

Supplementary Tables for  
“Differential expression analysis of multifactor RNA-Seq  
experiments with respect to biological variation”

Davis J. McCarthy, Yunshun Chen and Gordon K. Smyth

November 15, 2011

**List of Tables**

1	Top ten DE genes . . . . .	2
2	DE results for Yu gene set . . . . .	3
3	Counts per million for Yu gene set . . . . .	4
4	DE results for Tuch gene set . . . . .	5
5	Counts per million for Tuch gene set . . . . .	6
6	DE results for heterogeneous tumour effect . . . . .	7
7	Top 50 up-regulated gene sets in mean-rank gene-set enrichment test using the MSigDB database . . . . .	8
8	Top 50 down-regulated gene sets in mean-rank gene-set enrichment test using the MSigDB database . . . . .	9
9	Top 50 GO terms enriched for up-regulated DE genes . . . . .	10
10	Top 50 GO terms enriched for down-regulated DE genes . . . . .	11

Table 1: The ten most significantly DE genes from the **edgeR** analysis of the carcinoma data. “log-Conc” refers to “log concentration” our measure of the gene’s overall expression level, “logFC” refers to the log of the fold change difference in expression between tumour and normal samples (positive value means higher expression in tumour tissue), “LR” is the likelihood-ratio statistic, “PValue” is the raw  $P$  value from the LR test and “FDR” is the false discovery rate.

Gene Symbol	logConc	logFC	PValue	FDR
IL1F6	-10.06	-6.14	3.86e-23	4.04e-19
PYGM	-9.65	-5.47	1.88e-22	9.81e-19
TMPRSS11B	-8.51	-7.40	7.51e-22	2.62e-18
PTGFR	-10.52	-5.19	1.40e-21	3.65e-18
MAL	-8.95	-6.88	5.69e-20	1.19e-16
MYBPC2	-9.30	-5.46	7.08e-20	1.23e-16
KRT3	-9.30	-5.82	9.73e-20	1.46e-16
ATP2A1	-9.67	-4.62	1.40e-19	1.83e-16
ENO3	-9.42	-5.17	9.94e-19	1.16e-15
IGF1	-9.84	-3.99	3.23e-18	3.31e-15

Table 2: The edgeR DE results for the important head and neck squamous cell carcinoma genes as identified by Yu et al. ?. Column headings as per Table 1, omitting “logConc” and with the addition of the final column “DE Rank” which gives the ranking for DE for the gene in the whole dataset.

Gene	logFC	PValue	FDR	DE Rank
MAL	-6.88	5.69e-20	1.19e-16	5
SPP1	5.23	6.68e-15	2.06e-12	34
MMP12	4.32	4.98e-13	7.44e-11	70
LAMC2	3.79	8.29e-10	4.90e-08	177
ECM1	-2.67	2.27e-07	7.84e-06	303
TGM3	-4.50	9.59e-07	2.91e-05	345
MMP1	3.63	1.04e-06	3.11e-05	349
KRT4	-7.17	1.57e-06	4.34e-05	376
MMP10	2.98	7.84e-06	1.83e-04	449
EMP1	-2.91	3.24e-05	6.32e-04	536
KRT13	-5.59	6.51e-05	1.18e-03	579
PLAU	1.62	8.66e-05	1.52e-03	596
MMP3	2.02	2.44e-04	3.72e-03	687
IL8	2.43	2.84e-04	4.22e-03	703
POSTN	1.65	5.75e-04	7.57e-03	795
IFI6	1.36	5.36e-03	4.52e-02	1242
ITGA6	1.84	5.89e-03	4.86e-02	1270
KRT17	1.92	6.92e-03	5.50e-02	1317
TNC	1.19	1.05e-01	3.72e-01	2967
FN1	1.17	1.32e-01	4.22e-01	3266
COL5A2	0.56	4.55e-01	7.70e-01	6176
COL1A2	0.55	5.74e-01	8.42e-01	7131
SPARC	-0.17	7.93e-01	9.38e-01	8847
COL4A1	0.16	8.50e-01	9.59e-01	9274
KRT5	0.05	9.59e-01	9.91e-01	10125

Table 3: The counts per million for the important head and neck squamous cell carcinoma genes as identified by Yu et al. ?. Columns refer to the patient samples to which the counts correspond.

	N8	N33	N51	T8	T33	T51
MAL	299	378	80	0	20	0
SPP1	2	3	7	48	637	95
MMP12	1	3	1	35	79	6
LAMC2	35	36	64	570	337	1115
ECM1	333	698	115	28	105	36
TGM3	2865	2225	583	191	332	4
MMP1	5	6	397	159	116	1170
KRT4	8477	9498	2165	45	1533	2
MMP10	1	5	13	10	6	343
EMP1	1374	2123	771	292	143	126
KRT13	43661	34482	9678	291	7872	103
PLAU	27	28	89	81	96	251
MMP3	5	4	74	28	40	85
IL8	3	9	52	54	86	43
POSTN	21	51	151	112	97	455
IFI6	49	17	15	72	29	109
ITGA6	263	193	207	944	331	1553
KRT17	242	269	551	2211	706	1299
TNC	240	148	1292	458	938	1323
FN1	213	238	1957	543	1254	1606
COL5A2	144	201	1645	710	266	851
COL1A2	353	830	5878	2692	983	2328
SPARC	246	288	1492	447	223	750
COL4A1	207	356	1952	1030	351	589
KRT5	10501	9388	4115	7389	5975	10598

Table 4: DE results from **edgeR** for the nine genes of interest identified by Tuch et al. ?. Column headings as per Table 1 with the added final column “DE Rank” giving the ranking of the gene for DE with respect to the whole dataset.

Gene	logConc	logFC	PValue	FDR	DE Rank
CASQ1	-10.40	-6.742	8.38e-13	1.18e-10	74
INHBA	-9.84	3.624	1.33e-10	9.65e-09	144
MMP1	-8.08	3.626	1.04e-06	3.11e-05	349
SHANK2	-10.34	3.131	1.59e-06	4.37e-05	380
HMGA2	-10.18	3.825	1.38e-05	3.06e-04	474
WIF1	-11.97	-3.596	7.94e-05	1.41e-03	590
CCND1	-7.73	1.310	6.18e-02	2.70e-01	2401
CTTN	-7.58	1.006	1.78e-01	4.94e-01	3776
CDKN2B	-9.42	-0.475	6.66e-01	8.82e-01	7899

Table 5: Counts per million for the nine genes deemed to be of particular interest by Tuch et al. (?), rounded to nearest integer. Library sizes are taken to be the effective library sizes obtained by applying the TMM normalization factor to account for compositional bias between samples (?). Counts per million facilitate fairer comparisons of counts between libraries by eye and have been rounded to the nearest whole number.

Gene	N8	N33	N51	T8	T33	T51
CASQ1	58	82	38	0	0	5
INHBA	3	5	33	81	77	120
MMP1	5	6	397	159	116	1170
SHANK2	3	5	3	152	17	15
HMGA2	1	2	4	182	30	8
WIF1	9	20	3	5	1	0
CCND1	153	192	73	1939	139	141
CTTN	167	152	151	2326	114	166
CDKN2B	62	79	30	0	75	242

Table 6: DE results from **edgeR** for the heterogeneous tumour effect for the three genes of interest identified by Tuch et al. ? that were *not* found to be DE with the additive model (homogeneous tumour effect). Column headings as per Table 1, but with “logConc” omitted, “logFC Patient 33” indicating the log-fold change for the tumour effect for Patient 33 (relative to baseline, or homogeneous, tumour effect), “logFC Patient 51” indicating the log-fold change for the tumour effect for Patient 51 and the added final column “DE Rank” giving the ranking of the gene for DE with respect to the whole dataset.

Gene	logFC Patient 33	logFC Patient 51	PValue	FDR	DE Rank
CDKN2B	7.25	10.3	4.93e-13	5.16e-09	1
CTTN	-4.22	-3.66	2.78e-03	1.00e-01	290
CCND1	-4.14	-2.72	3.49e-03	1.19e-01	307

For Peer Review

Table 7: Top 50 of 417 up-regulated gene sets in mean-rank gene-set enrichment test using the MSigDB database.

GeneSet	P-Value	FDR
DODD NASOPHARYNGEAL CARCINOMA DN	9.85e-49	3.22e-45
CAIRO HEPATOBLASTOMA CLASSES UP	3.8e-44	6.21e-41
SHEDDEN LUNG CANCER POOR SURVIVAL A6	2.86e-42	3.12e-39
ROSTY CERVICAL CANCER PROLIFERATION CLUSTER	1.79e-41	1.46e-38
PUJANA BRCA1 PCC NETWORK	6.26e-40	4.09e-37
KINSEY TARGETS OF EWSR1 FLII FUSION UP	1.26e-38	6.88e-36
PUJANA CHEK2 PCC NETWORK	2.8e-38	1.31e-35
KOBAYASHI EGFR SIGNALING 24HR DN	3.87e-38	1.58e-35
VECCHI GASTRIC CANCER EARLY UP	6.53e-37	2.38e-34
RODRIGUES THYROID CARCINOMA POORLY DIFFERENTIATED UP	1.62e-32	5.31e-30
SOTIRIOU BREAST CANCER GRADE 1 VS 3 UP	5.92e-32	1.76e-29
WONG EMBRYONIC STEM CELL CORE	2.36e-29	6.43e-27
WEI MYCN TARGETS WITH E BOX	3.3e-28	8.3e-26
BERENJENO TRANSFORMED BY RHOA UP	8.59e-28	2.01e-25
FOURNIER ACINAR DEVELOPMENT LATE 2	2.75e-27	6.01e-25
RODRIGUES THYROID CARCINOMA ANAPLASTIC UP	4.87e-25	9.96e-23
BASAKI YBX1 TARGETS UP	3.43e-24	6.61e-22
MANALO HYPOXIA DN	2.62e-23	4.76e-21
WHITEFORD PEDIATRIC CANCER MARKERS	1.09e-20	1.88e-18
PUJANA BRCA2 PCC NETWORK	1.39e-20	2.25e-18
CHIANG LIVER CANCER SUBCLASS PROLIFERATION UP	1.45e-20	2.25e-18
BENPORATH PROLIFERATION	2.18e-20	3.25e-18
NUYTEN EZH2 TARGETS DN	2.3e-19	3.28e-17
NIKOLSKY BREAST CANCER 8Q23 Q24 AMPLICON	1.09e-18	1.49e-16
CROONQUIST IL6 DEPRIVATION DN	3.57e-17	4.67e-15
RHODES UNDIFFERENTIATED CANCER	1.74e-16	2.19e-14
LINDGREN BLADDER CANCER CLUSTER 3 UP	2.44e-16	2.93e-14
HORIUCHI WTAP TARGETS DN	2.51e-16	2.93e-14
ODONNELL TFRC TARGETS DN	3.95e-16	4.45e-14
KRIGE RESPONSE TO TOSEDOSTAT 24HR DN	8.04e-16	8.77e-14
MUELLER PLURINET	9.02e-16	9.52e-14
GRAHAM NORMAL QUIESCENT VS NORMAL DIVIDING DN	1.27e-15	1.3e-13
LIAO METASTASIS	1.81e-15	1.8e-13
KANG DOXORUBICIN RESISTANCE UP	2.13e-15	2.05e-13
BOYALT LIVER CANCER SUBCLASS G3 UP	2.22e-15	2.07e-13
GRADE COLON CANCER UP	2.46e-15	2.23e-13
LI WILMS TUMOR VS FETAL KIDNEY 1 DN	3.63e-15	3.21e-13
WINTER HYPOXIA UP	5.55e-15	4.78e-13
LOCKWOOD AMPLIFIED IN LUNG CANCER	5.96e-15	4.93e-13
GRAHAM CML DIVIDING VS NORMAL QUIESCENT UP	6.03e-15	4.93e-13
RUIZ TNC TARGETS DN	1.08e-14	8.63e-13
MITSIADES RESPONSE TO APLIDIN DN	2.47e-14	1.92e-12
BORCZUK MALIGNANT MESOTHELIOMA UP	2.71e-14	2.06e-12
NAKAYAMA SOFT TISSUE TUMORS PCA2 UP	3.17e-14	2.36e-12
WEST ADRENOCORTICAL TUMOR UP	7.99e-14	5.81e-12
REACTOME CELL CYCLE MITOTIC	8.97e-14	6.38e-12
WINNEPENNINGKX MELANOMA METASTASIS UP	1.07e-13	7.46e-12
CASORELLI ACUTE PROMYELOCYTIC LEUKEMIA DN	1.52e-13	1.04e-11
CROONQUIST NRAS SIGNALING DN	2.22e-13	1.49e-11
KRIGE RESPONSE TO TOSEDOSTAT 6HR DN	2.95e-13	1.93e-11



Table 8: Top 50 of 268 down-regulated gene sets in mean-rank gene-set enrichment test using the MSigDB database.

GeneSet	P-Value	FDR
SMID BREAST CANCER NORMAL LIKE UP	6.17e-54	2.02e-50
DODD NASOPHARYNGEAL CARCINOMA UP	7.09e-42	1.16e-38
VECCHI GASTRIC CANCER EARLY DN	1.31e-36	1.43e-33
RODRIGUES THYROID CARCINOMA ANAPLASTIC DN	1.02e-30	8.3e-28
SWEET LUNG CANCER KRAS DN	5.05e-30	3.3e-27
RICKMAN HEAD AND NECK CANCER F	1.41e-28	7.7e-26
LINDGREN BLADDER CANCER CLUSTER 2B	4.89e-28	2.29e-25
ONDER CDH1 TARGETS 2 UP	9.98e-28	4.08e-25
DELYS THYROID CANCER DN	3.59e-26	1.31e-23
SABATES COLORECTAL ADENOMA DN	1.69e-24	5.52e-22
PICCALUGA ANGIOIMMUNOBLASTIC LYMPHOMA UP	1.9e-23	5.67e-21
RODRIGUES THYROID CARCINOMA POORLY DIFFERENTIATED DN	1.64e-22	4.47e-20
WEST ADRENOCORTICAL TUMOR DN	7.03e-21	1.77e-18
SCHUETZ BREAST CANCER DUCTAL INVASIVE UP	1.25e-19	2.92e-17
RICKMAN HEAD AND NECK CANCER E	2.67e-18	5.82e-16
CROMER TUMORIGENESIS DN	8.73e-18	1.79e-15
SMID BREAST CANCER BASAL DN	2.05e-17	3.95e-15
CAIRO HEPATOBLASTOMA CLASSES DN	1.07e-16	1.95e-14
SMID BREAST CANCER LUMINAL A UP	3.65e-16	6.28e-14
WINTER HYPOXIA DN	9.6e-16	1.57e-13
LIU PROSTATE CANCER DN	6.57e-15	1.02e-12
REN ALVEOLAR RHABDOMYOSARCOMA UP	4e-13	5.96e-11
NAKAYAMA SOFT TISSUE TUMORS PCA2 DN	6.32e-13	8.99e-11
REACTOME STRIATED MUSCLE CONTRACTION	4.27e-12	5.82e-10
EBAUER TARGETS OF PAX3 FOXO1 FUSION UP	6.86e-12	8.98e-10
BENPORATH ES WITH H3K27ME3	1.75e-11	2.2e-09
DEURIG T CELL PROLYMPHOCYTIC LEUKEMIA DN	1.93e-11	2.34e-09
REACTOME MUSCLE CONTRACTION	4.96e-11	5.8e-09
EBAUER MYOGENIC TARGETS OF PAX3 FOXO1 FUSION	6.72e-11	7.58e-09
WALLACE PROSTATE CANCER RACE UP	7.91e-11	8.63e-09
ODONNELL TFRC TARGETS UP	1.4e-10	1.48e-08
RIGGI EWING SARCOMA PROGENITOR UP	1.81e-10	1.85e-08
WU SILENCED BY METHYLATION IN BLADDER CANCER	3.48e-10	3.45e-08
KUNINGER IGF1 VS PDGFB TARGETS UP	1.15e-09	1.11e-07
FULCHER INFLAMMATORY RESPONSE LECTIN VS LPS DN	2.13e-09	1.99e-07
GOZGIT ESR1 TARGETS DN	2.49e-09	2.26e-07
SHEDDEN LUNG CANCER GOOD SURVIVAL A4	4.29e-09	3.8e-07
LEE DIFFERENTIATING T LYMPHOCYTE	6.4e-09	5.51e-07
KINSEY TARGETS OF EWSR1 FLII FUSION DN	7.11e-09	5.97e-07
THUM SYSTOLIC HEART FAILURE UP	7.68e-09	6.29e-07
HOSHIDA LIVER CANCER SUBCLASS S3	8.56e-09	6.83e-07
RUTELLA RESPONSE TO HGF DN	9.18e-09	7.15e-07
ACEVEDO METHYLATED IN LIVER CANCER DN	9.86e-09	7.46e-07
SOTIRIOU BREAST CANCER GRADE 1 VS 3 DN	1e-08	7.46e-07
ACEVEDO LIVER CANCER WITH H3K27ME3 UP	1.11e-08	8.09e-07
YAO HOXA10 TARGETS VIA PROGESTERONE UP	1.54e-08	1.1e-06
THUM SYSTOLIC HEART FAILURE DN	1.73e-08	1.21e-06
BOYVAULT LIVER CANCER SUBCLASS G3 DN	1.92e-08	1.31e-06
RUIZ TNC TARGETS UP	2.09e-08	1.4e-06
BOQUEST STEM CELL CULTURED VS FRESH DN	2.39e-08	1.57e-06

Table 9: Top 50 of 146 GO terms enriched for up-regulated DE genes using gene ontology analysis.

GOID	Term	P-Value
GO:0032501	multicellular organismal process	1.19e-17
GO:0007275	multicellular organismal development	2.65e-14
GO:0032502	developmental process	6.9e-14
GO:0048731	system development	2.09e-13
GO:0048856	anatomical structure development	1.45e-12
GO:0009653	anatomical structure morphogenesis	2.09e-11
GO:0030154	cell differentiation	2.23e-10
GO:0005578	proteinaceous extracellular matrix	3.56e-10
GO:0031012	extracellular matrix	4.92e-10
GO:0009888	tissue development	4.98e-10
GO:0048869	cellular developmental process	8.61e-10
GO:0007399	nervous system development	9.32e-10
GO:0048513	organ development	2.58e-09
GO:0009887	organ morphogenesis	4.72e-09
GO:0007155	cell adhesion	7.94e-09
GO:0022610	biological adhesion	7.94e-09
GO:0031226	intrinsic to plasma membrane	9.25e-09
GO:0009605	response to external stimulus	1.5e-08
GO:0044421	extracellular region part	1.61e-08
GO:0005887	integral to plasma membrane	1.62e-08
GO:0040011	locomotion	2.06e-08
GO:0048468	cell development	4.18e-08
GO:0004872	receptor activity	4.8e-08
GO:0022008	neurogenesis	4.91e-08
GO:0030182	neuron differentiation	6.95e-08
GO:0031224	intrinsic to membrane	1.09e-07
GO:0004871	signal transducer activity	1.09e-07
GO:0060089	molecular transducer activity	1.09e-07
GO:0016021	integral to membrane	1.94e-07
GO:0008544	epidermis development	1.95e-07
GO:0048729	tissue morphogenesis	2.72e-07
GO:0042221	response to chemical stimulus	2.92e-07
GO:0005886	plasma membrane	3.17e-07
GO:0048699	generation of neurons	3.25e-07
GO:0005576	extracellular region	3.68e-07
GO:0071944	cell periphery	4.07e-07
GO:0007409	axonogenesis	4.64e-07
GO:0016020	membrane	5.32e-07
GO:0048667	cell morphogenesis involved in neuron differentiation	6.2e-07
GO:0050896	response to stimulus	6.67e-07
GO:0048666	neuron development	7.15e-07
GO:0048812	neuron projection morphogenesis	7.56e-07
GO:0000904	cell morphogenesis involved in differentiation	7.81e-07
GO:0031175	neuron projection development	1.25e-06
GO:0044425	membrane part	1.33e-06
GO:0009790	embryo development	1.63e-06
GO:0044459	plasma membrane part	2.01e-06
GO:0048562	embryonic organ morphogenesis	3.54e-06
GO:0005102	receptor binding	3.56e-06
GO:0048598	embryonic morphogenesis	4.84e-06

Table 10: Top 50 of 264 GO terms enriched for down-regulated DE genes using gene ontology analysis.

GOID	Term	P-Value
GO:0005576	extracellular region	6.66e-37
GO:0003012	muscle system process	3.37e-33
GO:0043292	contractile fiber	5.36e-31
GO:0006936	muscle contraction	2.31e-30
GO:0030016	myofibril	2.22e-29
GO:0044449	contractile fiber part	1.77e-28
GO:0003008	system process	1.37e-26
GO:0044421	extracellular region part	4.91e-26
GO:0030017	sarcomere	1.07e-25
GO:0032501	multicellular organismal process	4.06e-23
GO:0005886	plasma membrane	5.96e-20
GO:0071944	cell periphery	8.79e-20
GO:0005615	extracellular space	9.49e-20
GO:0030049	muscle filament sliding	2.61e-19
GO:0033275	actin-myosin filament sliding	2.61e-19
GO:0070252	actin-mediated cell contraction	2.61e-19
GO:0048856	anatomical structure development	7.43e-17
GO:0007517	muscle organ development	6.77e-16
GO:0030029	actin filament-based process	3.51e-15
GO:0061061	muscle structure development	3.97e-15
GO:0048731	system development	5.16e-15
GO:0031674	I band	3.04e-14
GO:0031012	extracellular matrix	3.18e-14
GO:0030048	actin filament-based movement	5.77e-14
GO:0044459	plasma membrane part	6.32e-14
GO:0044057	regulation of system process	7.75e-14
GO:0032502	developmental process	9.87e-14
GO:0030246	carbohydrate binding	1.26e-13
GO:0001871	pattern binding	1.38e-13
GO:0030247	polysaccharide binding	1.38e-13
GO:0007155	cell adhesion	2.54e-13
GO:0022610	biological adhesion	2.54e-13
GO:0008307	structural constituent of muscle	2.96e-13
GO:0048513	organ development	3.08e-13
GO:0006941	striated muscle contraction	3.81e-13
GO:0007275	multicellular organismal development	3.86e-13
GO:0005539	glycosaminoglycan binding	7.92e-13
GO:0030154	cell differentiation	1.14e-12
GO:0090257	regulation of muscle system process	1.29e-12
GO:0009653	anatomical structure morphogenesis	1.52e-12
GO:0009888	tissue development	2.34e-12
GO:0003013	circulatory system process	4.44e-12
GO:0008015	blood circulation	4.44e-12
GO:0048869	cellular developmental process	4.57e-12
GO:0005578	proteinaceous extracellular matrix	4.84e-12
GO:0008201	heparin binding	2.75e-11
GO:0014706	striated muscle tissue development	5.67e-11
GO:0006937	regulation of muscle contraction	7.82e-11
GO:0072376	protein activation cascade	1.19e-10
GO:0040011	locomotion	1.27e-10

## References

- Yu, Y.H., Kuo, H.K., Chang, K.W. and Rutherford, S. (2008) The evolving transcriptome of head and neck squamous cell carcinoma: A systematic review. *PLoS ONE*, **3**(9), e3215.
- Tuch, B.B., Laborde, R.R., Xu, X., Gu, J., Chung, C.B., Monighetti, C.K., Stanley, S.J., Olsen, K.D., Kasperbauer, J.L., Moore, E.J., Broomer, A.J., Tan, R., Brzoska, P.M., Muller, M.W., Siddiqui, A.S., Asmann, Y.W., Sun, Y., Kuersten, S., Barker, M.A., Vega, F.M.D.L. and Smith, D.I. (2010) Tumor transcriptome sequencing reveals allelic expression imbalances associated with copy number alterations. *PLoS ONE*, **5**(2), e9317.
- Robinson, M.D. and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.*, **11**(3), R25.

For Peer Review