

The New UCSC Cancer Genomics Browser

Brian Craft¹, Teresa Swatloski², Mary Goldman³, Kyle Ellrott⁴,
Singer Ma⁵, Chris Wilks⁶, Josh Stuart⁷, David Haussler⁸, and Jingchun Zhu⁹

UC Santa Cruz

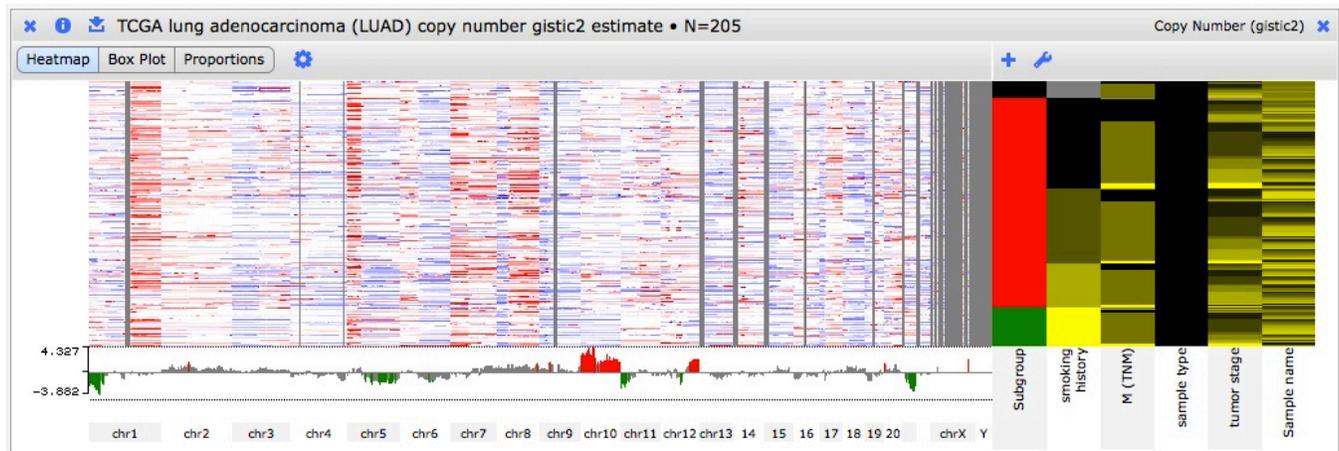


Figure 1: UCSC Cancer Genomics Browser.

ABSTRACT

The UCSC Cancer Genomics Browser comprises a suite of web-based tools to integrate, visualize, and analyze cancer genomics and clinical data. Users can order, filter, aggregate, classify, and display data interactively based on any given feature set including clinical features, genomic signatures/profiles, annotated biological pathways, and user-contributed collections of genes. Integrated standard statistical tools provide dynamic quantitative analysis within all available datasets. A new UCSC Cancer Genomics Browser was released July 2012 with immensely enhanced user interface and functionality, new user documentation, and an updated online tutorial.

Keywords: cancer, genomics, genome, human genome, gene expression, pathway, functional genomes.

Index Terms: H.5.2 [Information interfaces and presentation] User Interfaces; H.3.5 [Information storage and retrieval] Online Information Services— Web-based services

¹ e-mail: craft@soe.ucsc.edu

² e-mail: swat@soe.ucsc.edu

³ e-mail: mary@soe.ucsc.edu

⁴ e-mail: kellrott@soe.ucsc.edu

⁵ e-mail: singer@soe.ucsc.edu

⁶ e-mail: cwilks@soe.ucsc.edu

⁷ e-mail: jstuart@soe.ucsc.edu

⁸ e-mail: haussler@soe.ucsc.edu

⁹ e-mail: jzhu@soe.ucsc.edu

1 INTRODUCTION

Cancer has many different molecular mechanisms to disrupt cellular pathways, which result in uncontrolled cell proliferation [1-3]. Fortunately, development of high-throughput genomic technologies in recent years has greatly increased the amount of data available to researchers to investigate these mechanisms. Not only have the number of patients which now have genomic data increased, but the amount and type of data available per patient has also grown. In addition to the genomic information, each patient and their tumor typically have associated clinical information.

In spite of this wealth of data, analysis of the cancer genome can be a challenge due to the limitations in current technologies to visualize, integrate, compare and analyze cancer genomics data. These data, and the conclusions they support, must be presented in a coherent system for display and analysis as well as be accessible to the scientific and medical communities. The UCSC Cancer Genomics Browser was developed to display these expanding data sources in an integrative, interactive and versatile display and help facilitate comprehensive analysis of cancer genomics and its associated clinical data.

2 THE CANCER BROWSER

The UCSC Cancer Genomics Browser is a web-based tool to integrate, visualize and analyze genomics and clinical information (Figure 1). The browser displays whole-genome views of experimental measurements for multiple samples alongside their associated clinical information. Multiple datasets can be viewed simultaneously allowing comparison across studies and different data types, such as gene expression and copy number variation.

The browser provides interactive and dynamic views of the data from whole-genome to base-pair scale resolution, as well as zooming to a subset of samples. Users can interactively group samples by common clinical features such as response to chemotherapy, or by genomic signatures/profiles that predict response to a drug. Viewing genomic data by genes allows users to easily see functional changes to the genome as well as see trends across pathways of genes. Several statistical tools are available making it possible to obtain quantitative results dynamically. Additionally, the Tumor Image Viewer, based on Google Maps, allows users to interactively view slides of tumor tissue samples (Figure 2).

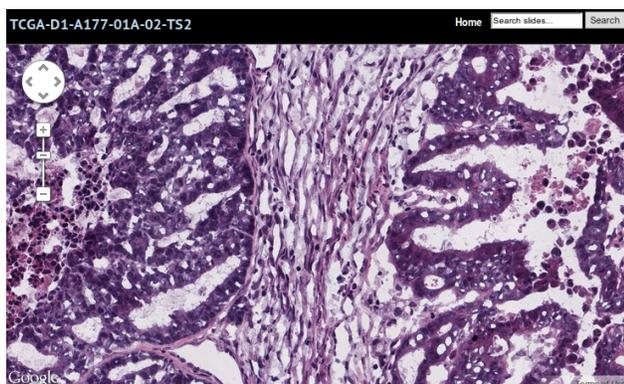


Figure 2: UCSC Tumor Image Viewer.

3 AVAILABLE DATA

The browser currently contains a total of 355 datasets, including 201 public-tier datasets from 22 TCGA (The Cancer Genome Atlas) cancer projects, including copy number variation, DNA methylation profiles, transcriptome expression levels, protein expression levels, and integrated gene activity level inferred using the PARADIGMLITE method [4].

We also have data from the Cancer Cell Line Encyclopedia project, public SU2C (Stand Up To Cancer) breast cancer data and 43 other published studies. A controlled access mechanism is also available for private data, restricting the display of this data to certain users. It currently supports the SU2C breast cancer dream team, ISPY 2 TRIAL (Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging And moLecular Analysis 2), LINCS (Library of Integrated Network-based Cellular Signatures) project, and other projects. Additionally, the Tumor Image Viewer has 2748 TCGA images from 3 cancer types.

4 NEWEST FEATURES

In the past year we have completely redesigned the interface in an effort to greatly increase the usability of our website. In particular we have focused on making the controls more obvious and intuitive, improving the readability of drawn text, and improving the responsiveness of the site by using more modern web technologies. It is also now possible for users to download the processed data used to display a dataset. In tandem with the new interface, the browser now has an interactive tutorial that highlights the new functionality.

5 FUTURE WORK

Our plans for the next year include providing users a way to bookmark a view both for themselves and to share with others.

We have already made progress on allowing users to visualize their own patient-identifiable or unpublished data in a secure manner. We will also develop new viewing capabilities that integrate data across tracks for multianalyte data, for example, view copy number, gene expression, DNA methylation, and clinical data from the same set of samples side by side. Lastly, we are planning on displaying mutation data in a new heatmap visualization.

ACKNOWLEDGEMENTS

We would like to thank NCI/NIH, NHGRI, American Association for Cancer Research (SU2C), UCSF Comprehensive Cancer Center, and California Institute for Quantitative Biosciences (QB3) for funding this project. We would also like to thank Eric Weiler and the UCSC Genome Browser for their support as well as the two anonymous reviewers for their feedback.

REFERENCES

- [1] W.C. Hahn and R.A. Weinberg. Modelling the molecular circuitry of cancer. *Nature Reviews Cancer*, 2: 331–341, 2002.
- [2] D. Hanahan and R.A. Weinberg. Hallmarks of Cancer: The Next Generation. *Cell*, 144(5): 646-674, 2011.
- [3] B. Vogelstein and K.W. Kinzler. Cancer genes and the pathways they control. *Nature Medicine*, 10:789–799, 2004.
- [4] C.J. Vaske, S.C. Benz, J.Z. Sanborn, D. Earl, C. Szeto, J. Zhu, D. Haussler and J.M. Stuart. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics*, 26(12):i237-245, 2010.