Genomics Data 5 (2015) 329–332

Contents lists available at ScienceDirect

Genomics Data

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

Data in Brief

Global gene expression profiling of a mouse model of ovarian clear cell carcinoma caused by ARID1A and PIK3CA mutations implicates a role for inflammatory cytokine signaling

Ronald L. Chandler [a,b], Jesse R. Raab [a,b], Mike Vernon [c], Terry Magnuson [a,b], Jonathan C. Schisler [b,d,e,*]

[a] Department of Genetics, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[b] Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[c] Functional Genomics Core, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[d] McAllister Heart Institute, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[e] Department of Pharmacology, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

ARTICLE INFO

Article history:
Received 17 June 2015
Accepted 24 June 2015
Available online 14 July 2015

Keywords:
Microarray, Gene expression, Ovarian clear-cell carcinoma
Ovarian clear-cell carcinoma

ABSTRACT

Ovarian clear-cell carcinoma (OCCC) is an aggressive form of epithelial ovarian cancer (EOC). OCCC represents 5–25% of all EOC incidences and is the second leading cause of death from ovarian cancer (Glasspool and McNeish, 2013) [1]. A recent publication by Chandler et al. reported the first mouse model of OCCC that resembles human OCCC both genetically and histologically by inducing a localized deletion of ARID1A and the expression of the PIK3CA^{H1047R} substitution mutation (Chandler et al., 2015) [2]. We utilized Affymetrix Mouse Gene 2.1 ST arrays for the global gene expression profiling of mouse primary OCCC tumor samples and animal-matched normal ovaries to identify cancer-dependent gene expression. We describe the approach used to generate the differentially expressed genes from the publicly available data deposited at the Gene Expression Omnibus (GEO) database under the accession number GSE57380. These data were used in cross-species comparisons to publically available human OCCC gene expression data and allowed the identification of coordinately regulated genes in both mouse and human OCCC and supportive of a role for inflammatory cytokine signaling in OCCC pathogenesis (Chandler et al., 2015) [2].

Specifications	
Organism/cell line/tissue	*Mus musculus*/ovarian clear cell carcinoma and healthy ovary
Strain	Outbred, Arid1a$^{fl/fl}$; (Gt)Rosa26PIK3CA*H1047R
Sex	Female
Sequencer or array type	Affymetrix Mouse Gene 2.1 ST Array
Data format	Affymetrix CEL files (raw), RMA normalized log base 2
Experimental factors	Ovarian clear cell carcinoma vs healthy ovary (paired)
Experimental features	Very brief experimental description
Consent	Level of consent allowed for reuse if applicable (typically for human samples)
Sample source location	Chapel Hill, NC, USA

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE57380

2. Experimental design, materials and methods

2.1. Mouse model of ovarian clear cell carcinoma

We maintained all mice used in this study at The University of North Carolina at Chapel Hill (UNC) Animal Facility using standard techniques in accordance with the University's Institutional Animal Care and Use Committee (IACUC). We performed all surgical procedures in accordance with the protocols approved by UNC IACUC. To induce genetic recombination in the ovarian surface epithelium (OSE), we employed a modified version of the ex-vivo AdCRE intrabursal delivery method [3]. First, AdCRE particles (obtained from the University of Iowa Gene Transfer Core) were diluted in sterile Dulbecco's phosphate buffered saline (dPBS) containing 8 μg/ml polybrene. We anesthetized 8–10 week old mice and administered a single 5 μl injection of AdCRE particles (2.5E7 plaque-forming units or pfu) into the right ovarian bursal cavity of the surgically exposed ovary using a sterile 31 gauge needle. Immediately following AdCRE injection, we washed the surgically exposed ovaries thoroughly with sterile 1× dPBS and placed the ovary back into the abdominal cavity. To control for experimental bias, we

* Corresponding author at: Department of Pharmacology, The University of North Carolina at Chapel Hill, 111 Mason Farm Road, MBRB 2340C, Chapel Hill, NC 27599-7126, USA.
E-mail address: schisler@unc.edu (J.C. Schisler).

http://dx.doi.org/10.1016/j.gdata.2015.06.027

performed surgeries on multiple occasions and randomized the age- and genotype-matched non-littermate animals prior to AdCRE injection. Additionally, we blinded the genotypes to our mouse surgeon prior to surgery.

2.2. RNA purification and quality control

Total RNA was extracted from pulverized tumor samples or matched, normal ovaries using the TRIzol method (Invitrogen), followed by an RNA cleanup step and on-column DNA digestion using the RNAeasy mini prep kit (Qiagen) according to the manufacturers' instructions. We measured the concentration of RNA using both spectroscopic (NanoDrop, Thermo Scientific) and fluorometric (Qubit, Life Technologies) methods (Table 1). All 18 samples were of high purity ($OD_{260}/OD_{280} \geq 2.00$) and integrity (RIN ≥ 7.0) and used for further processing.

2.3. Labeling protocol

We used the Beckman Coulter Biomek FXP Laboratory Automation Workstation with the Target Express setup to prepare the samples using the WT Expression HT Kit for Robotics (Ambion) to generate sense-strand cDNA from total RNA followed by fragmentation and the GeneChip HT Terminal Labeling Kit (Affymetrix) for cDNA labeling.

The fragmented and labeled cDNA was used to prepare a hybridization cocktail with the GeneTitan Hybridization Wash and Stain Kit for WT Arrays (Affymetrix). Hybridization, washing, staining and scanning of the Affymetrix Mouse Gene 2.1 ST peg plate arrays were carried out using the GeneTitan MC Instrument (Affymetrix) controlled with the GeneChip Command Console Software (AGCC). We used Expression Console Software (Affymetrix) for basic data extraction (CEL files) and quality control metrics.

2.4. Data processing

Affymetrix CEL files were normalized using the Robust Multichip Average (RMA) normalization method using Partek Genomics Suite v6.6 (Partek) using RMA Background Correction, quantile normalization, logging of probes using base 2, and median polishing for probeset summarization.

Table 1
RNA sample concentration and purity metrics.

#	Description	ng/μl	OD_{260}/OD_{280}	RIN
1	OV122 1T	8.28E+02	2.08	8.1
2	OV122 2c	5.54E+02	2.09	9.3
3	OV129 1T	1.68E+02	2.06	8.6
4	OV129 c	5.33E+02	2.11	9.3
5	OV137 1T	2.86E+02	2.06	8.4
6	OV137 c	8.52E+02	2.10	9.7
7	OV139 1T	3.90E+02	2.04	8.4
8	OV139 c	5.72E+02	2.10	9.3
9	OV153 1T	1.47E+03	2.09	8.8
10	OV153 2c	7.00E+02	2.09	9.6
11	OV166 1T	1.76E+03	2.05	8.8
12	OV166 2c	1.08E+03	2.09	9.4
13	OV179 1T	7.89E+02	2.08	9.6
14	OV179 2c	2.19E+02	2.05	8.3
15	OV190 1T	1.07E+03	2.08	8.9
16	OV190 2c	1.36E+02	2.03	7.0
17	OV127 1T	1.21E+02	2.09	7.5
18	OV127 c	3.86E+02	2.06	9.3

The table contains the sample identification numbers (#) and description in the format OV*XXX* YZ where XXX is the unique animal number, Y is the tissue sample number, and Z is the tissue classifier as either tumor (T) or matched normal ovary (c). We measured the RNA concentrations (ng/μl) as well as nucleic acid purity by a ratio of 1 cm pathlength optical density at 260 and 280 nm (OD_{260}/OD_{280}) and the RNA integrity number (RIN).

2.5. Sample quality control

We analyzed sample quality control by comparing the area under the curve (AUC) value for a receiver-operating characteristic (ROC), or ROC curve, plotting the detection of positive controls against the false detection of negative controls. A value of 1.0 represents a perfect value and a value of 0.5 illustrates no discernable difference between the positive and negative controls. All sample values fell within the expected range of 0.8 to 0.9 (mean = 0.881 ± 0.002, Fig. 1A) [4]. Additionally, we measured the mean absolute relative log expression (RLE) that measures the signal of each probe set compared to the median signal value of this probe set in the study. The RLE is useful to our study with similar sample types to detect outlier arrays. Using this metric, values greater than 0.5 indicate high sample variance. We found that the RLE mean of our samples fell within the expected range of 0.2 to 0.4 (mean = 0.284 ± 0.007, Fig. 1A) [4]. Furthermore, we did not detect any sample outliers via Tukey boxplots in analyzing the mean intensity of perfect match (pm) probes and background intensity, measured by the intensity of mismatch (mm) probes (Fig. 1B).

2.6. Hybridization and labeling quality control

We measured the hybridization bacterial spike-in controls *BioB*, *BioC*, *BioD*, and *CreX* (listed from the expected lowest to highest concentrations) and detected increasing concentrations of the spike-ins across all samples consistent with high-quality hybridization. Tukey boxplot analysis flagged sample #8 (Table 1, OV139 c) as a possible outlier due to overall decreased intensity (Fig. 1C). Likewise, we monitored

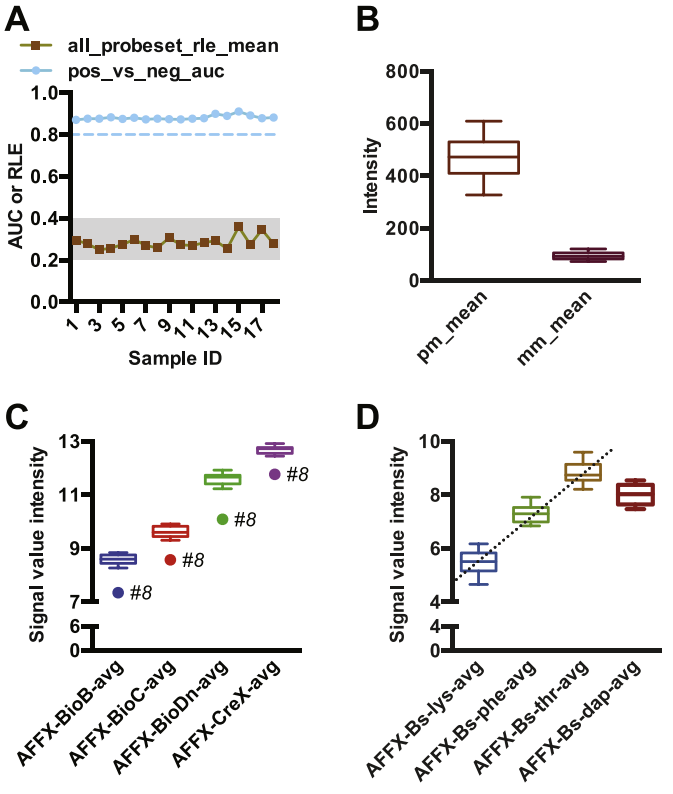

Fig. 1. Quality control metrics of microarrays. (A) The area under the curve (AUC) measuring the detection of positive controls versus false detection of negative controls; and the relative log expression (RLE) of each probe set across all 18 samples. The dashed line marks the lower cutoff of 0.8 for AUC outlier detection and the shaded region indicates the acceptable RLE range of 0.2–0.4. (B) The mean intensity of perfect match (pm) and mismatch (mm) probes. (C) Intensity of the bacterial gene spike-in controls *BioB*, *BioC*, *BioD*, and *CreX* represented by Tukey boxplots. The outlier, sample #8, is indicated. (**D**) Intensity of the polyA-control RNAs *lys*, *phe*, *thr*, and *dap* with the dashed line reflecting the linear increase between *lys*, *phe*, and *thr*.

labeling quality using the polyA-control RNAs *lys*, *phe*, *thr*, and *dap* (listed from the expected lowest to highest concentrations). We observed the *dap* signal out of rank order (Fig. 1D); however we did observe the expected linear trend across *lys*, *phe*, and *thr* [5] and did not identify any sample outliers (Fig. 1D). Further analysis of additional arrays run on the GeneTitan system identified a batch problem with the *dap* RNA in the polyA-control kit (Affymetrix, data not shown).

2.7. Outlier analysis and data structure

Our quality control measures flagged sample #8 as a potential outlier. To determine if #8 or other samples are outliers, we first compared the RMA normalized data distribution via box-whiskers plot (Fig. 2A) and principal component analysis (PCA, Fig. 2B). We did not observe any skewed samples or outliers using either visualization, respectively. Further, PCA clearly partitioned the samples by tissue classification (normal ovary or tumor) across the first principal component (PC1, 24%) indicating that the primary source of variance of these data likely consists of differences between healthy ovary and the tumor tissue rather than differences between animals (Fig. 2C). The remaining components shown here appear to reflect the variance within the classifications as seen in the second principal component (PC2, 10%) as well as the third and fourth components (PC3 and PC4, 8.0% and 6.6%, respectively) that capture sample variance across the normal ovary samples (Fig. 2B). In addition to PCA, unsupervised hierarchical clustering (HC) was used to reveal natural categories in gene expression data sets (Fig. 2D). Two primary clusters of microarray samples were comprised solely of either normal ovaries (red) or tumor tissue (blue) biological replicates. Given these results, we included sample #8 and the corresponding paired sample (#7) with the remainder of the downstream analyses. Given the clear partitioning of these unsupervised analyses we expected a large number of differentially expressed genes.

2.8. Detecting differential gene expression

The data as descried thus far was used in Chandler et al. to identify expression changes (normal versus tumor tissue) by using Linear Models for Microarray Data Analysis (LIMMA) or Significance Analysis of Microarray (SAM) analysis [2]. Probes with a false discovery rate (FDR) of 0% were considered statistically significant. As an example shown here, we removed probes near background or at low levels of detection, including probes that fell within the 20th and 100th percentiles that were present in at least 6 of the 9 samples per tissue type. SAM

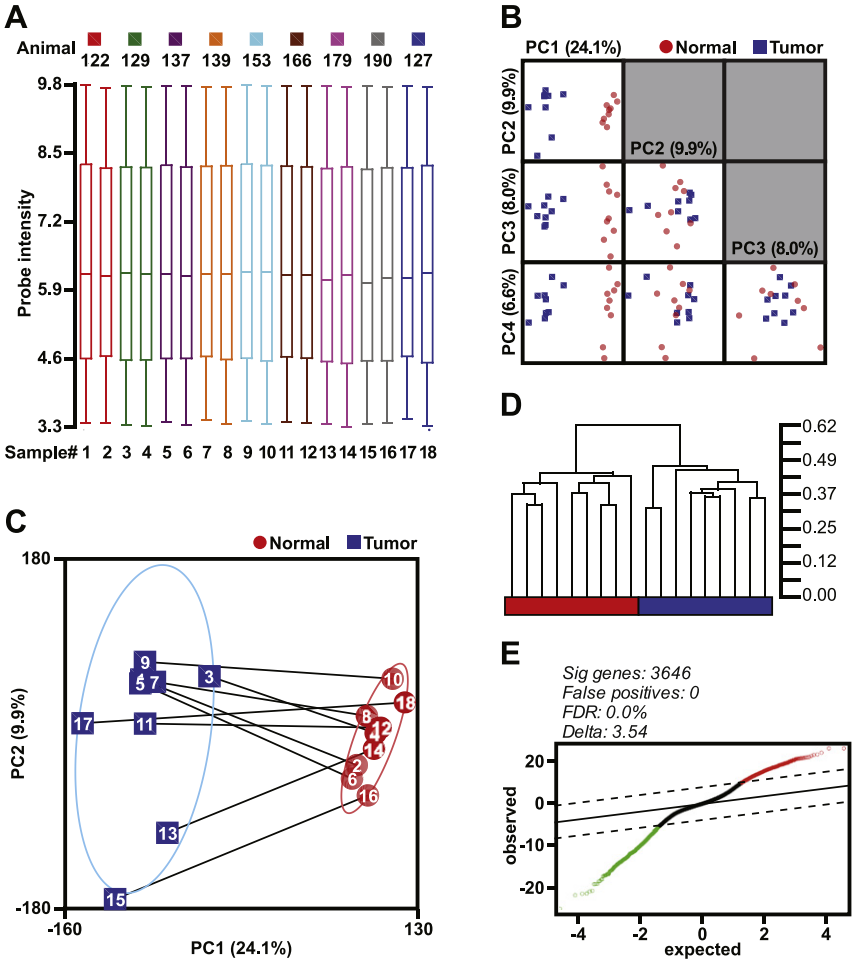

Fig. 2. Data structure, unsupervised gene clustering, and differential gene expression analysis. (A) RMA normalized data distribution via box-whiskers plot with the upper and lower 10th percentiles represented by the whiskers. (B) Principal component analysis visualized via matrix plots of the first four principal components (PC). Four eigenvectors were calculated for PCA and data represented are scaled to unit standard deviation. (C) PC1 versus PC2 scatterplot with the microarray samples identified both by number and paired samples (lines). Confidence ellipses categorized by tissue type represent 2 standard deviations. (D) Unsupervised hierarchical clustering using Pearson's Dissimilarity matrix with average linkage. The scale for the dendrogram represents the distance of clusters by Pearson's correlation coefficient. The tissue classification is colored red and blue representing normal and tumor tissue classification, respectively. (E) Significance analysis of microarrays plot of observed scores plotted against the expected scores. The solid line represents observed = expected, whereas the dashed lines indicate the significance threshold based on $\Delta = 3.54$. The genes identified as differentially expressed are indicated by red and green open circles, indicating higher and lower expression, respectively, of these genes in the mouse OCCC tumor tissue compared to normal ovaries. The number of differentially expressed genes, predicted false positives, and the false discovery rate (FDR) is provided.

identified 3646 probes that were differentially expressed at a false discovery rate of 0.0% (Fig. 2E).

3. Discussion

We utilized a normal versus tumorigenic tissue paired microarray analysis of gene expression from nine animals (Table 1) to produce a robust genetic signature of OCCC caused by ARID1A and PIK3CA mutations in a newly reported mouse model that genetically resembles the human disease [2]. Our mouse model and the transcriptional data set described herein represent a novel resource to study one of the most deadly forms of OCCC [1]. We described the sample quality (Fig. 1), sample processing (Fig. 2A), and technical details to reproduce the analysis of differential gene expression from the matched tissues. Additional analyses and comparisons to other cancer datasets can be used in future investigations examining molecular changes that promote OCCC tumorigenesis. The distance between tumor and healthy tissue both by PCA (Fig. 2B, C) and HC (Fig. 2D) revealed a strong differentiating transcriptional profile of this cancer model (Fig. 2E) that when combined with a human dataset (NCBI GEO: GSE6008) [6] implicated interleukin 6 signaling in OCCC pathophysiology [2]. Chandler et al. [2] and this *Data in Brief* demonstrate the importance of identifying and functionally analyzing tumor mutations and the pathways that are disrupted to uncovering new targets for cancer therapies.

Acknowledgments

R.L.C. was supported by an American Cancer Society Postdoctoral Fellowship (PF-09-116-01-CCG) and an Ann Schreiber Mentored Investigator Award (258831) from the Ovarian Cancer Research Fund. This work was supported by an NIH grant to T.M. (HD03665) and was supported in part by an NIH grant to J.C.S. (GM061728).

References

[1] R.M. Glasspool, I.A. McNeish, Clear cell carcinoma of ovary and uterus. Curr. Oncol. Rep. 15 (2013) 566–572, http://dx.doi.org/10.1007/s11912-013-0346-0.
[2] R.L. Chandler, J.S. Damrauer, J.R. Raab, J.C. Schisler, M.D. Wilkerson, J.P. Didion, et al., Coexistent ARID1A–PIK3CA mutations promote ovarian clear-cell tumorigenesis through pro-tumorigenic inflammatory cytokine signalling. Nat. Commun. 6 (2015) 6118, http://dx.doi.org/10.1038/ncomms7118.
[3] K.V. Clark-Knowles, K. Garson, J. Jonkers, B.C. Vanderhyden, Conditional inactivation of Brca1 in the mouse ovarian surface epithelium results in an increase in preneoplastic changes. Exp. Cell Res. 313 (2007) 133–145, http://dx.doi.org/10.1016/j.yexcr.2006.09.026.
[4] Affymetrix, Affymetrix GeneChip® Gene and Exon Array Whitepaper Collection, Httpmedia.Affymetrix.comsupporttechnicalwhitepapersexongenearraysqawhitepaper.Pdf. (n.d.) 18. http://media.affymetrix.com/support/technical/whitepapers/exon_gene_arrays_qa_whitepaper.pdf (accessed September 4, 2015).
[5] L.M.T. Eijssen, M. Jaillard, M.E. Adriaens, S. Gaj, P.J. de Groot, M. Müller, et al., User-friendly solutions for microarray quality control and pre-processing on ArrayAnalysis.org. Nucleic Acids Res. 41 (2013) W71–W76, http://dx.doi.org/10.1093/nar/gkt293.
[6] D.R. Schwartz, S.L.R. Kardia, K.A. Shedden, R. Kuick, G. Michailidis, J.M.G. Taylor, et al., Gene expression in ovarian cancer reflects both morphology and biological behavior, distinguishing clear cell from other poor-prognosis ovarian carcinomas. Cancer Res. 62 (2002) 4722–4729.