0.1 or > 0.1, signifying a methylation change of at least 10%. These analyses were performed on single loci and on differentially methylated regions (DMR) – a minimum of three significant consecutive probes. Pathway enrichment analysis was performed using Genomic Suite 6.6 (Partek). All significant loci were filtered to include loci that are within enhancer and/or promoter regions. These were then used for pathway enrichment analysis, a tool within Genomic Suite 6.6 (Partek). The enrichment score is the negative natural log of the enrichment *p*-value derived from the Fisher's exact test of the pathway enrichment analysis.

3. Discussion

Here we have described a dataset involving miRNA, gene and epigenome analysis of TNBC primary breast cancers, their matched normal adjacent tissues and matched lymph node metastases. This dataset also includes breast cancers of the triple negative breast cancer subtype with no known lymph node involvement, hence, providing a unique opportunity to dissect clinically relevant changes associated with lymph node metastasis in triple negative breast cancer. To the best of our knowledge, no other multi-platform dataset exists that includes matched lymph node metastases in this subtype. In breast cancer, the number of positive lymph nodes (LN) is known to have an inverse linear correlation with prognosis and survival [4]. This is not the case with TNBC, where it has been shown that any LN involvement is associated with worse disease-free and overall survival [5]. Hence, the use of this multi-platform dataset is likely to uncover novel markers and key pathways involved in progression to lymph node metastasis in TNBC.

Conflict of interest

The authors declare that there are no competing interests.

Acknowledgements

The authors would like to thank Dr. Ricardo Vilain for pathological review of all tumour and normal tissue specimens used in this analysis as well as Ms. Tina Hope and Ms. Sarah Nielsen for assistance with archival specimens. This work was supported by funding from the National Breast Cancer Foundation (CG-12-07) and the Hunter Medical Research Institute (HMRI 10-22).

References

[1] K.A. Avery-Kiejda, S.G. Braye, A. Mathe, J.F. Forbes, R.J. Scott, Decreased expression of key tumour suppressor microRNAs is associated with lymph node metastases in triple negative breast cancer, BMC Cancer 14 (1) (2014) 51.
[2] A. Mathe, M. Wong-Brown, W.J. Locke, C. Stirzaker, S.G. Braye, J.F. Forbes, S.J. Clark, K.A. Avery-Kiejda, R.J. Scott, DNA methylation profile of triple negative breast cancer-specific genes comparing lymph node positive patients to lymph node negative patients, Sci Rep 6 (2016) 33435.
[3] A. Mathe, M. Wong-Brown, B. Morten, J.F. Forbes, S.G. Braye, K.A. Avery-Kiejda, R.J. Scott, Novel genes associated with lymph node metastasis in triple negative breast cancer, Sci Rep 5 (2015) 15832.
[4] C.L. Carter, C. Allen, D.E. Henson, Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases, Cancer 63 (1) (1989) 181–187.
[5] L.F. Hernandez-Aya, M. Chavez-Macgregor, X. Lei, F. Meric-Bernstam, T.A. Buchholz, L. Hsu, A.A. Sahin, K.A. Do, V. Valero, G.N. Hortobagyi, A.M. Gonzalez-Angulo, Nodal status and clinical outcomes in a large cohort of patients with triple-negative breast cancer, J. Clin. Oncol. 29 (19) (2011) 2628–2634.