Effects of Mapping Algorithms on Gene Selection for RNA-Seq Analysis: Pulmonary Response to Acute Neonatal Hyperoxia.

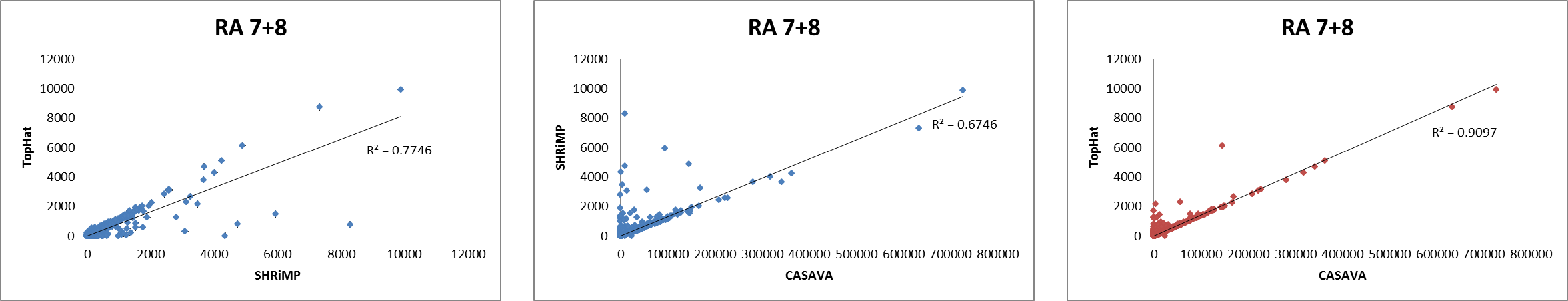
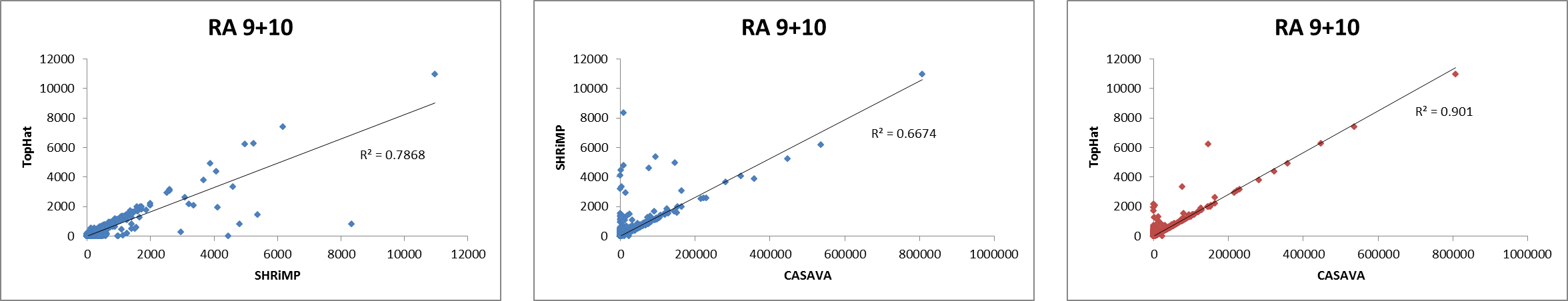
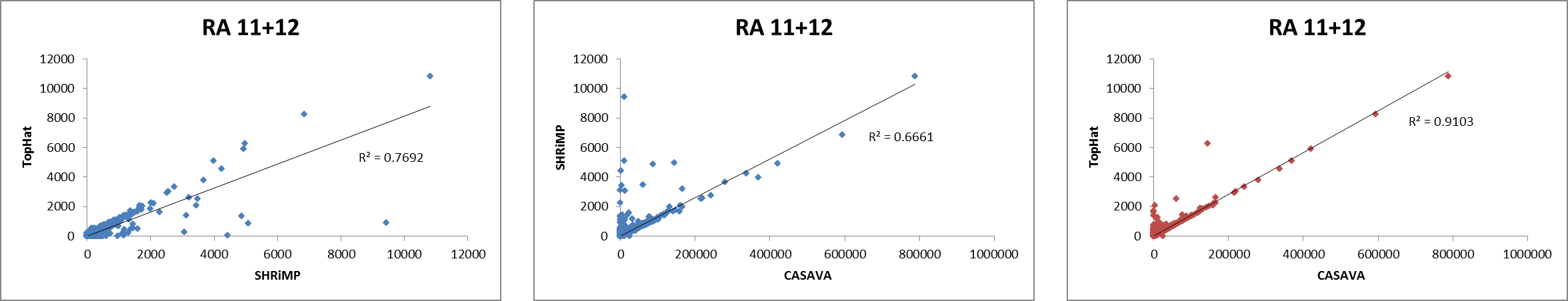
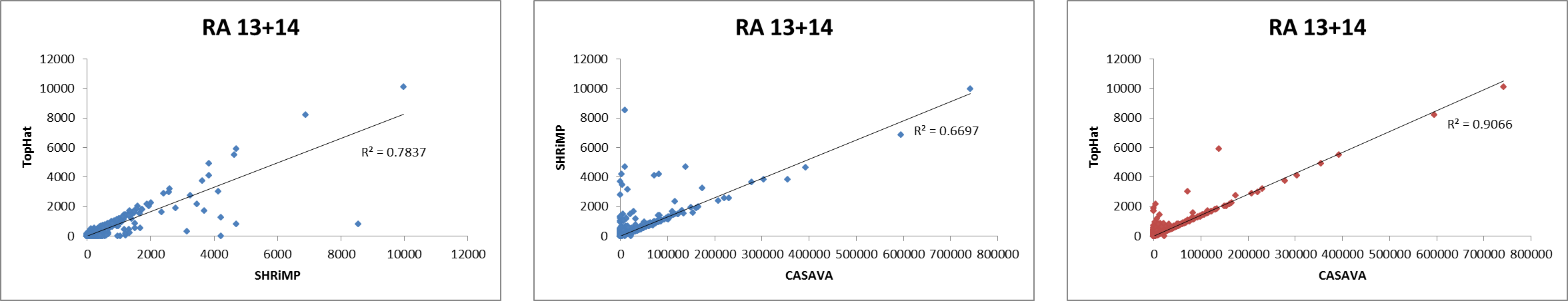
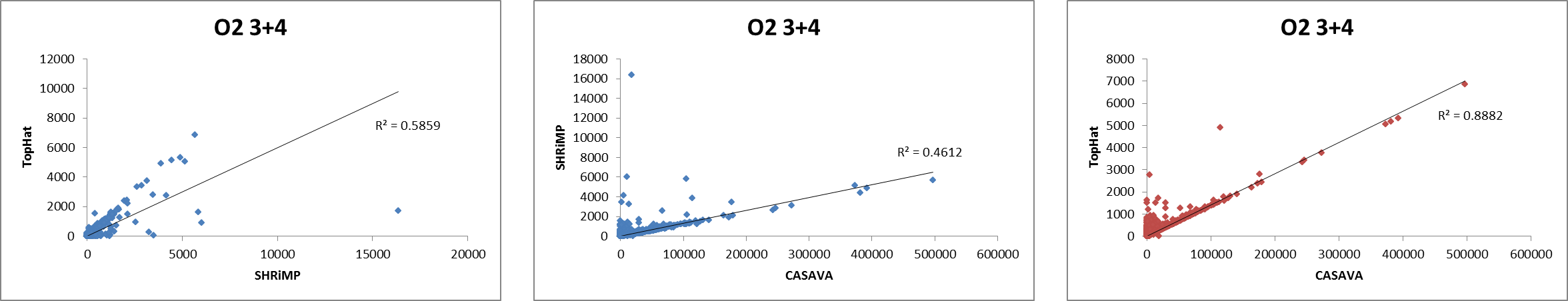
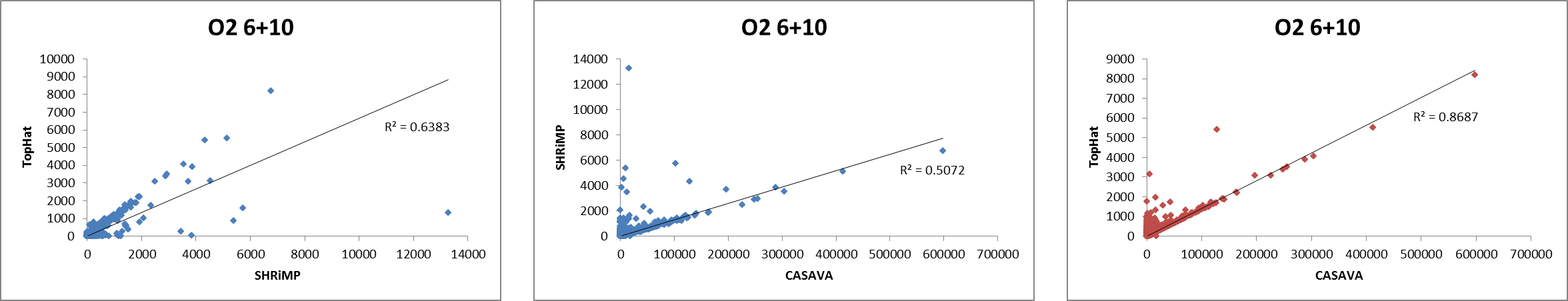
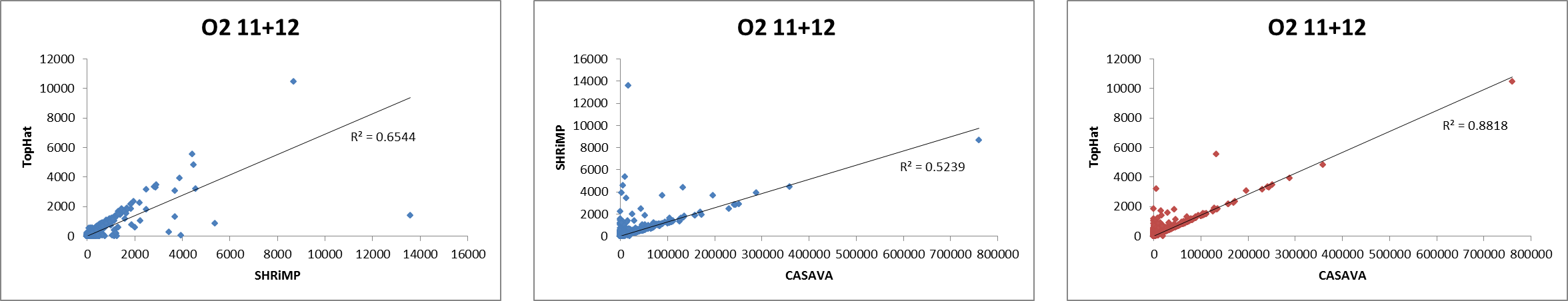
Chin-Yi Chu1,2,a, Soumyaroop Bhattacharya1,2,a, Zhongyang Zhou1, Min Yee1, Ashley M Lopez1, Valerie A Lunger1, Bradley W Buczynski1, Michael A O’Reilly1,3, Thomas J Mariani1,2

aThese authors contributed equally.

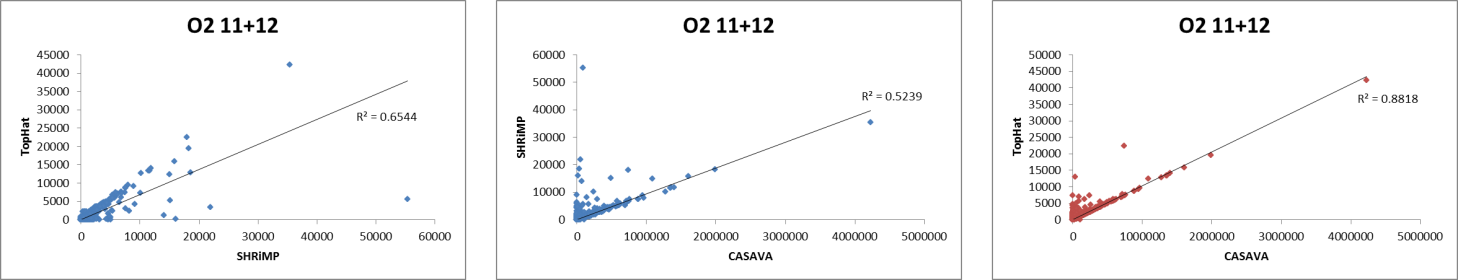
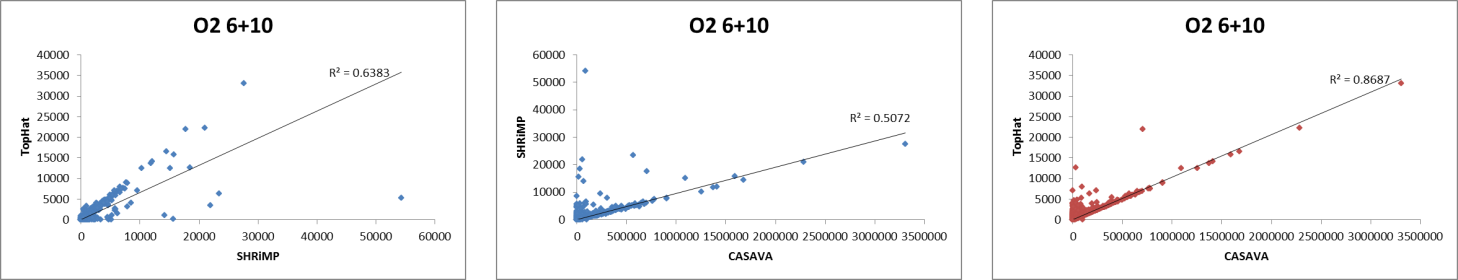
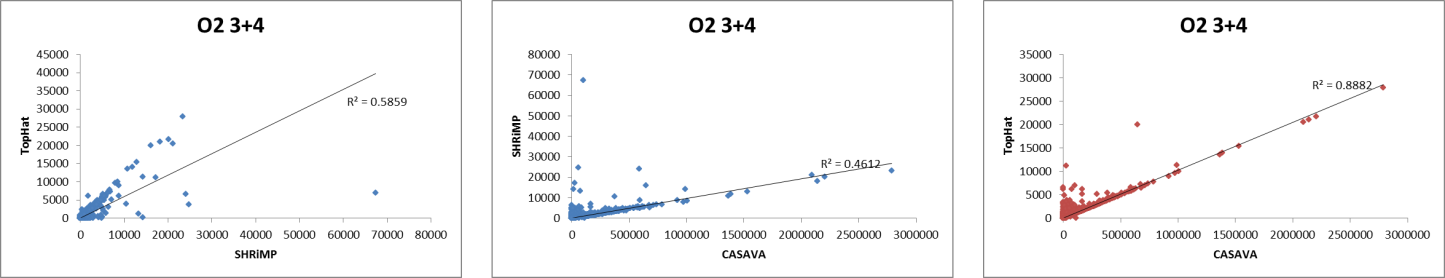
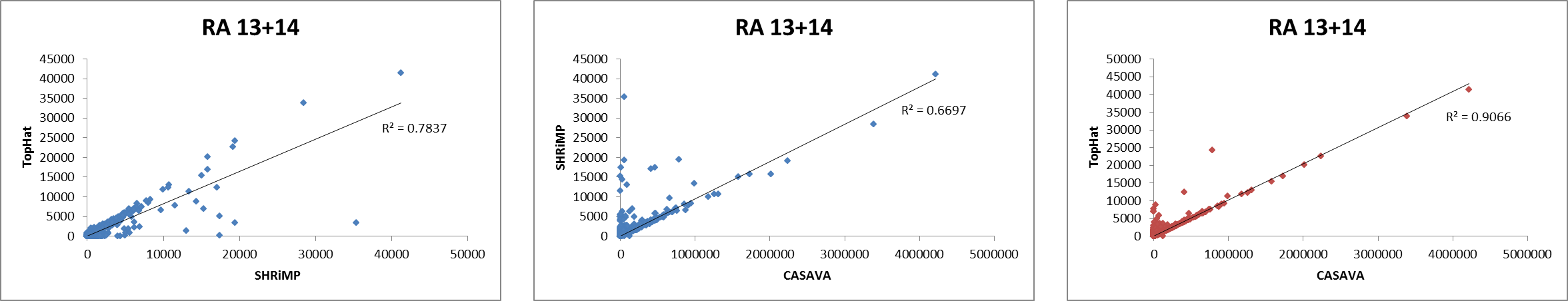
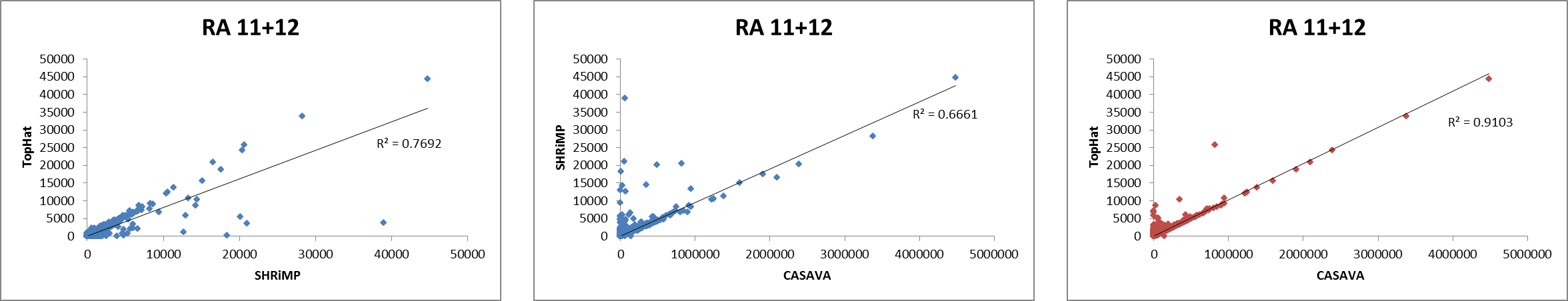
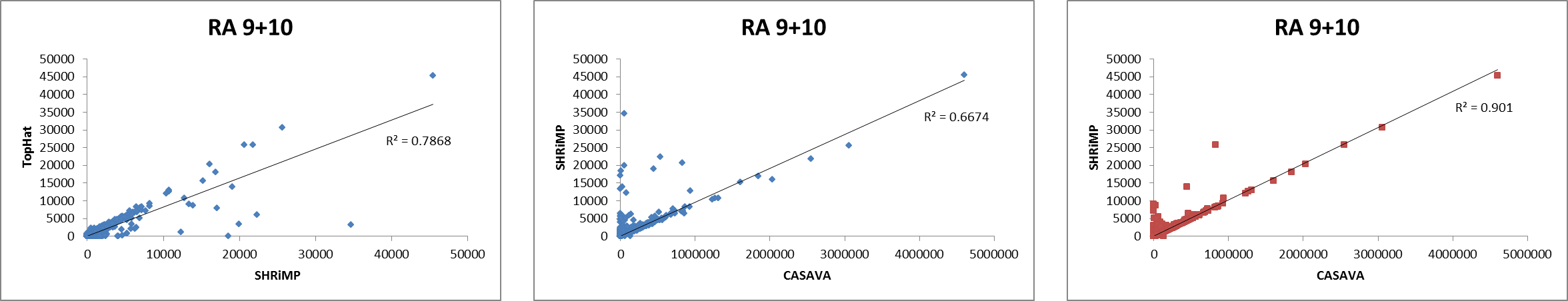
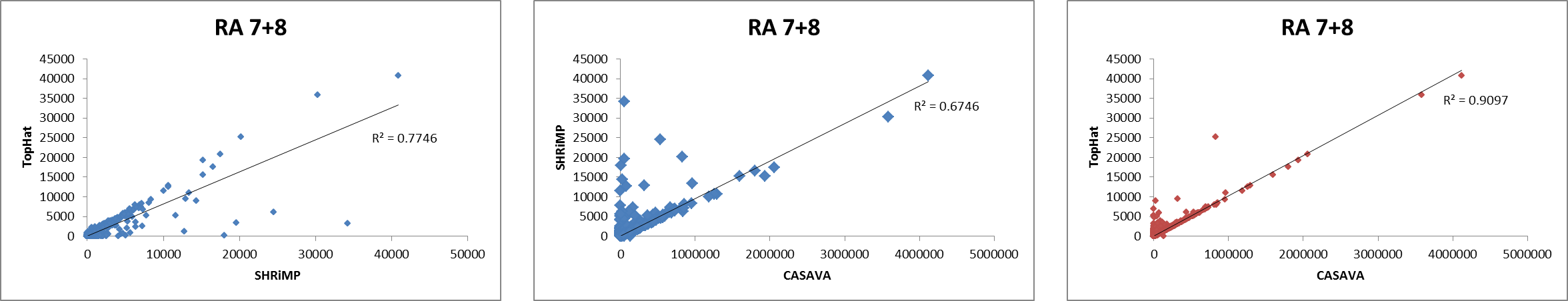
1Division of Neonatology, 2Pediatric Molecular and Personalized Medicine (PMPM) Program, and 3Perinatal and Pediatric Origins of Disease (PPOD) Program, Department of Pediatrics, University of Rochester Medical Center, Rochester NY

SUPPLEMENTAL DATA

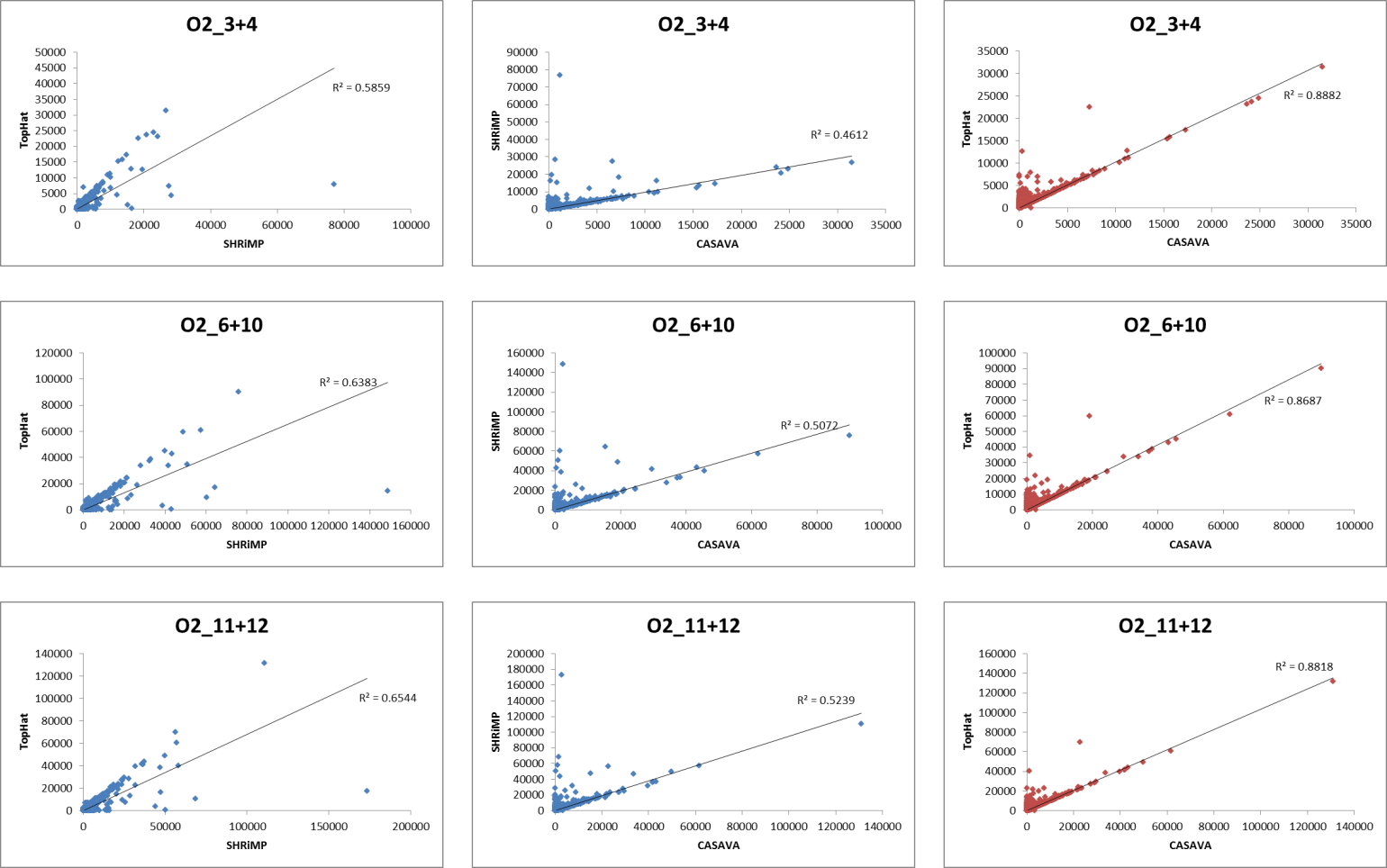
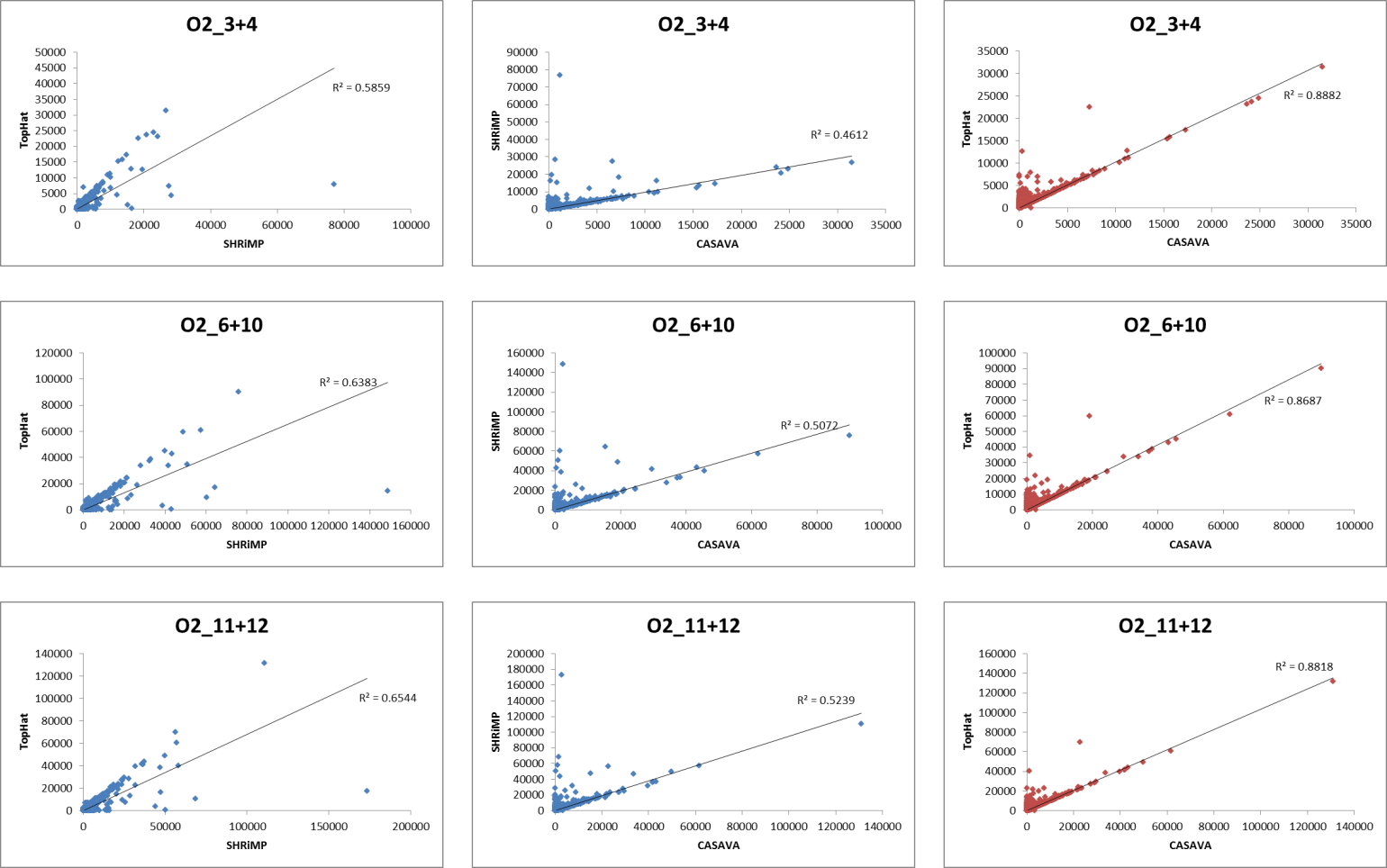
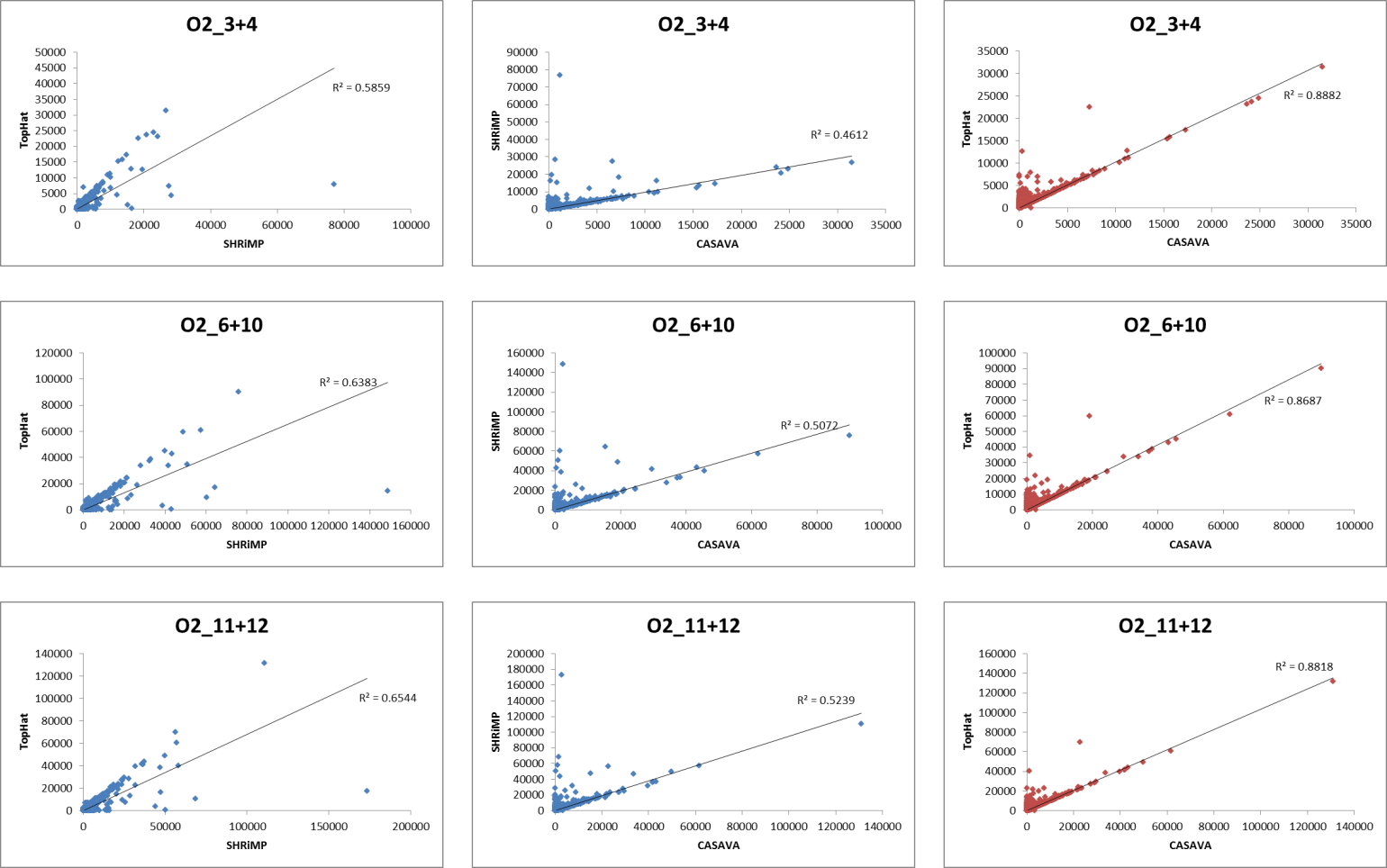
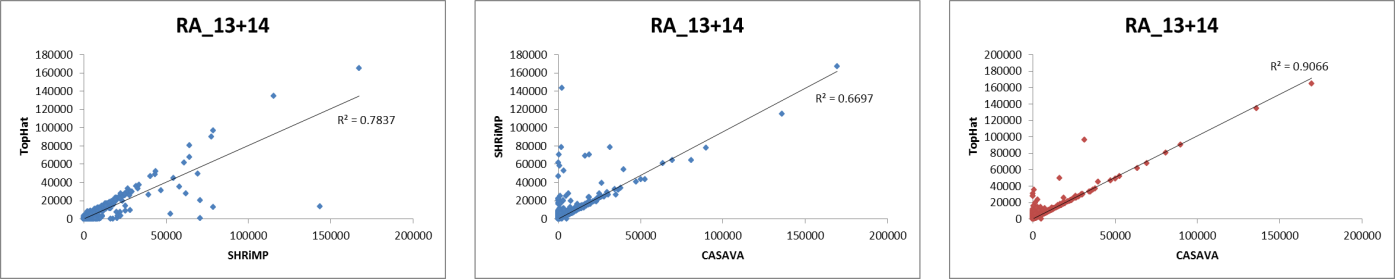
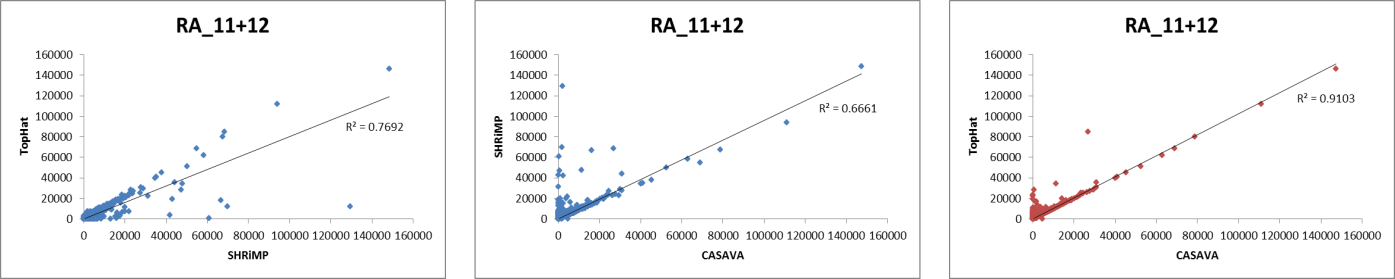
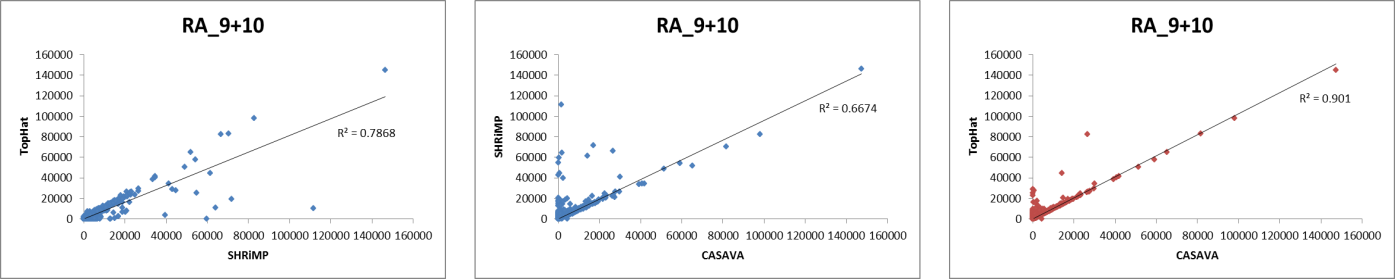
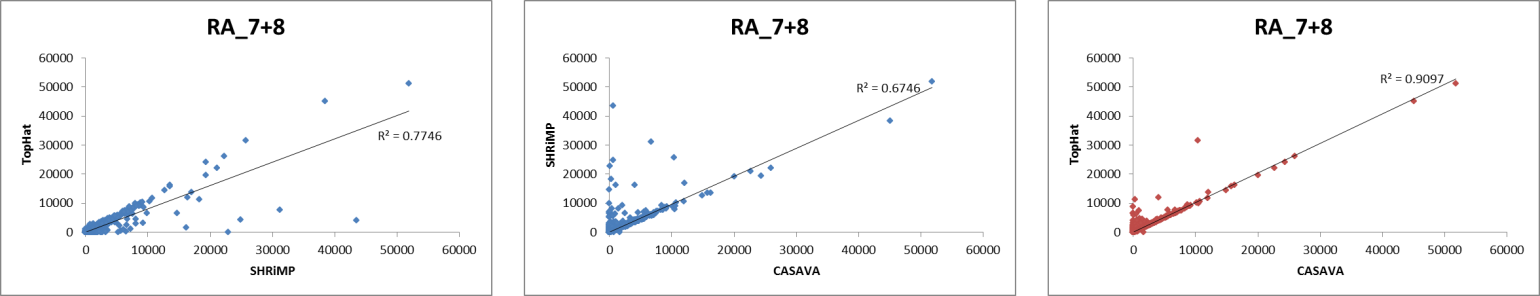
A. Correlation of gene expression of three mapping methods using RPM normalization



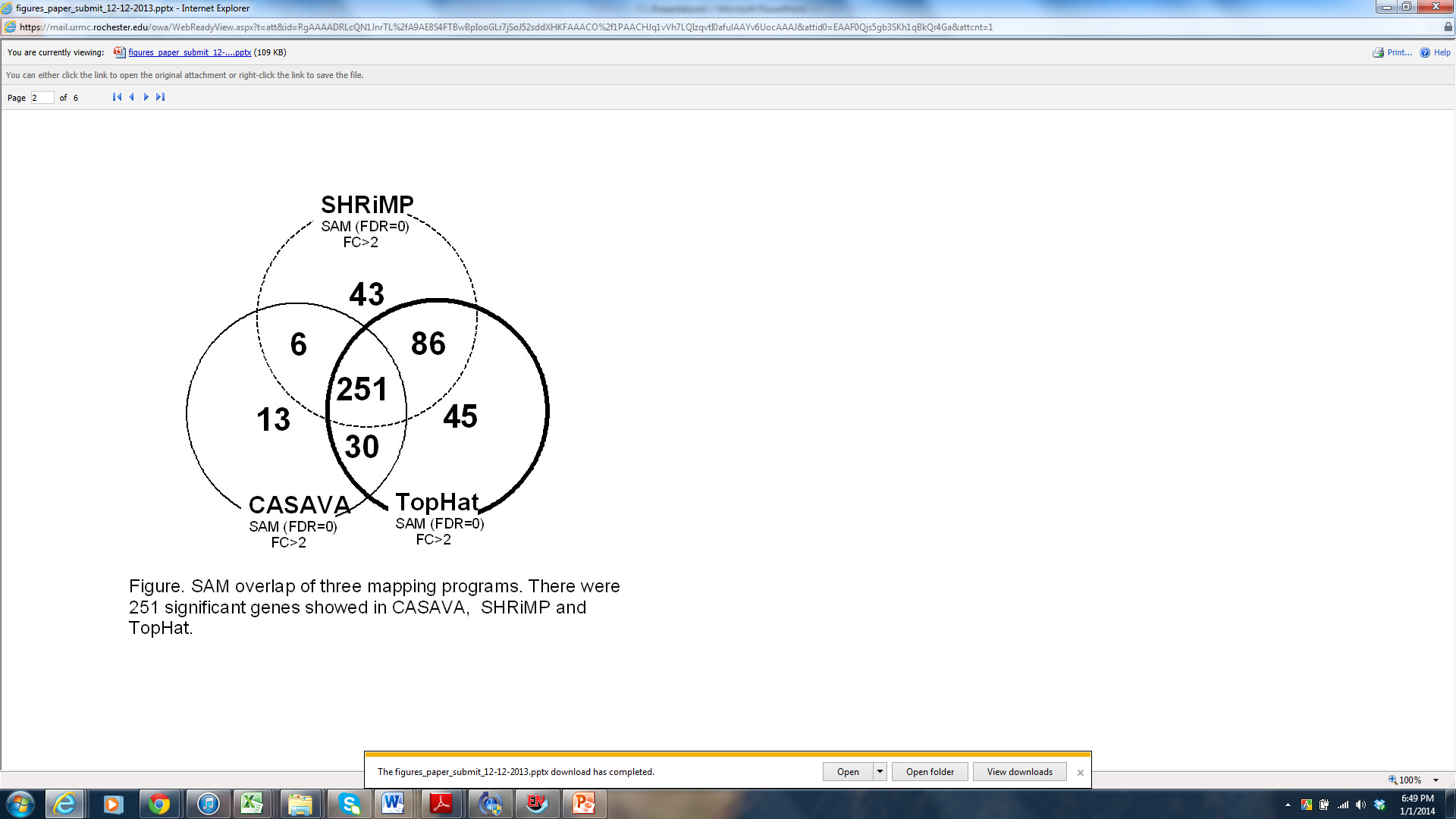
B. Correlation of gene expression of three mapping methods using TM normalization



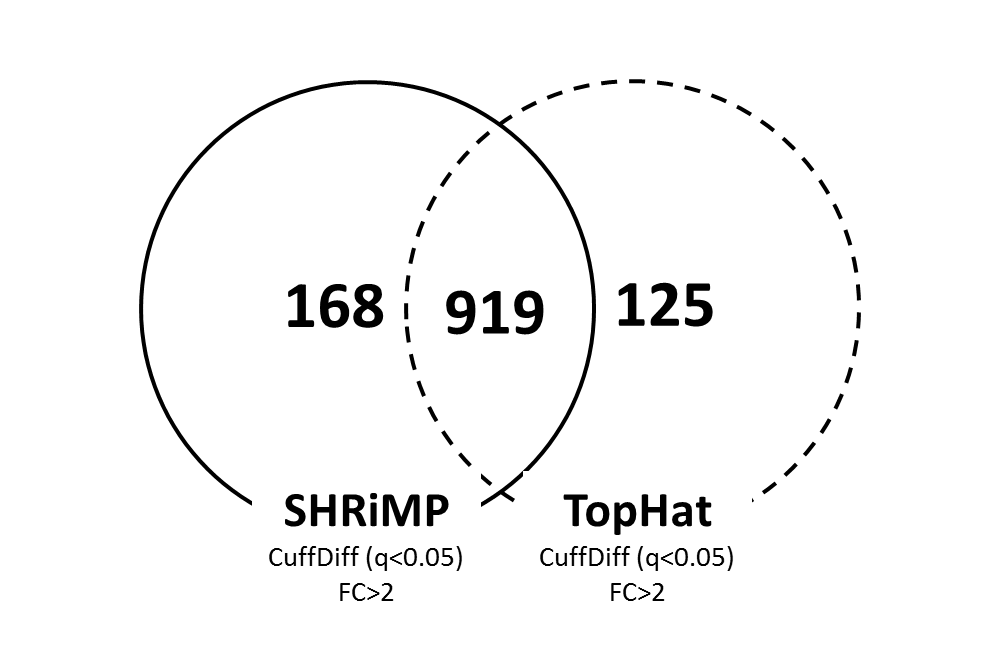
C. Correlation of gene expression of three mapping methods using raw counts



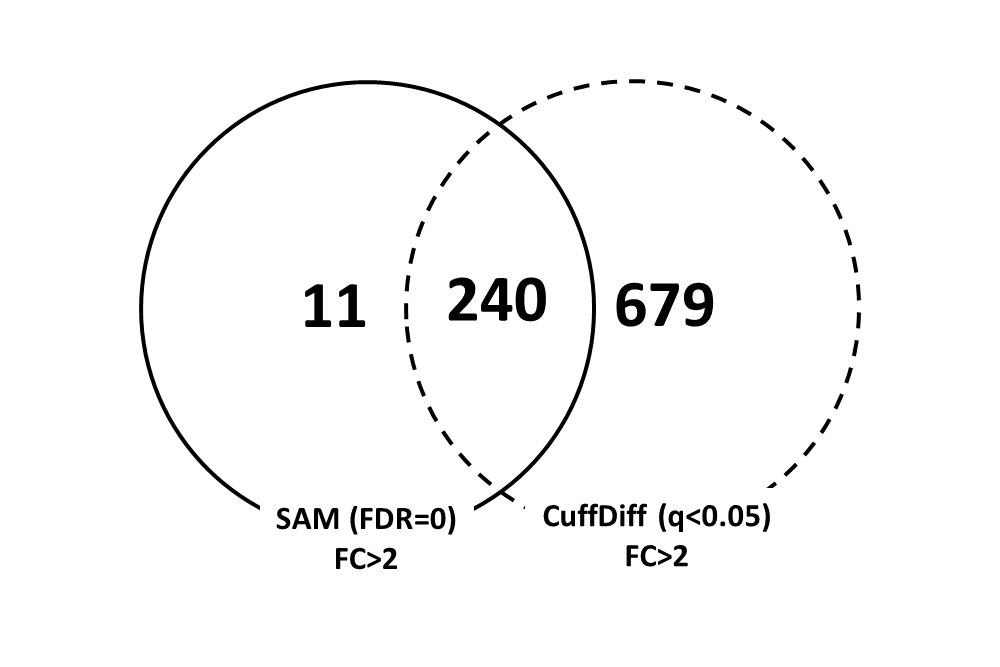
Supplemental Figure 1: Correlation between mapped reads of the sample across three mappers. Raw and normalized counts from individual mappers for the same sample were plotted to identify the correlation between the expression levels. Shown here are dot plots of all seven samples (4 Controls: RAs and 3 Hyperoxia: O2s) between TopHat and CASAVA counts, SHRiMP and CASAVA and TopHat and SHRiMP for RPM normalized counts (A), TM normalized counts (B) and raw un-normalized counts (C).



Supplemental Figure 2: Estimation of differential expression by SAM. Shown are the number of genes identified as differentially expressed by SAM and with fold change > 2, applied to each set of mapped data, independently. A total of 251 genes were identified by SAM using all three mapping approaches.



Supplemental Figure 3: Estimation of differential expression by Cuffdiff. Shown are the number of genes identified as differentially expressed by Cuffdiff and with fold change > 2, applied to SHRiMP and TopHat mapped data, independently. A total of 919 genes were identified by Cuffdiff using both mapping approaches.



Supplemental Figure 4. Estimation of differential expression by SAM and Cuffdiff. Shown are the number of genes identified as differentially expressed, and with fold change > 2, using SAM (applied to CASAVA, SHRiMP and TopHat) and Cuffdiff (SHRiMP and TopHat), independently. A total of 240 genes were consistently identified by both SAM and Cuffdiff using all mapping approaches.

Supplemental Table 1: Codes for mapping and Cuffdiff

|  |
| --- |
| *Codes for Alignment* |
| **TopHat** |
| tophat -p 8 --output /sample\_directory/sample\_1/tophat --library-type fr-unstranded --GTF /reference/Mus\_musculus/UCSC/mm10/Annotation/Archives/archive-2012-05-23-16-47-35/Genes/genes.gtf /reference/Mus\_musculus/UCSC/mm10/Sequence/Bowtie2Index/genome /sample\_location/001/sample-1\_hyperoxi.fastq |
|  |
| **SHRiMP** |
| gmapper -N 8 /sample\_location/sample-1\_hyperoxi.fastq /reference/Mus\_musculus/UCSC/mm10/Sequence/WholeGenomeFasta/genome.fa > /sample\_location/sample\_1/shrimp/hyperoxi.sample1.out >/sample\_location/sample\_1/shrimp/hyperoxi.sample1.sam |
|  |
| *Codes for Cuffdiff* |
| **TopHat** |
| cuffdiff -p 8 -L control,hypoxi -o /sample\_location/cuffdiff\_group2 -u /reference/Mus\_musculus/UCSC/mm10/Annotation/Archives/archive-2012-05-23-16-47-35/Genes/genes.gtf /sample\_location/sample\_1/tophat/accepted\_hits.bam,/sample\_location/sample\_2/tophat/accepted\_hits.bam,/sample\_location/sample\_3/tophat/accepted\_hits.bam,/sample\_location/sample\_4/tophat/accepted\_hits.bam /sample\_location/sample\_5/tophat/accepted\_hits.bam,/sample\_location/sample\_6/tophat/accepted\_hits.bam,/sample\_location/sample\_7/tophat/accepted\_hits.bam |
|  |
| **SHRiMP** |
| cuffdiff -p 8 -L control,hypoxi -o /sample\_location/cuffdiff\_shrimp -u /reference/Mus\_musculus/UCSC/mm10/Annotation/Archives/archive-2012-05-23-16-47-35/Genes/genes.gtf /sample\_location/sample\_1/shrimp/hyperoxi.sample1.sort.bam,/sample\_location/sample\_2/shrimp/hyperoxi.sample2.sort.bam,/sample\_location/sample\_3/shrimp/hyperoxi.sample3.sort.bam,/sample\_location/sample\_4/shrimp/hyperoxi.sample4.sort.bam /sample\_location/sample\_5/shrimp/hyperoxi.sample5.sort.bam,/sample\_location/sample\_6/shrimp/hyperoxi.sample6.sort.bam,/sample\_location/sample\_7/shrimp/hyperoxi.sample7.sort.bam |