



Data in Brief

Genomic analysis to define molecular basis of aggressiveness in a mouse model of oral cancer



Varun Chalivendra, Krishna Latha Kanchi, Michael D. Onken, Ashley E. Winkler, Elaine Mardis, Ravindra Uppaluri*

Department of Otolaryngology and The Genome Institute, Washington University School of Medicine in Saint Louis, Department of Otolaryngology, United States

ARTICLE INFO

Article history:

Received 7 November 2014

Received in revised form 11 November 2014

Accepted 11 November 2014

Available online 18 November 2014

Keywords:

Mouse models

Metastasis

Comparative microarrays

Gene signatures

ABSTRACT

To investigate the molecular basis underlying aggressive behavior in oral squamous cell carcinoma (OSCC), our laboratory developed a carcinogen-induced mouse oral cancer (MOC) cell line model that encompasses the growth and metastasis spectrum of its human counterpart. We performed next-generation sequencing (NGS) and gene expression microarray profiles to explore the genomic and transcriptional backgrounds of the differential MOC line phenotypes, as well as, the cross-species relevance of the model. Here we describe the comparative analysis of NGS (www.ncbi.nlm.nih.gov/biosample?LinkName=bioproject_biosample_all&from_uid=247825) and expression microarray (www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50041) data from the MOC lines and corresponding human data, as described in our recent publication [1].

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Specifications

Organism/cell line/tissue	<i>Mus musculus</i>
Sex	N/A
Sequencer or array type	NGS: Illumina HiSeq2000 Microarray: Illumina MouseRef-8
Data format	NGS: raw data in SRA files Microarray: raw data in TXT files
Experimental factors	NGS: mouse oral squamous cell carcinoma cell lines Microarray: mouse oral squamous cell carcinoma cell lines, metastatic lymph node, control primary C57BL/6 splenocytes
Experimental features	<ol style="list-style-type: none"> 1. NGS to determine mutational landscape of mouse oral cancer lines 2. Microarray to determine transcriptional landscape of mouse oral cancer lines 3. Comparative analysis of genomic and transcriptional data between aggressive and indolent mouse oral cancer lines 4. Comparative analysis between mouse model and published human sequencing and microarray data
Consent	N/A
Sample source location	Saint Louis, MO, United States

Direct link to deposited data

NGS: www.ncbi.nlm.nih.gov/biosample?LinkName=bioproject_biosample_all&from_uid=247825

Microarray: www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50041

Experimental design, materials and methods

Cell lines

We have previously described a 7,12 dimethylbenzanthracene (DMBA)-induced mouse cell line model of OSCC where individual lines display fixed *in vivo* phenotypes [2]. We classified these lines into indolent or aggressive subtypes based on their *in vivo* growth patterns in syngeneic mice. Indolent lines formed primary tumors at orthotopic and heterotopic injection sites but did not display evidence of metastasis. By contrast the aggressive lines all displayed regional lymph node as well as occasional lung metastases. Further genotyping analysis revealed that the three aggressive lines were related with MOC2-7 and MOC2-10 being derived from the MOC2 line.

Sequencing

MOC lines were grown in defined media and used for genomic DNA preparation, while CXCR3^{-/-} DNA was generated from mouse

* Corresponding author at: Washington University School of Medicine, Department of Otolaryngology, Box 8115, 660 South Euclid Avenue, St. Louis, Missouri 63110, USA. Tel.: +1314 362 6599; fax: +1314 362 7522.

E-mail addresses: uppalurr@wustl.edu (R. Uppaluri).

splenocytes (Qiagen DNeasy Blood & Tissue Kit). NGS was performed at The Genome Institute at Washington University School of Medicine with their standard platforms and resulted in excellent depth of coverage. A critical point in the analysis was due to the lack of control tissue available from the original mice bearing the parental tumors. Thus, mutation reads were normalized to the reference C57BL/6 genome and a genomic database of 17 commonly used strains of inbred laboratory mice.

Human mutational data from all oral squamous cell carcinoma (OSCC) cases were acquired from published data, The Cancer Genome Atlas (TCGA), The Broad Institute platform, and from cBioPortal (<http://www.cbioportal.org/public-portal/>).

Sequencing quality control and analysis

Analysis of sequencing data is described in detail in our publication [1]. We excluded the common and well-known variants using the dbSNP filter, and the novel variants were manually reviewed to decrease the false discovery rate. With the goal of further understanding the signatures of metastasis and identifying immune targets, we performed VarScan copy number analysis and recurrent gene analysis. We also performed clonality analysis using the point mutations to reveal the subclonal architecture within the cell lines.

Growth phenotype-specific mutations, and comparative analysis between aggressive versus indolent MOC lines, and lymph node metastasis negative versus lymph node positive TCGA OSCC samples, were established in Microsoft Excel using IF and VLOOKUP commands. The prevalence of specific mutations in human OSCC cases from TCGA was determined using cBioPortal.

Microarray

Primary oral keratinocytes were generated by harvesting oral mucosa from C57BL/6 mice. The mucosa was digested overnight in Dispase (Roche) and epithelial layers were then microdissected. Epithelium was then briefly digested with trypsin (HyClone) and single cell suspensions were generated. These cells were then grown in specialized media (CellNTec) until near confluent cultures were obtained. Cells were subsequently grown for 24 h in MOC cell line media. Total RNA was generated from 3 to 4 independent cultures of MOC lines and two independent cultures of primary oral keratinocytes using the RNeasy Kit (Qiagen). RNA quality was assessed using Agilent Nano Chips and samples with RNA integrity values of 8–10 were subsequently submitted for gene expression profiling using Illumina MouseRef-8 Expression BeadChips. Raw expression data were normalized using the cubic spline method in GenomeStudio 2011.1.

Human OSCC expression data were acquired from four independent published databases (UW/FHCRC, MD Anderson [3], University of Pennsylvania [4] and TCGA). Only proven human papilloma virus negative cases were included. Tumor stage or regional lymph node metastasis status were used to categorize aggressive versus indolent cases.

Microarray analysis

The normalized expression data were imported into Partek Genomics Suite 6.6. Further analysis is described in our recent publication and resulted in the identification of gene sets representing a mouse signature of aggressiveness [1]. Iterative and tandem gene set enrichment analysis (GSEA) was used to find overlap of the mouse signature genes in three human datasets, and to eliminate mouse-specific transcripts. This process trimmed the number of cross-species relevant transcripts to 188 genes.

Discussion

We describe here our two-fold genomic approach, of exome sequencing and expression microarray, to expose mutational and transcriptional alterations and to delineate the genetic basis of aggressive growth in our murine OSCC model. By performing a comparative analysis with human genomic and gene expression data, we found (1) overlap with the most commonly mutated genes in human HNSCC, (2) new potential driver mutations, (3) novel candidate genes with transcriptional changes associated with aggressive growth, and (4) a cross-species conserved 118-gene signature predictive of aggressive disease and metastasis in OSCC. These data support the use of carcinogen-induced mouse oral cancer cell lines as a high-fidelity model of the human counterpart.

Conflict of interest

The authors have no conflicts of interest.

Acknowledgments

This work was supported by The Genome Institute at Washington University for sequencing and analysis via a center-initiated project (NIH/NHGRI 5U54HG003079; Richard Wilson, principal investigator). NIH P30DC04665 supports the RCAVS Histology Core (Brian Faddis and Dorothy Edwards). Additional support was provided by the Siteman Cancer Frontier Fund, Tissue Procurement Core (NCI Cancer Center Support Grant #P30 CA91842) and GTAC (NCCR ICTS/CTSA UL1TR000448 and NCI-CA91842). M.D. Onken is supported by NIGMS #R01GM38542 (John Cooper, principal investigator).

References

- [1] M.D. Onken, et al., A surprising cross-species conservation in the genomic landscape of mouse and human oral cancer identifies a transcriptional signature predicting metastatic disease. *Clin. Cancer Res.* 20 (11) (2014).
- [2] N.P. Judd, A.E. Winkler, O. Murillo-Sauca, J.J. Brotman, J.H. Law, J.S. Lewis Jr., et al., ERK1/2 regulation of CD44 modulates oral cancer aggressiveness. *Cancer Res.* 72 (2012) 365–374.
- [3] P. Lohavanichbutr, J. Houck, D.R. Doody, P. Wang, E. Mendez, N. Futran, et al., Gene expression in uninvolved oral mucosa of OSCC patients facilitates identification of markers predictive of OSCC outcomes. *PLoS ONE* 7 (2012) e46575.
- [4] R.K. O'Donnell, M. Kupferman, S.J. Wei, S. Singhal, R. Weber, B. O'Malley, et al., Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity. *Oncogene* 24 (2005) 1244–1251.