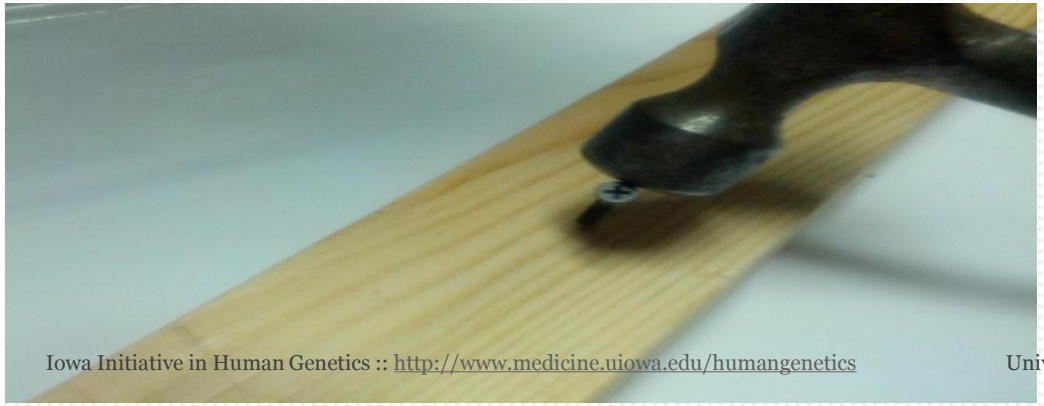
Other applications of MPS

Appropriate uses for MPS

Tom Bair
thomas-bair@uiowa.edu



Iowa Initiative in Human Genetics :: http://www.medicine.uiowa.edu/humangenetics

University of Iowa Bioinformatics :: http://bioinform-div.healthcare.uiowa.edu/

Types of workflows

- reduce
  - assemble
  - align
- identify significant areas:
  - peak-finding/depth
  - snp calling
- Annotate areas of interest
  - Coordinates -> information
- display
  - Understand the information

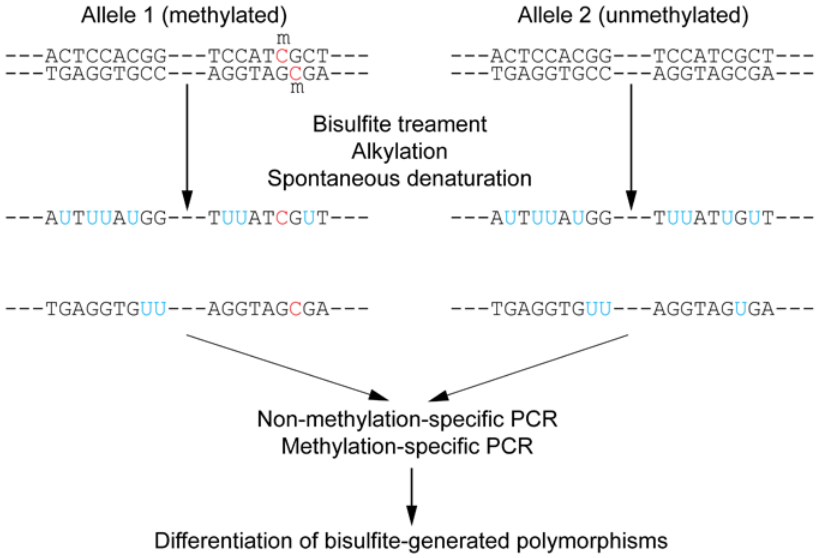
Data reduction

- Alignment
  - Bowtie, bwa …
- Assemble
  - Velvet, Mira …
- Some small genome assembly
  - High memory requirements
  - Probably will shift as longer reads are economical
  - Source data critical
    - Long paired end or just long reads make everything easy.

Identification of significant areas

- SNP/deletion discovery
  - --the focus of this course
- Chip-seq
  - Find DNA protein interactions
  - Depth of alignment vs background
- RNA-seq
  - Find differentially expressed transcripts
  - Depth of one sample vs another
  - Linkage of areas
    - Gene definition

- Methylation patterns
  - Somewhat unique alignment protocol
  - Look at the alignment of a couple different genome versions to see which matches the best, this gives the methylation status of the area
  - Bismark does this but downstream outputs limited



- Metagenomics
  - Align to a specific database
    - Greengenes or similar of rRNA sequences
  - Align to database of all sequences
  - Find distribution of genus/species

Display approaches

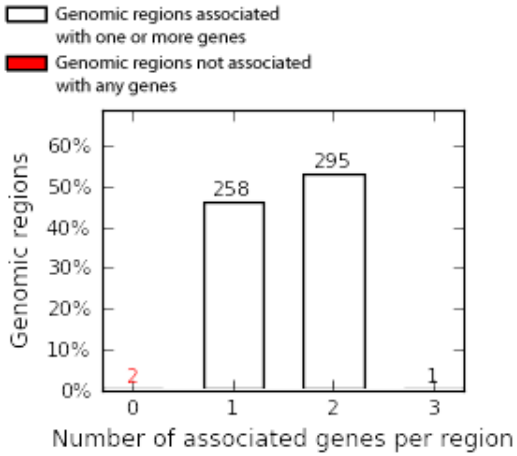
- Level of detail
- IGV genome browser
  - High level of detail, can potentially see every read, local so annotation may lag
- UCSC genome browser
  - Difficult to get all your read information uploaded
  - Very current annotation
- Trackster
  - Galaxy specific, developing

Chip Seq workflow

- Sample prep and antibody critical
- Align with bowtie2
- Peak calling with MACs
  - Mosaics – deeper sequenced samples
  - Dbchip – merge experiments
- Annotation
  - Cistrome
  - GREAT
  - Typically need adjacent genes and distance to TSS

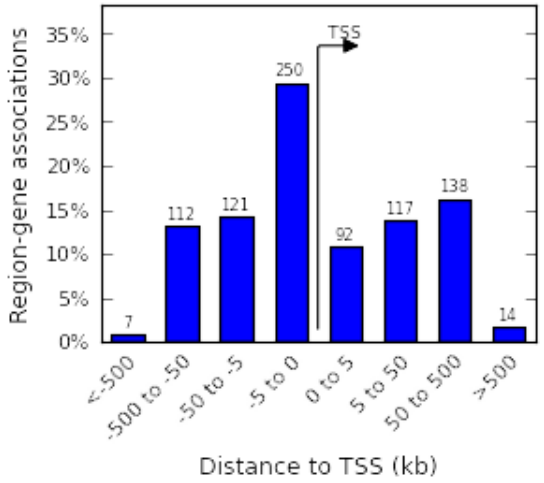
Number of associated genes per region

Download as PDF.



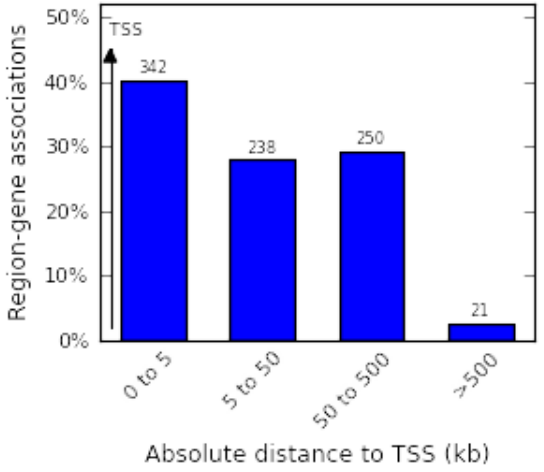
Binned by orientation and distance to TSS

Download as PDF.



Binned by absolute distance to TSS

Download as PDF.



GO Cellular Component (5 terms)

Global controls

Table controls: Export | Shown top rows in this table: 20 Set | Term annotation count: Min: 1 Max: Inf Set | Visualize this table: [select one]

Term Name	Binom Rank	Binom Raw P-Value	*Binom FDR Q-Val*	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	*Hyper FDR Q-Val*	Hyper Fold Enrichment	Hyper Observed Gene Hits	Hyper Total Genes	Hyper Gene Set Coverage
actin cytoskeleton	1	9.3775e-11	1.0512e-7	2.8281	50	8.99%	2	1.1912e-5	2.6396	40	337	5.04%
stress fiber	2	1.5680e-7	8.7888e-5	6.6040	13	2.34%	22	1.7569e-2	4.4477	8	40	1.01%
actin filament bundle	5	2.5918e-7	5.8107e-5	6.3139	13	2.34%	23	2.8131e-2	4.1374	8	43	1.01%
actomyosin	13	3.1716e-6	2.7349e-4	5.0198	13	2.34%	21	1.8282e-2	4.0029	9	50	1.13%
ruffle	33	1.4147e-4	4.8056e-3	2.7582	18	3.24%	16	9.2660e-3	3.0052	15	111	1.89%

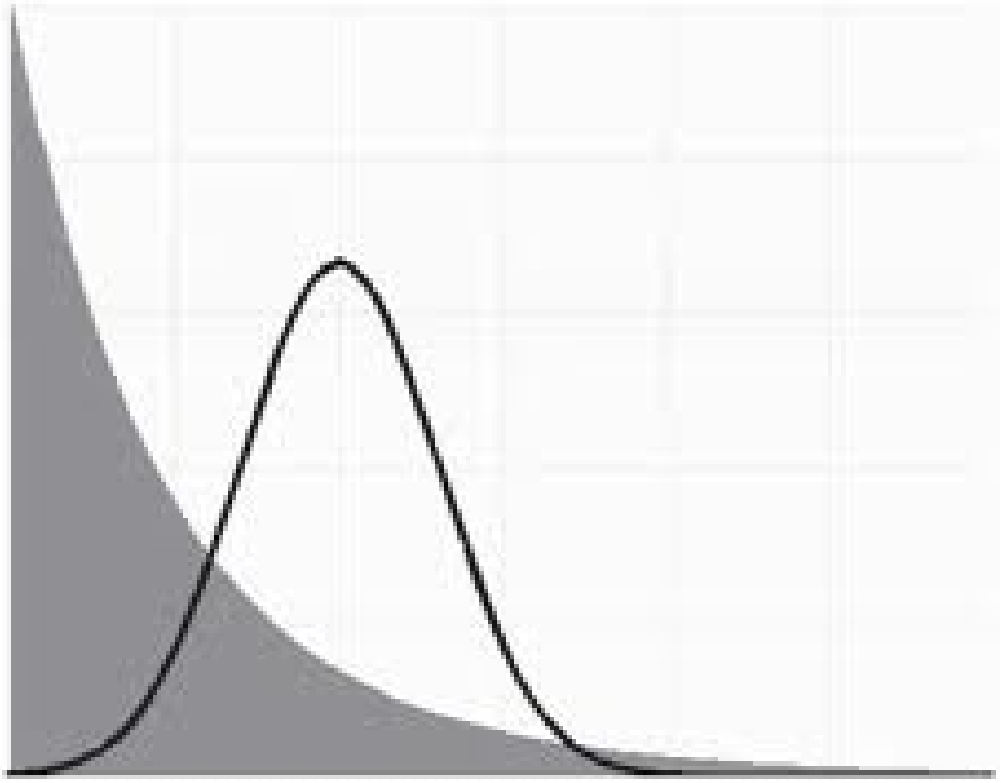
Exon re-sequencing

- BWA alignment
- GATK calibration and SNP identification
- Annotation SIFT, SVA, ANNOVAR
- IGV for visualization
- Trending to genomic rather than exon capture
  - Still may filter down to subset
  - Less bias for CNV
  - Cost, capture designed to reduce sequencing but sequencing droping in cost
  - Still have dark spots

RNA-Seq Workflow

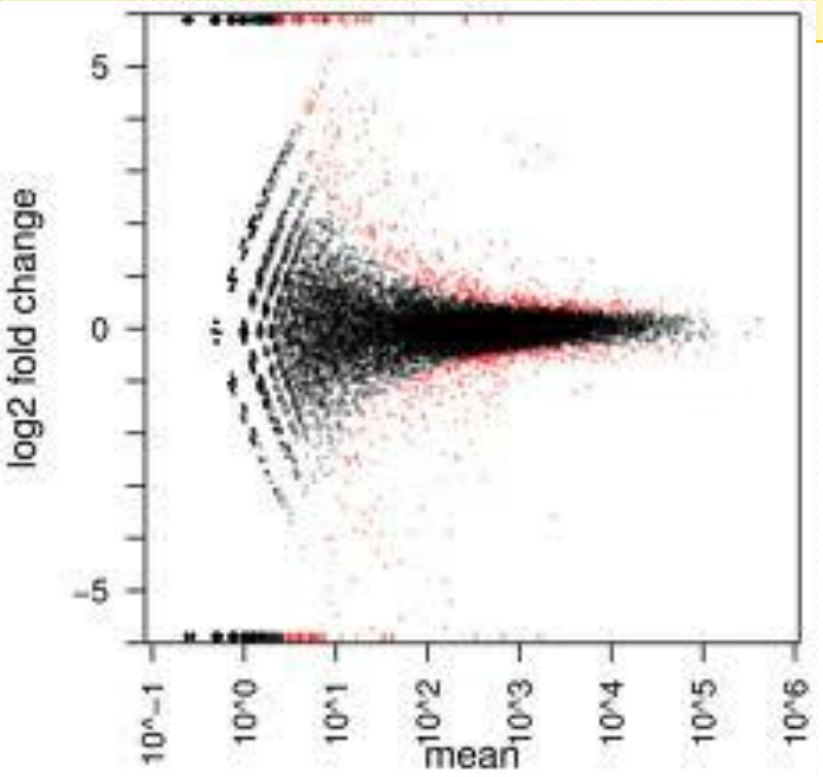
- Tophat alignment
- Htseq-cout or cufflinks for counts
- EdgeR, DESeq, or cuffdiff for differential expression
  - ANOVA, t-test do not work well

Distributions



RNA-Seq length bias

- It is easier (higher statistical power) to find changes in genes that have high expression
- You may find that pathways predicted by microarray tools are simply composed of high numbers of long genes
- This is a caution that applies to many MPS techniques that have a similar array based technique.



RNA-seq w/o Reference

- Trinity
  - 16G of RAM minimum start
  - More is better
  - Active community and optimization
  - Easy paths to differential expression
  - Easy paths to ORF characterization

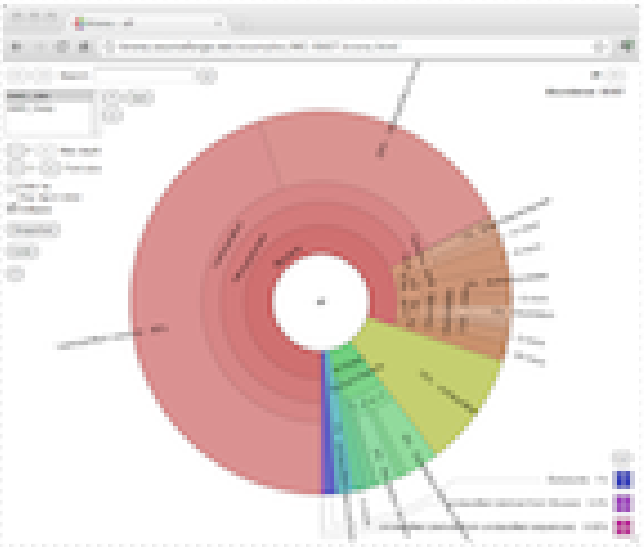
Methylation

- Specific alignment requirements
  - Use bowtie with specialized references
- Differential modification a similar problem to RNA-Seq
- Annotation similar to CHIP-Seq'
- Bismark a popular option for Display as well as analysis

Metagenomics

- Longer reads make everything easier/more accurate
- First data reduction typically a clustering to merge identical reads
  - The clusters are then assigned an OTU (Operational taxonomic units)

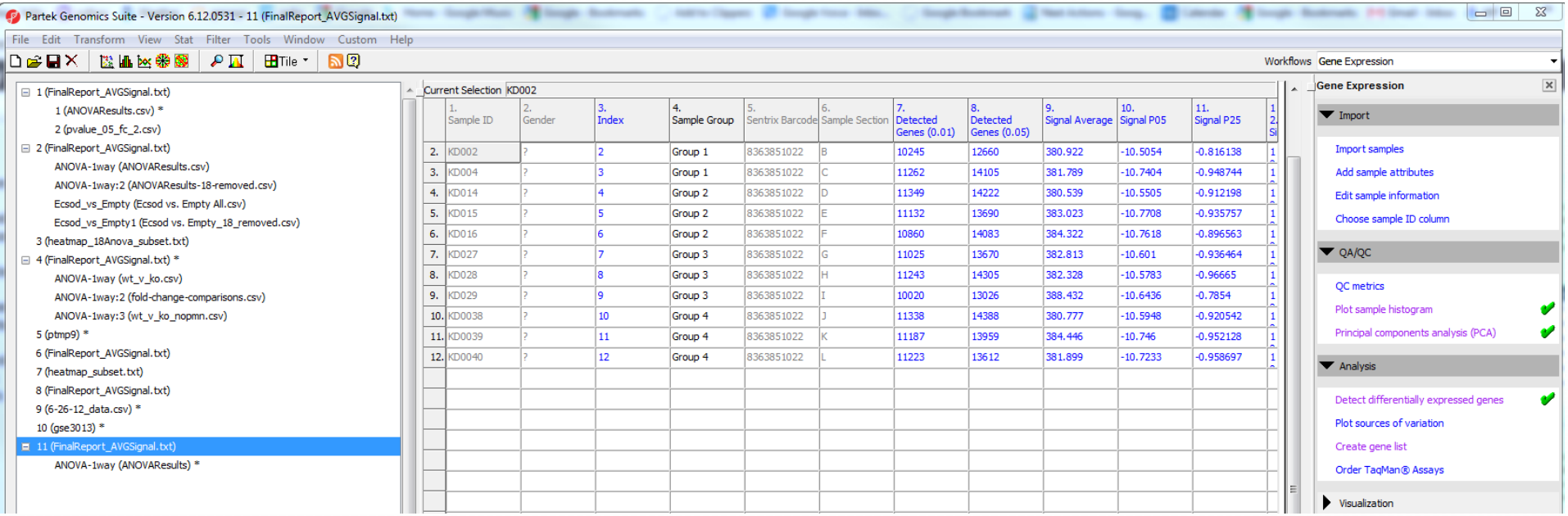
- QIIME - nice tool to assign and display output
- Krona – nice display tool



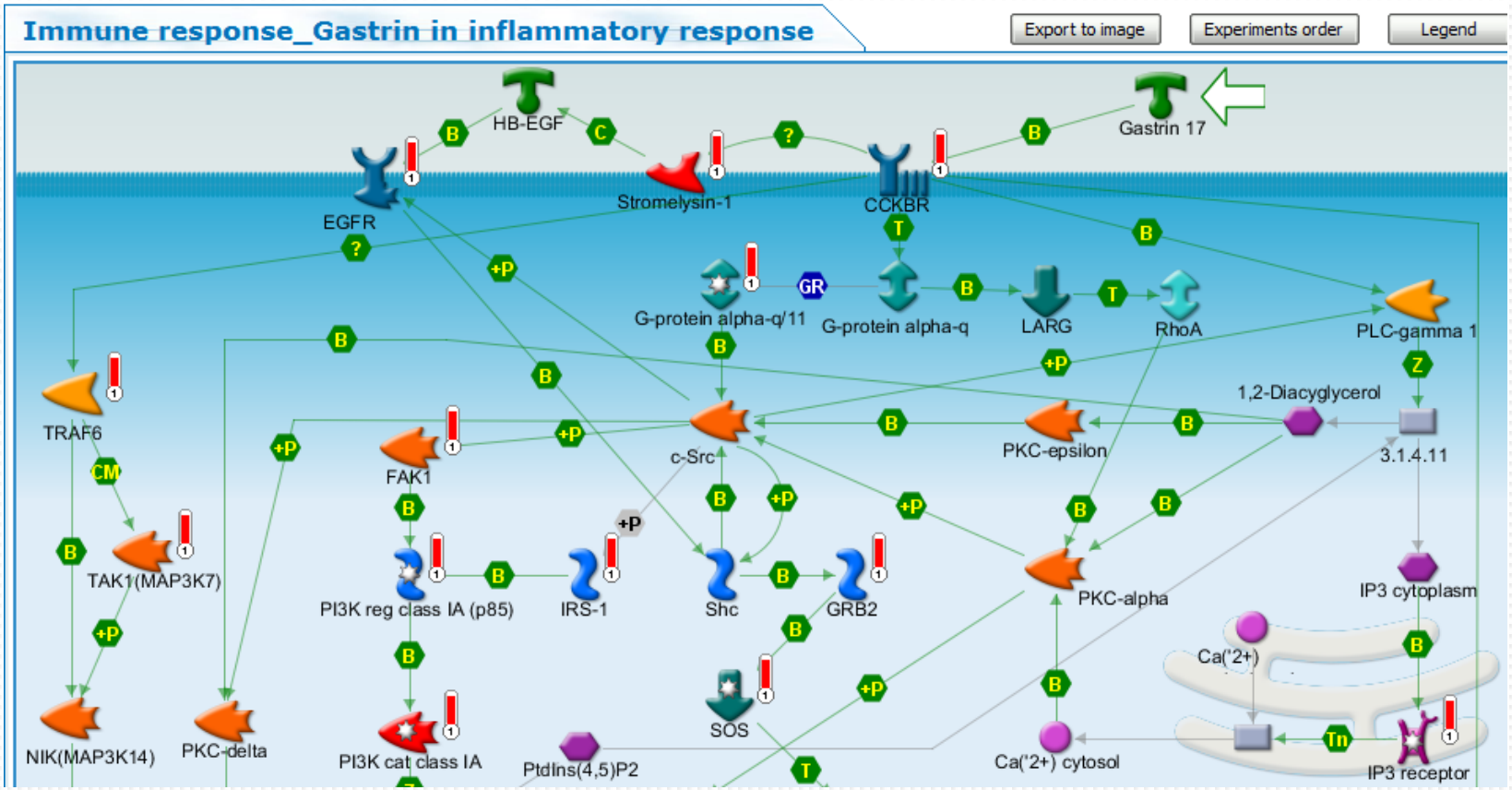
Local resources

- Galaxy
  - Harness local compute cluster helium with web page ease and interface
- Partek
  - Statistical analysis package
- Genego
  - Pathway annotation and mining tool
- Laser gene- NGS module may be soon
  - http://cs.its.uiowa.edu/software/lasergene.shtml
- IIHG bioinformatics core
  - http://bioinform-div.healthcare.uiowa.edu/

Partek


Partek Genomics Suite - Version 6.12.0531 - 11 (FinalReport_AVGSignal.txt)
File Edit Transform View Stat Filter Tools Window Custom Help
Tile
Workflows Gene Expression
1 (FinalReport_AVGSignal.txt)
1 (ANOVAResults.csv) *
2 (pvalue_05_fc_2.csv)
2 (FinalReport_AVGSignal.txt)
ANOVA-1way (ANOVAResults.csv)
ANOVA-1way:2 (ANOVAResults-18-removed.csv)
Ecsod_vs_Empty (Ecsod vs. Empty All.csv)
Ecsod_vs_Empty1 (Ecsod vs. Empty_18_removed.csv)
3 (heatmap_18Anova_subset.txt)
4 (FinalReport_AVGSignal.txt) *
ANOVA-1way (wt_v_ko.csv)
ANOVA-1way:2 (fold-change-comparisons.csv)
ANOVA-1way:3 (wt_v_ko_nopmn.csv)
5 (ptmp9) *
6 (FinalReport_AVGSignal.txt)
7 (heatmap_subset.txt)
8 (FinalReport_AVGSignal.txt)
9 (6-26-12_data.csv) *
10 (gse3013) *
11 (FinalReport_AVGSignal.txt)
ANOVA-1way (ANOVAResults) *
Current Selection KD002
1. Sample ID
2. Gender
3. Index
4. Sample Group
5. Sentrix Barcode
6. Sample Section
7. Detected Genes (0.01)
8. Detected Genes (0.05)
9. Signal Average
10. Signal P05
11. Signal P25
2. KD002 ? 2 Group 1 8363851022 B 10245 12660 380.922 -10.5054 -0.816138
3. KD004 ? 3 Group 1 8363851022 C 11262 14105 381.789 -10.7404 -0.948744
4. KD014 ? 4 Group 2 8363851022 D 11349 14222 380.539 -10.5505 -0.912198
5. KD015 ? 5 Group 2 8363851022 E 11132 13690 383.023 -10.7708 -0.935757
6. KD016 ? 6 Group 2 8363851022 F 10860 14083 384.322 -10.7618 -0.896563
7. KD027 ? 7 Group 3 8363851022 G 11025 13670 382.813 -10.601 -0.936464
8. KD028 ? 8 Group 3 8363851022 H 11243 14305 382.328 -10.5783 -0.96665
9. KD029 ? 9 Group 3 8363851022 I 10020 13026 388.432 -10.6436 -0.7854
10. KD0038 ? 10 Group 4 8363851022 J 11338 14388 380.777 -10.5948 -0.920542
11. KD0039 ? 11 Group 4 8363851022 K 11187 13959 384.446 -10.746 -0.952128
12. KD0040 ? 12 Group 4 8363851022 L 11223 13612 381.899 -10.7233 -0.958697
Gene Expression
Import
Import samples
Add sample attributes
Edit sample information
Choose sample ID column
QA/QC
QC metrics
Plot sample histogram
Principal components analysis (PCA)
Analysis
Detect differentially expressed genes
Plot sources of variation
Create gene list
Order TaqMan® Assays
Visualization

GeneGo


Immune response_Gastrin in inflammatory response
Export to image
Experiments order
Legend
HB-EGF
Gastrin 17
EGFR
Stromelysin-1
CCKBR
G-protein alpha-q/11
G-protein alpha-q
LARG
RhoA
PLC-gamma 1
1,2-Diacyglycerol
3.1.4.11
TRAF6
FAK1
c-Src
PKC-epsilon
TAK1(MAP3K7)
PI3K reg class IA (p85)
IRS-1
Shc
GRB2
PKC-alpha
IP3 cytoplasm
Ca('2+)
NIK(MAP3K14)
PKC-delta
PI3K cat class IA
SOS
PtdIns(4,5)P2
Ca('2+) cytosol
IP3 receptor

MPS – other applications

- Many new uses for this technology
- Continue to be a driver as costs go down
- Need to be careful to test and understand as the techniques may be similar but the analysis may have different assumptions/conditions

Links

- http://wiki.g2.bx.psu.edu/Events/GCC2012/TrainingDay/VMs -- galaxy on your laptop
- http://trinityrnaseq.sourceforge.net/ -- trinity
- http://www.stat.wisc.edu/~keles/Software/mosaics/ -- chip seq w/ mosaics
- http://www.bioinformatics.babraham.ac.uk/projects/bismark/ -- methylation
- http://sourceforge.net/p/krona/home/krona/ -- krona metagenomics display

Discussion

- Proposed projects??
  - Feedback, discussion, experiences