

# TABI Report

*Isobel Beasley*

*Updated 15/02/2020*

## Contents

<b>(1) Rationale:</b>	<b>1</b>
<b>(2) Methodology:</b>	<b>1</b>
<b>(3) Summary Statistics:</b>	<b>3</b>
Distribution of Calculated Parameter Values . . . . .	4
Relationship between calculated parameters and log fold gene change . . . . .	5
<b>(4) Results / Plots:</b>	<b>6</b>
<b>(5) Examples of Trends Detected</b>	<b>9</b>
Examples of Genes with Localised Inflections (i.e. Sigmoidal-like Trends) ( <i>Calculated Inflection standard deviation &lt;0.7375, in the first quartile</i> ) . . . . .	9
Examples of Genes with Wide Inflections (i.e. Linear-like Trend) ( <i>Inflection sd &gt;0.9749, in the fourth quartile</i> ) .	10
Examples of trends initially caused by outliers . . . . .	10
<b>(6) Biological Results and Interpretation:</b>	<b>11</b>
<b>(7) Appendix:</b>	<b>13</b>
Tables from overrepresentation tests . . . . .	13

## **(1) Rationale:**

Presently, when approaching the analysis of differential gene expression from RNA seq data in Prostate Cancer tumours, samples are split into two arbitrary groups; low and high risk of recurrence. A linear equation is then fitted to the read counts and used to distinguish between genes that increase, decrease or have a constant level of transcription throughout the progression of the disease. Whilst this approach is quite simple to both understand and apply, it results in the loss of a considerable amount of information through compressing several risk scores into just two groups.

In order to increase the amount of information able to be extracted from TCGA data on Prostate Cancer (and potentially other cancers), our aim was to develop and improve a Bayesian inference model for differential gene transcript abundance on continuous covariates (TABI). This model would utilise the pseudo-continuous nature of CAPRA-S (Cancer of the Prostate Risk Assessment Postsurgical; a risk of recurrence score from 0 to 7, where 0 is benign tissue) in order to calculate more precise stages of disease progression for which gene expression change occurs. This in turn would enable the creation of a more accurate picture of the transcriptomic tumour microenvironment at different stages of the disease than has previously been possible.

Furthermore such a model may be more representative of the underlying biology as gene transcription often has lower and upper limits as a result of biological requirements and constraints, which liner equation is unable to represent (for e.g. there may be a basal level of gene transcription of an essential gene product required for a cell to survive, or an upper limit where too much of a gene product becomes toxic or where there is insufficient enzymatic activity to further catalyse gene transcription). Hence, our approach may capture more information about the causitive biological processes involved in disease progression than current practices.

## (2) Methodology:

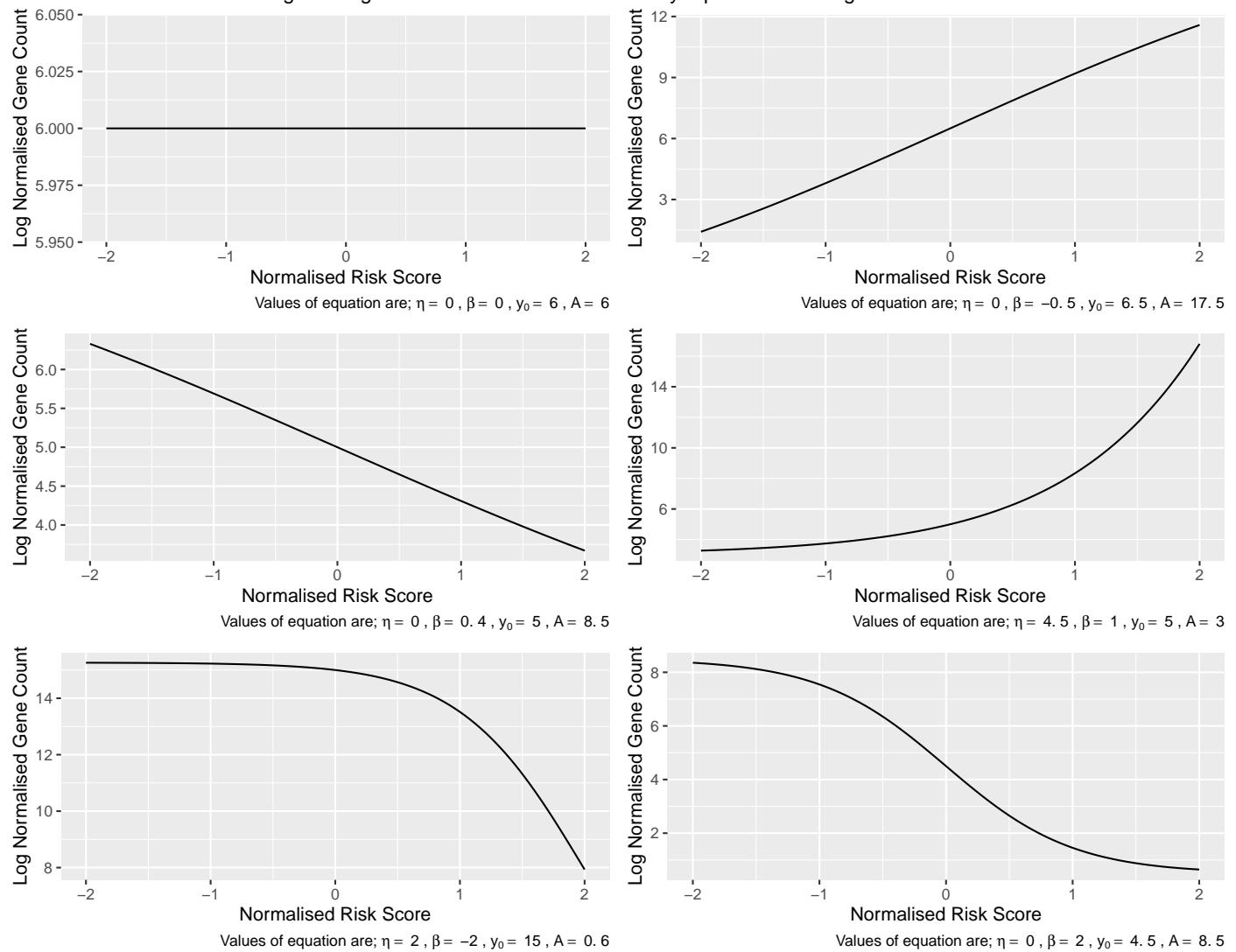
Raw transcript abundance from the TCGA data set on Prostate Cancer was normalised using the TMM algorithm (through tidyTranscriptomics package ttBulk, GitHub: stemangiola/ttBulk@dev). The effect of known and unknown confounding factors were considered and resulted in the incorporation of purity score as an secondary continuous covariate in the running of the model. The model itself was encoded in the probabilistic programming language Stan, with the dataset partitioned and run in blocks of 5000 genes.

The model assumes that the normalised gene read counts follow a negative binomial distribution (defined by mean and overdispersion parameters). The log of the expected value normalised gene read ( $\log(\hat{Y})$ ) was fitted to a reparametrised sigmoidal curve;

$$(1) GLA(X, y_0, \beta, \eta, A) = A + \frac{(y_0 - A)(1 + e^{\eta\beta_1})}{(1 + e^{\eta\beta_1 - X\beta})}$$

In this equation;  $X$  is the design matrix containing scaled CAPRA-S risk scores and purity scores,  $y_0$  is the log of the expected read count at the middle CAPRA-S risk score,  $\beta$  is the matrix of slopes (such that  $\beta_1$  refers to the slope corresponding to the main factor of interest, in our case CAPRA-S),  $\eta$  refers to the x coordinate of the highest rate of change in the equation (i.e. the x coordinate of the halfway point between the upper and lower plateaus) which is used to infer the stage at which gene expression changes, and  $A$  is a vertical translation parameter which alters the height of either the lower or upper plateau (depending on whether the slope is negative or positive). This equation is able to model a variety of gene expression trends with respect to CAPRA-S score including; constitutive, linear increase or decrease, exponential increase or decrease, increase or decrease leading to or from plateau, and two plateaus with an intermediate rising or falling phase (demonstrated in plots below).

Fig 1. Range of trends able to be modelled by reparametrised sigmoidal curve



In comparison to previous versions of the model, parameter  $A$  was introduced to better model situations where gene expression has a non-zero basal level (without  $A$  the lower plateau is by definition 0). Additionally, after the incorporation of  $A$  into the model  $y_0$  was restricted by introducing the parameter  $y_{0\text{raw}}$  such that;

$$(2) \quad y_0 = y_{0\text{raw}} + A \quad (y_{0\text{raw}} \geq 0)$$

in order to prevent problems with instances of multiple solutions. A linear relationship was fitted between  $\log(\hat{Y})$  and  $\phi$  (log of the reciprocal of overdispersion) over all genes such that the slope coefficient obtained ( $-0.3186$ ) was input into the model so the gene specific value of  $\phi$  was defined to be

$$(3) \quad \phi = -0.3186 \cdot \log(\hat{Y}) + \sigma$$

(where  $\sigma$  was added to account for variation between genes in this relationship). The statistical model is hence defined by the above equations and the following (where  $Y$  refers to the normalised read count, and  $\hat{Y}$ ) is the expected normalised transcript abundance);

$$(4) \quad Y_{t,g} \sim NB(Y_{t,g}, \frac{1}{e^\phi})$$

$$(5) \quad \eta \sim N(0, 1)$$

$$(6) \quad y_{0\text{raw}} \sim N(0, 1)$$

$$(7) \quad A \sim N(0, 2)$$

$$(8) \quad \sigma \sim N(0, 1)$$

$$(9) \quad \beta_r \sim N(0, 1)$$

Furthermore, before fitting this equation to the sigmoidal equation, the covariate data (in our case CAPRA-S and purity score) were both scaled to have a mean of 0 and a standard deviation of 1 (this step was performed by TABI). Henceforth, any discussion of the value of  $\eta$  (which may be considered the CAPRA-S value at which the largest difference in gene transcription between CAPRA-S scores occurs) is in reference to its scaled value (as this is what the model uses to make predictions) and not its raw unscaled value.

### (3) Summary Statistics:

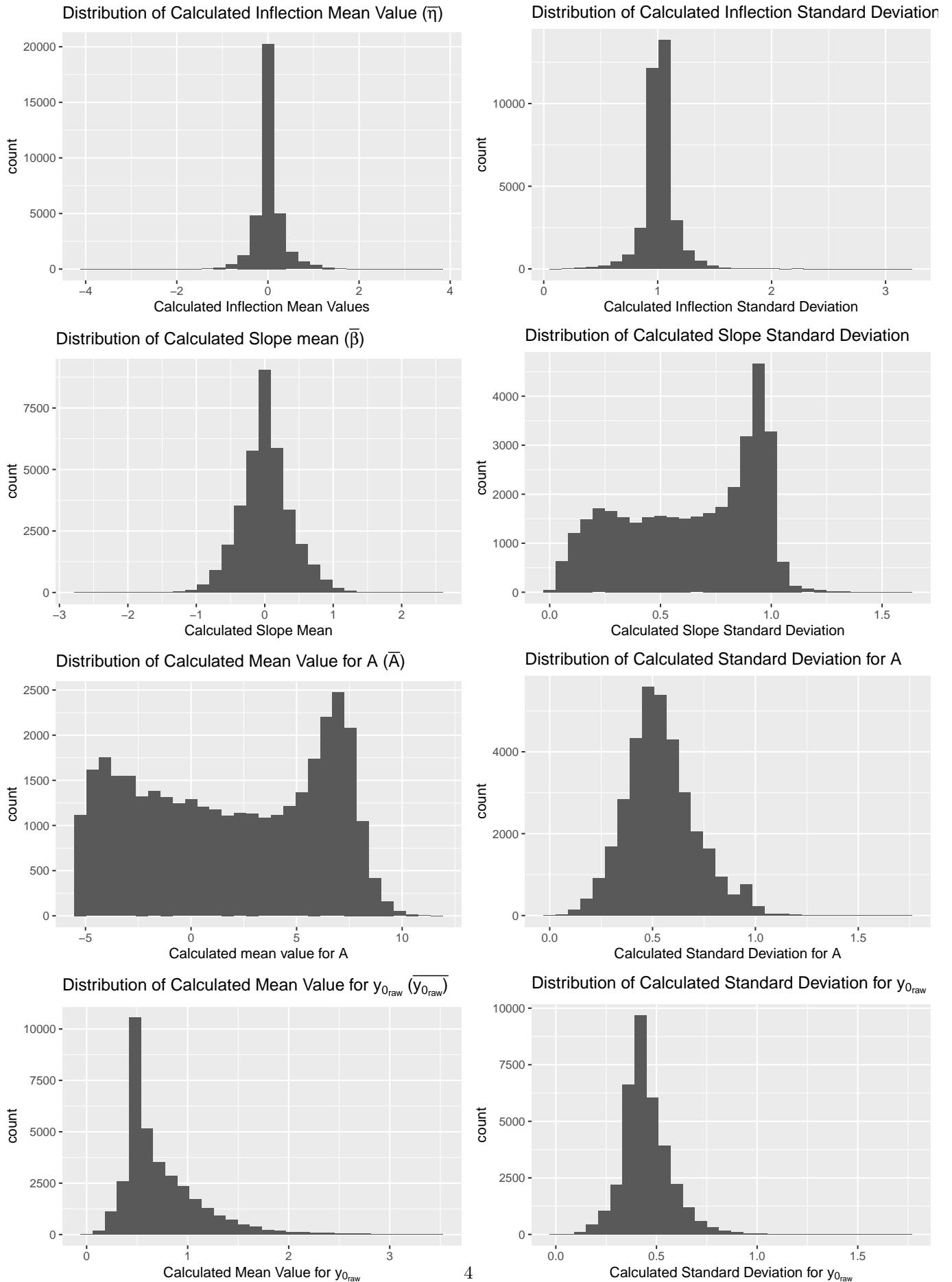
Any genes with more than one non-zero normalised read count were passed to TABI ( $n = 34862$  of a total 37318 measured in the TCGA data set). An initial run of 500 iterations (250 warm-up, 250 sampling) resulted in 3695 genes (9.9 % of all genes) being assigned as having a statistically significant slope coefficient with respect to CAPRA-S (proxy for differential gene expression through the progression of the disease). Of all analysed parameters (total  $n = 418356$ ) the majority ( $n = 402203$ ) had an

$$Rhat \leq 1.1$$

indicating that in most cases chains had mixed well in the analysis and effective sample sizes were sufficient (average effective sample size  $n = 403$ ). Once this first passage was complete, generated quantities for  $Y$  (with 95 width intervals) were used to identify and remove outliers for a second pass. A larger number of iterations were utilised in the second passage (5000, 2500 warm-up, 2500 sampling) in order to better define \$95% credible intervals for our parameters. (\*\*Add when second run is complete)

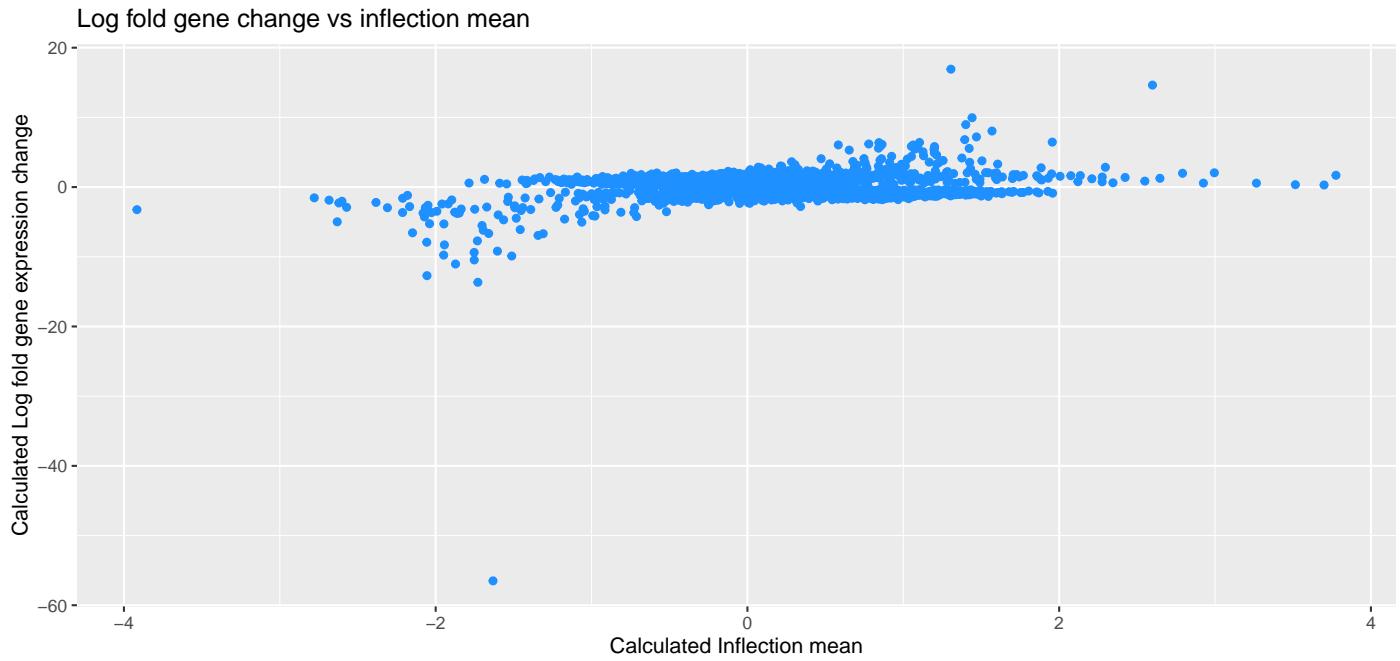
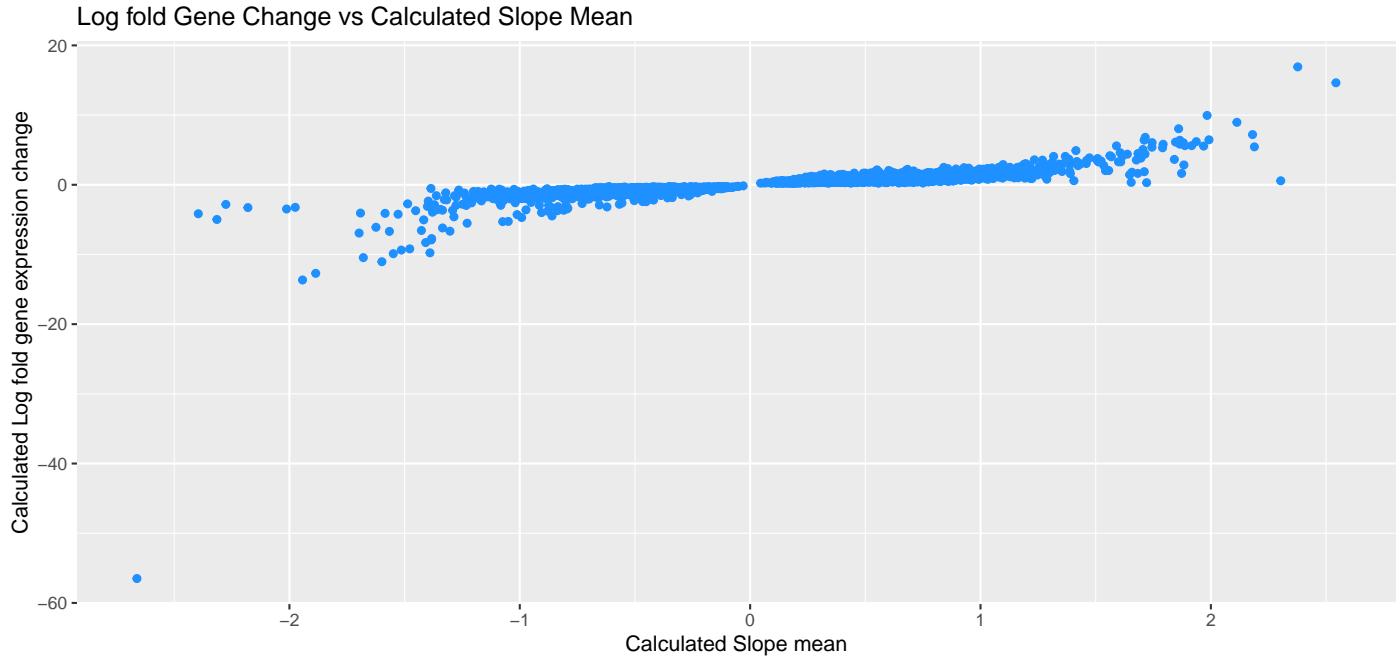
## Distribution of Calculated Parameter Values

Fig 2. Distribution of calculated mean and standard deviation of model parameters



\*\*Add

## Relationship between calculated parameters and log fold gene change

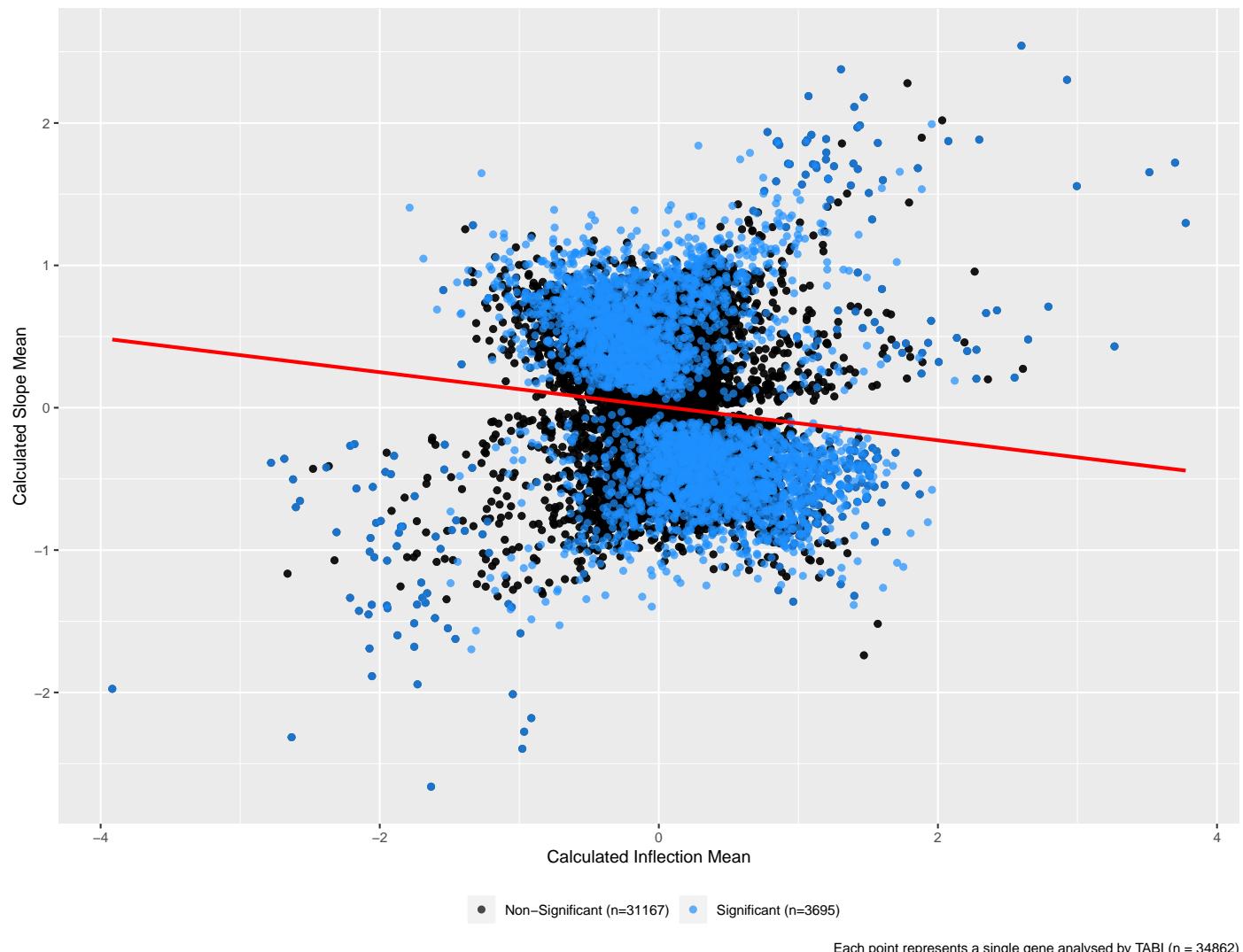


xpression change calculated as the log of the expected value of read count at a CAPRA-S score of 7 (estimated using mean parameter modeled sigmoidal equation) minus the same calculation at a CAPRA-S score of 0

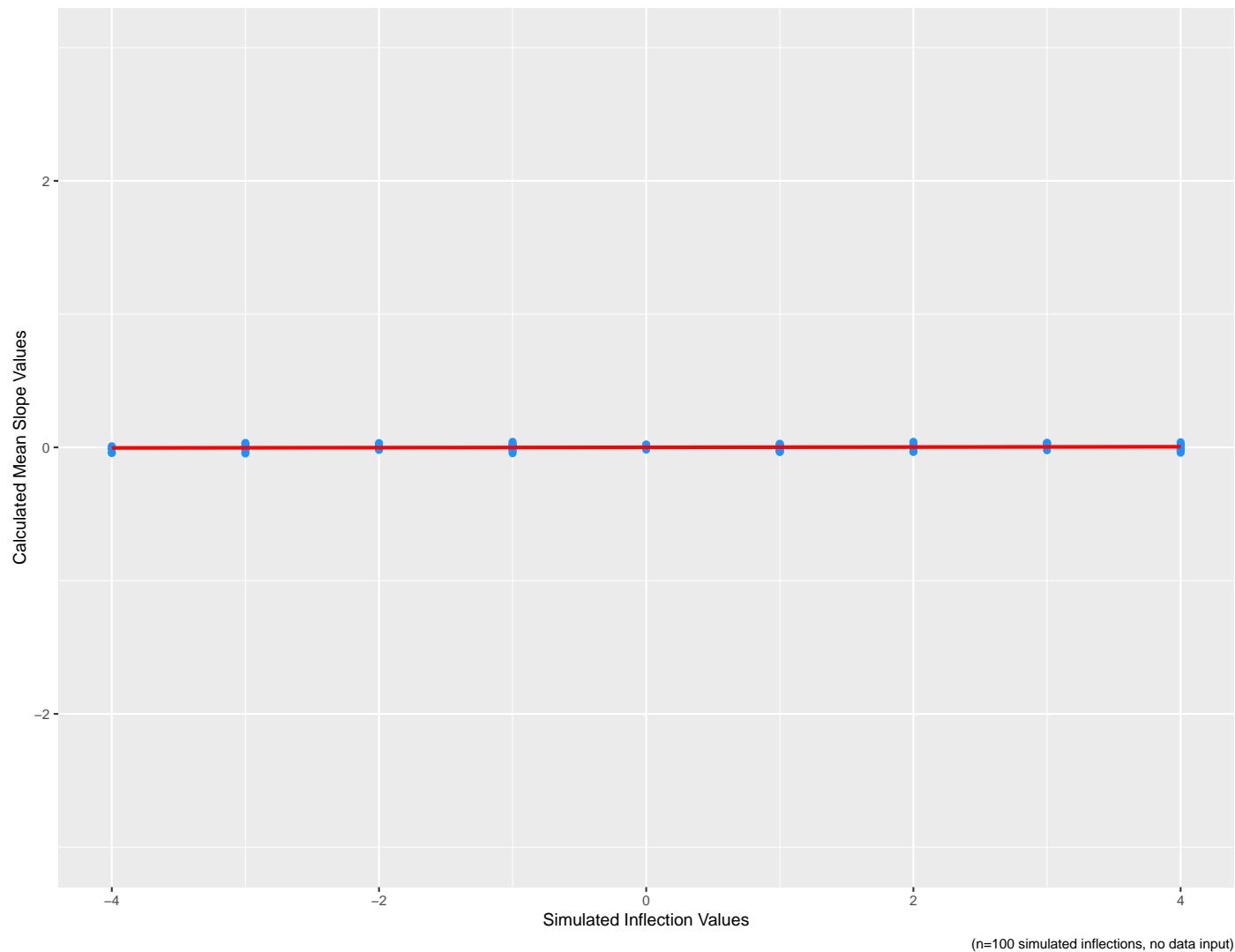
The above graph shows that whilst the sign of the slope and the direction of gene change is consistent (as we would expect) there is not a precise relationship between log fold gene change and slope. This is because parameters  $A$  and  $y_0$  impact the calculated upper and lower plateaus, which potentially impact the initial value or final value of the equation and hence - log fold gene change. \*\*Add

## (4) Results / Plots:

Fig 2. TABI Calculated Slope mean vs Inflection mean

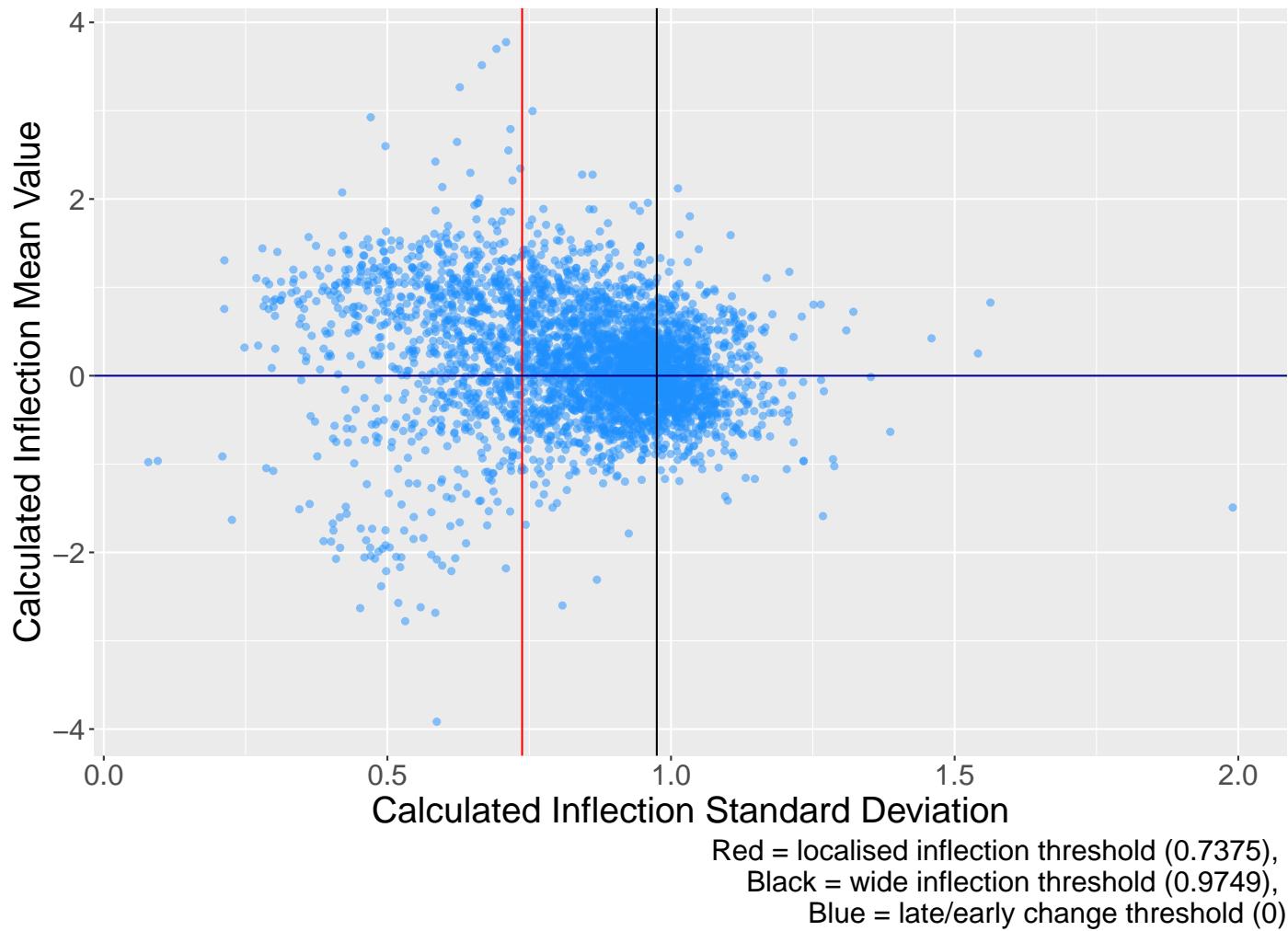


### TABI Calculated Slope mean vs Inputted Inflection



In general there is a negative trend between the calculated slope mean and inflection mean (i.e. genes which change expression levels late in disease progression tend to be those which decrease in gene expression over disease progression). Potential outliers (genes with large absolute values of mean slope, in the top right hand corner of the graph and the bottom left hand corner of the graph) may be skewing the plotting of a linear relationship between mean slope and inflection, hence the true trend may be larger and stronger than that calculated above.

## Calculated Inflection Mean Value vs Standard Deviation (n=34862)

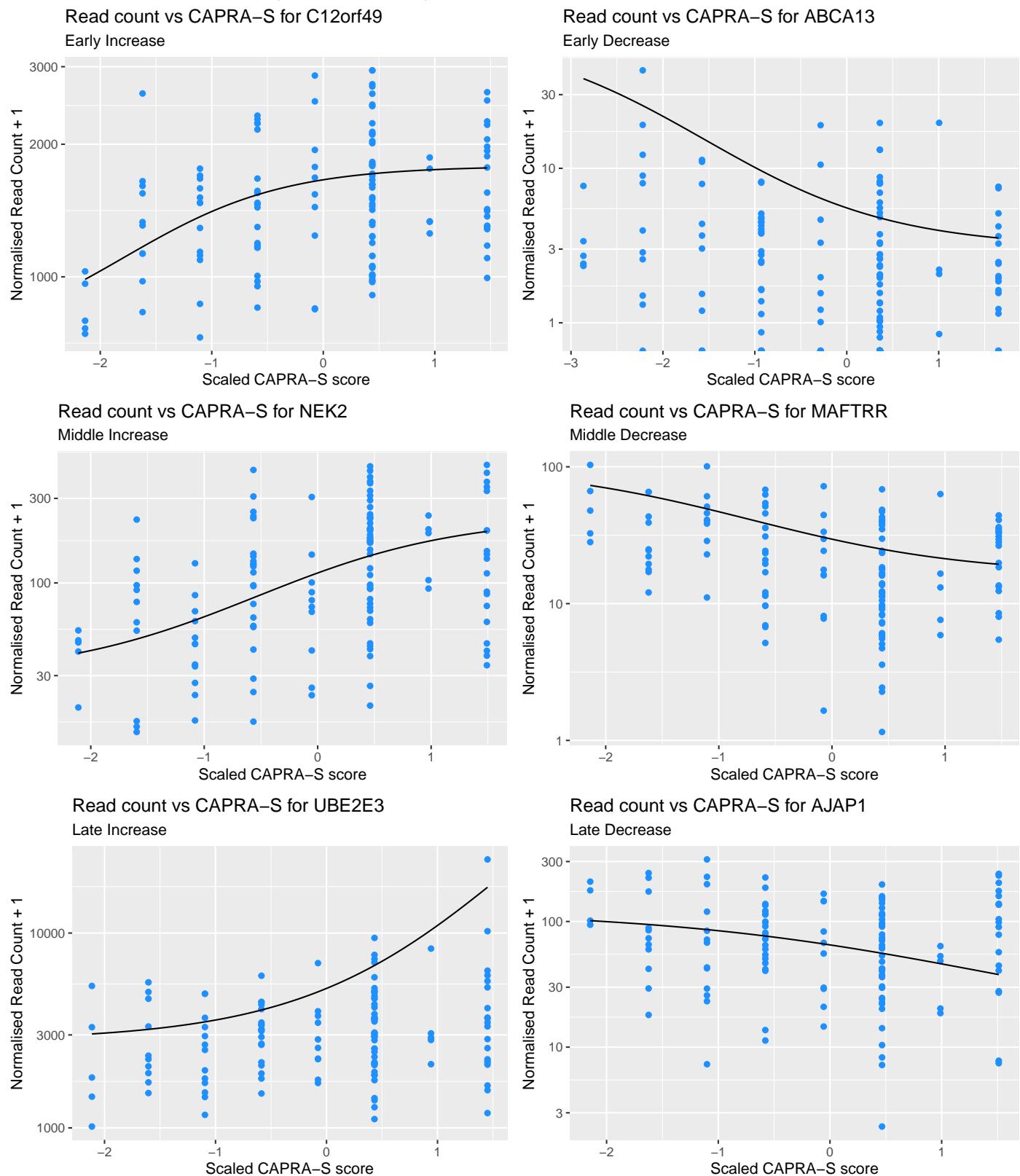


As can be seen in the above plot, the mean and standard deviation of inflection cluster around zero and one respectively (may be a reflection of our prior on inflection of  $N(0,1)$ ). Low standard deviation for inflection is a indication of a sharper inflection (a gene with a more sigmoidal trend), so points towards the left of the plot are good candidates for sigmoidal trends. From this point on, localised inflection refers to genes with an inflection sd in the first quartile (i.e.  $<0.7375$ , generally genes with more sigmoidal like trends) and wide inflection refers to genes with an inflection sd in the fourth quartile (i.e.  $>0.9749$ , generally genes with more linear like trends).

## (5) Examples of Trends Detected

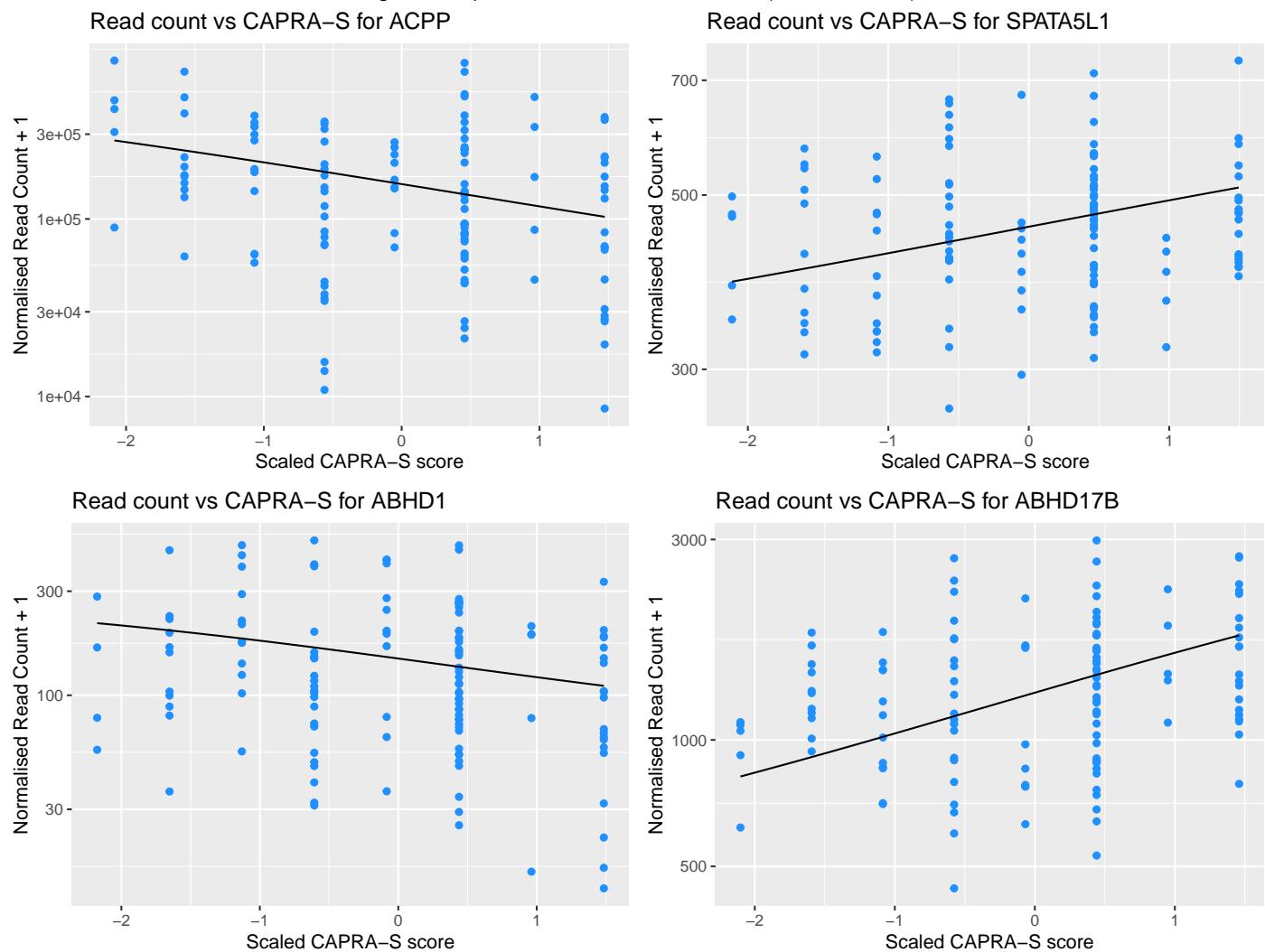
Examples of Genes with Localised Inflections (i.e. Sigmoidal-like Trends) (*Calculated Inflection standard deviation <0.7375, in the first quartile*)

Fig 3. Examples of Sigmodial Trends Detected (Localised inflection)



## Examples of Genes with Wide Inflections (i.e. Linear-like Trend) ( $Inflection\ sd > 0.9749$ , in the fourth quartile)

Fig 4. Examples of Linear Trends Detected (Wide inflection)

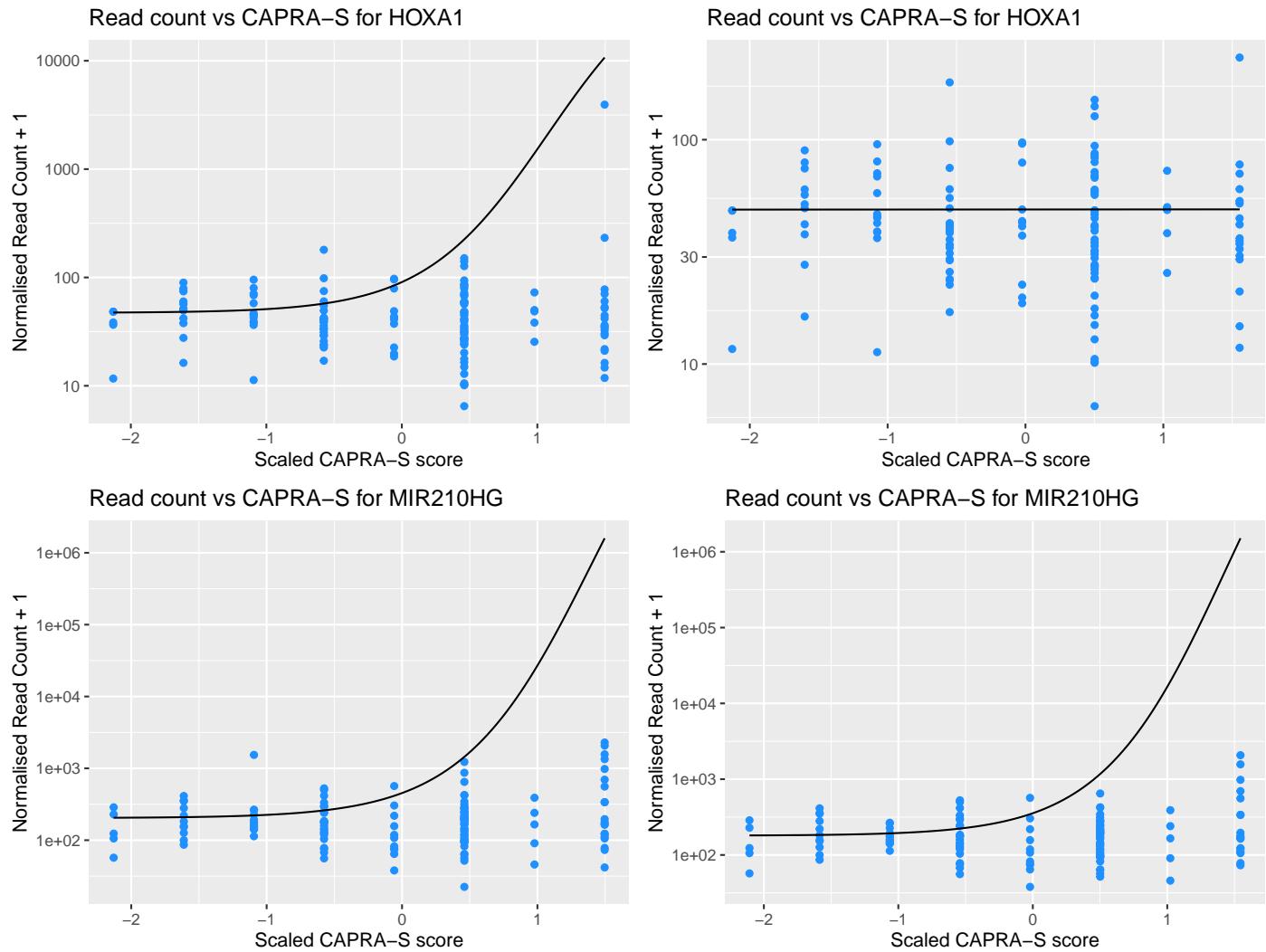


Each point represents a recorded result from a different sample. The overlayed black line is the curve fitted by the model.

## Examples of trends initially caused by outliers

\*\*\*To add comment when second passage has finished running

Fig 5. Examples of Trends Before (Left) and After (Right) Outlier Removal



Each point represents a recorded result from a different sample. The overlayed black line is the curve fitted by the model.

## (6) Biological Results and Interpretation:

Of all statistically significant genes, those with localised inflections (calculated inflection standard deviations  $<0.7375$ ) were taken, and two further groups were created; genes with early changes (defined as calculated inflection mean  $<0$ ) and genes with late changes (defined as calculated inflection mean  $>0$ ). Two different overrepresentation test were then undertaken (PANTHER Overrepresentation Test using GO Ontology database Released 2020-01-03, From: <http://geneontology.org/> and Computed overlap with Molecular Signatures Database (MSigDB)\* From: <https://www.gsea-msigdb.org/gsea/msigdb/index.jsp>) and the functional gene groups which were significantly overrepresented in these two groups were collected (see tables in appendix).

Of all declared statistically significant overlap between genes with localised early changes and genes sets noticeable is the recurrent instance of larger than expected numbers of genes from gene sets involved with **cell cycle and division** (*e.g mitotic spindle assembly, sister chromatid segregation, organelle fission, cell cycle gene groups etc.*) as well as those involved in **immunological processes** (*phagocytosis, response to bacterium, B cell receptor signalling pathway, complement activation etc.*).

Table 1: Significant results (P-value <0.05) from PANTHER Overrepresentation Test using GO Ontology database Released 2020-01-03 (Genes with early changes) - Top 4 Broad Categories

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
phagocytosis, recognition (GO:0006910)	106	13	1.02	+	12.68	1.24e-10	3.96e-07
B cell receptor signaling pathway (GO:0050853)	126	13	1.22	+	10.67	8.76e-10	1.27e-06
mitotic sister chromatid segregation (GO:0000070)	106	12	1.02	+	11.71	1.51e-09	1.72e-06
mitotic spindle assembly (GO:0090307)	37	5	0.36	+	13.98	4.89e-05	1.59e-02

In comparison, genes with localised late changes were recurrently significantly associated with overlap in genes sets involved in **cellular, muscular and neural development and contraction** (*muscle filament sliding, sarcomere organization, actomyosin structure organization, neurotransmitter transport, neuron differentiation, neurogenesis, long-term synaptic potentiation etc.*).

Table 2: Significant Results (P-value <0.05) from PANTHER Overrepresentation Test using GO Ontology database Released 2020-01-03 (Genes with late changes) - Top 4 Broad Categories

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
muscle filament sliding (GO:0030049)	38	10	0.94	+	10.58	1.85e-07	8.94e-05
chemical synaptic transmission (GO:0007268)	427	31	10.62	+	2.92	3.16e-07	1.36e-04
skeletal muscle thin filament assembly (GO:0030240)	9	4	0.22	+	17.88	2.06e-04	3.12e-02
cardiac myofibril assembly (GO:0055003)	21	5	0.52	+	9.58	3.55e-04	4.68e-02

When this analysis was reapplied using Molecular Signatures Database (MSigDB, with collections C1, C3, C4, C5, C6, C7, H) a similar trend of gene set overlap was found, now with localised early changes also showing associations with genes sets of hallmark changes found in other cancers (such as lung and breast) and late changes showed association between known genes sets of hallmarks of prostate cancer.

## (7) Appendix:

### Tables from overrepresentation tests

Table 3: Significant results (P-value <0.05) from PANTHER Over-representation Test using GO Ontology database Released 2020-01-03 (Genes with early changes) - Sorted by P-value

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
organelle fission (GO:0048285)	311	24	3.01	+	7.98	1.88e-14	3.00e-10
nuclear division (GO:0000280)	282	23	2.73	+	8.44	2.30e-14	1.83e-10
nuclear chromosome segregation (GO:0098813)	217	18	2.10	+	8.58	1.33e-11	7.09e-08
mitotic nuclear division (GO:0140014)	142	15	1.37	+	10.93	3.14e-11	1.25e-07
phagocytosis, recognition (GO:0006910)	106	13	1.02	+	12.68	1.24e-10	3.96e-07
cell cycle process (GO:0022402)	1040	35	10.06	+	3.48	1.87e-10	4.97e-07
mitotic cell cycle (GO:0000278)	734	29	7.10	+	4.09	2.43e-10	5.54e-07
sister chromatid segregation (GO:0000819)	139	14	1.34	+	10.42	2.59e-10	5.16e-07
mitotic cell cycle process (GO:1903047)	655	27	6.33	+	4.26	4.43e-10	7.85e-07
chromosome segregation (GO:0007059)	276	18	2.67	+	6.75	5.27e-10	8.40e-07
B cell receptor signaling pathway (GO:0050853)	126	13	1.22	+	10.67	8.76e-10	1.27e-06
phagocytosis, engulfment (GO:0006911)	128	13	1.24	+	10.50	1.05e-09	1.39e-06
cell cycle (GO:0007049)	1363	39	13.18	+	2.96	1.41e-09	1.73e-06
mitotic sister chromatid segregation (GO:0000070)	106	12	1.02	+	11.71	1.51e-09	1.72e-06
complement activation, classical pathway (GO:0006958)	161	14	1.56	+	8.99	1.53e-09	1.63e-06

Table 3: Significant results (P-value <0.05) from PANTHER Over-representation Test using GO Ontology database Released 2020-01-03 (Genes with early changes) - Sorted by P-value (*continued*)

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
humoral immune response mediated by circulating immunoglobulin (GO:0002455)	166	14	1.60	+	8.72	2.21e-09	2.20e-06
plasma membrane invagination (GO:0099024)	137	13	1.32	+	9.81	2.25e-09	2.11e-06
membrane invagination (GO:0010324)	144	13	1.39	+	9.34	3.93e-09	3.48e-06
complement activation (GO:0006956)	176	14	1.70	+	8.23	4.46e-09	3.74e-06
cell recognition (GO:0008037)	233	15	2.25	+	6.66	1.80e-08	1.44e-05
positive regulation of B cell activation (GO:0050871)	169	13	1.63	+	7.96	2.34e-08	1.78e-05
immunoglobulin mediated immune response (GO:0016064)	204	14	1.97	+	7.10	2.58e-08	1.87e-05
B cell mediated immunity (GO:0019724)	207	14	2.00	+	7.00	3.06e-08	2.12e-05
meiotic nuclear division (GO:0140013)	155	12	1.50	+	8.01	7.69e-08	5.11e-05
meiotic cell cycle (GO:0051321)	224	14	2.17	+	6.46	7.72e-08	4.93e-05
meiotic cell cycle process (GO:1903046)	169	12	1.63	+	7.34	1.86e-07	1.14e-04
regulation of B cell activation (GO:0050864)	213	13	2.06	+	6.31	2.95e-07	1.74e-04
defense response to bacterium (GO:0042742)	338	16	3.27	+	4.90	3.35e-07	1.91e-04
meiotic chromosome segregation (GO:0045132)	90	9	0.87	+	10.34	4.66e-07	2.56e-04
lymphocyte mediated immunity (GO:0002449)	262	14	2.53	+	5.53	4.72e-07	2.51e-04

Table 3: Significant results (P-value <0.05) from PANTHER Over-representation Test using GO Ontology database Released 2020-01-03 (Genes with early changes) - Sorted by P-value (*continued*)

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains (GO:0002460)	270	14	2.61	+	5.36	6.65e-07	3.42e-04
cell division (GO:0051301)	504	18	4.87	+	3.69	3.04e-06	1.52e-03
adaptive immune response (GO:0002250)	612	20	5.92	+	3.38	3.11e-06	1.50e-03
gene expression (GO:0010467)	2043	3	19.75	-	0.15	3.60e-06	1.69e-03
positive regulation of leukocyte activation (GO:0002696)	410	16	3.96	+	4.04	3.76e-06	1.71e-03
positive regulation of lymphocyte activation (GO:0051251)	368	15	3.56	+	4.22	4.61e-06	2.04e-03
humoral immune response (GO:0006959)	371	15	3.59	+	4.18	5.06e-06	2.18e-03
phagocytosis (GO:0006909)	325	14	3.14	+	4.46	5.28e-06	2.22e-03
chromosome organization (GO:0051276)	1060	27	10.25	+	2.63	5.53e-06	2.26e-03
positive regulation of cell activation (GO:0050867)	424	16	4.10	+	3.90	5.66e-06	2.25e-03
cellular component organization (GO:0016043)	5600	84	54.14	+	1.55	6.84e-06	2.66e-03
antigen receptor-mediated signaling pathway (GO:0050851)	294	13	2.84	+	4.57	8.83e-06	3.35e-03

Table 3: Significant results (P-value <0.05) from PANTHER Over-representation Test using GO Ontology database Released 2020-01-03 (Genes with early changes) - Sorted by P-value (*continued*)

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
microtubule cytoskeleton organization involved in mitosis (GO:1902850)	175	10	1.69	+	5.91	1.23e-05	4.55e-03
response to bacterium (GO:0009617)	708	20	6.85	+	2.92	2.45e-05	8.87e-03
mitotic spindle organization (GO:0007052)	152	9	1.47	+	6.12	2.58e-05	9.14e-03
cellular component organization or biogenesis (GO:0071840)	5793	84	56.01	+	1.50	2.62e-05	9.09e-03
import into cell (GO:0098657)	670	19	6.48	+	2.93	3.75e-05	1.27e-02
immune response-activating cell surface receptor signaling pathway (GO:0002429)	446	15	4.31	+	3.48	4.06e-05	1.35e-02
mitotic spindle assembly (GO:0090307)	37	5	0.36	+	13.98	4.89e-05	1.59e-02
spindle organization (GO:0007051)	216	10	2.09	+	4.79	6.78e-05	2.16e-02
regulation of cell activation (GO:0050865)	645	18	6.24	+	2.89	7.37e-05	2.30e-02
reproductive process (GO:0022414)	1437	30	13.89	+	2.16	7.37e-05	2.26e-02
reproduction (GO:0000003)	1440	30	13.92	+	2.15	7.60e-05	2.29e-02
immune response-regulating cell surface receptor signaling pathway (GO:0002768)	481	15	4.65	+	3.23	9.23e-05	2.72e-02
mitotic cell cycle phase transition (GO:0044772)	277	11	2.68	+	4.11	1.10e-04	3.19e-02

Table 3: Significant results (P-value <0.05) from PANTHER Overrepresentation Test using GO Ontology database Released 2020-01-03 (Genes with early changes) - Sorted by P-value (*continued*)

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
endocytosis (GO:0006897)	549	16	5.31	+	3.01	1.15e-04	3.29e-02
biological_process (GO:0008150)	17890	191	172.97	+	1.10	1.33e-04	3.71e-02
Unclassified (UNCLASSIFIED)	3106	12	30.03	-	0.40	1.33e-04	3.65e-02
cell cycle phase transition (GO:0044770)	285	11	2.76	+	3.99	1.40e-04	3.79e-02
meiosis I (GO:0007127)	113	7	1.09	+	6.41	1.59e-04	4.22e-02
response to external stimulus (GO:0009605)	2478	43	23.96	+	1.79	1.73e-04	4.53e-02
meiosis I cell cycle process (GO:0061982)	115	7	1.11	+	6.30	1.76e-04	4.53e-02
regulation of nuclear division (GO:0051783)	198	9	1.91	+	4.70	1.79e-04	4.53e-02

Table 4: Significant Results (P-value <0.05) from PANTHER Overrepresentation Test using GO Ontology database Released 2020-01-03 (Genes with late changes) - sorted by P-value

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
muscle contraction (GO:0006936)	246	38	6.12	+	6.21	1.10e-17	1.75e-13
muscle system process (GO:0003012)	295	40	7.33	+	5.45	8.55e-17	6.82e-13
system development (GO:0048731)	4451	188	110.66	+	1.7	2.21e-14	1.17e-10
multicellular organism development (GO:0007275)	5074	203	126.15	+	1.61	2.17e-13	8.66e-10
muscle structure development (GO:0061061)	481	45	11.96	+	3.76	2.45e-13	7.81e-10

Table 4: Significant Results (P-value <0.05) from PANTHER Overrepresentation Test using GO Ontology database Released 2020-01-03 (Genes with late changes) - sorted by P-value (*continued*)

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
anatomical structure development (GO:0048856)	5463	213	135.82	+	1.57	4.07e-13	1.08e-09
developmental process (GO:0032502)	5912	223	146.98	+	1.52	2.94e-12	6.68e-09
cell differentiation (GO:0030154)	3745	156	93.11	+	1.68	3.56e-11	7.09e-08
cellular developmental process (GO:0048869)	3838	158	95.42	+	1.66	5.68e-11	1.01e-07
macromolecule metabolic process (GO:0043170)	6206	90	154.29	-	0.58	1.84e-10	2.93e-07
nitrogen compound metabolic process (GO:0006807)	7041	108	175.05	-	0.62	2.22e-10	3.22e-07
animal organ development (GO:0048513)	3219	137	80.03	+	1.71	3.04e-10	4.04e-07
cellular nitrogen compound metabolic process (GO:0034641)	3480	37	86.52	-	0.43	3.08e-10	3.77e-07
myofibril assembly (GO:0030239)	61	15	1.52	+	9.89	3.54e-10	3.76e-07
cellular macromolecule metabolic process (GO:0044260)	5092	68	126.60	-	0.54	3.54e-10	4.03e-07
multicellular organismal process (GO:0032501)	7022	245	174.58	+	1.4	3.60e-10	3.59e-07
muscle organ development (GO:0007517)	294	30	7.31	+	4.1	4.06e-10	3.81e-07
nucleobase-containing compound metabolic process (GO:0006139)	2822	26	70.16	-	0.37	5.00e-10	4.42e-07
nucleic acid metabolic process (GO:0090304)	2309	18	57.41	-	0.31	7.75e-10	6.51e-07

Table 4: Significant Results (P-value <0.05) from PANTHER Overrepresentation Test using GO Ontology database Released 2020-01-03 (Genes with late changes) - sorted by P-value (*continued*)

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
actomyosin structure organization (GO:0031032)	110	18	2.73	+	6.58	2.04e-09	1.63e-06
cellular aromatic compound metabolic process (GO:0006725)	3054	32	75.93	-	0.42	3.03e-09	2.30e-06
heterocycle metabolic process (GO:0046483)	3008	31	74.78	-	0.41	3.34e-09	2.42e-06
muscle cell differentiation (GO:0042692)	247	26	6.14	+	4.23	3.35e-09	2.32e-06
striated muscle cell differentiation (GO:0051146)	204	23	5.07	+	4.53	8.42e-09	5.60e-06
muscle cell development (GO:0055001)	146	19	3.63	+	5.23	2.13e-08	1.36e-05
cellular metabolic process (GO:0044237)	7703	131	191.51	-	0.68	2.84e-08	1.74e-05
striated muscle cell development (GO:0055002)	133	18	3.31	+	5.44	2.91e-08	1.72e-05
primary metabolic process (GO:0044238)	7512	127	186.76	-	0.68	3.15e-08	1.79e-05
muscle tissue development (GO:0060537)	301	27	7.48	+	3.61	3.75e-08	2.06e-05
striated muscle tissue development (GO:0014706)	287	26	7.14	+	3.64	5.61e-08	2.98e-05
cellular component assembly involved in morphogenesis (GO:0010927)	102	15	2.54	+	5.92	1.58e-07	8.13e-05
actin-myosin filament sliding (GO:0033275)	38	10	0.94	+	10.58	1.85e-07	9.22e-05
muscle filament sliding (GO:0030049)	38	10	0.94	+	10.58	1.85e-07	8.94e-05

Table 4: Significant Results (P-value <0.05) from PANTHER Overrepresentation Test using GO Ontology database Released 2020-01-03 (Genes with late changes) - sorted by P-value (*continued*)

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
organic cyclic compound metabolic process (GO:1901360)	3288	41	81.75	-	0.5	2.17e-07	1.02e-04
gene expression (GO:0010467)	2043	19	50.79	-	0.37	2.38e-07	1.08e-04
synaptic signaling (GO:0099536)	472	33	11.73	+	2.81	2.93e-07	1.30e-04
chemical synaptic transmission (GO:0007268)	427	31	10.62	+	2.92	3.16e-07	1.36e-04
anterograde trans-synaptic signaling (GO:0098916)	427	31	10.62	+	2.92	3.16e-07	1.33e-04
RNA metabolic process (GO:0016070)	1702	14	42.31	-	0.33	4.51e-07	1.84e-04
striated muscle contraction (GO:0006941)	112	15	2.78	+	5.39	4.68e-07	1.86e-04
anatomical structure morphogenesis (GO:0009653)	2182	93	54.25	+	1.71	5.15e-07	2.00e-04
system process (GO:0003008)	1975	86	49.10	+	1.75	5.35e-07	2.03e-04
muscle fiber development (GO:0048747)	56	11	1.39	+	7.9	5.80e-07	2.15e-04
trans-synaptic signaling (GO:0099537)	446	31	11.09	+	2.8	7.58e-07	2.74e-04
metabolic process (GO:0008152)	8461	155	210.36	-	0.74	7.61e-07	2.70e-04
intracellular transport (GO:0046907)	1544	12	38.39	-	0.31	7.99e-07	2.77e-04
RNA processing (GO:0006396)	877	3	21.80	-	0.14	9.70e-07	3.29e-04
actin-mediated cell contraction (GO:0070252)	90	13	2.24	+	5.81	1.27e-06	4.22e-04
organic substance metabolic process (GO:0071704)	7891	143	196.19	-	0.73	1.40e-06	4.55e-04

Table 4: Significant Results (P-value <0.05) from PANTHER Overrepresentation Test using GO Ontology database Released 2020-01-03 (Genes with late changes) - sorted by P-value (*continued*)

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
actin filament-based process (GO:0030029)	580	36	14.42	+	2.5	1.44e-06	4.59e-04
cellular localization (GO:0051641)	2437	28	60.59	-	0.46	1.88e-06	5.87e-04
regulation of multicellular organismal process (GO:0051239)	3189	121	79.28	+	1.53	2.03e-06	6.22e-04
nervous system development (GO:0007399)	2384	96	59.27	+	1.62	2.95e-06	8.89e-04
cell development (GO:0048468)	1632	71	40.57	+	1.75	7.57e-06	2.24e-03
cellular protein metabolic process (GO:0044267)	3713	55	92.31	-	0.6	8.64e-06	2.50e-03
protein transport (GO:0015031)	1536	14	38.19	-	0.37	9.96e-06	2.84e-03
synapse organization (GO:0050808)	294	22	7.31	+	3.01	1.06e-05	2.96e-03
actin filament-based movement (GO:0030048)	112	13	2.78	+	4.67	1.12e-05	3.07e-03
peptide transport (GO:0015833)	1567	15	38.96	-	0.39	1.28e-05	3.45e-03
smooth muscle contraction (GO:0006939)	52	9	1.29	+	6.96	1.53e-05	4.08e-03
cellular macromolecule biosynthetic process (GO:0034645)	1726	18	42.91	-	0.42	2.05e-05	5.36e-03
tissue development (GO:0009888)	1786	74	44.40	+	1.67	2.51e-05	6.45e-03
cardiac muscle tissue development (GO:0048738)	161	15	4.00	+	3.75	2.70e-05	6.83e-03
skeletal myofibril assembly (GO:0014866)	11	5	0.27	+	18.28	2.88e-05	7.18e-03

Table 4: Significant Results (P-value <0.05) from PANTHER Overrepresentation Test using GO Ontology database Released 2020-01-03 (Genes with late changes) - sorted by P-value (*continued*)

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
regulation of system process (GO:0044057)	591	33	14.69	+	2.25	2.96e-05	7.27e-03
generation of neurons (GO:0048699)	1564	66	38.88	+	1.7	3.65e-05	8.83e-03
neurogenesis (GO:0022008)	1667	69	41.44	+	1.66	4.30e-05	1.02e-02
muscle tissue morphogenesis (GO:0060415)	76	10	1.89	+	5.29	4.40e-05	1.03e-02
ribonucleoprotein complex biogenesis (GO:0022613)	431	0	10.72	-	< 0.01	4.45e-05	1.03e-02
protein localization (GO:0008104)	2218	28	55.14	-	0.51	4.47e-05	1.02e-02
amide transport (GO:0042886)	1601	17	39.80	-	0.43	4.91e-05	1.10e-02
mRNA metabolic process (GO:0016071)	700	3	17.40	-	0.17	5.03e-05	1.11e-02
long-term synaptic potentiation (GO:0060291)	47	8	1.17	+	6.85	5.08e-05	1.11e-02
cell-cell signaling (GO:0007267)	1127	51	28.02	+	1.82	5.72e-05	1.23e-02
supramolecular fiber organization (GO:0097435)	456	27	11.34	+	2.38	6.30e-05	1.34e-02
macromolecule modification (GO:0043412)	3305	50	82.17	-	0.61	6.71e-05	1.41e-02
tissue morphogenesis (GO:0048729)	566	31	14.07	+	2.2	7.11e-05	1.47e-02
actin cytoskeleton organization (GO:0030036)	501	29	12.46	+	2.33	7.21e-05	1.47e-02
cardiac muscle tissue morphogenesis (GO:0055008)	65	9	1.62	+	5.57	7.40e-05	1.49e-02
learning or memory (GO:0007611)	267	19	6.64	+	2.86	7.85e-05	1.56e-02

Table 4: Significant Results (P-value <0.05) from PANTHER Overrepresentation Test using GO Ontology database Released 2020-01-03 (Genes with late changes) - sorted by P-value (*continued*)

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
positive regulation of synapse assembly (GO:0051965)	66	9	1.64	+	5.48	8.23e-05	1.62e-02
muscle organ morphogenesis (GO:0048644)	83	10	2.06	+	4.85	8.63e-05	1.68e-02
neurotransmitter transport (GO:0006836)	180	15	4.48	+	3.35	8.78e-05	1.69e-02
DNA metabolic process (GO:0006259)	757	4	18.82	-	0.21	9.37e-05	1.78e-02
cellular nitrogen compound biosynthetic process (GO:0044271)	1675	19	41.64	-	0.46	9.85e-05	1.85e-02
intracellular protein transport (GO:0006886)	1011	8	25.14	-	0.32	1.01e-04	1.86e-02
protein metabolic process (GO:0019538)	4303	72	106.98	-	0.67	1.11e-04	2.03e-02
cellular component morphogenesis (GO:0032989)	833	40	20.71	+	1.93	1.15e-04	2.08e-02
cellular protein localization (GO:0034613)	1652	19	41.07	-	0.46	1.28e-04	2.29e-02
sarcomere organization (GO:0045214)	40	7	0.99	+	7.04	1.29e-04	2.28e-02
cellular macromolecule localization (GO:0070727)	1660	19	41.27	-	0.46	1.29e-04	2.26e-02
regulation of muscle system process (GO:0090257)	233	17	5.79	+	2.93	1.39e-04	2.42e-02
chromosome organization (GO:0051276)	1060	9	26.35	-	0.34	1.41e-04	2.42e-02
regulation of synapse assembly (GO:0051963)	107	11	2.66	+	4.13	1.42e-04	2.39e-02

Table 4: Significant Results (P-value <0.05) from PANTHER Overrepresentation Test using GO Ontology database Released 2020-01-03 (Genes with late changes) - sorted by P-value (*continued*)

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
nucleobase-containing compound catabolic process (GO:0034655)	378	0	9.40	-	< 0.01	1.42e-04	2.41e-02
mitotic cell cycle process (GO:1903047)	655	3	16.28	-	0.18	1.46e-04	2.42e-02
ncRNA processing (GO:0034470)	386	0	9.60	-	< 0.01	1.48e-04	2.42e-02
organonitrogen compound metabolic process (GO:1901564)	5328	95	132.46	-	0.72	1.53e-04	2.49e-02
cognition (GO:0050890)	307	20	7.63	+	2.62	1.58e-04	2.55e-02
circulatory system development (GO:0072359)	875	41	21.75	+	1.88	1.63e-04	2.60e-02
regulation of developmental process (GO:0050793)	2632	96	65.44	+	1.47	1.79e-04	2.83e-02
regulation of cellular metabolic process (GO:0031323)	6315	118	157.00	-	0.75	1.88e-04	2.94e-02
behavior (GO:0007610)	599	31	14.89	+	2.08	1.94e-04	3.01e-02
neuron differentiation (GO:0030182)	1019	46	25.33	+	1.82	1.99e-04	3.04e-02
skeletal muscle thin filament assembly (GO:0030240)	9	4	0.22	+	17.88	2.06e-04	3.12e-02
ncRNA metabolic process (GO:0034660)	475	1	11.81	-	0.08	2.09e-04	3.15e-02
macromolecule biosynthetic process (GO:0009059)	1785	22	44.38	-	0.5	2.18e-04	3.25e-02
establishment of protein localization (GO:0045184)	1628	19	40.48	-	0.47	2.21e-04	3.26e-02

Table 4: Significant Results (P-value <0.05) from PANTHER Overrepresentation Test using GO Ontology database Released 2020-01-03 (Genes with late changes) - sorted by P-value (*continued*)

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
animal organ morphogenesis (GO:0009887)	975	44	24.24	+	1.82	2.24e-04	3.28e-02
skeletal muscle contraction (GO:0003009)	31	6	0.77	+	7.78	2.43e-04	3.52e-02
negative regulation of smooth muscle cell proliferation (GO:0048662)	45	7	1.12	+	6.26	2.47e-04	3.55e-02
musculoskeletal movement (GO:0050881)	45	7	1.12	+	6.26	2.47e-04	3.51e-02
multicellular organismal movement (GO:0050879)	45	7	1.12	+	6.26	2.47e-04	3.48e-02
protein localization to organelle (GO:0033365)	768	5	19.09	-	0.26	2.91e-04	4.06e-02
negative regulation of cell population proliferation (GO:0008285)	691	34	17.18	+	1.98	2.93e-04	4.07e-02
anatomical structure formation involved in morphogenesis (GO:0048646)	898	41	22.33	+	1.84	2.96e-04	4.06e-02
regulation of presynapse organization (GO:0099174)	33	6	0.82	+	7.31	3.27e-04	4.46e-02
regulation of synapse organization (GO:0050807)	229	16	5.69	+	2.81	3.42e-04	4.62e-02
biological_process (GO:0008150)	17890	473	444.78	+	1.06	3.43e-04	4.60e-02
Unclassified (UNCLASSIFIED)	3106	49	77.22	-	0.63	3.43e-04	4.56e-02
cardiac myofibril assembly (GO:0055003)	21	5	0.52	+	9.58	3.55e-04	4.68e-02

Table 4: Significant Results (P-value <0.05) from PANTHER Overrepresentation Test using GO Ontology database Released 2020-01-03 (Genes with late changes) - sorted by P-value (*continued*)

Gene Function Set	Total Number Genes Set	Number of Overlap	Expected Number to overlap	Fewer or Greater Than Expected	Fold Enrichment	Raw P-value	FDR
regulation of nitrogen compound metabolic process (GO:0051171)	5922	111	147.23	-	0.75	3.77e-04	4.93e-02

Table 5: Computing overlap with Molecular Signatures Database (MSigDB) Early Changes

Functional Gene set	Total Genes In Gene Set	Genes In Overlap	Expected Number to Overlap	P-value	FDR	q value
GSE15750 DAY6 VS DAY10	200	25	0.1250	3.84e-23	3.28e-19	
EFF CD8 TCELL UP						
GSE39110 DAY3 VS DAY6	200	25	0.1250	3.84e-23	3.28e-19	
POST IMMUNIZATION						
CD8 TCELL DN						
MODULE 54	262	27	0.1031	1.22e-22	6.94e-19	
GSE15750 DAY6 VS DAY10	200	23	0.1150	1.53e-20	6.52e-17	
TRAF6KO EFF CD8 TCELL						
UP						
GNF2 CDC2	62	15	0.2419	6.46e-19	2.21e-15	
GNF2 CCNA2	68	15	0.2206	2.96e-18	8.43e-15	
GNF2 CDC20	56	14	0.2500	5.83e-18	1.42e-14	
GNF2 CCNB2	57	14	0.2456	7.68e-18	1.64e-14	
GSE13547 CTRL VS ANTI	193	19	0.0984	6.55e-16	1.24e-12	
IGM STIM BCELL 12H UP						
GNF2 CENPF	62	13	0.2097	1.11e-15	1.90e-12	
HALLMARK G2M	200	19	0.0950	1.27e-15	1.98e-12	
CHECKPOINT						
GSE30962 PRIMARY VS	200	18	0.0900	1.87e-14	2.66e-11	
SECONDARY ACUTE						
LCMV INF CD8 TCELL UP						
GO ORGANELLE FISSION	463	25	0.0540	2.15e-14	2.82e-11	
GSE13547 2H VS 12 H ANTI	184	17	0.0924	6.54e-14	7.99e-11	
IGM STIM BCELL UP						
GSE14415 NATURAL	190	17	0.0895	1.11e-13	1.27e-10	
TREG VS TCONV DN						
GNF2 PCNA	68	12	0.1765	1.29e-13	1.38e-10	
GSE27241 WT VS RORGT	175	16	0.0914	4.26e-13	4.28e-10	
KO TH17 POLARIZED CD4						
TCELL UP						
GSE14415 INDUCED VS	188	16	0.0851	1.29e-12	1.23e-09	
NATURAL TREG DN						

Table 5: Computing overlap with Molecular Signatures Database (MSigDB) Early Changes (*continued*)

Functional Gene set	Total Genes In Gene Set	Genes In Overlap	Expected Number to Overlap	P-value	FDR q value
GO SISTER CHROMATID SEGREGATION	189	16	0.0847	1.40e-12	1.26e-09
GO NUCLEAR CHROMOSOME SEGREGATION	262	18	0.0687	1.93e-12	1.62e-09
GNF2 HMMR	47	10	0.2128	1.99e-12	1.62e-09
GO MITOTIC NUCLEAR DIVISION	282	18	0.0638	6.62e-12	5.14e-09
GO MITOTIC CELL CYCLE	1009	32	0.0317	9.11e-12	6.77e-09
GO MITOTIC SISTER CHROMATID SEGREGATION	151	14	0.0927	1.05e-11	7.48e-09
GO ANTIGEN BINDING	192	15	0.0781	2.25e-11	1.48e-08
GSE45365 HEALTHY VS MCMV INFECTION CD11B DC DN	192	15	0.0781	2.25e-11	1.48e-08
GSE29614 CTRL VS DAY7 TIV FLU VACCINE PBMC DN	193	15	0.0777	2.43e-11	1.48e-08
GSE40274 CTRL VS EOS TRANSDUCED ACTIVATED CD4 TCELL UP	193	15	0.0777	2.43e-11	1.48e-08
GNF2 MKI67	28	8	0.2857	2.52e-11	1.48e-08
GO CELL CYCLE PROCESS	1383	37	0.0268	2.69e-11	1.53e-08
GO CHROMOSOME SEGREGATION	316	18	0.0570	4.35e-11	2.40e-08
GO CELL CYCLE	1847	43	0.0233	4.85e-11	2.59e-08
GO MEIOTIC CELL CYCLE	249	16	0.0643	9.07e-11	4.70e-08
GNF2 BUB1B	49	9	0.1837	1.07e-10	5.36e-08
GNF2 ESPL1	35	8	0.2286	1.82e-10	8.90e-08
GSE45365 WT VS IFNAR KO BCELL DN	189	14	0.0741	2.14e-10	1.02e-07
GSE13547 CTRL VS ANTI IGM STIM BCELL 2H UP	190	14	0.0737	2.30e-10	1.06e-07
GSE21063 WT VS NFATC1 KO 8H ANTI IGM STIM BCELL UP	199	14	0.0704	4.23e-10	1.86e-07
GSE36476 CTRL VS TSST ACT 72H MEMORY CD4 TCELL YOUNG DN	199	14	0.0704	4.23e-10	1.86e-07
GSE24634 TEFF VS TCONV DAY7 IN CULTURE UP	200	14	0.0700	4.52e-10	1.93e-07
GNF2 RRM1	88	10	0.1136	1.32e-09	5.52e-07
GO IMMUNOGLOBULIN COMPLEX	69	9	0.1304	2.58e-09	1.02e-06
GO IMMUNOGLOBULIN RECEPTOR BINDING	69	9	0.1304	2.58e-09	1.02e-06

Table 5: Computing overlap with Molecular Signatures Database (MSigDB) Early Changes (*continued*)

Functional Gene set	Total Genes In Gene Set	Genes In Overlap	Expected Number to Overlap	P-value	FDR q value
GSE45365 WT VS IFNAR KO BCELL MCMV INFECTION DN	193	13	0.0674	3.05e-09	1.19e-06
GSE13411 PLASMA CELL VS MEMORY BCELL UP	195	13	0.0667	3.46e-09	1.31e-06
GSE24634 IL4 VS CTRL TREATED NAIVE CD4 TCELL DAY7 UP	198	13	0.0657	4.16e-09	1.48e-06
GSE25088 WT VS STAT6 KO MACROPHAGE DN	198	13	0.0657	4.16e-09	1.48e-06
GSE25088 WT VS STAT6 KO MACROPHAGE IL4 STIM DN	198	13	0.0657	4.16e-09	1.48e-06
GSE2405 S AUREUS VS UNTREATED NEUTROPHIL DN	199	13	0.0653	4.42e-09	1.49e-06
GNF2 SMC2L1	33	7	0.2121	4.55e-09	1.49e-06

Table 6: Computing overlap with Molecular Signatures Database (MSigDB) Late Changes

Functional Gene set	Total Genes In Gene Set	Genes In Overlap	Expected Number to Overlap	P-value	FDR q value
GO MUSCLE CONTRACTION	360	43	0.1194	1.68e-23	2.87e-19
GO MUSCLE SYSTEM PROCESS	473	48	0.1015	5.37e-23	4.59e-19
CAGCTG AP4 Q5	1537	82	0.0534	7.73e-20	4.40e-16
GO MUSCLE STRUCTURE DEVELOPMENT	659	48	0.0728	4.77e-17	2.03e-13
GO CONTRACTILE FIBER	229	29	0.1266	5.94e-17	2.03e-13
HALLMARK MYOGENESIS	200	26	0.1300	1.28e-15	3.66e-12
SRF C	215	26	0.1209	7.50e-15	1.83e-11
GO MUSCLE ORGAN DEVELOPMENT	410	35	0.0854	8.65e-15	1.85e-11
SRF Q6	247	27	0.1093	2.80e-14	5.32e-11
GO MYOFIBRIL ASSEMBLY	70	16	0.2286	4.62e-14	7.90e-11
GO SUPRAMOLECULAR COMPLEX	946	51	0.0539	7.48e-13	1.08e-09
TATAAA TATA 01	1313	62	0.0472	7.59e-13	1.08e-09
MODULE 220	333	29	0.0871	1.07e-12	1.40e-09
GO MUSCLE TISSUE DEVELOPMENT	408	32	0.0784	1.23e-12	1.50e-09
GO CELL SURFACE	868	48	0.0553	1.48e-12	1.69e-09
SRF Q4	229	24	0.1048	1.90e-12	2.04e-09

Table 6: Computing overlap with Molecular Signatures Database (MSigDB) Late Changes (*continued*)

Functional Gene set	Total Genes In Gene Set	Genes In Overlap	Expected Number to Overlap	P-value	FDR q value
GO MUSCLE CELL DIFFERENTIATION	378	30	0.0794	4.64e-12	4.66e-09
GO ACTIN FILAMENT BASED PROCESS	744	43	0.0578	5.43e-12	5.16e-09
GO INTRINSIC COMPONENT OF PLASMA MEMBRANE	1708	71	0.0416	6.04e-12	5.43e-09
GO NEUROGENESIS	1599	68	0.0425	6.45e-12	5.51e-09
SRF Q5 01	224	23	0.1027	8.42e-12	6.85e-09
MODULE 176	230	23	0.1000	1.45e-11	1.13e-08
GO NEURON DIFFERENTIATION	1348	60	0.0445	2.03e-11	1.51e-08
MODULE 112	260	24	0.0923	2.86e-11	2.03e-08
GO STRIATED MUSCLE CELL DIFFERENTIATION	286	25	0.0874	3.61e-11	2.47e-08
MODULE 1	367	28	0.0763	6.03e-11	3.97e-08
GO MUSCLE CELL DEVELOPMENT	184	20	0.1087	6.88e-11	4.18e-08
GO CELLULAR COMPONENT ASSEMBLY INVOLVED IN MORPHOGENESIS	110	16	0.1455	6.92e-11	4.18e-08
GO ACTIN BINDING	422	30	0.0711	7.10e-11	4.18e-08
MODULE 117	722	40	0.0554	1.08e-10	6.14e-08
GO CYTOSKELETAL PROTEIN BINDING	950	47	0.0495	1.12e-10	6.18e-08
GO STRUCTURAL CONSTITUENT OF MUSCLE	44	11	0.2500	1.53e-10	8.16e-08
GO ACTOMYOSIN STRUCTURE ORGANIZATION	193	20	0.1036	1.63e-10	8.46e-08
GO ANIMAL ORGAN MORPHOGENESIS	1034	49	0.0474	1.93e-10	9.49e-08
MODULE 88	833	43	0.0516	1.94e-10	9.49e-08
TGANTCA AP1 C	1146	52	0.0454	2.44e-10	1.16e-07
GO TISSUE MORPHOGENESIS	661	37	0.0560	4.12e-10	1.90e-07
GO ANATOMICAL STRUCTURE FORMATION INVOLVED IN MORPHOGENESIS	1172	52	0.0444	5.32e-10	2.39e-07
GO ACTIN CYTOSKELETON	491	31	0.0631	6.38e-10	2.80e-07
MODULE 2	383	27	0.0705	7.64e-10	3.26e-07
GO MUSCLE FILAMENT SLIDING	39	10	0.2564	8.03e-10	3.35e-07
GO CELL CELL SIGNALING	1644	64	0.0389	1.01e-09	4.12e-07

Table 6: Computing overlap with Molecular Signatures Database  
 (MSigDB) Late Changes (*continued*)

Functional Gene set	Total Genes In Gene Set	Genes In Overlap	Expected Number to Overlap	P-value	FDR q value
GO BEHAVIOR	594	34	0.0572	1.21e-09	4.80e-07
GO STRUCTURAL MOLECULE ACTIVITY	826	41	0.0496	1.63e-09	6.32e-07
GO MUSCLE FIBER DEVELOPMENT	68	12	0.1765	1.70e-09	6.45e-07
MODULE 55	830	41	0.0494	1.87e-09	6.95e-07
GO CELLULAR COMPONENT	1115	49	0.0439	2.33e-09	8.37e-07
MORPHOGENESIS					
GO CYTOSKELETON ORGANIZATION	1298	54	0.0416	2.35e-09	8.37e-07
MODULE 47	225	20	0.0889	2.46e-09	8.57e-07
TGGNNNNNNKCCAR	434	28	0.0645	2.68e-09	9.17e-07
UNKNOWN					