St. Geme transposon library mapping project

John K. Everett, Ph.D.

November 2019, draft 3

This analysis describes the creation of a sequencing library created from *Kingella kingae* DNA samples provided by the St. Geme research group and the subsequent mapping of identified transpon insertions. The sequencing library was created by shearing genomic DNA and the subsequent ligation of adapter sequences followed by a nested PCR where the first set of primers bound within the body of the experimental transposon while the second set of primers bound within the transposon ITR segments. The library was sequenced with the Illumina MiSeq platform and transposon insertions were identified by searching for the 8 terminal ITR nucleotides followed by a TA sequence (CAACCTGTTA). The number of insertions recovered from each sample is shown in Table 1. Sequences were aligned to the *Kingella kingae* strain *KWG1*. 63.0% of insertions were detected by sequencing out of both ITRs. Insertions identified via a single ITR were typically from less abundant clones compared to the dually detected insertions.

For the purpose of visualizing the data, the number of recovered insertions were normalized by dividing the number of sites within 10KB genomic blocks by the total number of sites recovered in each sample (Figure 1).

Figure 1. Visualization of recovered insertions within *Kingella kingae*.

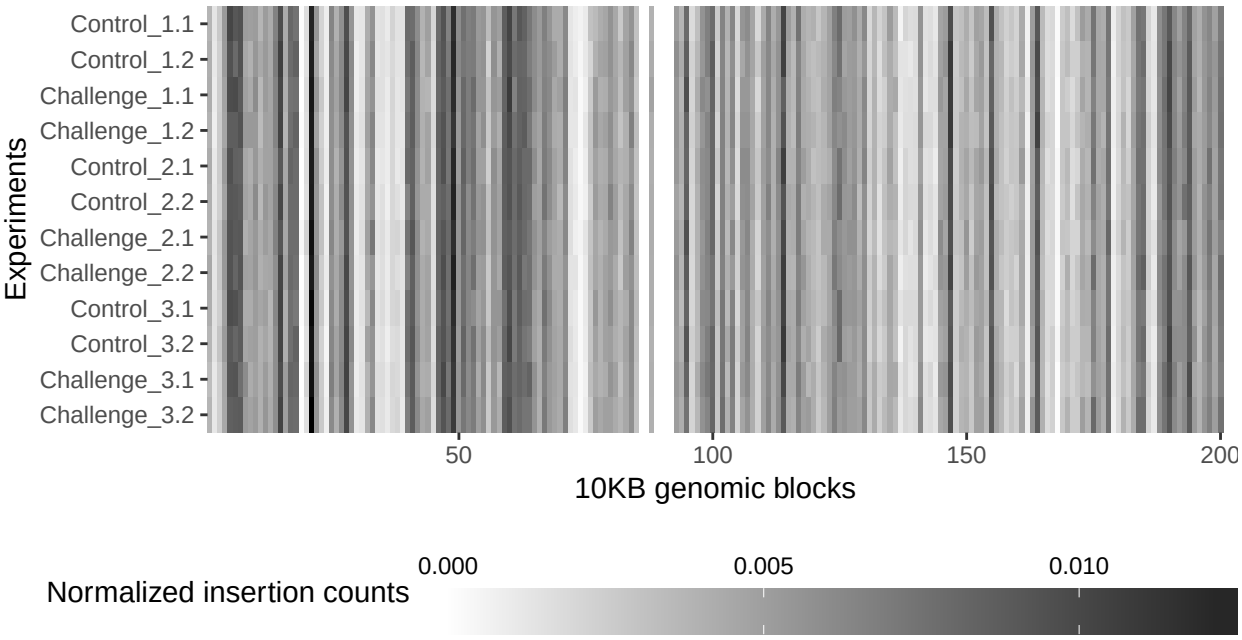


Table 1. Number of recovered insertions.

Sample	Insertions	Sample	Insertions
Control_1.1	7,588	Challenge_1.1	7,097
Control_1.2	7,001	Challenge_1.2	7,145
Control_2.1	7,270	Challenge_2.1	7,284
Control_2.2	7,363	Challenge_2.2	7,142
Control_3.1	7,190	Challenge_3.1	6,777
Control_3.2	6,931	Challenge_3.2	7,332

The number of insertions within transcription units (TUs) was gauged using two approaches. The first approach considered the number of insertions within each TU divided by the total number of insertions recovered in the sample. The second approach considered the total number of inferred cells (unique genomic break points) associated with insertions within each TU divided by the total number of inferred cells in the sample. The site count approach showed a fair degree of variation between techinical replicates (Figure 2) while the abundance method showed less variation between replicates and averaged samples (Figure 3).

Figure 2. Distriubtions of differences between technical replicate insertion counts within TUs using the site count normalization approach.

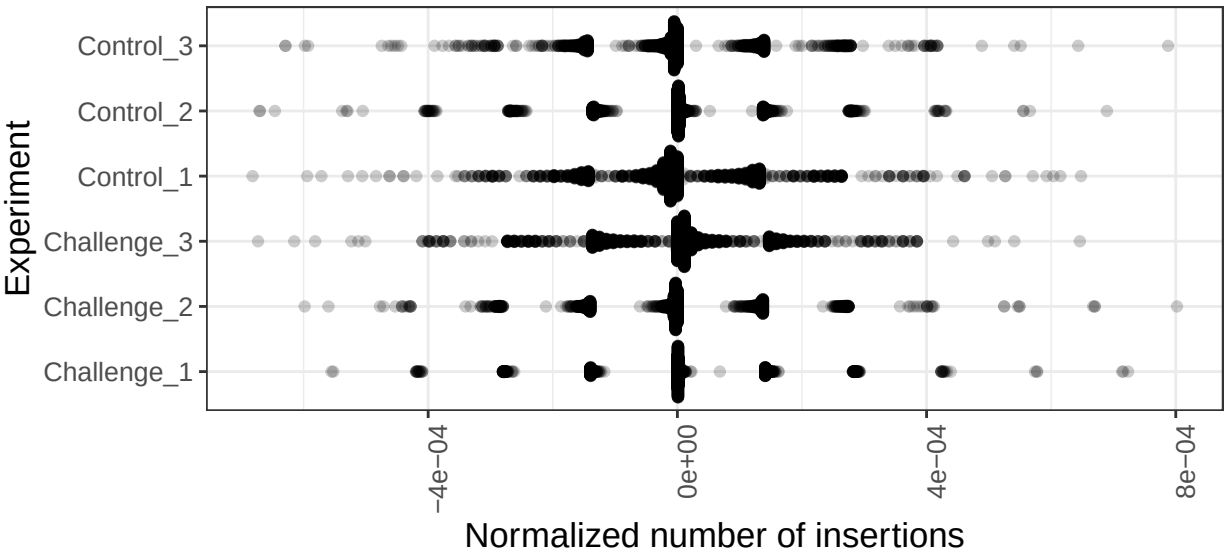
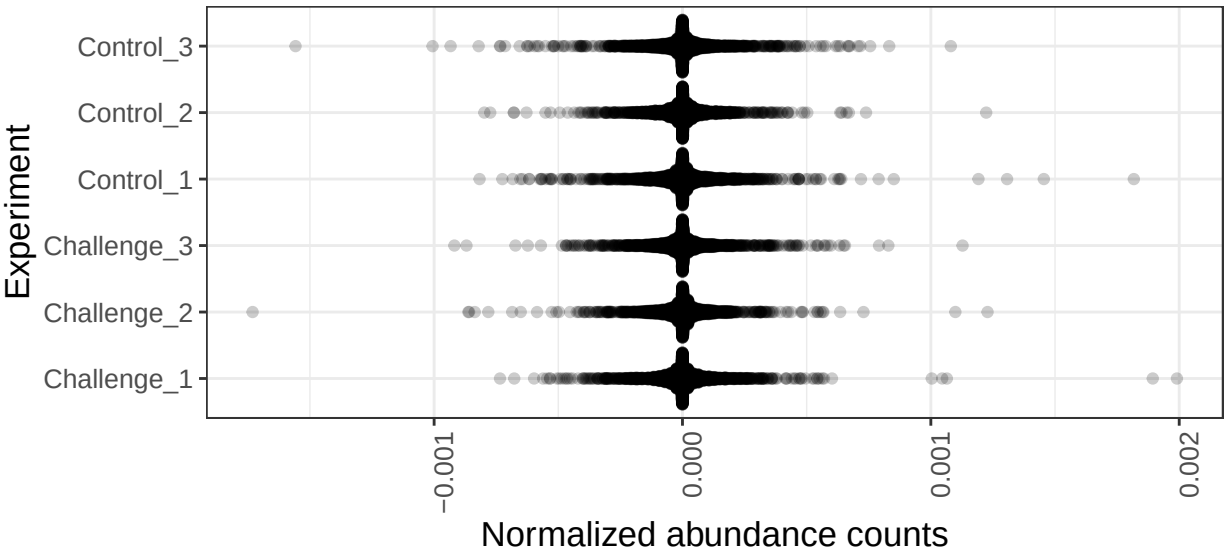


Figure 3. Distriubtions of differences between technical replicate insertion counts within TUs using the abundance normalization approach.



Using the abundance approach, clear clustering of biological samples was found though there was not remarkable separation between control and challenge samples whithin biological sample clusters (Figures 4 & 5). The normalized site count approach provided less distinctive clustering (Supp. Figures S1 & S2).

Figure 4. Principle component analysis of all samples using the abundance normalization approach.

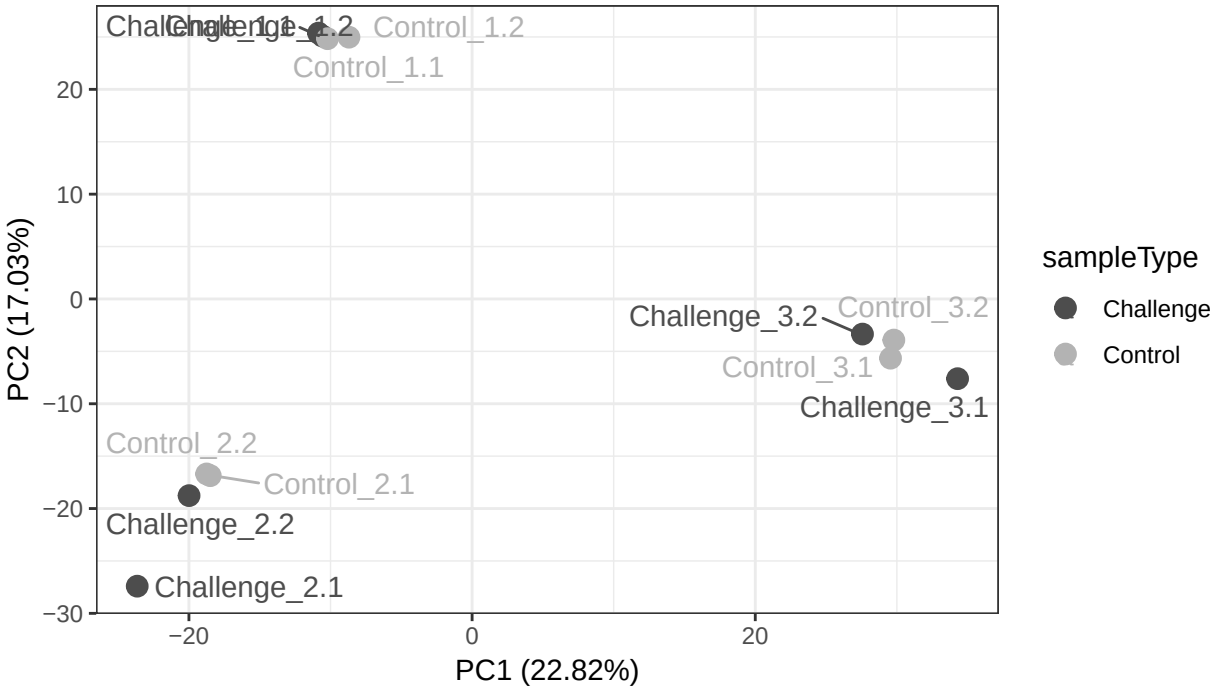
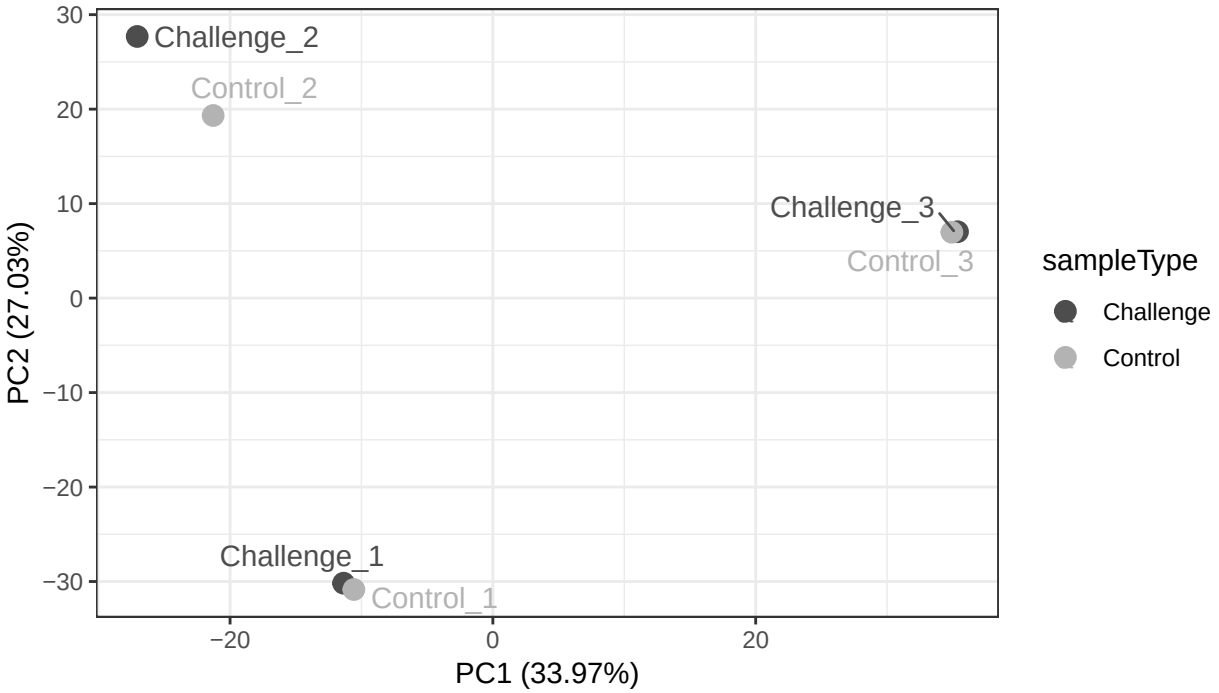


Figure 5. Principle component analysis of averaged technical replicates using the abundance normalization approach.



For each transcription unit, t-tests were used to test for differences between control and challenge insertion frequencies. Transcription units with significant uncorrected p-values are shown in Tables 2 & 3. Gene names followed by 'PRO' represent potential promoter regions 1-50 NTs upstream of genes. Full gene tables are available on-line via this **link**.

Table 2a. Genes with significant uncorrected p-values using the **abundance correction method**.

nearestFeature	geneDesc	pVal	pVal.adj	higherInChallenge
RS06325	hypothetical protein	0.0002257	0.5177234	FALSE
RS06320	membrane protein	0.0005052	1.0000000	FALSE
RS05655 PRO	tRNA guanosine(34) transglycosylase Tgt PRO	0.0025402	1.0000000	TRUE
RS06315	bifunctional glutamine synthetase adenylyltransferase/deadenyltransferase	0.0078044	1.0000000	TRUE
RS02950	membrane protein	0.0116743	1.0000000	TRUE
RS02760	hypothetical protein	0.0118791	1.0000000	FALSE
RS05360	chromosome partitioning protein ParB	0.0128667	1.0000000	TRUE
RS01290 PRO	octaprenyl diphosphate synthase PRO	0.0174566	1.0000000	TRUE
RS07855	hydrolase	0.0175507	1.0000000	TRUE
RS02465 PRO	cysB PRO	0.0180220	1.0000000	TRUE
RS07575 PRO	twin-arginine translocation pathway signal PRO	0.0186341	1.0000000	FALSE
RS06380	amine oxidase	0.0202342	1.0000000	TRUE
RS03045	ABC transporter ATP-binding protein	0.0203126	1.0000000	FALSE
RS08120	DDE transposase family protein	0.0226922	1.0000000	TRUE
RS05485 PRO	aspartate carbamoyltransferase PRO	0.0229357	1.0000000	FALSE
RS01340	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	0.0243480	1.0000000	TRUE
RS08760	CDP-6-deoxy-delta-3,4-glucoseen reductase	0.0277319	1.0000000	FALSE
RS09165	hypothetical protein	0.0289160	1.0000000	FALSE
RS10785	segregation and condensation protein A	0.0293165	1.0000000	FALSE
RS09110	ushA	0.0298085	1.0000000	TRUE
RS02995 PRO	tetrapyrrole methylase PRO	0.0309888	1.0000000	FALSE
RS03145 PRO	regulator of pilE expression PRO	0.0334814	1.0000000	FALSE
RS05285 PRO	peptidoglycan-binding protein LysM PRO	0.0343454	1.0000000	FALSE
RS06575	AbrB/MazE/SpoVT family DNA-binding domain-containing protein	0.0355805	1.0000000	TRUE
RS07240 PRO	tRNA dihydrouridine synthase DusB PRO	0.0379355	1.0000000	TRUE
RS10000	hypothetical protein	0.0395066	1.0000000	FALSE
RS00845 PRO	dihydroxy-acid dehydratase PRO	0.0404018	1.0000000	TRUE
RS01275	RhtB family homoserine/homoserine lactone efflux pump	0.0412109	1.0000000	TRUE
RS08020	hypothetical protein	0.0413579	1.0000000	FALSE
RS10200	hypothetical protein	0.0418517	1.0000000	FALSE
RS10240 PRO	phosphogluconate dehydratase PRO	0.0423961	1.0000000	TRUE
RS07750 PRO	serine O-acetyltransferase PRO	0.0429211	1.0000000	TRUE
RS00490,RS00495	hypothetical protein, hypothetical protein	0.0449955	1.0000000	TRUE
RS09790 PRO	3-isopropylmalate dehydratase large subunit PRO	0.0451630	1.0000000	FALSE
RS00260	peptidylprolyl isomerase	0.0483947	1.0000000	TRUE
RS06960	phage morphogenesis protein	0.0494974	1.0000000	TRUE

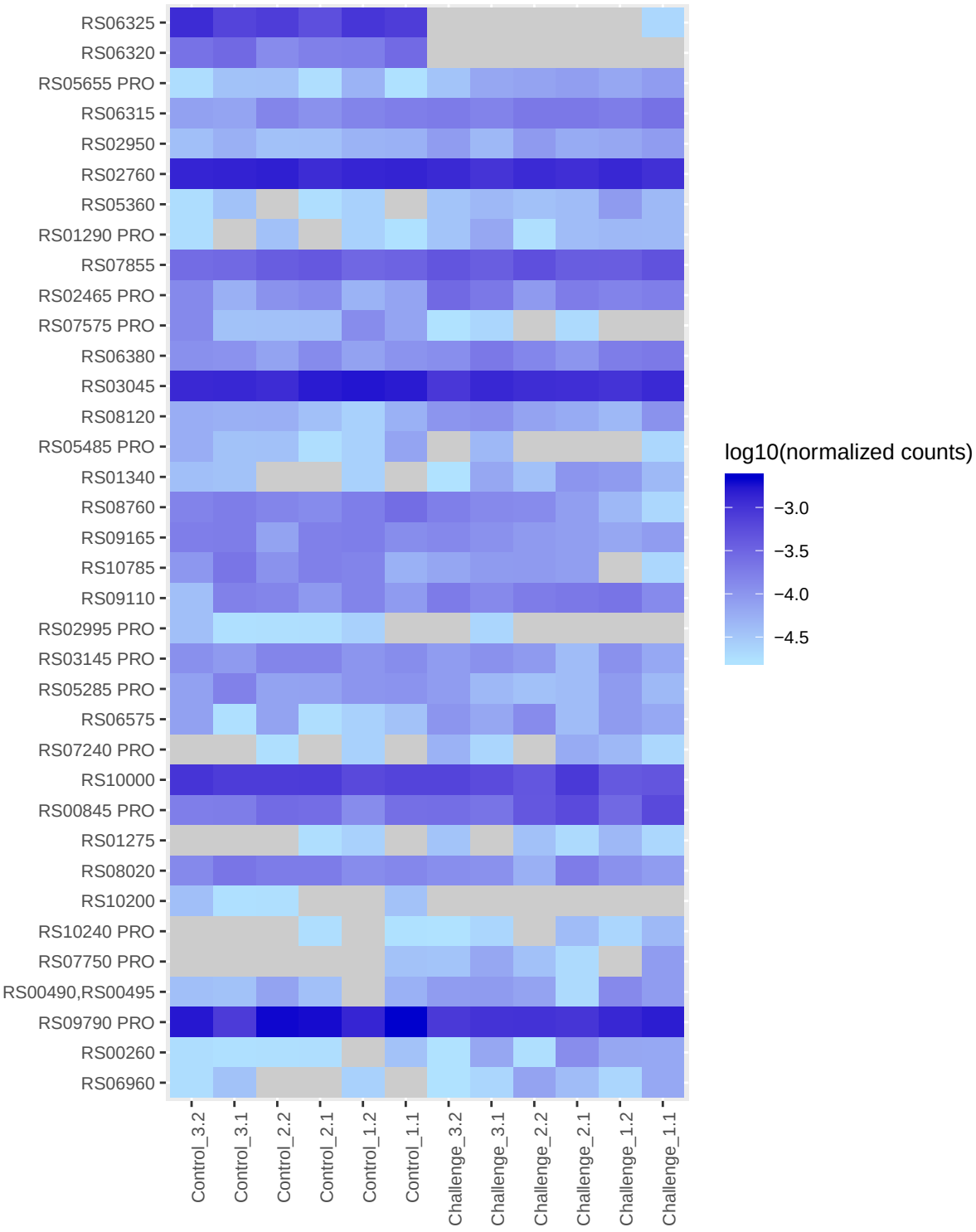

RS06325
RS06320
RS05655 PRO
RS06315
RS02950
RS02760
RS05360
RS01290 PRO
RS07855
RS02465 PRO
RS07575 PRO
RS06380
RS03045
RS08120
RS05485 PRO
RS01340
RS08760
RS09165
RS10785
RS09110
RS02995 PRO
RS03145 PRO
RS05285 PRO
RS06575
RS07240 PRO
RS10000
RS00845 PRO
RS01275
RS08020
RS10200
RS10240 PRO
RS07750 PRO
RS00490,RS00495
RS09790 PRO
RS00260
RS06960
Control_3.2
Control_3.1
Control_2.2
Control_2.1
Control_1.2
Control_1.1
Challenge_3.2
Challenge_3.1
Challenge_2.2
Challenge_2.1
Challenge_1.2
Challenge_1.1
log10(normalized counts)
−3.0
−3.5
−4.0
−4.5

Table 2b. Genes with significant uncorrected p-values using the **abundance correction method** where replicates have been averaged.

nearestFeature	geneDesc	pVal_avgReps	pVal.adj_avgReps	higherInChallenge_avgReps
RS02760	hypothetical protein	0.0017685	1	FALSE
RS03580	SAM-dependent methyltransferase	0.0052306	1	TRUE
RS08745 PRO	AsmA family protein PRO	0.0055018	1	TRUE
RS00635	membrane protein insertase YidC	0.0067260	1	TRUE
RS09110	ushA	0.0067420	1	TRUE
RS10260	phosphocarrier protein HPr	0.0070279	1	TRUE
RS06325	hypothetical protein	0.0082137	1	FALSE
RS06575	AbrB/MazE/SpoVT family DNA-binding domain-containing protein	0.0088640	1	TRUE
RS02950	membrane protein	0.0090342	1	TRUE
RS07240 PRO	tRNA dihydrouridine synthase DusB PRO	0.0092688	1	TRUE
RS04380	hypothetical protein	0.0167800	1	FALSE
RS07750 PRO	serine O-acetyltransferase PRO	0.0168313	1	TRUE
RS03245 PRO	acetylglutamate kinase PRO	0.0209137	1	TRUE
RS06320	membrane protein	0.0239137	1	FALSE
RS01185	transcriptional regulator	0.0240406	1	TRUE
RS03450	membrane protein	0.0285038	1	FALSE
RS05830	membrane protein	0.0294757	1	FALSE
RS10240 PRO	phosphogluconate dehydratase PRO	0.0302785	1	TRUE
RS05485 PRO	aspartate carbamoyltransferase PRO	0.0320363	1	FALSE
RS09180	hypothetical protein	0.0325594	1	TRUE
RS01290 PRO	octaprenyl diphosphate synthase PRO	0.0368176	1	TRUE
RS06770	hypothetical protein	0.0375086	1	TRUE
RS01275	RhtB family homoserine/homoserine lactone efflux pump	0.0383973	1	TRUE
RS01445 PRO	hypothetical protein PRO	0.0400657	1	FALSE
RS01340	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	0.0406549	1	TRUE
RS05655 PRO	tRNA guanosine(34) transglycosylase Tgt PRO	0.0426308	1	TRUE
RS07520 PRO	long-chain-fatty-acid–CoA ligase	0.0428092	1	TRUE
RS07685	hypothetical protein	0.0458854	1	FALSE
RS06230	thymidine phosphorylase	0.0483467	1	TRUE
RS02085	preprotein translocase subunit SecA	0.0485071	1	FALSE

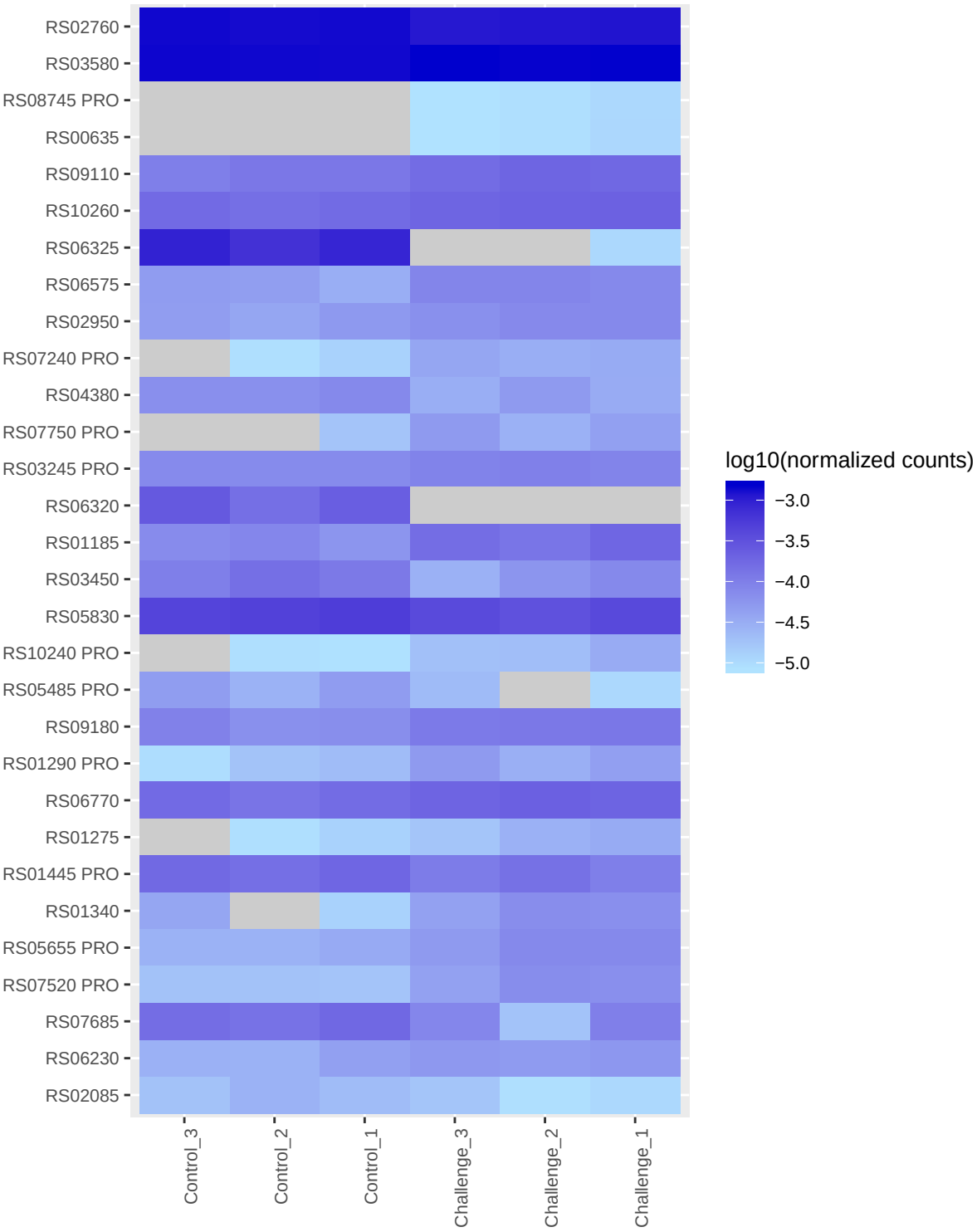

RS02760
RS03580
RS08745 PRO
RS00635
RS09110
RS10260
RS06325
RS06575
RS02950
RS07240 PRO
RS04380
RS07750 PRO
RS03245 PRO
RS06320
RS01185
RS03450
RS05830
RS10240 PRO
RS05485 PRO
RS09180
RS01290 PRO
RS06770
RS01275
RS01445 PRO
RS01340
RS05655 PRO
RS07520 PRO
RS07685
RS06230
RS02085
Control_3
Control_2
Control_1
Challenge_3
Challenge_2
Challenge_1
log10(normalized counts)
−3.0
−3.5
−4.0
−4.5
−5.0

Table 3a. Genes with significant uncorrected p-values using the **site count correction method**.

nearestFeature	geneDesc	pVal	pVal.adj	higherInChallenge
RS07175	beta-ketoacyl-[acyl-carrier-protein] synthase II	0.0000001	0.0002461	TRUE
RS06325	hypothetical protein	0.0000136	0.0312984	FALSE
RS06320	membrane protein	0.0003398	0.7787999	FALSE
RS04110	hypothetical protein	0.0034665	1.0000000	FALSE
RS07165	fabG	0.0042780	1.0000000	FALSE
RS02695	acid phosphatase/phosphotransferase	0.0070759	1.0000000	TRUE
RS09585	beta-phosphoglucomutase	0.0076343	1.0000000	TRUE
RS06315	bifunctional glutamine synthetase adenylyltransferase/deadenyltransferase	0.0117527	1.0000000	TRUE
RS03410	bacterioferritin	0.0118347	1.0000000	FALSE
RS07415	spermidine synthase	0.0122965	1.0000000	TRUE
RS09225	transposase	0.0139049	1.0000000	TRUE
RS07750 PRO	serine O-acetyltransferase PRO	0.0144021	1.0000000	TRUE
RS06060	iron ABC transporter substrate-binding protein	0.0152173	1.0000000	TRUE
RS10835	cysteine synthase A	0.0170845	1.0000000	TRUE
RS08780	tRNA (uridine(54)-C5)-methyltransferase TrmA	0.0183050	1.0000000	FALSE
RS00715	thiol:disulfide interchange protein	0.0188796	1.0000000	TRUE
RS00225	DDE transposase	0.0208518	1.0000000	FALSE
RS02440	oligopeptide transporter, OPT family	0.0211885	1.0000000	TRUE
RS02995 PRO	tetrapyrrole methylase PRO	0.0218830	1.0000000	FALSE
RS05640 PRO	transposase PRO	0.0226680	1.0000000	FALSE
RS00325	prepilin-type cleavage/methylation domain-containing protein	0.0238197	1.0000000	FALSE
RS09790 PRO	3-isopropylmalate dehydratase large subunit PRO	0.0249653	1.0000000	FALSE
RS03750	transposase	0.0250973	1.0000000	TRUE
RS10335,RS10335 PRO	hypothetical protein, hypothetical protein PRO	0.0251696	1.0000000	FALSE
RS10200	hypothetical protein	0.0251783	1.0000000	FALSE
RS10820	hypothetical protein	0.0272716	1.0000000	FALSE
RS01105	murein transglycosylase	0.0281385	1.0000000	TRUE
RS05725	hypothetical protein	0.0292943	1.0000000	FALSE
RS09390	laccase	0.0296422	1.0000000	TRUE
RS06540	twitching motility protein PilT	0.0297873	1.0000000	FALSE
RS05530	DNA translocase FtsK	0.0301050	1.0000000	FALSE
RS07955	TonB-dependent copper receptor	0.0307645	1.0000000	TRUE
RS05485 PRO	aspartate carbamoyltransferase PRO	0.0309073	1.0000000	FALSE
RS10170	glucose/galactose MFS transporter	0.0309840	1.0000000	TRUE
RS04015	membrane protein	0.0325873	1.0000000	FALSE
RS04380 PRO	hypothetical protein PRO	0.0333539	1.0000000	TRUE
RS07085	hypothetical protein	0.0391338	1.0000000	TRUE
RS10305	helicase	0.0414531	1.0000000	TRUE
RS01325 PRO	membrane protein PRO	0.0430577	1.0000000	FALSE
RS03085	ychF	0.0431731	1.0000000	TRUE
RS00315	fimb protein	0.0433840	1.0000000	TRUE
RS09075	cyclophilin	0.0436596	1.0000000	FALSE
RS00405	hypothetical protein	0.0438704	1.0000000	TRUE
RS09925	PqiA family protein	0.0440591	1.0000000	FALSE
RS00495 PRO	hypothetical protein PRO	0.0440661	1.0000000	FALSE
RS01350 PRO	hypothetical protein PRO	0.0445367	1.0000000	FALSE
RS10165	MFS transporter	0.0463183	1.0000000	FALSE

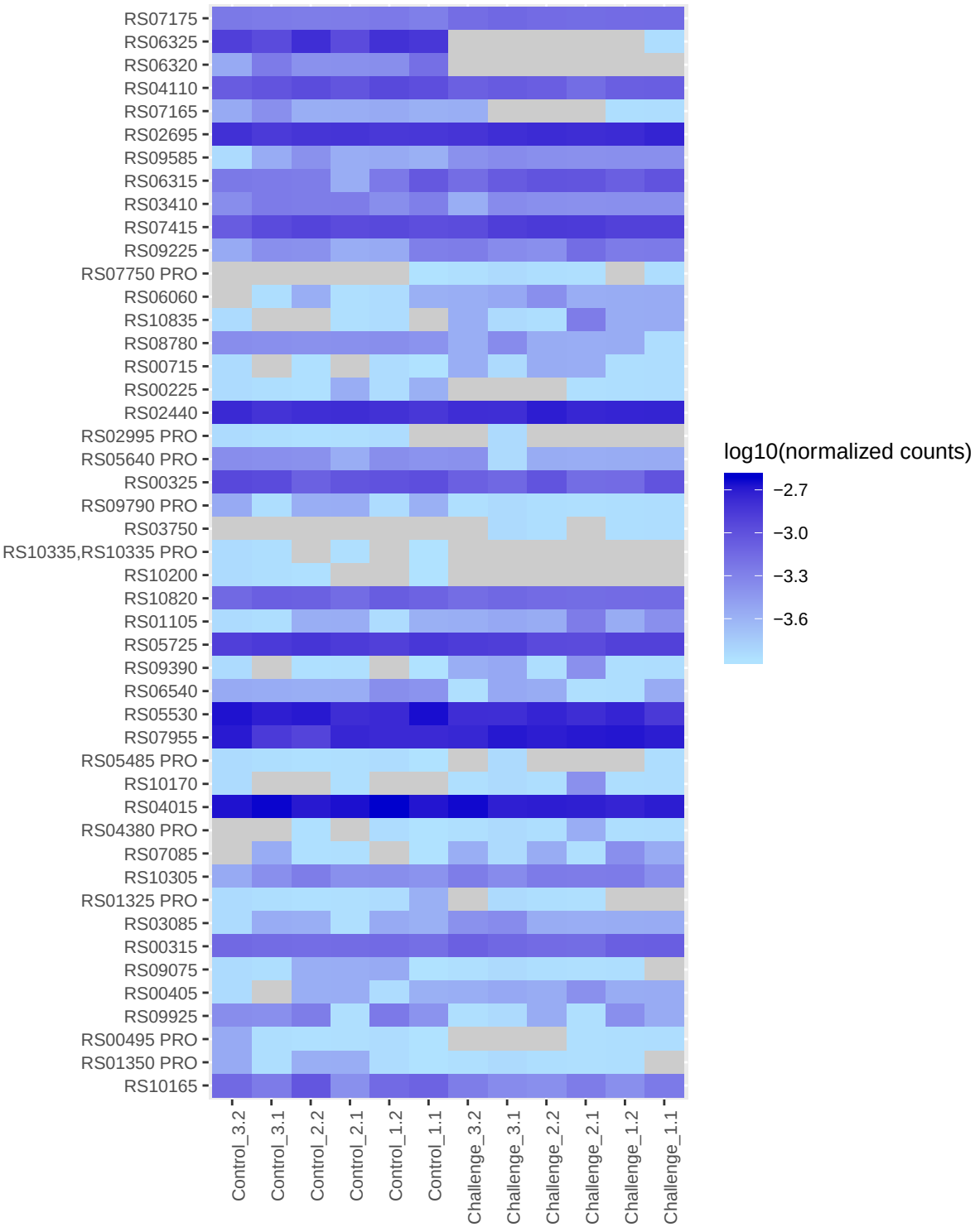
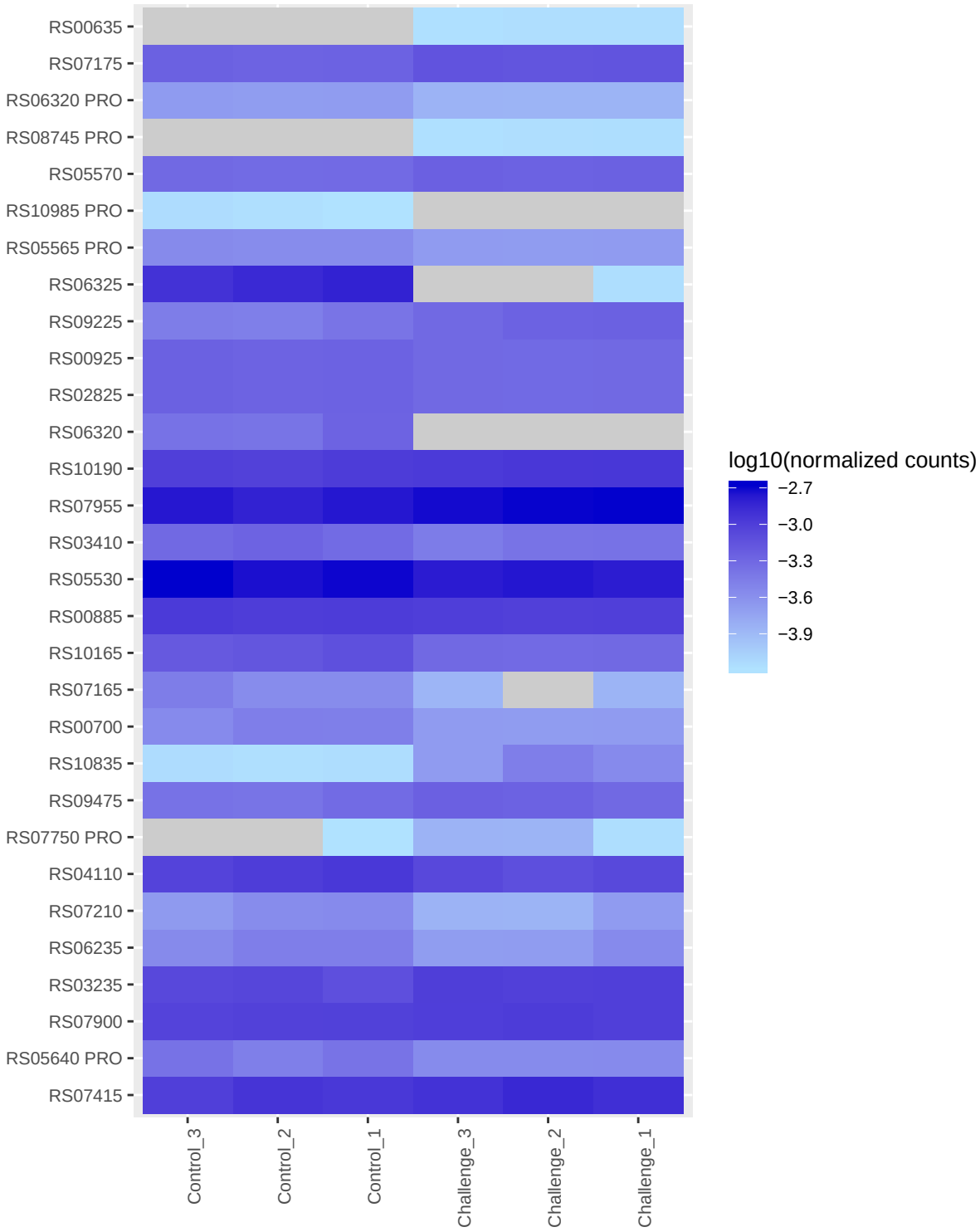

RS07175
RS06325
RS06320
RS04110
RS07165
RS02695
RS09585
RS06315
RS03410
RS07415
RS09225
RS07750 PRO
RS06060
RS10835
RS08780
RS00715
RS00225
RS02440
RS02995 PRO
RS05640 PRO
RS00325
RS09790 PRO
RS03750
RS10335,RS10335 PRO
RS10200
RS10820
RS01105
RS05725
RS09390
RS06540
RS05530
RS07955
RS05485 PRO
RS10170
RS04015
RS04380 PRO
RS07085
RS10305
RS01325 PRO
RS03085
RS00315
RS09075
RS00405
RS09925
RS00495 PRO
RS01350 PRO
RS10165
Control_3.2
Control_3.1
Control_2.2
Control_2.1
Control_1.2
Control_1.1
Challenge_3.2
Challenge_3.1
Challenge_2.2
Challenge_2.1
Challenge_1.2
Challenge_1.1
log10(normalized counts)
−2.7
−3.0
−3.3
−3.6

Table 3b. Genes with significant uncorrected p-values using the **site count correction method** where replicates have been averaged.

nearestFeature	geneDesc	pVal_avgReps	pVal.adj_avgReps	higherInChallenge_avgReps
RS00635	membrane protein insertase YidC	0.0000747	0.1714131	TRUE
RS07175	beta-ketoacyl-[acyl-carrier-protein] synthase II	0.0000791	0.1814739	TRUE
RS06320 PRO	membrane protein PRO	0.0000850	0.1949334	FALSE
RS08745 PRO	AsmA family protein PRO	0.0000986	0.2259064	TRUE
RS05570	N-acetyltransferase	0.0003516	0.8050824	TRUE
RS10985 PRO	hypothetical protein PRO	0.0006848	1.0000000	FALSE
RS05565 PRO	hypothetical protein PRO	0.0016750	1.0000000	FALSE
RS06325	hypothetical protein	0.0026199	1.0000000	FALSE
RS09225	transposase	0.0044454	1.0000000	TRUE
RS00925	septal ring lytic transglycosylase RlpA family lipoprotein	0.0045922	1.0000000	FALSE
RS02825	hypothetical protein	0.0045922	1.0000000	FALSE
RS06320	membrane protein	0.0085305	1.0000000	FALSE
RS10190	RNase adaptor protein RapZ	0.0182590	1.0000000	TRUE
RS07955	TonB-dependent copper receptor	0.0201903	1.0000000	TRUE
RS03410	bacterioferritin	0.0205324	1.0000000	FALSE
RS05530	DNA translocase FtsK	0.0211981	1.0000000	FALSE
RS00885	hypothetical protein	0.0222886	1.0000000	FALSE
RS10165	MFS transporter	0.0254068	1.0000000	FALSE
RS07165	fabG	0.0266884	1.0000000	FALSE
RS00700	IS5/IS1182 family transposase	0.0308223	1.0000000	FALSE
RS10835	cysteine synthase A	0.0331074	1.0000000	TRUE
RS09475	glycosyl transferase	0.0350758	1.0000000	TRUE
RS07750 PRO	serine O-acetyltransferase PRO	0.0413951	1.0000000	TRUE
RS04110	hypothetical protein	0.0438521	1.0000000	FALSE
RS07210	hypothetical protein	0.0443253	1.0000000	FALSE
RS06235	hypothetical protein	0.0469268	1.0000000	FALSE
RS03235	DNA polymerase III subunit epsilon	0.0473629	1.0000000	TRUE
RS07900	branched-chain amino acid ABC transporter permease	0.0480385	1.0000000	TRUE
RS05640 PRO	transposase PRO	0.0481931	1.0000000	FALSE
RS07415	spermidine synthase	0.0486415	1.0000000	TRUE


RS00635
RS07175
RS06320 PRO
RS08745 PRO
RS05570
RS10985 PRO
RS05565 PRO
RS06325
RS09225
RS00925
RS02825
RS06320
RS10190
RS07955
RS03410
RS05530
RS00885
RS10165
RS07165
RS00700
RS10835
RS09475
RS07750 PRO
RS04110
RS07210
RS06235
RS03235
RS07900
RS05640 PRO
RS07415
Control_3
Control_2
Control_1
Challenge_3
Challenge_2
Challenge_1
log10(normalized counts)
−2.7
−3.0
−3.3
−3.6
−3.9

Supplimental

Figure S1. Principle component analysis of all samples using the normalized site count method.

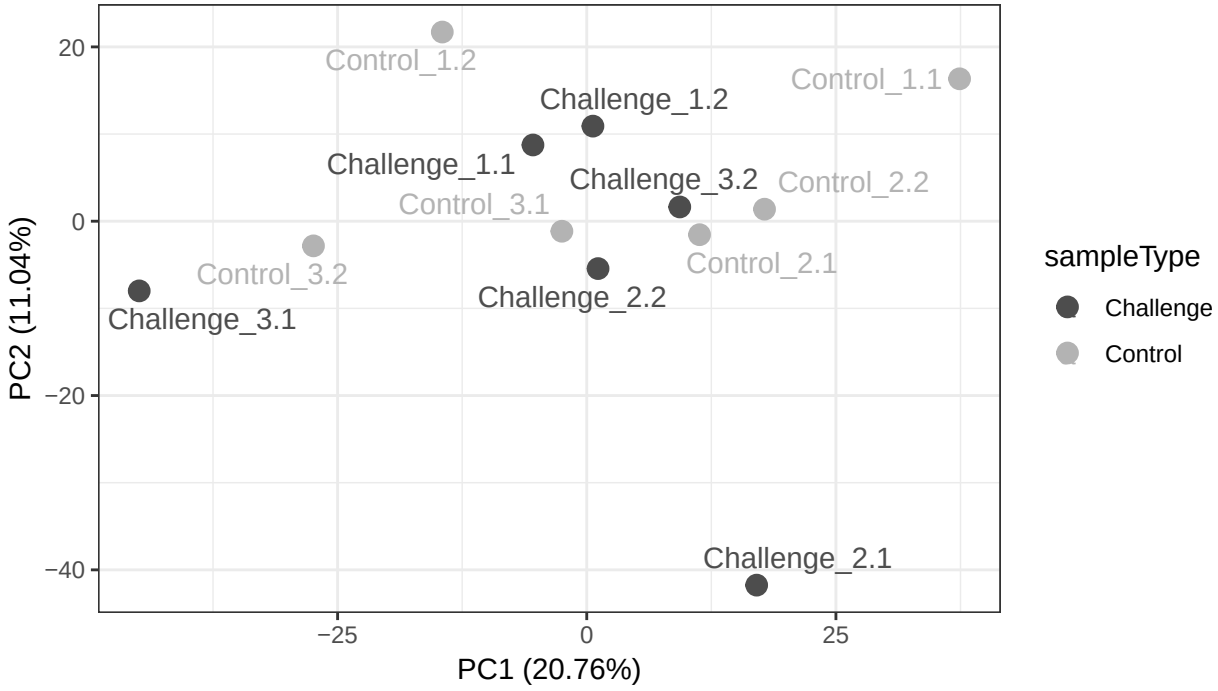


Figure S2. Principle component analysis of averaged technical replicates using the normalized site count method.