Vials: A Novel Tool for Isoform Visualization

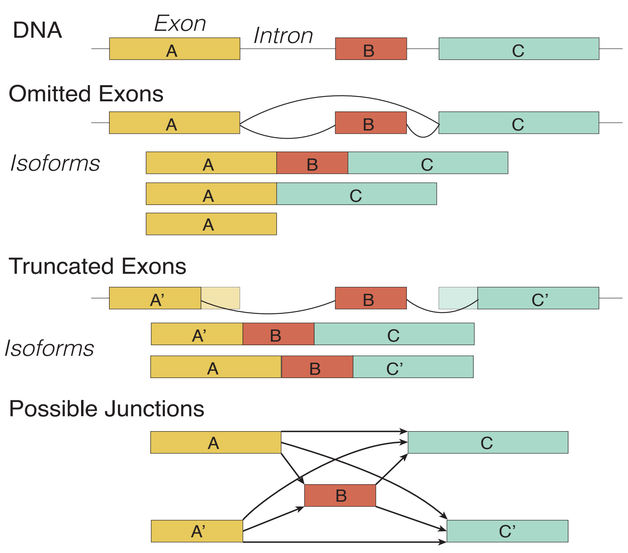
Hendrik Strobelt, Bilal Alsallakh, Joseph Botros, Brant Peterson, Mark Borowsky, Hanspeter Pfister, Alexander Lex

What are Isoforms?

Genes consist of exons and introns. The exons code for proteins, the introns are spliced out.

Exons can be truncated and omitted in a process called **alternative splicing**. These different combinations of exons are called Isoforms.

Isoforms are a natural way to diversify functional gene products but are also associated with diseases.

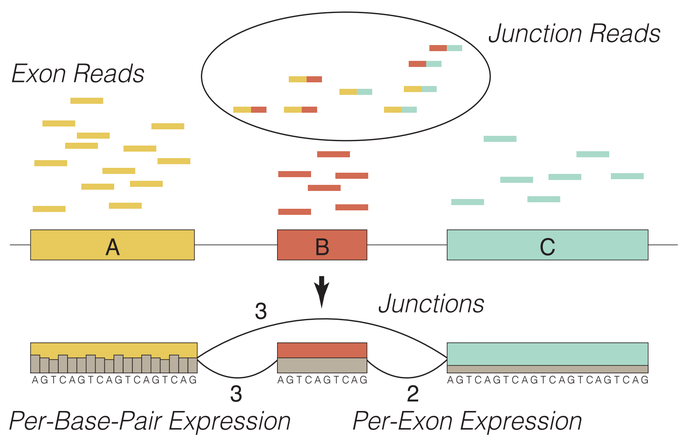


How do we measure Isoforms?

We can estimate how frequent isoforms are using **mRNA-seq**. We get short reads that we can match either to an exon or to a "junction" between two exons. Counting junction reads and calculating expression of exons let us reason about inclusions and exclusion of exons from which we can estimate isoform frequency.

A typical isoform dataset comes in three parts:

- **Expression** values on a per-base-pair or per-exon level
- The number of **junction reads**
- An estimate for **isoform abundances**



Why Visualize Isoforms and Splicing Data?

Our collaborators at a big pharmaceutical company have four goals when investigating isoforms:

1. Explore differences between samples and groups.
Are samples with e.g., different phenotypes different with respect to their isoform use?

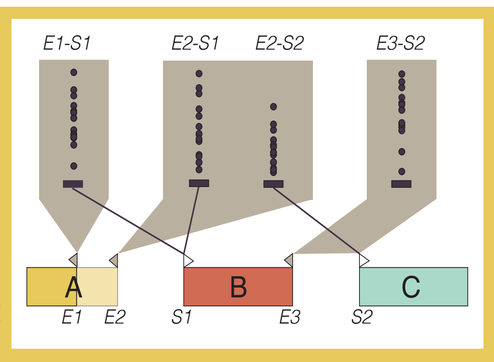
2. Discover Novel Isoforms.
Does a sample contain isoforms that is not yet described in the reference databases?

3. Evaluate Isoforms.
Judge, e.g., how similar two isoforms are.

4. Control Data Quality
mRNA-seq and Isoform abundance estimation are noisy processes. It is important to judge the data quality, e.g., by judging if an isoform thought to be common in a sample also shows frequent use of its junctions.

Junction View

Junction reads tell us which exons are connected. We have designed a view that shows us the number of reads for each sample and each junction by plotting them as dot and box plots.



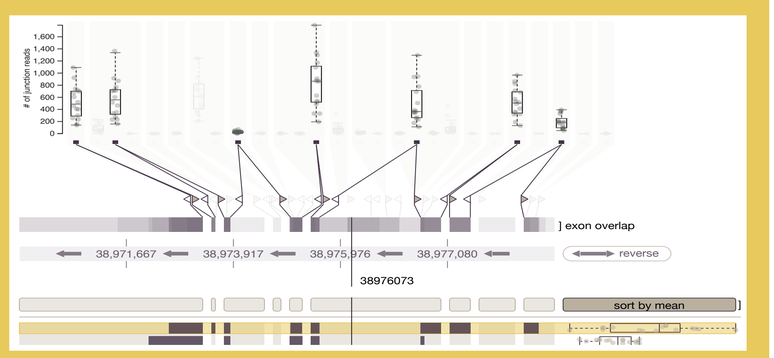
The detached triangles on top of the exon show a junction start side (S1, S2) or a junction end side (E1-E3).

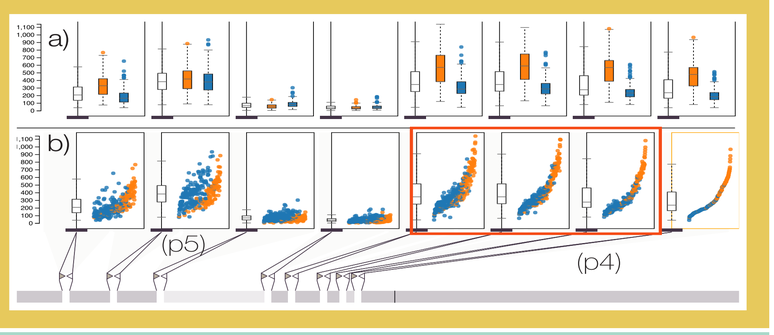
The exons in the junction view are rendered on top of each other using transparency, resulting in an exon heat map. The genomic coordinates and the reading direction are shown below the heat map.

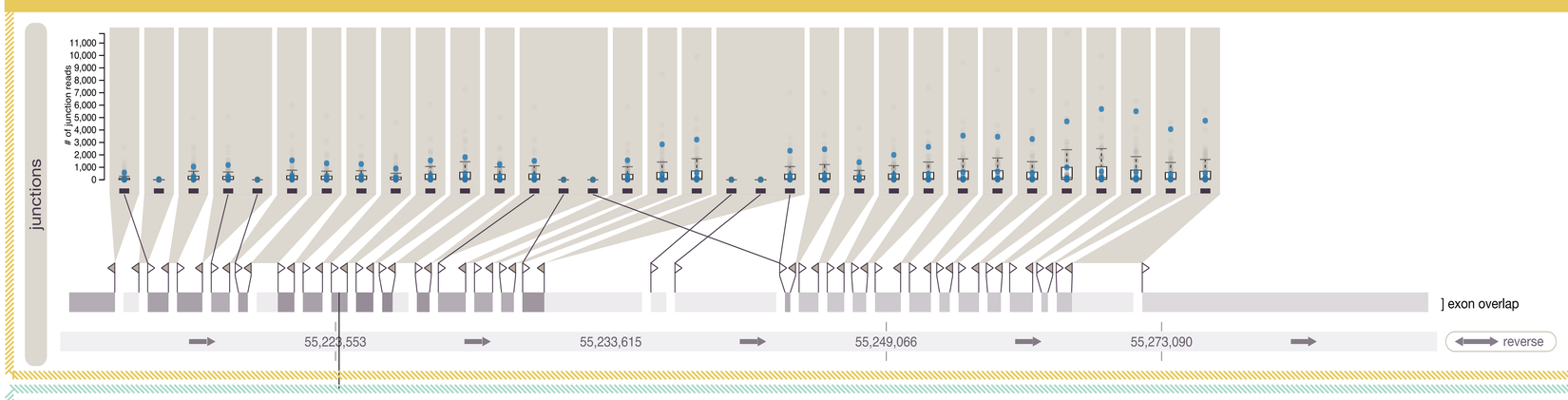
When an isoform is selected, we show only the junctions that are relevant for that isoform.

To compare groups of samples the dots are color-coded. We can also separate groups into one plot per group for accurate comparison.

To see trends in the data we can select one junction to sort the samples by.





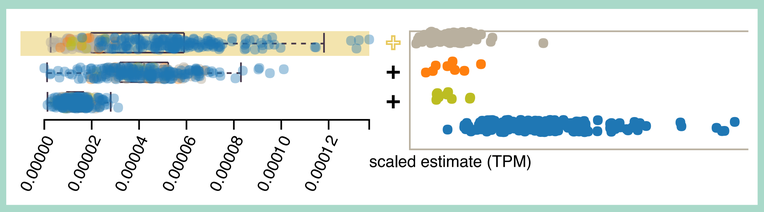


Isoform View

This view visualizes the composition and the estimated abundance of isoforms. It shows the exon composition of all known isoforms from a reference database on the left and the estimated abundances of these isoforms in all samples on the right.

The isoforms can be sorted by the average abundance or by exon containment.

To visualize the abundances we again use dot and box plots. To better compare the abundance distribution of samples a detail view can be opened that shows each group in a separate line.

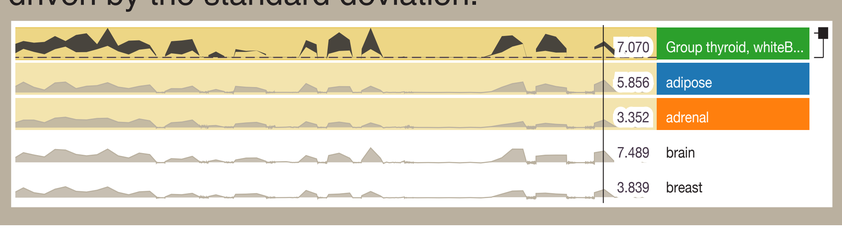


Reads View

The reads view shows the expression of exons and introns for each sample. Expression data is either available for each base-pair or only for each exon. The expression data is shown as an area chart for each sample. The crosshair shows the expression values at each track.

The reads view is also used for selecting and grouping samples. Samples can be grouped manually or automatically, based on meta-data.

Grouped samples can be collapsed. The collapsed view shows a band centered at the average value, the width is driven by the standard deviation.



http://vcglab.org/vials

Contact: Alexander Lex
http://alexander-lex.net
alex@sci.utah.edu

Acknowledgments

Vials is built with Caleydo Web, the new web-based version of the Caleydo Visualization Framework. We would like to thank Samuel Gratzl for help with the framework, Nils Gehlenborg for comments and Sebastian Hörsch and other members of the SDA team for their expertise and feedback.

This work was supported by a grant by Novartis Institutes for Bio-Medical Research, the Austrian Science Fund (J 3437-N15) and the Air Force Research Laboratory and DARPA grant FA8750-12-C-0300.

HARVARD School of Engineering and Applied Sciences · TU WIEN · NOVARTIS NOVARTIS INSTITUTES FOR BIOMEDICAL RESEARCH · THE UNIVERSITY OF UTAH · CALEYDO