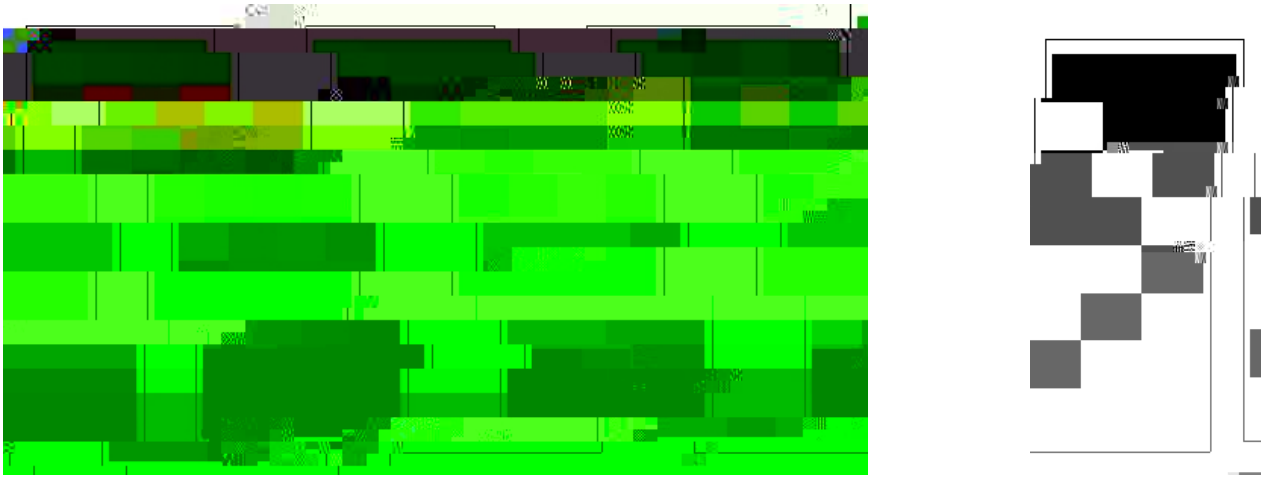


Figure 2: (Left three plots) Color matrix plot of the re-rordered toy data, using three di□erent color mappings. Di□erent structure can be perceived from each mapping. How many di□erent interpretations can you produce? (Right plot) Can you recognize this simple 3D geometric shape? The answer is in the appendix.

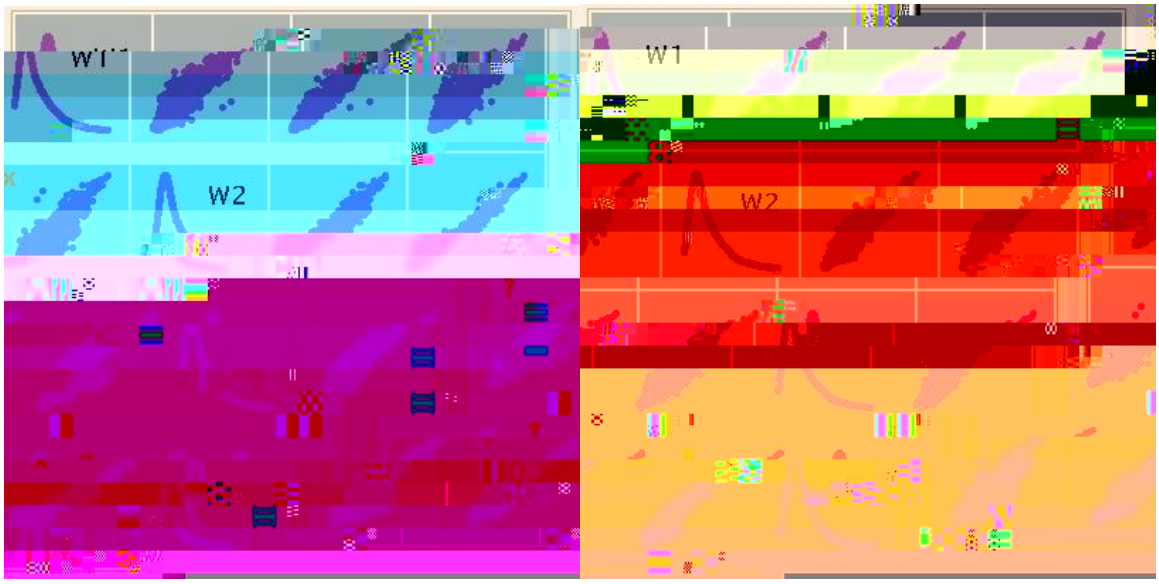
color and intensity for the toy data. Quite di□erent perceptions of the structure are possible with the

work of gene expression data analysis, experiments with more replicates, more treatments, and time-dependent measurements. The Appendix lists the software packages used for the analysis and explains how to prepare the data for interactive graphical exploration.

2 Two-Factor, Single Replicate Data

2.1 Data description

GeneID	M		MT		W		WT	
	Rep 1	Rep 2	Rep 1	Rep 2	Rep 1	Rep 2	Rep 1	Rep 2
1								
.								


W1
W2
W1
W2

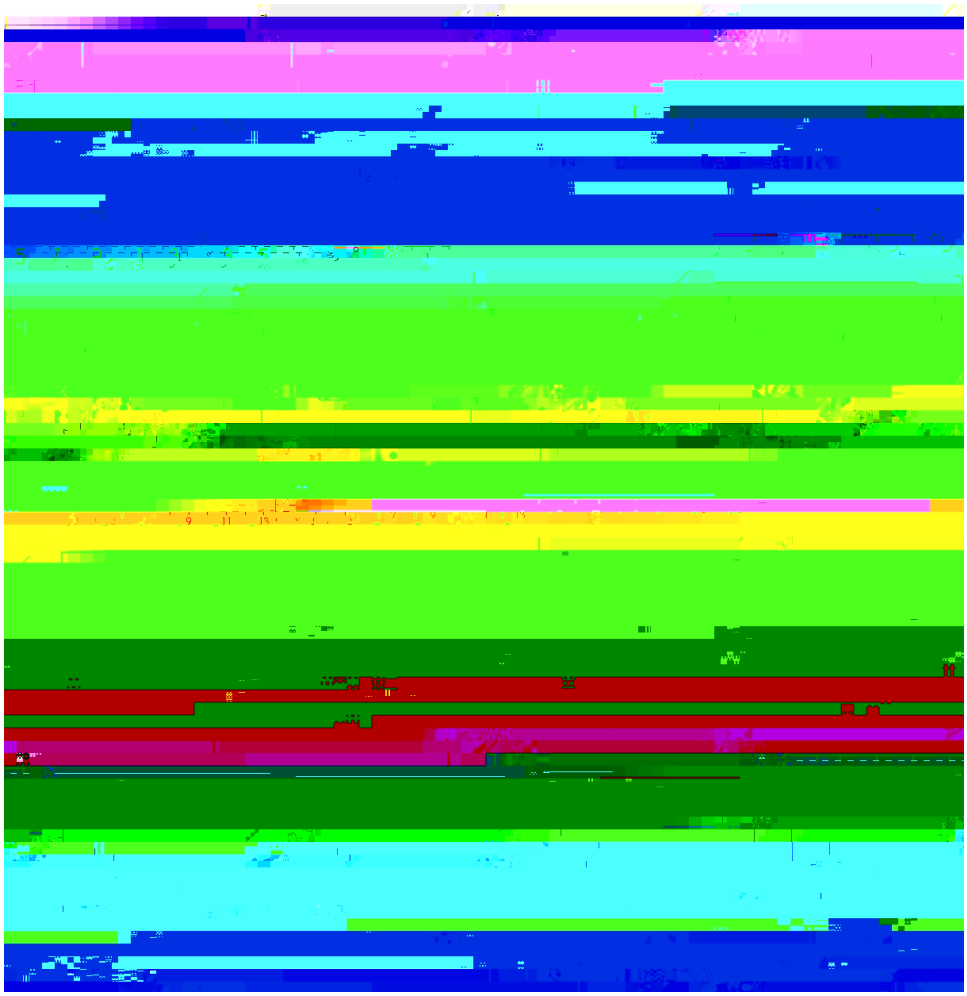


Figure 7: Plots of the treatments against each other. The treatment pair where the genes behave the most similarly are the two genotypes with treatment added (MT, WT). The mutant genotype without treatment

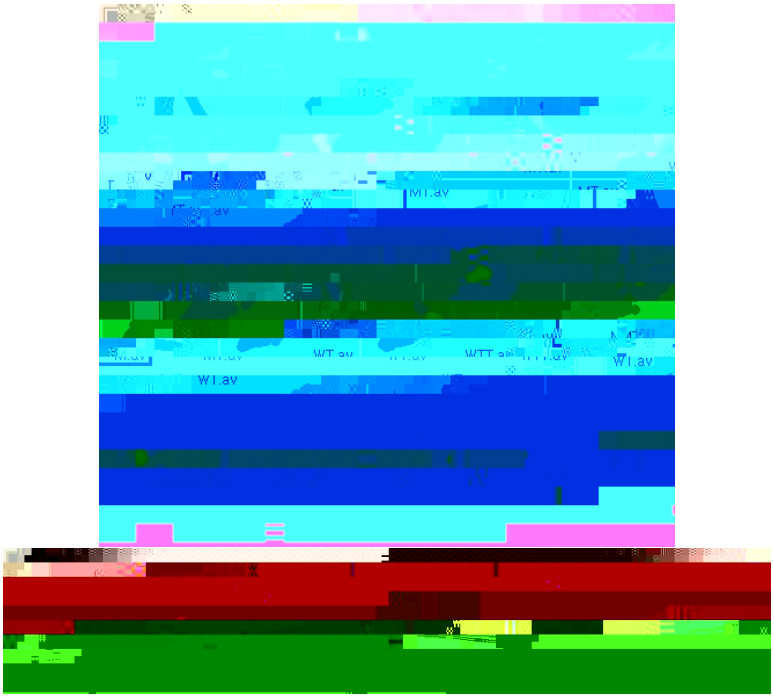


Figure 8: Scatterplot matrix of the treatment averages, linked by brushing. A profile plot of the 8 treat-

Figure 9: Profiles of the gene 15160_s_at and 12044_at: (left) genotype e□ect, (middle) treatment added

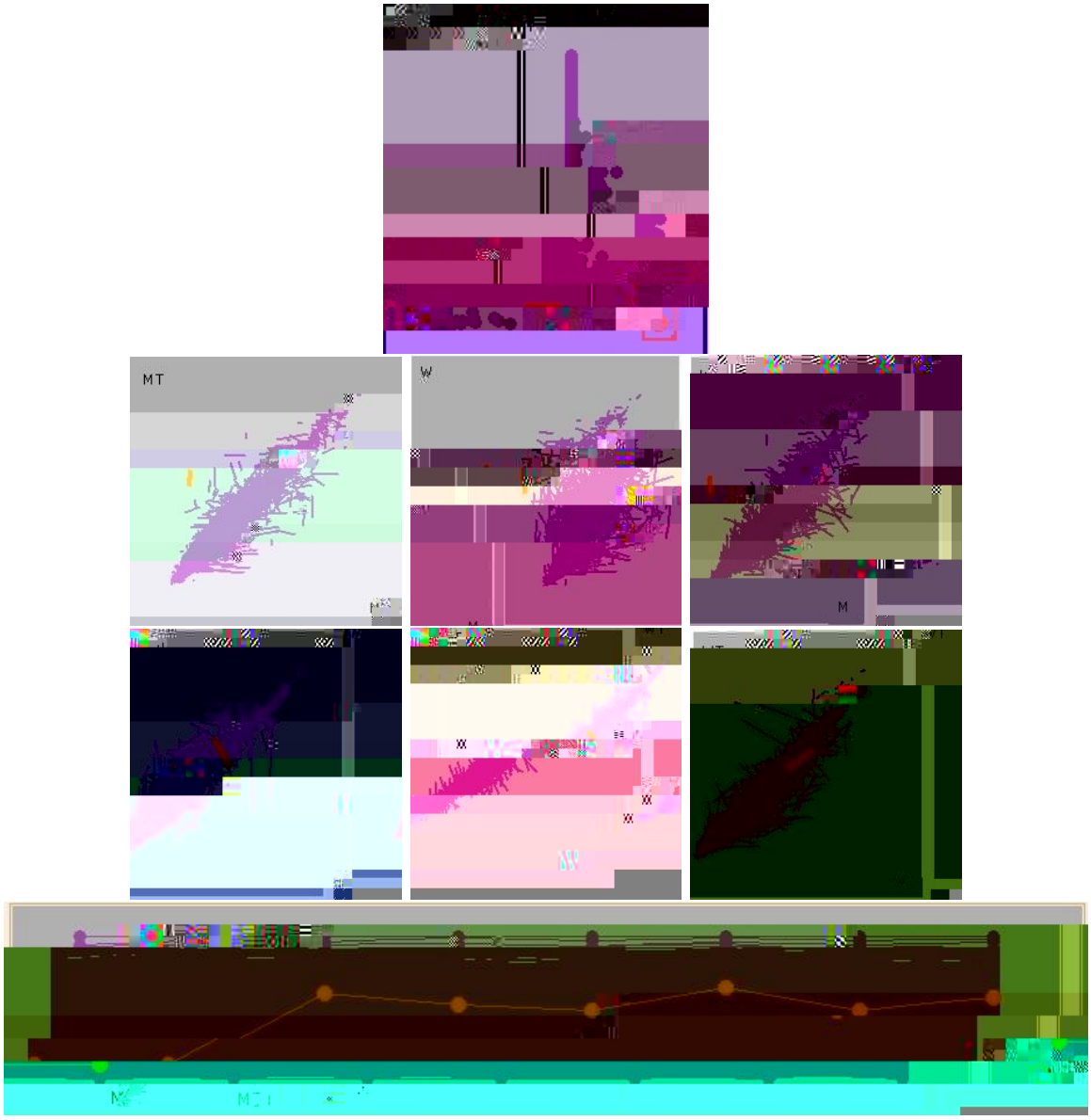


Figure 11: Searching for interesting genes: genes that have large MS treatment value and small *p*-value are considered interesting. Here the gene 16016_at is highlighted. It has a large MS interaction value, and relatively small *p*-value. Examining the plots of the treatment pairs,and the profile plot of this gene, it can be seen that this gene has much smaller expression on the mutant without treatment than the other three treatments.

these line traces to plots of measures calculated on the replicates. More than pairwise comparisons can be

The approach described here is manual acd labor-intensive, requiricg the analyst to specd a lot of time looking at plots acd probing and querying di□erent aspects of the plot. This might be considered the learning phase, the stage where we spend time to ucderstand the scientific problem, how di□erent approaches compare, how data manifests from the experiments. As the problems and methods are understood in detail it will be easier to build automatic processors to analyze the ecsuing flood of data fr(e)-333(m)-1(i)1(c)-1(r)1(oarra)28(ys.)TJ/F2914.

```
<realvariable name="MT2" nickname="M2"/>
<realvariable name="W1" nickname="w1"/>
<realvariable name="W2" nickname="w2"/>
<realvariable name="WT1" nickname="W1"/>
<realvariable name="WT2" nickname="W2"/>
<realvariable name="M.av" nickname="A1"/>
<realvariable name="MT.av" nickname="A2"/>
<realvariable name="W.av" nickname="A3"/>
<realvariable name="WT.av" nickname="A4"/>
<realvariable name="MS.Geno" nickname="MG"/>
<realvariable name="MS.Treat" nickname="MT"/>
<realvariable name="MS.Interaction" nickname="MI"/>
<realvariable name="q.Geno" nickname="q1"/>
```

Buja, A., Cook, D. & Swayne, D. (1996), 'Interactive High-Dimensional Data Visualization', *Journal of Computational and Graphical Statistics* 5(1), 78–99. See also www.research.att.com/ andreas/xgobi/heidel/.