Contents lists available at ScienceDirect

Data in Brief

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

Data Article

A dataset of chronic nicotine-induced genes in breast cancer cells



Samson Mugisha, Shreyas Labhsetwar, Devam Dave, Richard Klemke, Jay S. Desgrosellier*

Department of Pathology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

ARTICLE INFO

Article history: Received 12 March 2025 Revised 14 April 2025 Accepted 15 April 2025 Available online 22 April 2025

Dataset link: A dataset of chronic nicotine-induced genes in HCC38 breast cancer cells (Original data)

Keywords: Tobacco products Smoking Nicotine Breast cancer Single-cell RNA-seq Gene expression Inflammation

ABSTRACT

These data show the differentially expressed genes (DEG) from HCC38 breast cancer cell line chronically exposed to nicotine versus vehicle control. Additional data is also provided from dynamic trajectory analysis, identifying the most dynamic genes due to chronic nicotine treatment. To produce this dataset, we first performed single cell RNA sequencing from HCC38 cells chronically treated with vehicle or nicotine, followed by scanpy analysis to yield 6 discrete cell clusters at conservative resolution. We then evaluated differential gene expression between chronic nicotine and control cells for each individual cluster or in the whole sample using PyDESeq2. For dynamic trajectory analysis, Velocyto (0.6) was used to estimate the spliced and unspliced counts for each gene between chronic nicotine-treated cells and vehicle, allowing computation of gene velocities. These data are useful for analysing the expression of individual genes or gene velocities either in the whole sample or in the different clusters identified. Since the HCC38 cell line used in these experiments is heterogeneous, including cells with features of stem-like, luminal progenitor-like and more differentiated cells, this dataset allows examination of the conserved as well as disparate gene expression effects of nicotine in different breast cancer cell types. Our dataset has a great potential

* Corresponding author.

E-mail address: jdesgrosellier@ucsd.edu (J.S. Desgrosellier).

https://doi.org/10.1016/j.dib.2025.111573



^{2352-3409/© 2025} The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

for re-use given the recent surge in interest surrounding the role tobacco-use plays in breast cancer progression. © 2025 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY license

(http://creativecommons.org/licenses/by/4.0/)

Specifications Table

Subject	Cancer Research
Specific subject area	Our work focuses on the role of tobacco products in breast cancer progression
	and metastasis.
Data format	Raw
Type of data	Excel spreadsheet
Data collection	Data were collected after comparing expression of each gene in the chronic
	nicotine treated cells to the levels found in the corresponding cell cluster or
	whole sample from the vehicle-treated cells with PyDESeq2. The provided Log2
	fold changes for each gene (DEG) are listed in the Excel spreadsheet deposited.
Data source location	Institution: University of California, San Diego
	City/Town/Region: La Jolla, CA
	Country: USA
	Latitude and longitude: 32.876328, -117.236067
Data accessibility	Repository name: UC San Diego Library Digital Collections
	Raw data in NCBI: https://www.ncbi.nlm.nih.gov/sra/SRX25630993
	https://www.ncbi.nlm.nih.gov/sra/SRX25630994
	Data identification number: 10.6075/J04×585F
	Direct URL to data: https://library.ucsd.edu/dc/object/bb71464553
Related research article	Mugisha, S., Baba, S.A., Labhsetwar, S. et al. S100A8/A9 innate immune
	signaling as a distinct mechanism driving progression of smoking-related
	breast cancers. Oncogene (2025). 10.1038/s41388-025-03276-5

1. Value of the Data

- The data we provided here added value to our original research article by providing an unbiased view of the genes upregulated by nicotine in HCC38 breast cancer cells. This allowed us to assess potential associations between nicotine's effects and particular cell types. Using these findings, we were able to narrow our focus to cytokine expression as a particular class of genes impacted by nicotine and assess their effects on tumor progression.
- The Log2 fold changes (Log2FC) provided in this dataset are useful for comparing the relative effects of nicotine on different breast cancer cell populations.
- This dataset may be useful to anyone associated with breast cancer research. The mechanisms underpinning smoking's effects on breast cancer progression are little understood. Additionally, other sources of nicotine, such as e-cigarettes are growing in use and popularity, particularly among young adults. Thus, there is a great need to further elucidate the mechanisms associated with nicotine's effects on tumor progression, also this dataset may aid future studies of this topic.
- These data can also be used for further insights into associations between chronic nicotine exposure and numerous additional variables, including cell-of-origin, mutational status, copy number alterations (CNA), subtype, etc.

2. Background

Tobacco-use is associated with an increased risk of breast cancer progression. In addition to enhancing the risk of developing breast cancer, strong clinical evidence also links smoking with a significantly worse outcome in patients already diagnosed with the disease [1,2]. Despite the

discovered clinical link between tobacco-use and breast cancer progression, the cause is still relatively unknown. This is mainly due to a lack of understanding of the mechanisms responsible for smoking's effects. In fact, the level at which nicotine directly alters breast cancer cell behavior remains largely unexplored. Previous studies of nicotine's effect in breast cancer cells have yielded only modest and transient alterations on EMT markers and cell proliferation [3–6], leading some to speculate that its primary impact may be on the tumor microenvironment [7–9]. In our published original article, we used unbiased approach to generated a significant difference in innate immune gene expression reprogramming due to chronic nicotine, providing a novel molecular basis of smoking-related breast cancers.

3. Data Description

The deposited dataset consists of one Excel spreadsheet listing 8 different worksheet tabs labelled and described as follow:

DEG – The top of each column of the sheet-tabs #2–7 indicates Log2FC and associated adjusted P-values of corresponding downregulated genes in vehicle compared to chronic nicotinetreated cells in each cluster or the overall sample. Presented Log2FC scores for respective genes were obtained via differential gene expression by using PyDESeq2, and data are ranked by Log2FC with a cut-off of 0.5. Additional columns include baseMean, representing the average of the normalized count values across all samples, log2 fold change standard error (lfcSE), Wald statistic (stat) representing Log2FC divided by its standard error, *p*-value and adjusted p-value (padj) generated using the Wald statistical test. The resulting data were processed into Excel format.

Trajectory analysis – The worksheet tab #8 identifies the list of the most dynamic genes ranked from high to low with their corresponding clusters indicated on the top of each column in chronic nicotine-treated cells versus control-vehicle. For dynamic trajectory analysis, Velocyto was used to generate gene-specific velocities and associated genes, before drafting the resulting data into Excel spreadsheet.

Genes are ranked in order of velocities from high to low.

4. Experimental Design, Materials and Methods

HCC38 cells used in this study were purchased from ATCC (Manassas, VA, USA). The cells were cultured in complete DMEM medium and maintained in a humid incubator at 37 0 C supplied with 5 % CO2. We seeded 3 × 10⁶ cells into 10 cm petri-dishes before adding 100 nM of nicotine (N3876-5ML, Sigma-Aldrich, Saint Louis, MO, USA) every other day for 14 days, with passages performed every 3 days. In a second dish of cells, we concurrently administered an equal volume of vehicle (ethanol) to generate appropriate controls. Cells were passaged 3 times post nicotine treatment prior to library preparation. To prepare sc-RNA-seq libraries, we used 1 × 10⁴ HCC38 cells chronically treated with vehicle or nicotine following the 10X Genomics Single Cell 3' V2 Reagent Kits User Guide PN-120233 (10X Genomics, Pleasanton, CA, USA). Libraries were then submitted to the UCSD Institute for Genomic Medicine (IGM) Genomics Core for validation of quality and sequencing on a NovaSeq 6000 (Illumina, San Diego, CA, USA).

To process the sc-RNA-seq data, we used the Cell Ranger toolkit (version 3.1.0) provided by 10x Genomics. We performed dimension reduction and unsupervised clustering according to the standard workflow in Scanpy [10]. Principal component analysis (PCA) was then performed on top 8000 highly variable genes (HVGs) to reduce noise, followed by cell clustering using Leiden algorithm [11]. For differential gene expression analysis, we applied PyDESeq2, a re-implementation of DESeq2 in Python [12]. PyDESeq2 estimates variance-mean dependence in count data from sequencing assays and tests for differential expression based on a model using the negative binomial distribution. P < 0.05 was considered statistically significant. Additionally, we performed dynamic trajectory analysis using Velocyto (0.6) to estimate the spliced and unspliced counts from the pre-aligned bam files. The dynamical velocity model from scVelo (0.2.2) was applied to calculate RNA velocity, latent time, root, and terminal states. The output was then processed into Excel format and saved as final results files.

Limitations

None.

Ethics Statement

We confirm that the authors have read and follow the ethical requirements for publication in Data in Brief and that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Data Availability

A dataset of chronic nicotine-induced genes in HCC38 breast cancer cells (Original data) (University of California San Diego Library Digital Collections)

CRediT Author Statement

Samson Mugisha: Conceptualization, Methodology, Formal analysis, Writing – original draft; Shreyas Labhsetwar: Methodology, Formal analysis, Software; Devam Dave: Methodology, Formal analysis, Software; Richard Klemke: Methodology, Formal analysis, Software; Jay S. Desgrosellier: Conceptualization, Methodology, Formal analysis, Visualization, Investigation, Supervision, Funding acquisition.

Acknowledgements

This work was supported by funding from the Tobacco-Related Disease Research Program [Grant #T32IR4741 (to J.S.D.)]; and the California Breast Cancer Research Program [Grant #B28IB5479 (to J.S.D.)].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- N. Saquib, M.L. Stefanick, L. Natarajan, J.P. Pierce, Mortality risk in former smokers with breast cancer: pack-years vs. smoking status, Int. J. Cancer 133 (2013) 2493–2497, doi:10.1002/ijc.28241.
- [2] J.P. Pierce, R.E. Patterson, C.M. Senger, S.W. Flatt, B.J. Caan, L. Natarajan, S.J. Nechuta, E.M. Poole, X.O. Shu, W.Y. Chen, Lifetime cigarette smoking and breast cancer prognosis in the after breast cancer pooling project, J. Natl. Cancer Inst. 106 (2014) djt359, doi:10.1093/jnci/djt359.
- [3] K. Kumari, B. Das, A. Adhya, S. Chaudhary, S. Senapati, S.K. Mishra, Nicotine associated breast cancer in smokers is mediated through high level of EZH2 expression which can be reversed by methyltransferase inhibitor DZNepA, Cell Death Dis. 9 (2018) 152, doi:10.1038/s41419-017-0224-z.

- [4] P. Guha, G. Bandyopadhyaya, S.K. Polumuri, S. Chumsri, P. Gade, D.V. Kalvakolanu, H. Ahmed, Nicotine promotes apoptosis resistance of breast cancer cells and enrichment of side population cells with cancer stem cell-like properties via a signaling cascade involving galectin-3, alpha9 nicotinic acetylcholine receptor and STAT3, Breast Cancer Res. Treat. 145 (2014) 5–22, doi:10.1007/s10549-014-2912-z.
- [5] P. Dasgupta, W. Rizwani, S. Pillai, R. Kinkade, M. Kovacs, S. Rastogi, S. Banerjee, M. Carless, E. Kim, D. Coppola, E. Haura, S. Chellappan, Nicotine induces cell proliferation, invasion and epithelial-mesenchymal transition in a variety of human cancer cell lines, Int. J. Cancer 124 (2009) 36–45, doi:10.1002/ijc.23894.
- [6] C.H. Lee, C.S. Huang, C.S. Chen, S.H. Tu, Y.J. Wang, Y.J. Chang, K.W. Tam, P.L. Wei, T.C. Cheng, J.S. Chu, L.C. Chen, C.H. Wu, Y.S. Ho, Overexpression and activation of the alpha9-nicotinic receptor during tumorigenesis in human breast epithelial cells, J. Natl. Cancer Inst. 102 (2010) 1322–1335, doi:10.1093/jnci/djq300.
- [7] P.C. Chen, W.Y. Lee, H.H. Ling, C.H. Cheng, K.C. Chen, C.W. Lin, Activation of fibroblasts by nicotine promotes the epithelial-mesenchymal transition and motility of breast cancer cells, J. Cell Physiol. 233 (2018) 4972–4980, doi:10. 1002/jcp.26334.
- [8] A. Tyagi, S. Sharma, K. Wu, S.Y. Wu, F. Xing, Y. Liu, D. Zhao, R.P. Deshpande, R.B. D'Agostino Jr., K. Watabe, Nicotine promotes breast cancer metastasis by stimulating N2 neutrophils and generating pre-metastatic niche in lung, Nat. Commun. 12 (2021) 474, doi:10.1038/s41467-020-20733-9.
- [9] K. Takada, S. Kashiwagi, Y. Asano, W. Goto, K. Takahashi, H. Fujita, T. Takashima, S. Tomita, K. Hirakawa, M. Ohira, Clinical verification of the relationship between smoking and the immune microenvironment of breast cancer, J. Transl. Med. 17 (2019) 13, doi:10.1186/s12967-019-1773-y.
- [10] F.A. Wolf, P. Angerer, F.J. Theis, SCANPY: large-scale single-cell gene expression data analysis, Genome Biol. 19 (2018) 15, doi:10.1186/s13059-017-1382-0.
- [11] V.A. Traag, L. Waltman, N.J. van Eck, From Louvain to Leiden: guaranteeing well-connected communities, Sci. Rep. 9 (2019) 5233, doi:10.1038/s41598-019-41695-z.
- [12] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550, doi:10.1186/s13059-014-0550-8.