

Vials: Visualizing Alternative Splicing of Genes

Supplementary Material

Hendrik Strobel

hstrobel@seas.harvard.edu
Harvard University

Bilal Alsallakh

alsallakh@cvast.tuwien.ac.at
Vienna University of Technology

Joseph Botros

Harvard University

Brant Peterson

brant.peterson@novartis.com
Novartis Institute of BioMedical Research

Mark Borowsky

mark.borowsky@novartis.com
Novartis Institute of BioMedical Research

Hanspeter Pfister

pfister@seas.harvard.edu
Harvard University

Alexander Lex

alex@sci.utah.edu
Harvard University, University of Utah

Supplementary Figure 1

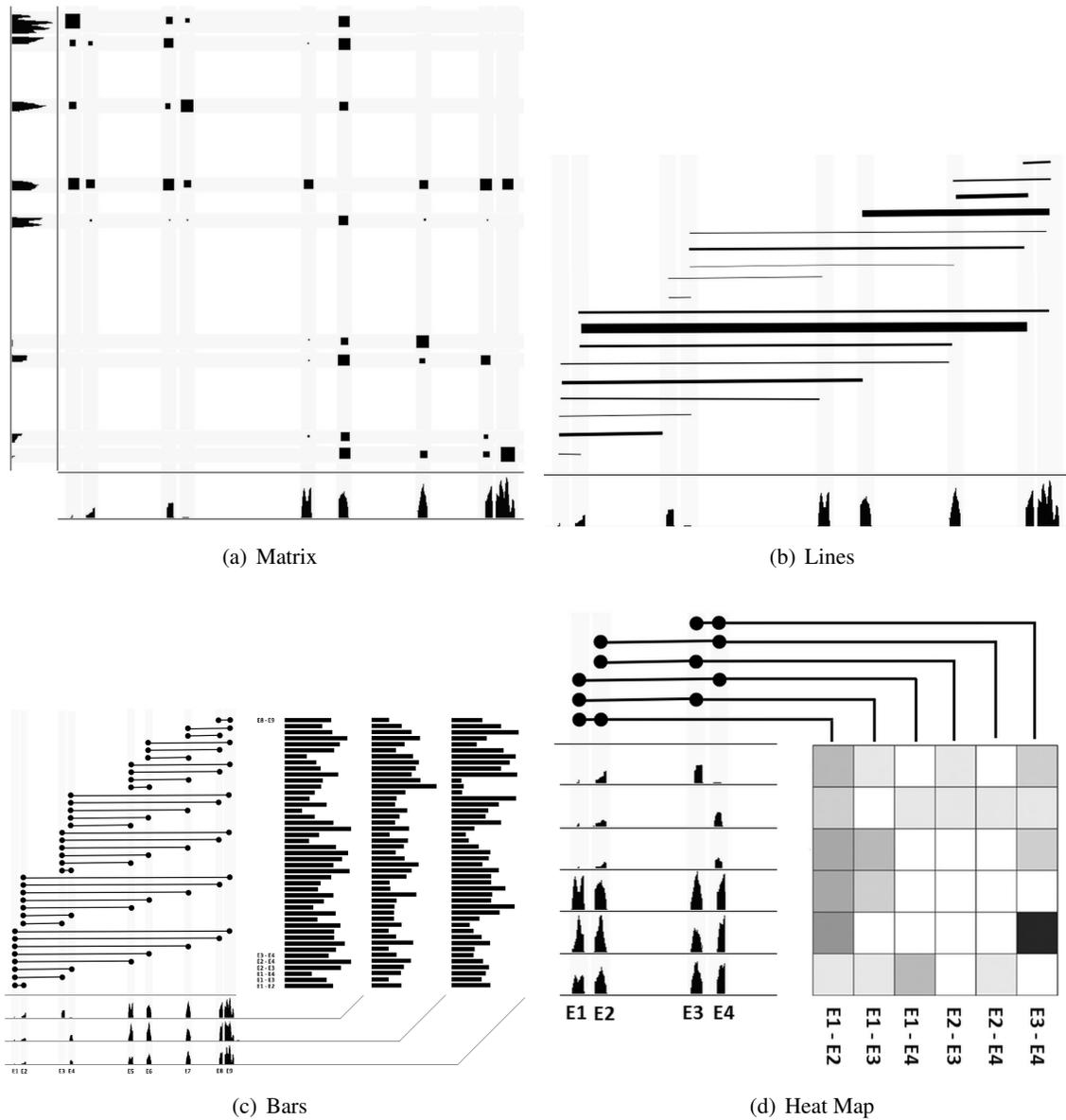


Figure S1: Enlarged reproduction of Figure 9: Selected design alternatives for the junction view. The figures are reproduced in a larger format in the supplemental material. (a) Two sequences are juxtaposed and a matrix view visualizes the strengths of the connections. (b) Lines directly show the edges, where the edge weight encodes the connection strength. (c) Bars show the weight of the edges, the source and target of an edge are identified using a matrix to the left. Each bar column is associated with a single sample. (d) A heat map's rows are placed next to samples, the columns represent edges that are traced to the exons using connection lines.

Supplementary Case Study: Alternative Splicing in Different Tissues

To explore the utility of interactive visualization in genes with complex alternative splicing in many tissues, our collaborators chose to explore the *Illumina BodyMap 2.0 (Bodymap)* dataset¹. Specifically, they selected the *ABI1* gene, which is subject to diverse alternative splicing across several exons in multiple diseases and healthy tissues. *ABI1* encodes a protein that is involved in regulating a suite of cellular processes involving cell motility, division, and morphology. Alternatively spliced isoforms of *ABI1* have been shown to impact its function in these processes [1].

The Illumina Bodymap 2.0 provides an ideal dataset for visualizing and exploring the diversity of isoforms produced from this gene (Fig. S2). For example, the isoforms labeled A and B are inversely expressed in brain and muscle: isoform A is abundant in brain (blue) and low in muscle (orange), whereas the inverse is true for isoform B. These two isoforms differ by several structural features including the inclusion of two exons indicated “isoform B-specific exons” in Figure S2. The junctions view (top) shows that all four relevant junctions required for inclusion of the two isoform B-specific exons are elevated in muscle (orange) compared to brain (blue).

Our collaborators noted that this is a real-world example of the complexity of alternative splicing in which Vials permits direct observation of both the structural differences underpinning isoform diversity (indicated exon inclusion differentiates brain and muscle isoforms) and the experimental support for these structures (junction counts and exon abundance) through interactive visualization.

¹<http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-513/>

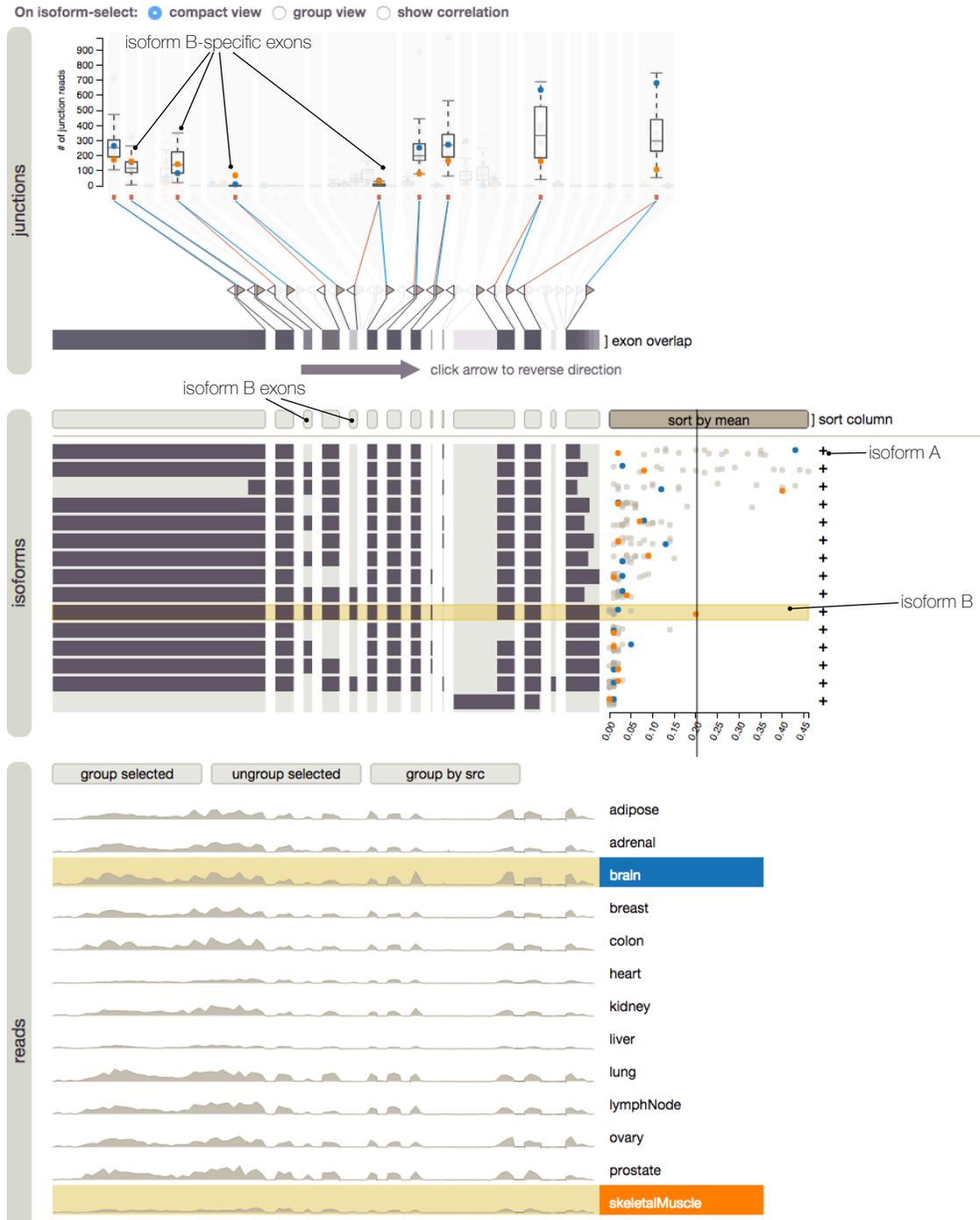


Figure S2: Case Study ABI1: an example of complex alternative splicing. Vials permits direct observation of structural differences and experimental support for these structures through interactive visualization. Brain (orange) and muscle (blue) tissue have inverse isoform abundances for Isoforms A and B, which is also evident in the junction and reads view.

Data Sources and Preprocessing

The two sources of data we currently handle in our preprocessing pipeline² are TCGA³ and Bodymap⁴. These sources differ in their data formats. For TCGA we derive the abundances for exons, isoforms, and junctions directly from the downloaded files. We restructure the data into projects and filter subsets of genes related to the project. When processing Bodymap data we operate on raw RNA-seq data (BAM files) and derive isoform and exon features from utilizing and running the misopy package [2]. To allow the display of reads in our visualization we sample the reads, as the resolution of the data makes both data transfer and rendering slow, while the resolution of the data far exceeds the screen resolution. We choose the target resolution such that we have the appropriate level of detail for the zoomed exons.

As reference genome we use the Homo Sapiens GRCh38 dataset.⁵

The server component of Vials offers a REST API to deliver data for each gene to the client. The API response is in JSON format and can be stored as file. This file can be used as input for Vials in a client-only web page, as demonstrated at <http://vcglab.org/vials>.

References

- [1] P. M. Dubielecka, P. Cui, X. Xiong, S. Hossain, S. Heck, L. Angelov, and L. Kotula. Differential regulation of macropinocytosis by Abi1/Hssh3bp1 isoforms. *PloS one*, 5(5):e10430, 2010.
- [2] Y. Katz, E. T. Wang, E. M. Airoidi, and C. B. Burge. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods*, 7(12):1009–1015, 2010.

²<https://github.com/Caleydo/vials-project-builder>

³<https://tcga-data.nci.nih.gov/tcga/>

⁴http://www.illumina.com/science/data_library.html

⁵<http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/data.shtml>