UNDERGRADUATE GENOMICS DATA ANALYSIS

Exploratory Analysis of DDX3Y and XIST Expression by Reported Sex in Cancer Cell Lines

Chelsea Doop

Arizona State University | School of Life Sciences

Data source: Cancer Cell Line Encyclopedia (CCLE)
Tools: R, RStudio, dplyr, ggplot2, knitr, and TinyTeX
Project type: Exploratory undergraduate genomics analysis

> **Archive note.** This report documents an analysis originally completed for an undergraduate genomics research course. The original R code, source data file, and CCLE release identifier are no longer available. The report therefore presents only the workflow, figures, and findings documented in the archived course project. It is not presented as a fully reproducible or publication-ready study.

Project Summary

This exploratory project compared expression of DDX3Y and XIST across cancer cell lines in the Cancer Cell Line Encyclopedia (CCLE). The analysis examined whether expression of these two sex-linked genes corresponded with the sex reported in the CCLE metadata. In the surviving visualizations, high DDX3Y expression occurred predominantly among male-labeled cell lines, while high XIST expression occurred predominantly among female-labeled cell lines. Both visualizations also contained low-expression groups and observations that did not follow the predominant pattern.

Background

The original project used the CCLE as a source of cancer cell-line gene-expression data (Barretina et al., 2012). The project was motivated by published research on sex differences in cancer mechanisms and on the possible relevance of X-chromosome inactivation and Y-chromosome loss to cancer biology (Rubin et al., 2020; Dunford et al., 2017; Abdel-Hafiz et al., 2023).

DDX3Y was selected as the Y-linked gene of interest, while XIST was selected because of its association with X-chromosome inactivation. The project compared their expression with the reported sex of the CCLE cell lines and explored whether the observed patterns might be useful when examining cell lines with missing or unexpected sex-associated expression.

Research Question

How do DDX3Y and XIST expression patterns differ between CCLE cancer cell lines categorized as male or female, and where do the observed expression profiles diverge from the reported metadata?

Methods

Data source and variables

Gene-expression and cell-line metadata were obtained from the CCLE. The archived figures indicate that the analysis included DDX3Y expression, XIST expression, reported sex, reported age, and a pathology category distinguishing primary from metastatic samples. Expression values were displayed as log counts in the original analysis. The specific CCLE release, exact normalization procedure, and sample inclusion criteria were not preserved and therefore cannot be reported with greater precision.

Computational workflow

The analysis was conducted in RStudio. The dplyr package was used to organize and subset the CCLE data, and ggplot2 was used to create scatterplots and violin plots. knitr and TinyTeX supported R Markdown report generation. The archived report does not include formal statistical testing or the original code. The results below therefore describe the visual patterns reported in the original project.

Results

DDX3Y expression across reported age

DDX3Y expression formed two visually distinct ranges. Female-labeled cell lines were concentrated in the lower-expression range. Male-labeled cell lines appeared in both the lower- and higher-expression ranges, with the higher range composed almost entirely of male-labeled observations. The plot also displays reported age and whether the cell line was categorized as primary or metastatic. Because the original report did not include a formal comparison of these categories, the figure is presented descriptively.

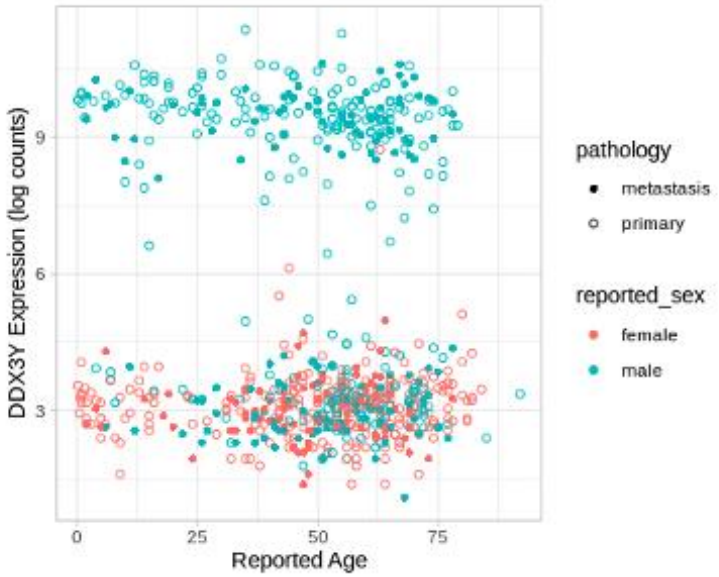

Figure 1. DDX3Y expression plotted against reported age, with color indicating reported sex and point style indicating primary or metastatic pathology. High DDX3Y expression was concentrated among male-labeled cell lines, while lower expression occurred in both groups.

DDX3Y expression by reported sex

The violin plot reinforced the sex-associated pattern observed in the scatterplot. Female-labeled cell lines were concentrated at lower DDX3Y expression levels, although a small number of higher observations were present. Male-labeled cell lines showed one lower-expression group overlapping the female-labeled distribution and a second group at substantially higher expression. The original project considered whether low DDX3Y expression in male-labeled cell lines might be related to Y-chromosome loss, but chromosome loss was not measured in this analysis.

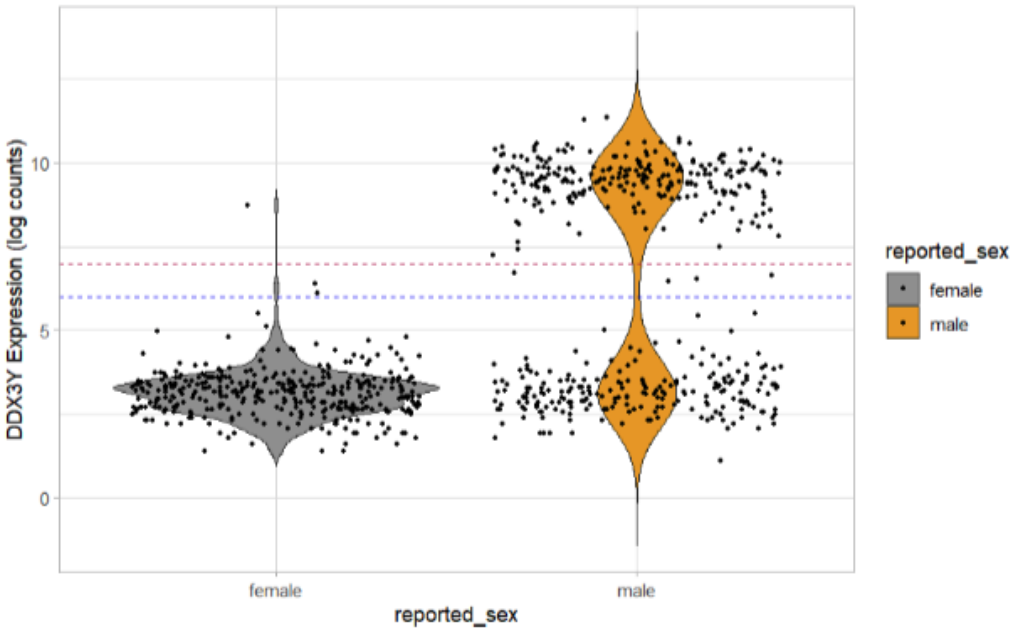

Figure 2. Distribution of DDX3Y expression by reported sex. Male-labeled cell lines displayed both low- and high-expression modes, whereas female-labeled cell lines were concentrated in the lower range. Dashed reference lines are retained from the original visualization; their derivation was not preserved.

XIST expression by reported sex

XIST expression showed a complementary but similarly incomplete pattern. Many female-labeled cell lines formed a higher-expression group, while most male-labeled cell lines were concentrated at lower expression. A second low-expression group was also apparent among female-labeled cell lines, and several male-labeled observations extended into the higher range. The groups therefore showed a predominant pattern as well as overlap and exceptions.

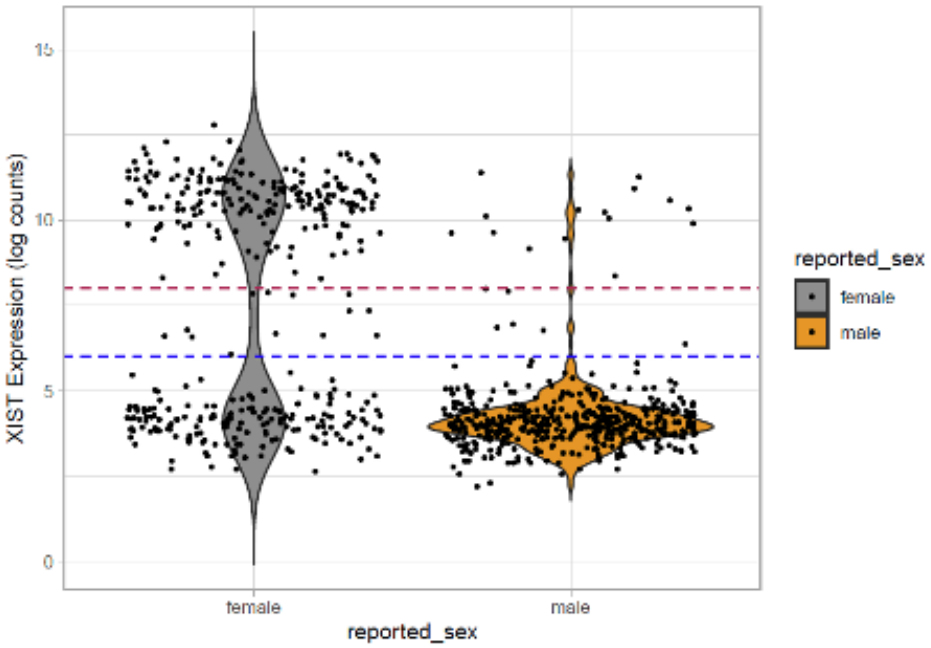

Figure 3. Distribution of XIST expression by reported sex. Higher XIST expression occurred predominantly among female-labeled cell lines, but both categories included discordant observations. Dashed reference lines are retained from the original visualization; their derivation was not preserved.

Discussion

Together, the DDX3Y and XIST visualizations showed sex-associated expression patterns in the CCLE cell lines. High DDX3Y expression appeared primarily among male-labeled cell lines, whereas high XIST expression appeared more often among female-labeled cell lines. Neither plot separated the reported groups completely.

The original project proposed that reduced DDX3Y expression might indicate loss of Y-chromosome material. Because the project analyzed gene expression and did not directly measure chromosome loss, that possibility was not tested. The results are therefore limited to the expression patterns displayed in the figures.

Limitations

- The original R code, processed data, dataset release, and filtering decisions are unavailable, so the analysis cannot currently be reproduced exactly.
- The archived report does not provide sample counts, formal statistical testing, or the calculations used to create the dashed reference lines.
- The project measured gene expression and did not directly test for Y-chromosome loss.

- The results are descriptive and are limited to the surviving CCLE figures.

Conclusion

This undergraduate project used R and CCLE data to explore how expression of two sex-linked genes varied across cancer cell lines. In the surviving figures, DDX3Y and XIST expression differed by reported sex, but the groups also overlapped. These findings document the visual patterns identified in the original course project and should be interpreted as an exploratory analysis.

Sources Cited in the Original Course Report

Abdel-Hafiz, H. A., Schafer, J. M., Chen, X., et al. (2023). Y chromosome loss in cancer drives growth by evasion of adaptive immunity. Nature, 619(7970), 624-631. https://doi.org/10.1038/s41586-023-06234-x

Barretina, J., Caponigro, G., Stransky, N., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature, 483(7391), 603-607. https://doi.org/10.1038/nature11003

Cechova, M., Vegesna, R., Tomaszkiewicz, M., et al. (2020). Dynamic evolution of great ape Y chromosomes. Proceedings of the National Academy of Sciences, 117(42), 26273-26280. https://doi.org/10.1073/pnas.2001749117

Dunford, A., Weinstock, D. M., Savova, V., et al. (2017). Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. Nature Genetics, 49(1), 10-16. https://doi.org/10.1038/ng.3726

Ross, M. T., Grafham, D. V., Coffey, A. J., et al. (2005). The DNA sequence of the human X chromosome. Nature, 434(7031), 325-337. https://doi.org/10.1038/nature03440

Rubin, J. B., Lagas, J. S., Broestl, L., et al. (2020). Sex differences in cancer mechanisms. Biology of Sex Differences, 11, 17. https://doi.org/10.1186/s13293-020-00291-x