Interactive Visualization of GRT and BioHTS Data

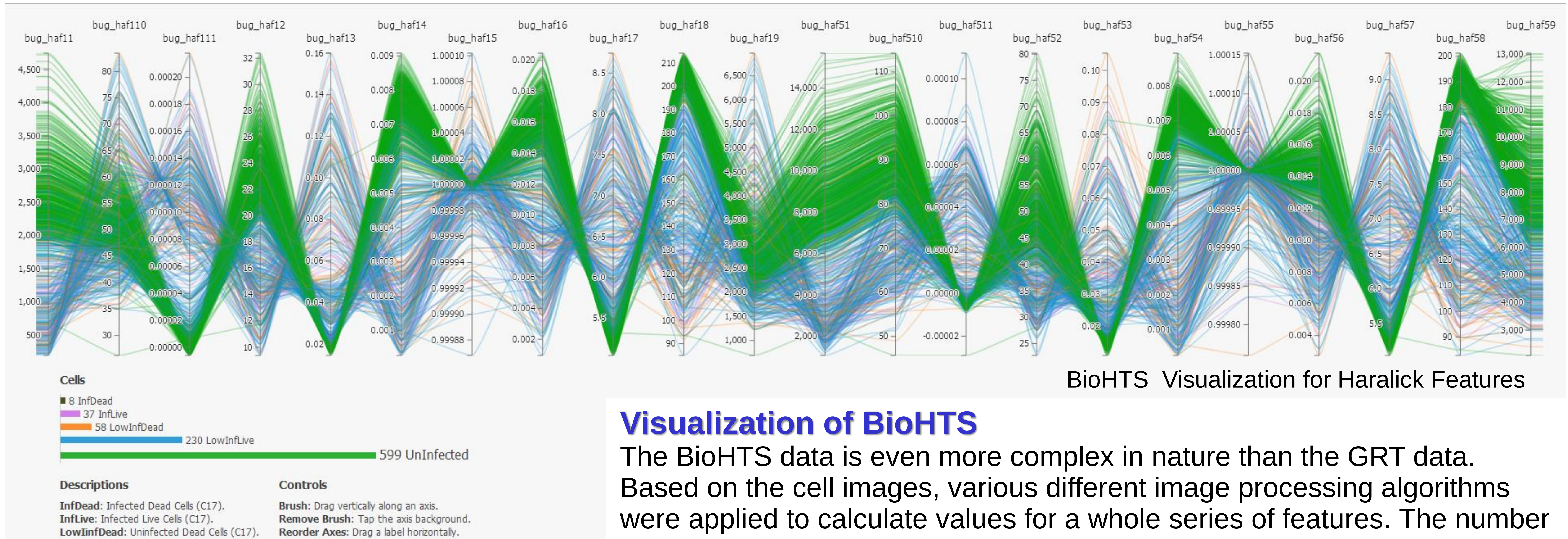
Sara Gharabaghi (WSU), Thomas Wischgoll (WSU), Rhonda Vickery (Engility), Ross Smith (Engility),
Leslie Blaha (711 HPW/RHCV), Thomas Lamkin (AFRL/XPR), Steven Kawamoto (UES), Robert Trevino (ASU),
Eric Bardes (CHMCC), Scott Tabar (CHMCC)

Introduction

The scope of this project is to provide better tools for statistical and informational visual analysis for High Throughput Screening of Biological Infectious Agents (BioHTS), General Recognition Theory (GRT) modeling, and areas where pipelines of unstructured datasets of all types must be analyzed. A parallel coordinates plot is one of the more effective visualization methods for visualizing multi variant data.

Parallel Coordinates

Parallel coordinates plots are used to show a set of points in an n-dimensional space. It uses nparallel vertical axes. A point in n-dimensional space is represented as a ploy-line with vertices on the parallel axes; the position of the vertex on the i-th axis corresponds to the i-th coordinate of the point [1]. In this project, the D3 [2] library was used to generate the plots. Based on established principles, such as parallel coordinates, these algorithms were amended specifically to the needs of the GRT and BioHTS researchers. The specific approaches are outlined in the following subsections.



Descriptions

InfDead: Infected Dead Cells (C17).
InfLive: Infected Live Cells (C17).
LowIinfDead: Uninfected Dead Cells (C17).
LowInfLive: Uninfected Live Cells (C17).
Uninfected: Uninfected Cells (E13).

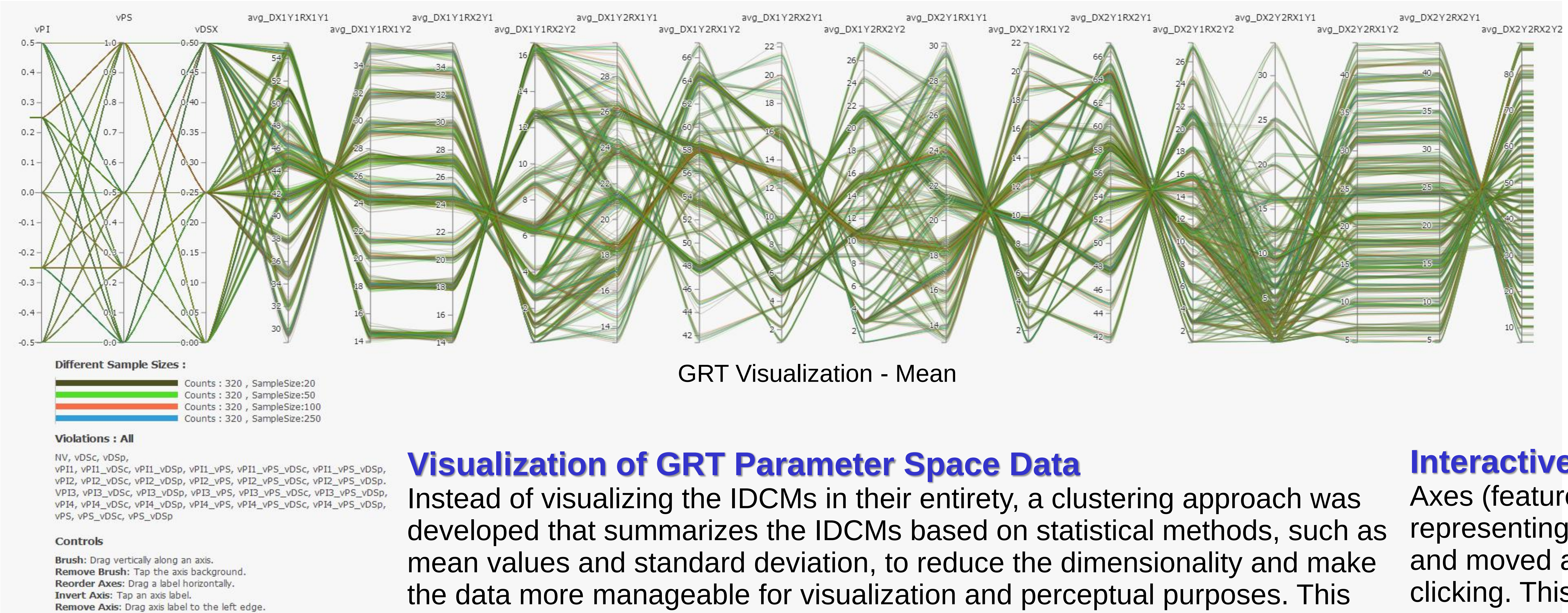
Controls

Brush: Drag vertically along an axis.
Remove Brush: Tap the axis background.
Reorder Axes: Drag a label horizontally.
Invert Axis: Tap an axis label.
Remove Axis: Drag axis label to the left edge.

BioHTS Visualization for Haralick Features

Visualization of BioHTS

The BioHTS data is even more complex in nature than the GRT data. Based on the cell images, various different image processing algorithms were applied to calculate values for a whole series of features. The number of features is large (around 11,000), which makes the visualization of all of those values for all cells nearly impossible. Instead of representing all feature values at the same time, groups of related feature values were chosen and then visualized in parallel coordinate plots.



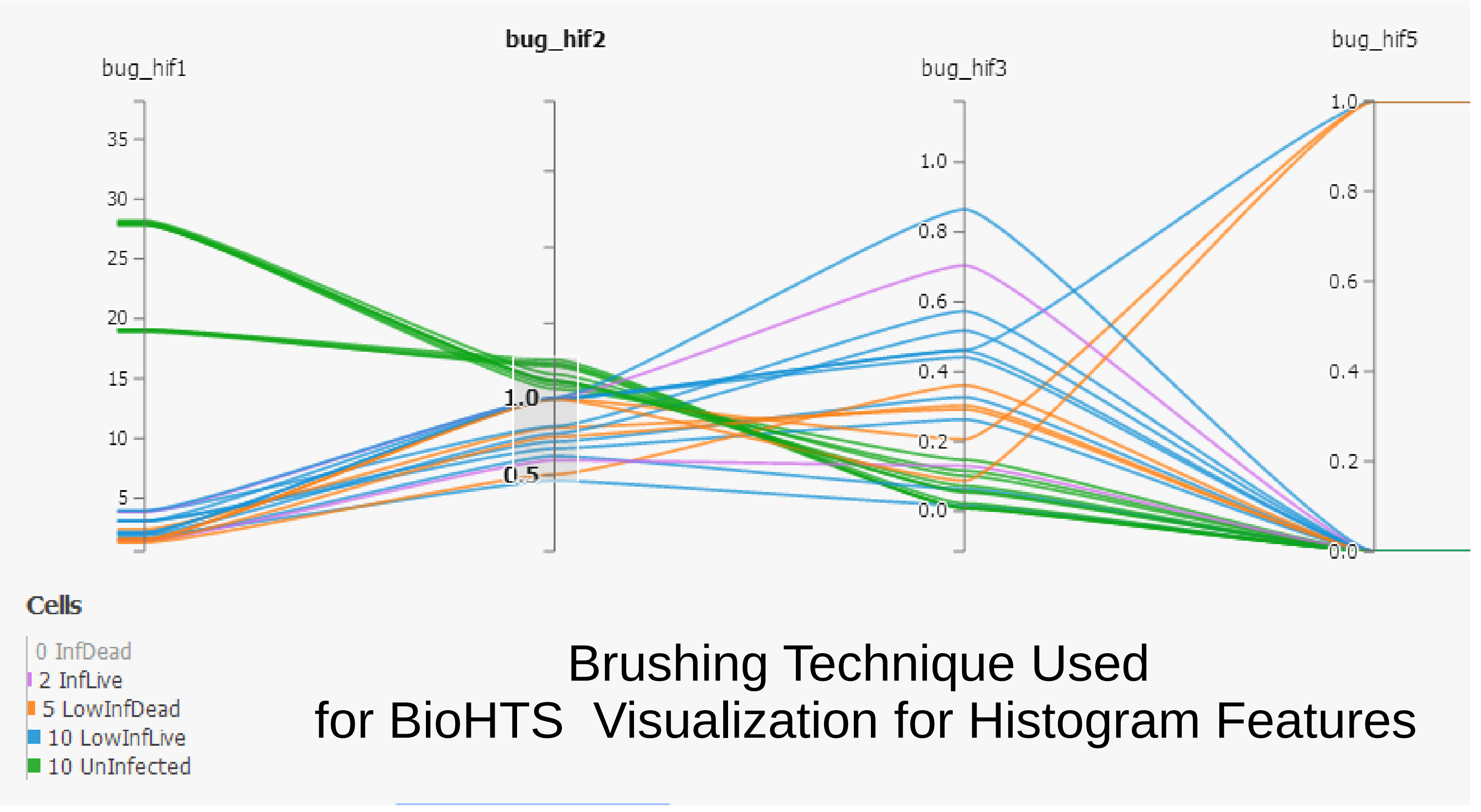
Violations : All

NV, vDSc, vDSp,
vPI1, vPI1_vDSc, vPI1_vDSp, vPI1_vPS, vPI1_vPS_vDSc, vPI1_vPS_vDSp,
vPI2, vPI2_vDSc, vPI2_vDSp, vPI2_vPS, vPI2_vPS_vDSc, vPI2_vPS_vDSp.
VPI3, vPI3_vDSc, vPI3_vDSp, vPI3_vPS, vPI3_vPS_vDSc, vPI3_vPS_vDSp,
vPI4, vPI4_vDSc, vPI4_vDSp, vPI4_vPS, vPI4_vPS_vDSc, vPI4_vPS_vDSp,
vPS, vPS_vDSc, vPS_vDSp

Controls

Brush: Drag vertically along an axis.
Remove Brush: Tap the axis background.
Reorder Axes: Drag a label horizontally.
Invert Axis: Tap an axis label.
Remove Axis: Drag axis label to the left edge.

GRT Visualization - Mean



Brushing Technique Used
for BioHTS Visualization for Histogram Features

Visualization of GRT Parameter Space Data

Instead of visualizing the IDCMs in their entirety, a clustering approach was developed that summarizes the IDCMs based on statistical methods, such as mean values and standard deviation, to reduce the dimensionality and make the data more manageable for visualization and perceptual purposes. This then allows the visualization of the IDCMs on just 16 axes, which means a significant reduction in complexity within the visualization.

References:
[1] Parallel Coordinates in Wikipedia.
[2] Data-Driven Documents, http://d3js.org.

Interactive Tools for Parallel Coordinates

Axes (features) that are not desirable can be eliminated by dragging the axis representing that feature all the way to the left. Similarly, axes can be swapped and moved as needed. The orientation of an axis can be flipped easily by double clicking. This can make interpretation easier if many values cross near the same point between two axes. In addition, the brushing technique is available for selecting a specific range on the parallel coordinates to focus on only this selected range of values. This level of interactivity is very important as it allows the BioHTS researcher to better mine the data and drill down on specific aspects hidden within the data.