plot_exp_by_delabelsygroups

Kate Weaver

4/7/2022

Plot 1: Expression of all genes across all samples

n=8646000
150 samples and 57640 genes
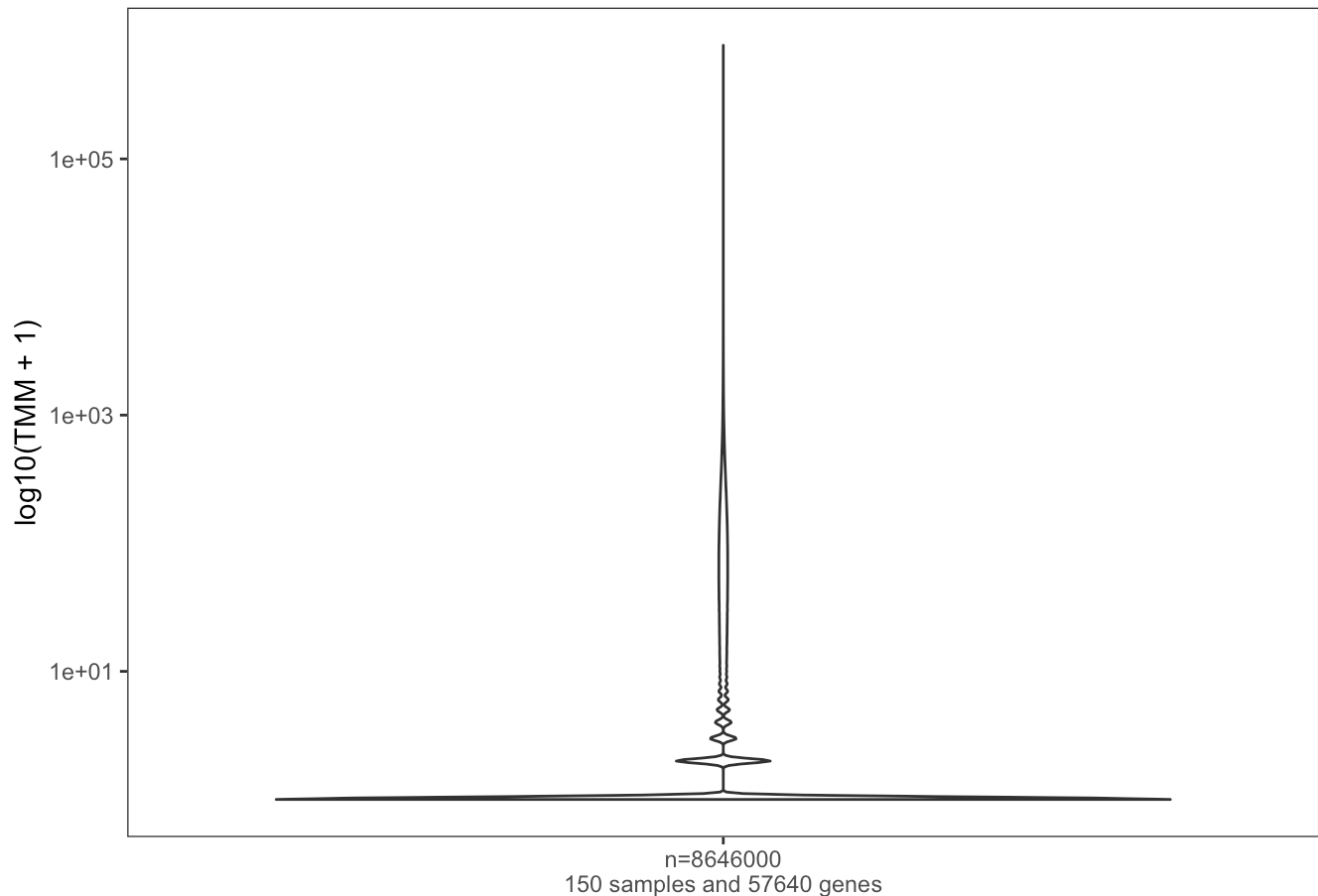
Plot 2: Expression of all genes across all samples, splitting genes into two groups based on filtering prior to DE

- A. those which were filtered out prior to differential expression analysis because they had low mean expression across samples (leaving them in the data would decrease the power to identify differentially expressed genes – see here for a description (https://docs.google.com/document/d/1tXxy3q4yHtB3VQ4bV0TAwJ_cmcbxXTdk9cnN0x4bP-o/edit#bookmark=id.ljuz75se4fyx))
- B. those which were assessed in differential expression analysis
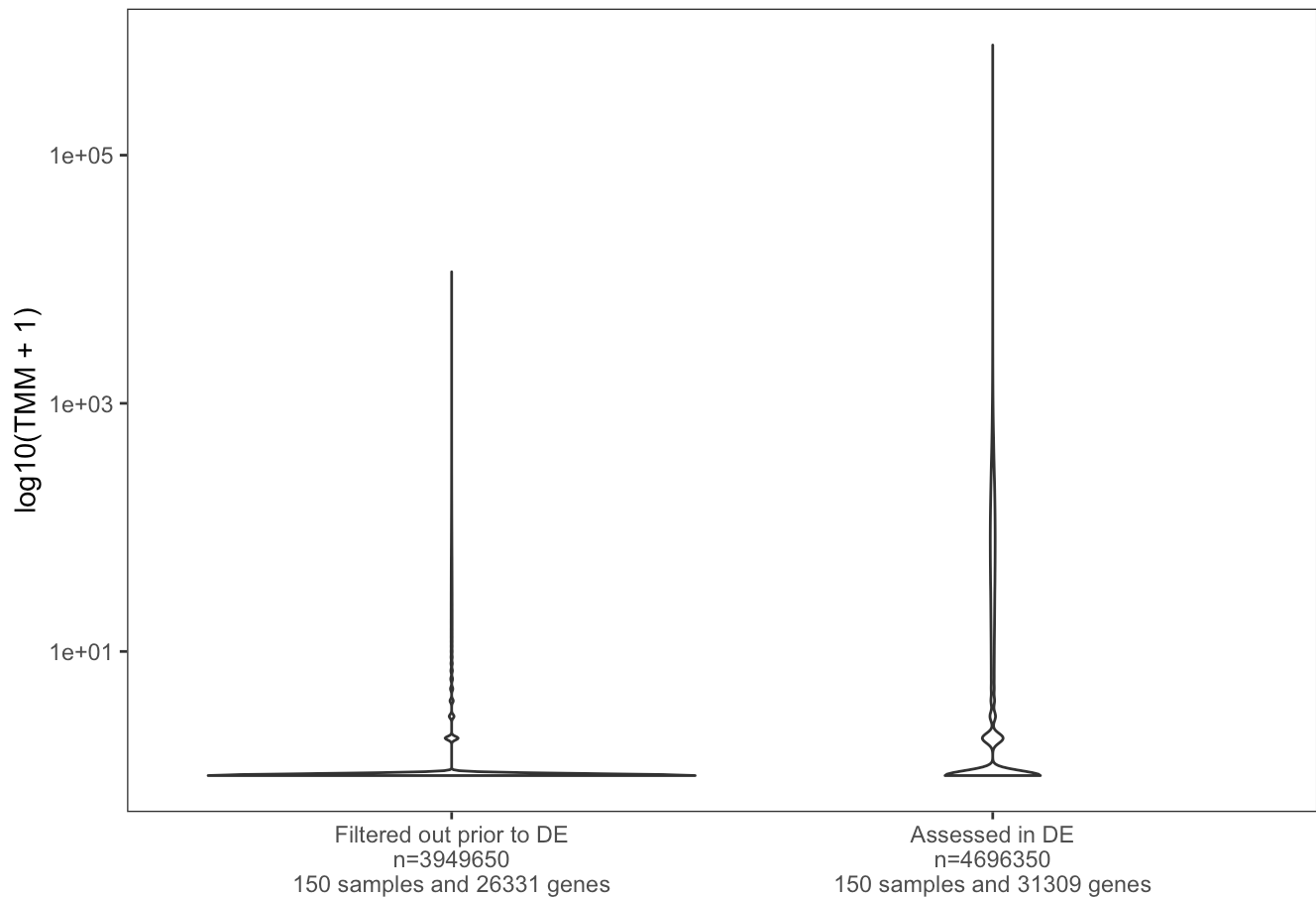
Plot 3: Focusing on the expression across all samples of just the genes which were assessed in differential expression analysis, split into 3 groups

- • A. Those not differentially expressed
- • B. Those upregulated in cases of implantation
- • C. Those downregulated in cases of immplantation

Note that differential expression analysis was performed with just the training data while all samples are plotted here.
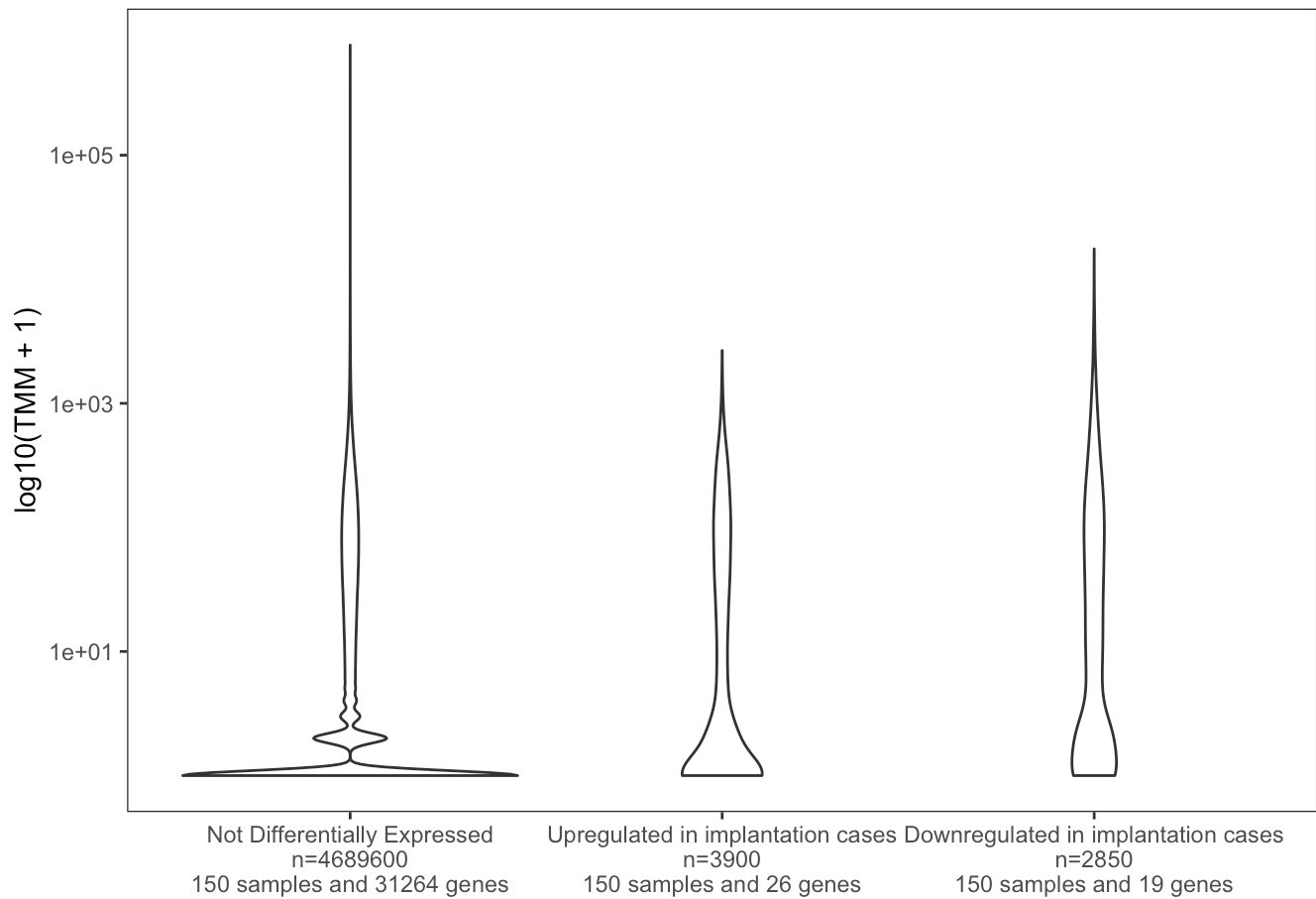
Plot 4 (Violin): Focusing on just the upregulated genes, specifically for each gene, distribution of expression for samples which successfully implanted vs those that did not implant
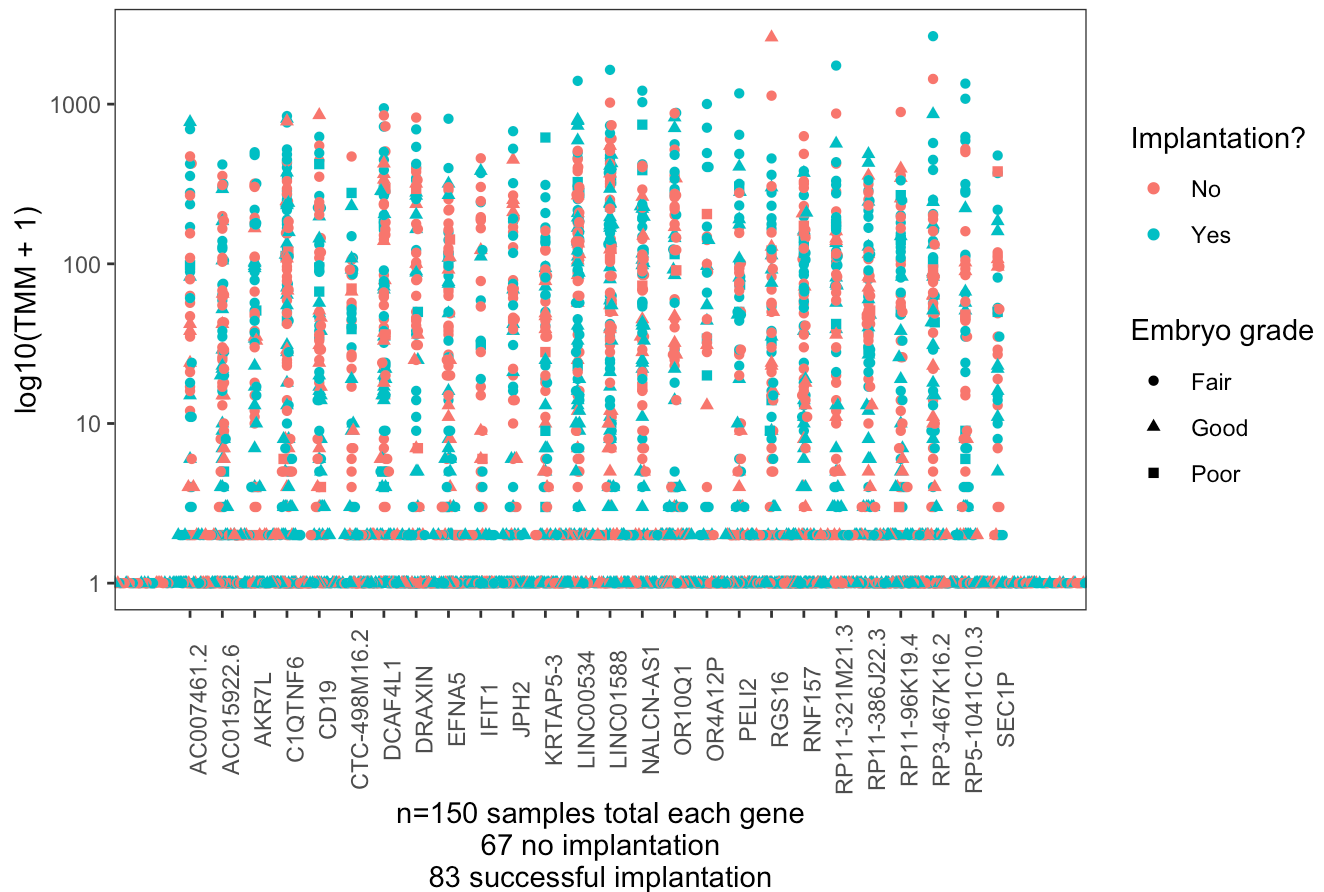
Upregulated Genes in Cases of Implantation

n=150 samples total each gene
67 no implantation
83 successful implantation
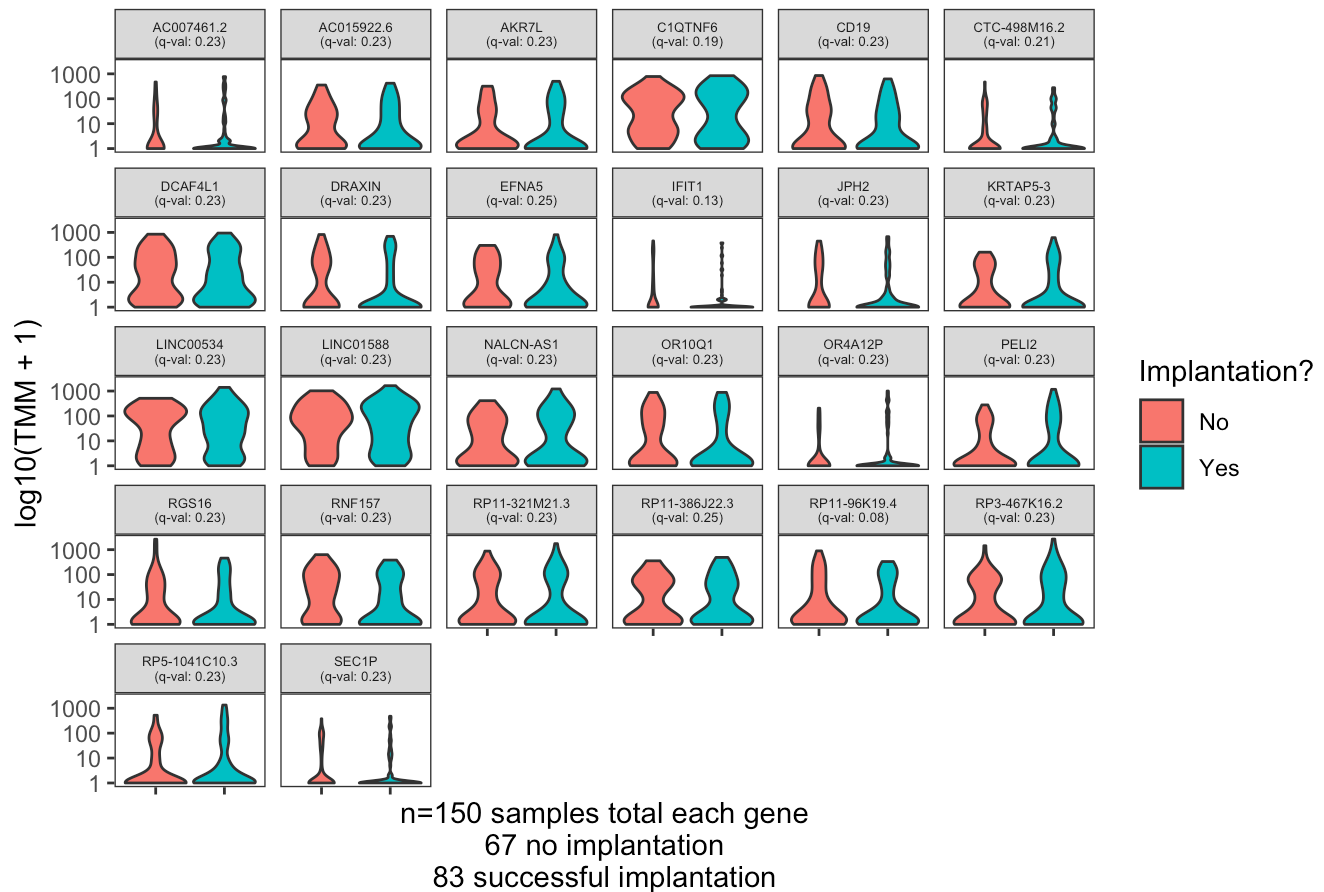
Plot 5 (Beeswarm): Focusing on just the upregulated genes, specifically for each gene, distribution of expression

- genes separated on x-axis
- distribution of expression for samples which successfully implanted vs those that did not implant is visualized, separating using color
- shape of points reflects embryo grade at freezing
- Don't love that it's actually difficult to separate the genes from each other, specifically in the more lowly expressed samples section with this visualization. Also difficult to see the differences between the two implantation groups based on color alone
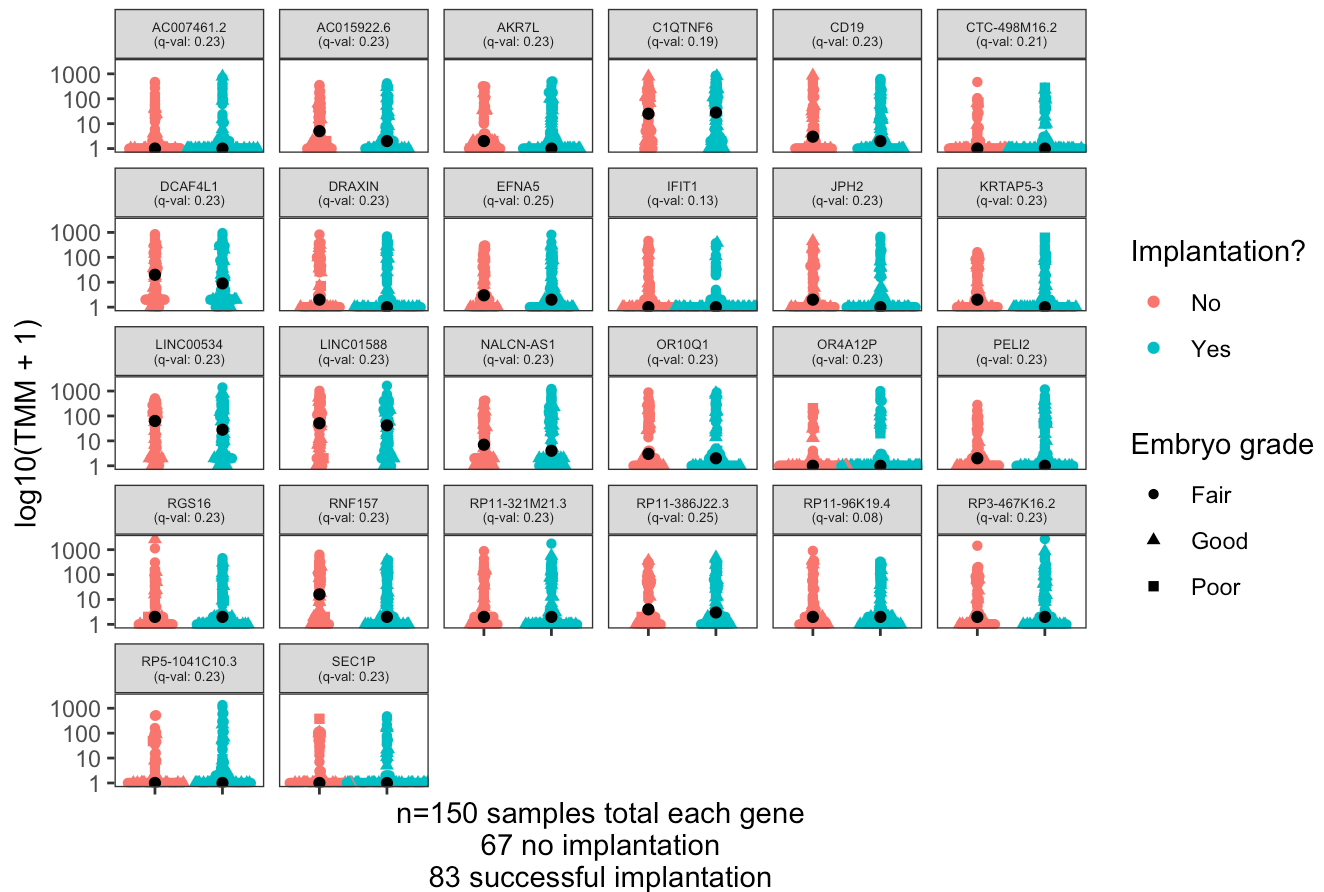
Upregulated Genes in Cases of Implantation

n=150 samples total each gene
67 no implantation
83 successful implantation

Plot 6 (Beeswarm, faceted): Focusing on just the upregulated genes, specifically for each gene, distribution of expression

- genes separated by facets to better separate them from one another
- distribution of expression for samples which successfully implanted vs those that did not implant is visualized, separating using color and position on the x-axis
- shape of points reflects embryo grade at freezing
- Also difficult to see the differences between the two implantation groups – not sure how to add median lines, but that would be great

Upregulated Genes in Cases of Implantation

n=150 samples total each gene
67 no implantation
83 successful implantation

Plot 7 (Beeswarm, faceted, no embryo grade): Focusing on just the upregulated genes, specifically for each gene, distribution of expression

- genes separated by facets to better separate them from one another
- distribution of expression for samples which successfully implanted vs those that did not implant is visualized, separating using color and position on the x-axis
- Also difficult to see the differences between the two implantation groups – not sure how to add median lines, but that would be great
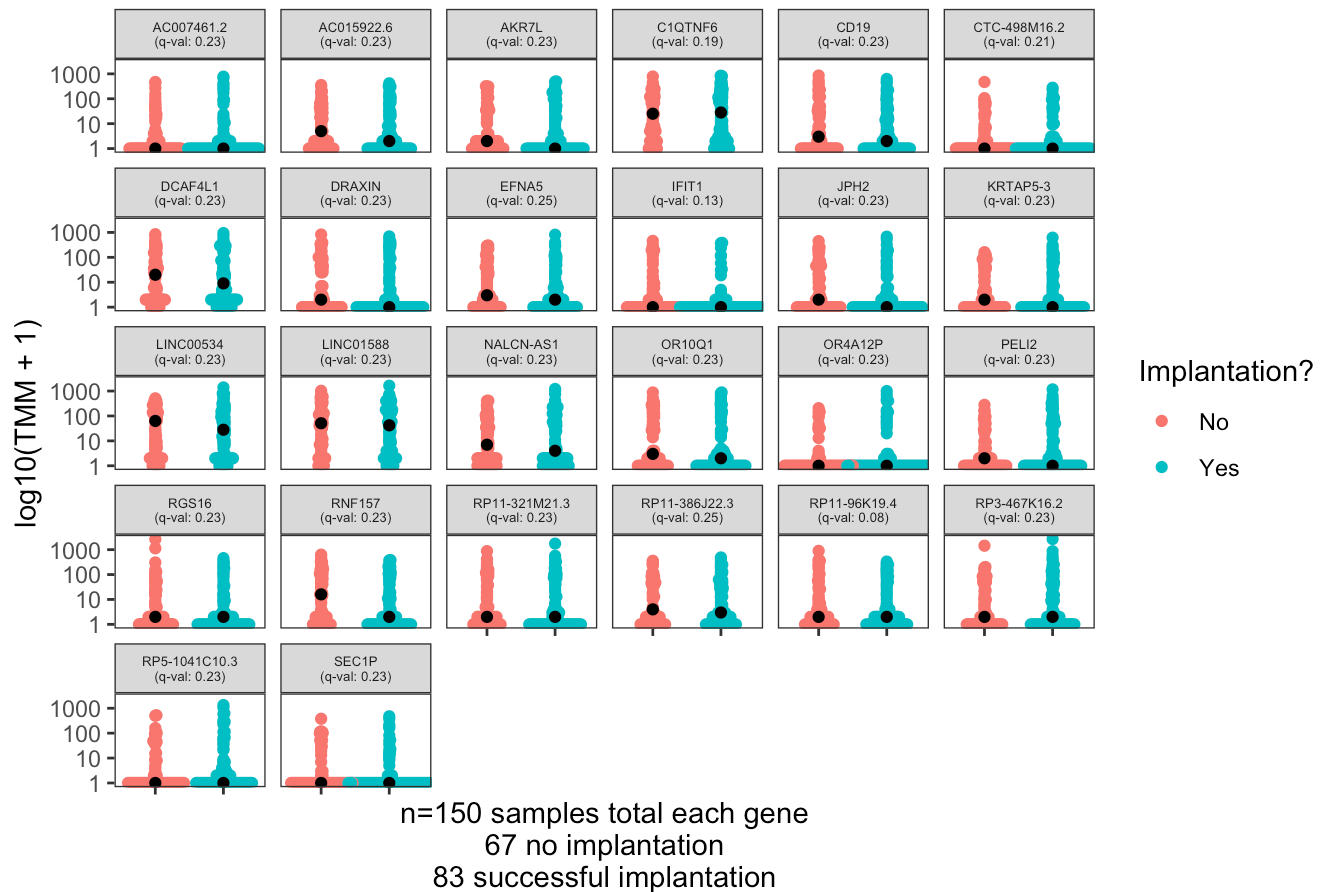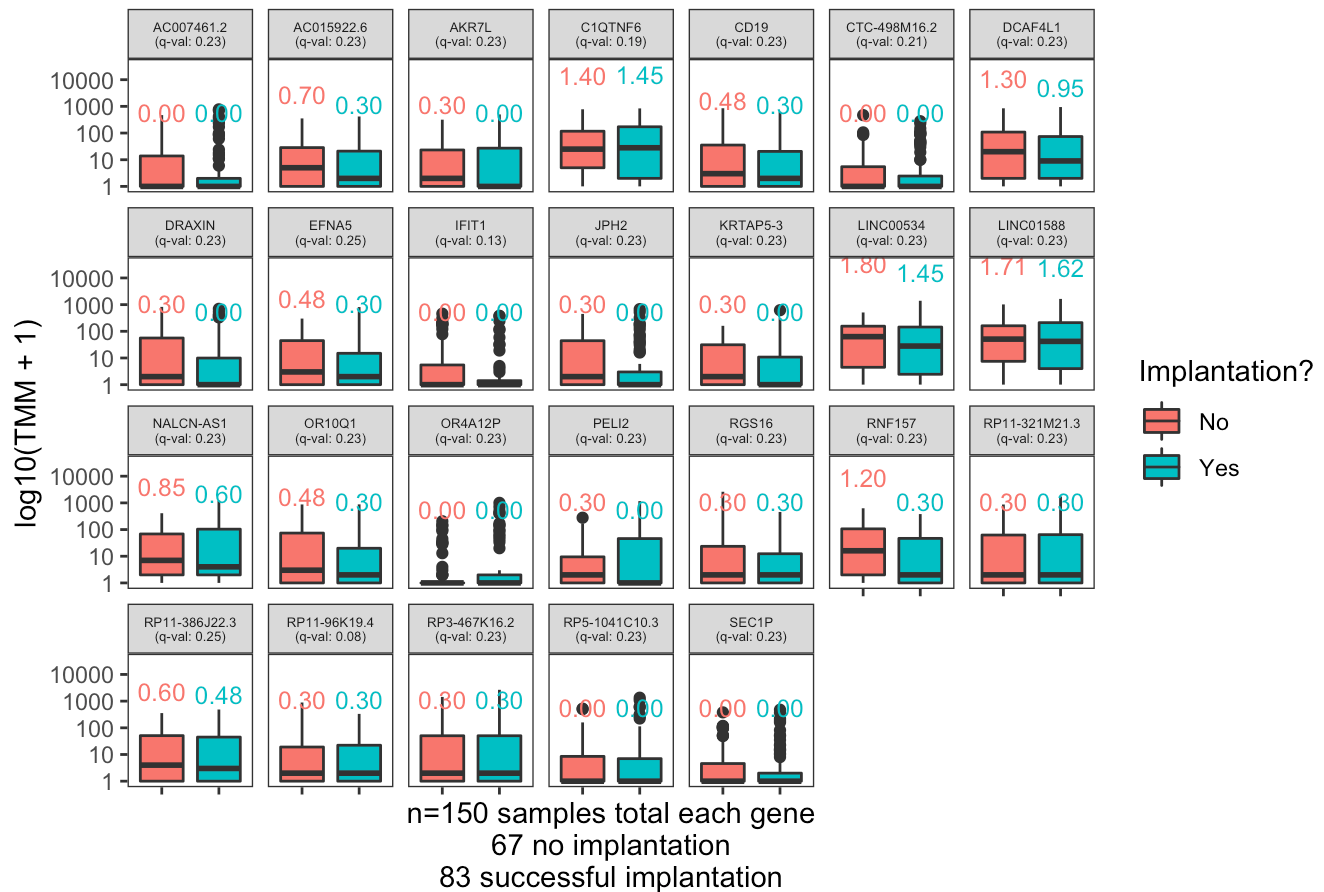
Upregulated Genes in Cases of Implantation

n=150 samples total each gene
67 no implantation
83 successful implantation

Plot 8 (Boxplot): Focusing on just the upregulated genes, specifically for each gene, distribution of expression

- genes separated by facets to better separate them from one another
- distribution of expression for samples which successfully implanted vs those that did not implant is visualized, separating using color and position on the x-axis
- Using the boxplot to visualize the quartiles and medians of the two groups
- Medians sometimes go down from no implantation to implantation – this could be due to: visualizing all data, not just training data; not controlling for embryo grade which de analysis did do; or use of the very permissive q-value (0.25); de analysis comparing more than medians
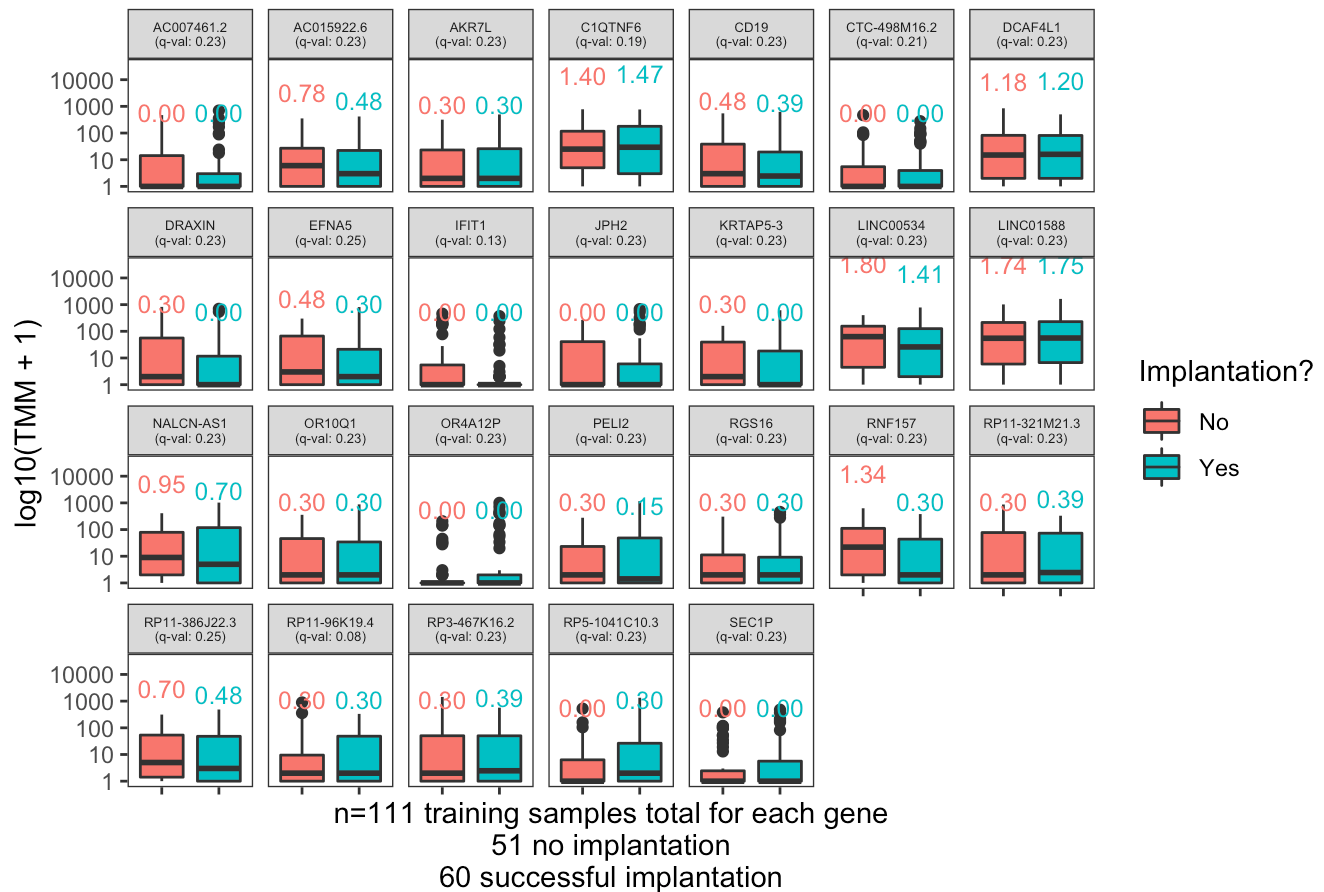
Upregulated Genes in Cases of Implantation

log10(TMM + 1)

| AC007461.2 (q-val: 0.23) | AC015922.6 (q-val: 0.23) | AKR7L (q-val: 0.23) | C1QTNF6 (q-val: 0.19) | CD19 (q-val: 0.23) | CTC-498M16.2 (q-val: 0.21) | DCAF4L1 (q-val: 0.23) |

0.00 0.00 · 0.70 0.30 · 0.30 0.00 · 1.40 1.45 · 0.48 0.30 · 0.00 0.00 · 1.30 0.95

| DRAXIN (q-val: 0.23) | EFNA5 (q-val: 0.25) | IFIT1 (q-val: 0.13) | JPH2 (q-val: 0.23) | KRTAP5-3 (q-val: 0.23) | LINC00534 (q-val: 0.23) | LINC01588 (q-val: 0.23) |

0.30 0.00 · 0.48 0.30 · 0.00 0.00 · 0.30 0.00 · 0.30 0.00 · 1.80 1.45 · 1.71 1.62

| NALCN-AS1 (q-val: 0.23) | OR10Q1 (q-val: 0.23) | OR4A12P (q-val: 0.23) | PELI2 (q-val: 0.23) | RGS16 (q-val: 0.23) | RNF157 (q-val: 0.23) | RP11-321M21.3 (q-val: 0.23) |

0.85 0.60 · 0.48 0.30 · 0.00 0.00 · 0.30 0.00 · 0.30 0.30 · 1.20 0.30 · 0.30 0.30

| RP11-386J22.3 (q-val: 0.25) | RP11-96K19.4 (q-val: 0.08) | RP3-467K16.2 (q-val: 0.23) | RP5-1041C10.3 (q-val: 0.23) | SEC1P (q-val: 0.23) |

0.60 0.48 · 0.30 0.30 · 0.30 0.30 · 0.00 0.00 · 0.00 0.00

Implantation?
- No
- Yes

n=150 samples total each gene
67 no implantation
83 successful implantation

Plot 9 (Boxplot, just training data): Focusing on just the upregulated genes, specifically for each gene, distribution of expression
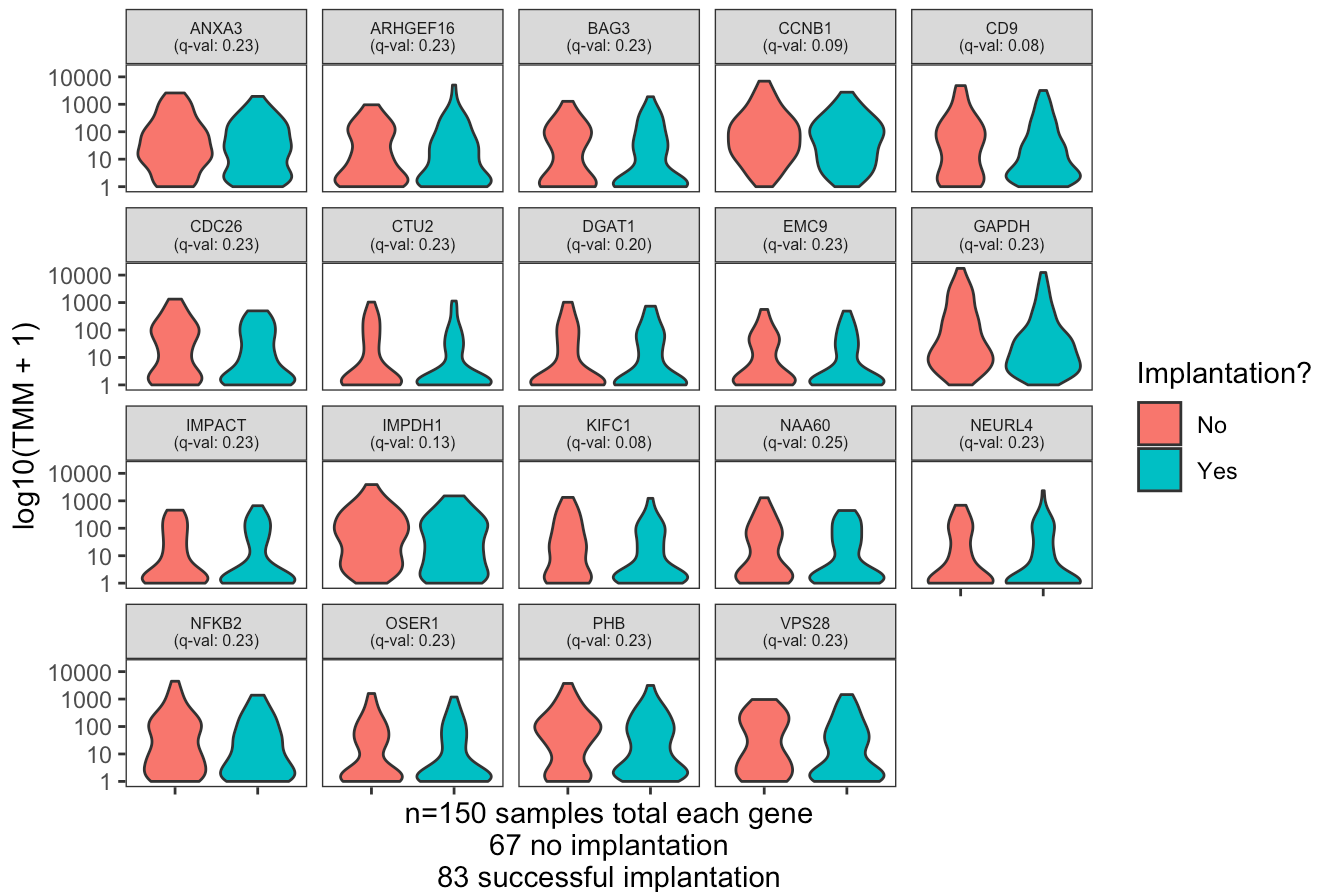
- genes separated by facets to better separate them from one another
- distribution of expression for samples which successfully implanted vs those that did not implant is visualized, separating using color and position on the x-axis
- Using the boxplot to visualize the quartiles and medians of the two groups
- Only using the training samples since those were the only samples used in the differential expression analysis. The idea with this was that perhaps separations in medians might be more stark than compared to looking at all the samples, or maybe some of the medians wouldn't go down
- some medians still go down

Upregulated Genes in Cases of Implantation (only training data)

n=111 training samples total for each gene
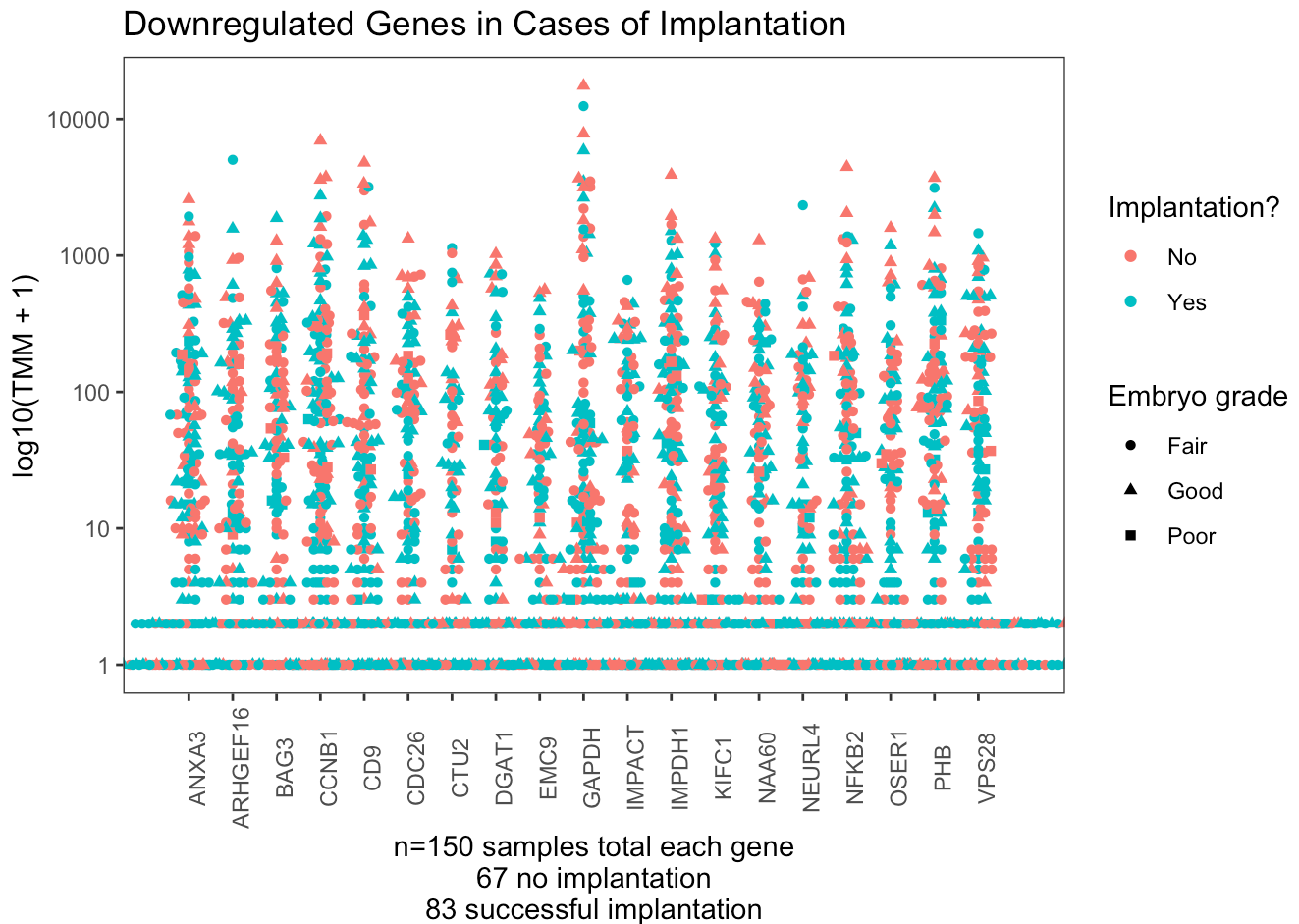51 no implantation
60 successful implantation

Plot 10 (Violin): Focusing on just the downregulated genes, specifically for each gene, distribution of expression for samples which successfully implanted vs those that did not implant

Downregulated Genes in Cases of Implantation

n=150 samples total each gene
67 no implantation
83 successful implantation

Plot 11 (Beeswarm): Focusing on just the downregulated genes, specifically for each gene, distribution of expression
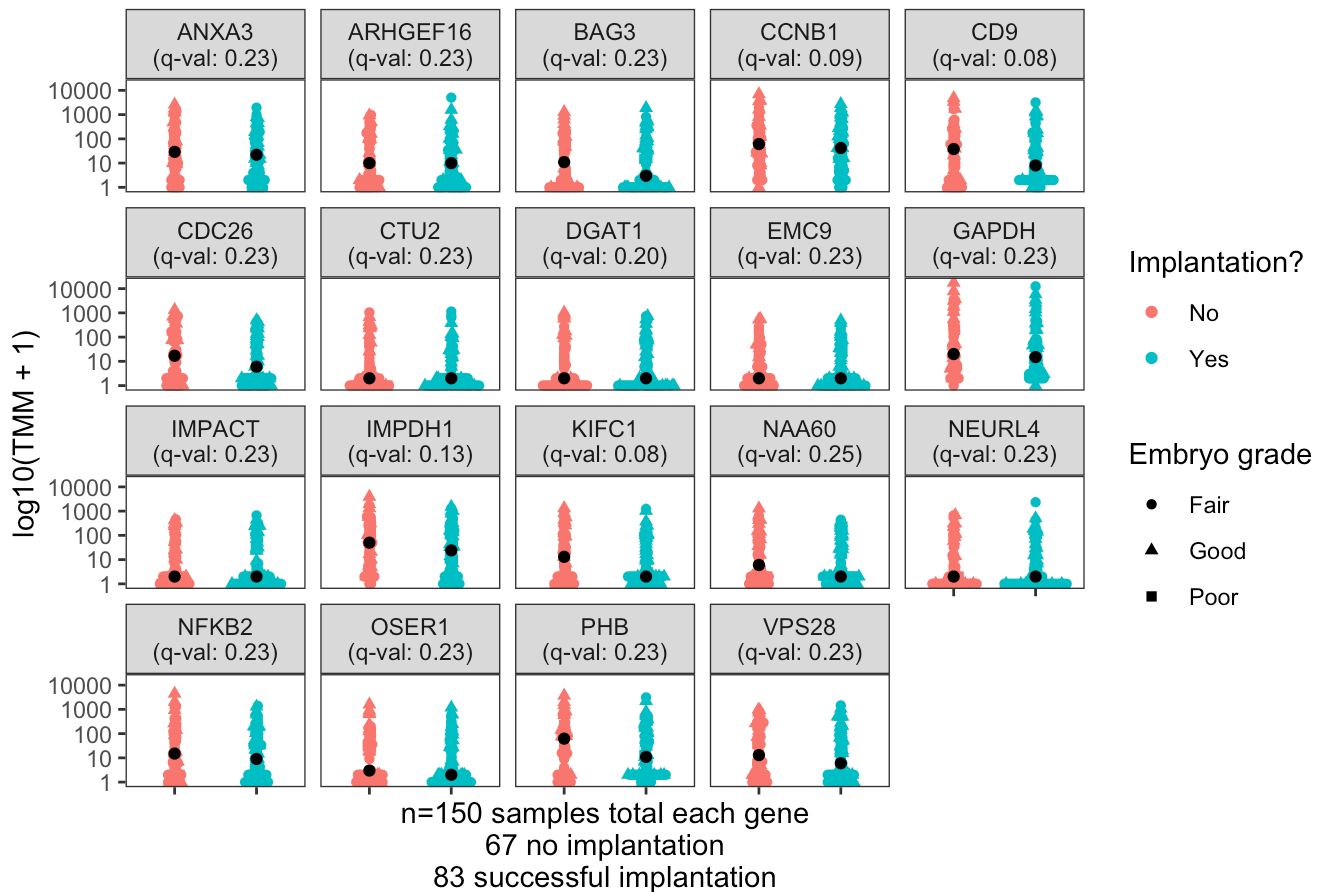
- genes separated on x-axis
- distribution of expression for samples which successfully implanted vs those that did not implant is visualized, separating using color
- shape of points reflects embryo grade at freezing
- Don't love that it's actually difficult to separate the genes from each other, specifically in the more lowly expressed samples section with this visualization. Also difficult to see the differences between the two implantation groups based on color alone

Downregulated Genes in Cases of Implantation

n=150 samples total each gene
67 no implantation
83 successful implantation

Plot 12 (Beeswarm, faceted): Focusing on just the downregulated genes, specifically for each gene, distribution of expression

- genes separated by facets to better separate them from one another
- distribution of expression for samples which successfully implanted vs those that did not implant is visualized, separating using color and position on the x-axis
- shape of points reflects embryo grade at freezing
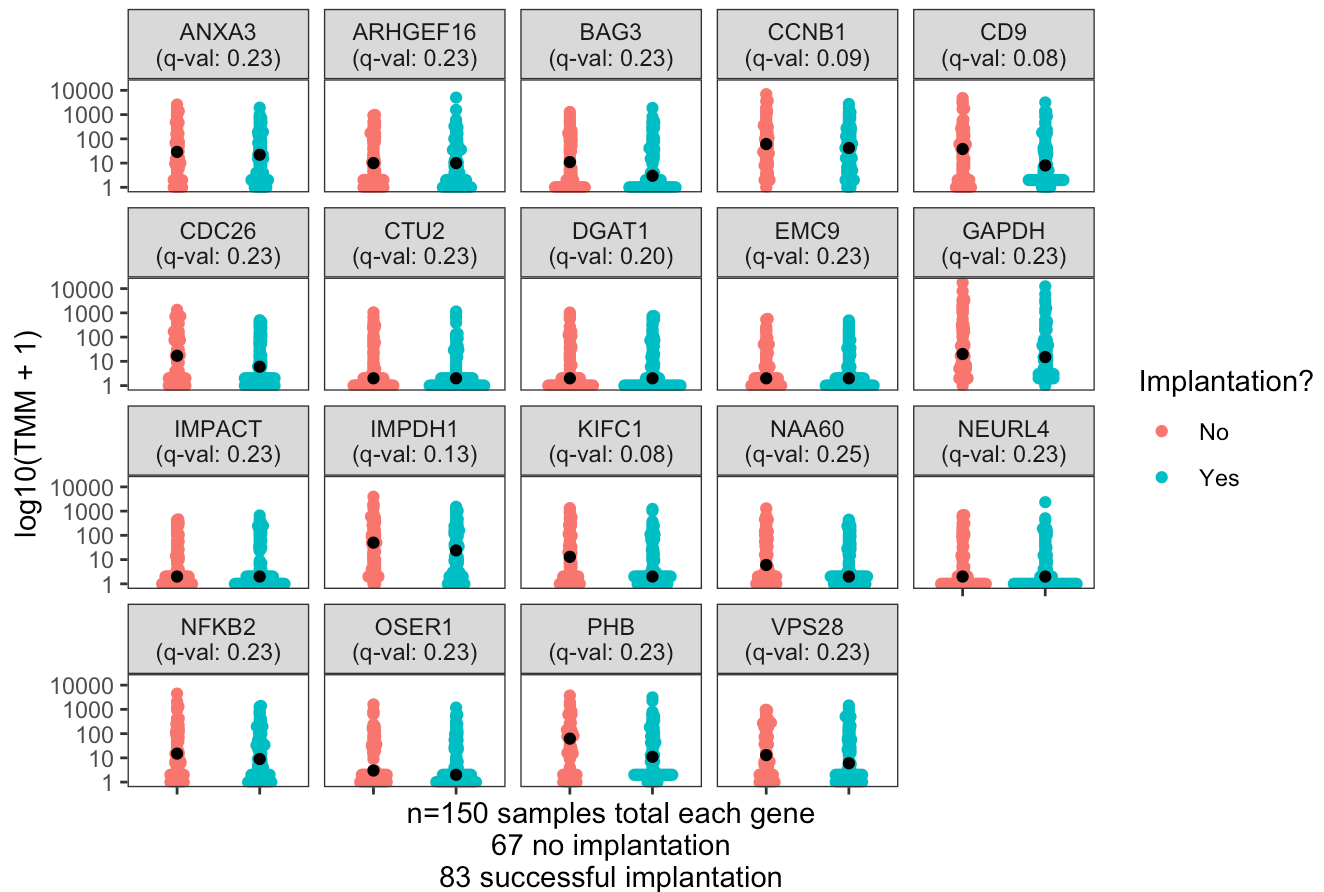- Also difficult to see the differences between the two implantation groups – so I added points for the medians

Downregulated Genes in Cases of Implantation

n=150 samples total each gene
67 no implantation
83 successful implantation

Plot 13 (Beeswarm, faceted, no embryo grade): Focusing on just the downregulated genes, specifically for each gene, distribution of expression

- genes separated by facets to better separate them from one another
- distribution of expression for samples which successfully implanted vs those that did not implant is visualized, separating using color and position on the x-axis
- Also difficult to see the differences between the two implantation groups so I added points for the medians

Downreguated Genes in Cases of Implantation

n=150 samples total each gene
67 no implantation
83 successful implantation

Plot 14 (Boxplot): Focusing on just the downregulated genes, specifically for each gene, distribution of expression

- genes separated by facets to better separate them from one another
- distribution of expression for samples which successfully implanted vs those that did not implant is visualized, separating using color and position on the x-axis
- Using the boxplot to visualize the quartiles and medians of the two groups

Downregulated Genes in Cases of Implantation
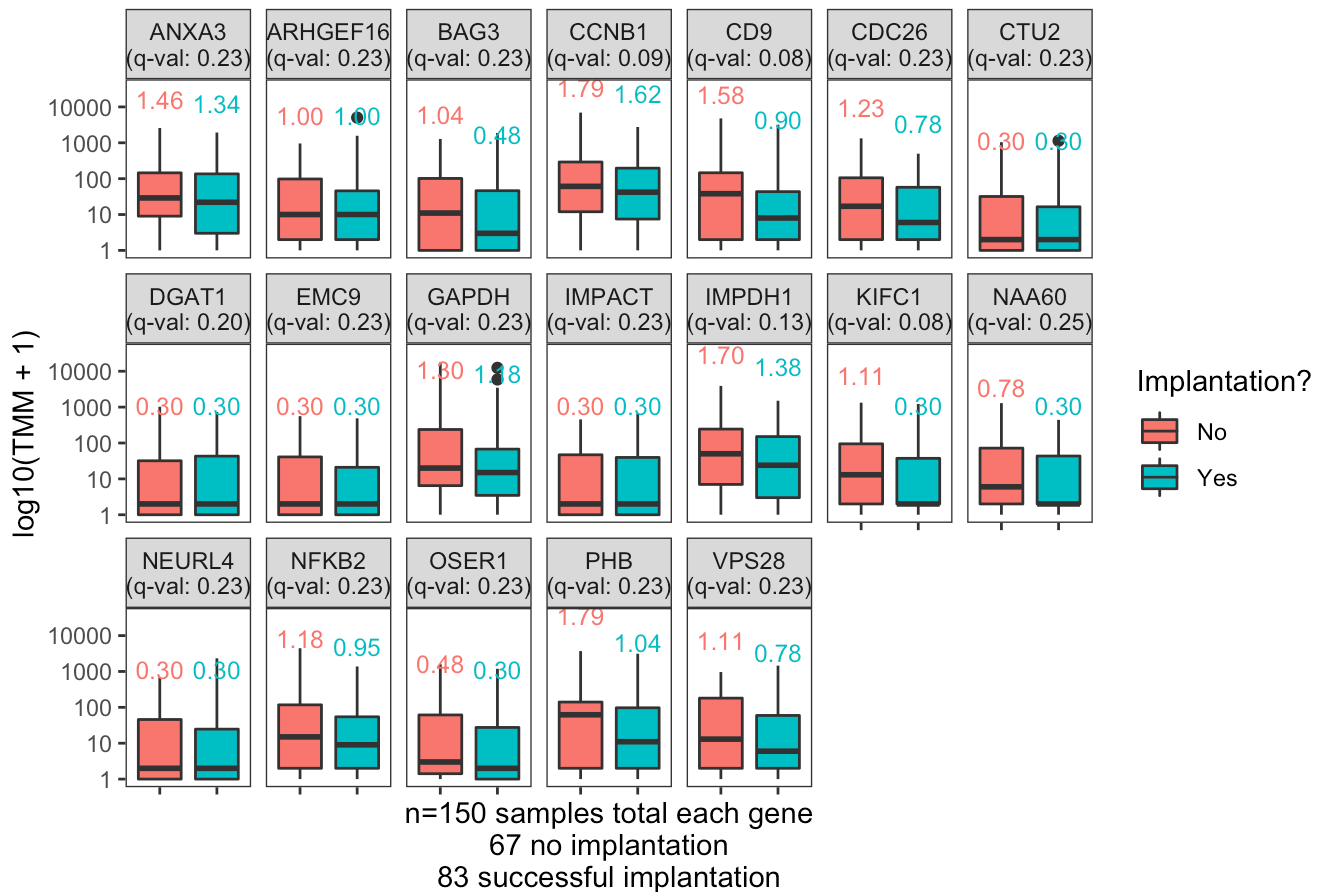
Plot 15 (Boxplot, just training data): Focusing on just the downregulated genes, specifically for each gene, distribution of expression

- genes separated by facets to better separate them from one another
- distribution of expression for samples which successfully implanted vs those that did not implant is visualized, separating using color and position on the x-axis
- Using the boxplot to visualize the quartiles and medians of the two groups
- Only using the training samples since those were the only samples used in the differential expression analysis. The idea with this was that perhaps separations in medians might be more stark than compared to looking at all the samples

Downregulated Genes in Cases of Implantation (only training data)

n=111 samples total each gene
51 no implantation
60 successful implantation