nature | **methods**

Cell type–specific gene expression differences in complex tissues

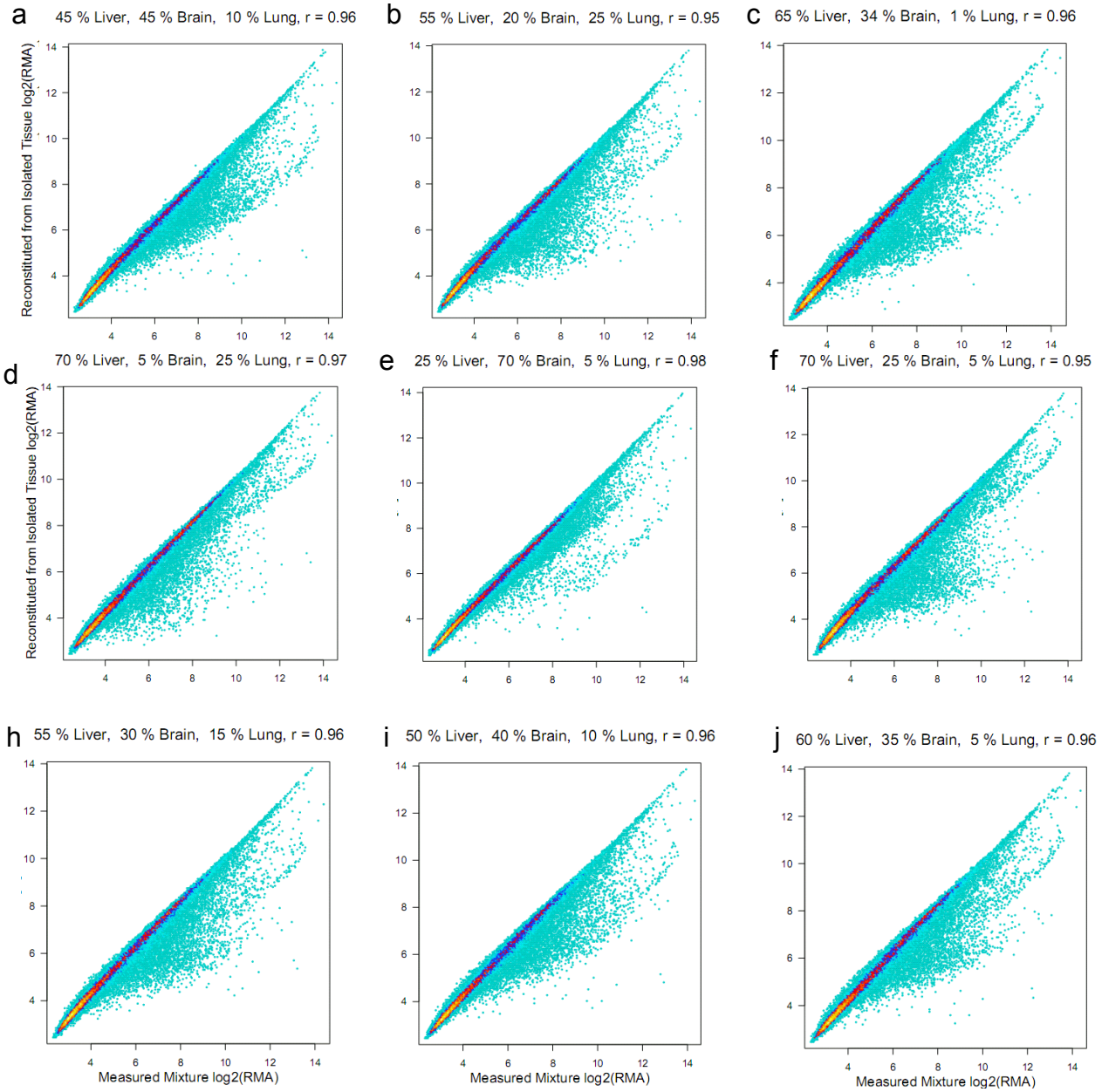
Shai S Shen-Orr, Robert Tibshirani, Purvesh Khatri, Dale L Bodian, Frank Staedtler, Nicholas M Perry, Trevor Hastie, Minnie M Sarwal, Mark M Davis & Atul J Butte

Supplementary figures and text:

Supplementary Figure 1	Total measured gene expression in a heterogeneous tissue sample is the sum of gene expression in individual tissue subsets.
Supplementary Figure 2	Fraction of probes misestimated by statistical deconvolution as a function of fold change difference from their measured abundance in pure tissue.
Supplementary Figure 3	Off diagonal probe effects are reproducible and are not due to biological inter-tissue interactions but rather to technical considerations.
Supplementary Figure 4	Differential expression analysis by statistical deconvolution is specific and sensitive.
Supplementary Figure 5	csSAM reveals cell-type specific differential expression undetectable at heterogeneous tissue level from whole genome data.
Supplementary Figure 6	Restricting csSAM analysis to reduce the number of multiple hypothesis tested and separately assaying for up and down regulation yields improved performance of csSAM.
Supplementary Figure 7	Differential expression analysis of adjusted cell-type subset frequency invariant whole blood samples.
Supplementary Figure 8	Differential expression analysis across all cell-type subset expression profiles. Deconvolution of group specific cell-type expression enables sophisticated differential expression analysis tests.
Supplementary Table 1	Experimental design for rat microarray experiment.
Supplementary Table 2	Off diagonal probes are highly overlapping between mixtures.
Supplementary Table 3	Cell-type subset frequency and abundance for each kidney transplant patient.
Supplementary Note 1	Determination of source and downstream affects of off-diagonal probes in rat experiment.
Supplementary Note 2	Justification for treating adjusted data as real

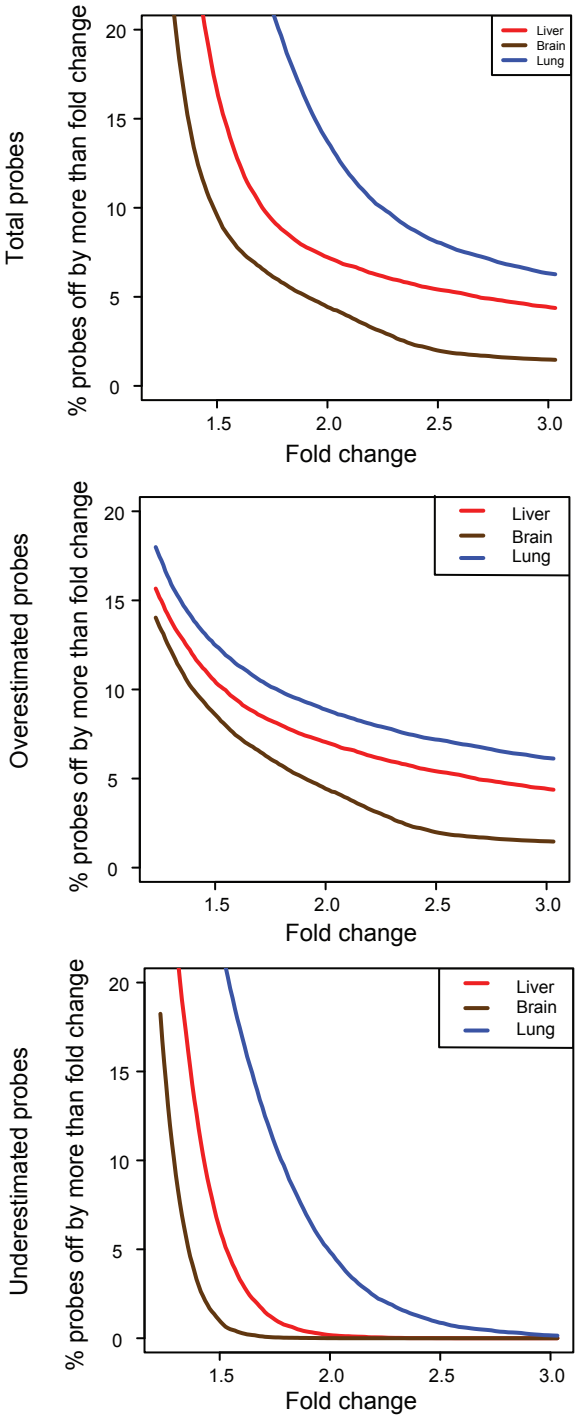
Note: The Supplementary Data file is available on the Nature Methods website.

Supplementary Figure 1 Total measured gene expression in a heterogeneous tissue sample is the sum of gene expression in individual tissue subsets. (**a-j**) Density plots of measured gene expression from a mixed sample plotted against a reconstituted expression pattern simulated from pure tissue samples multiplied by the frequency of the pure tissue in the sample. Color represents point density from a single probe (cyan) to 100 probes (yellow). Mixture ratios and r values appear above each figure. A small fraction of probes are higher in measured mixtures than would be expected by their expression in pure tissues.



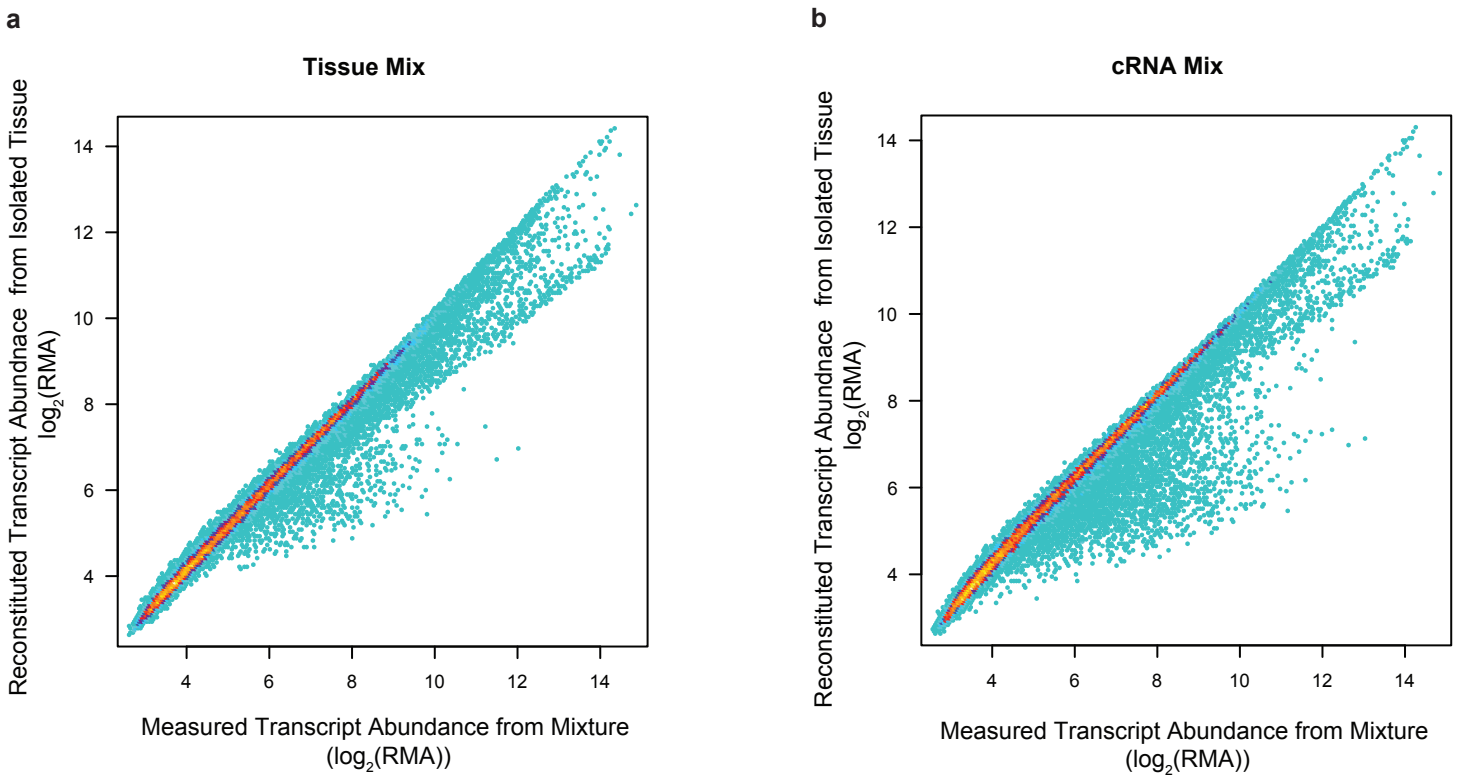
Nature Methods: doi: 10.1038/nmeth.1439

Supplementary Figure 2



Fraction of probes misestimated by statistical deconvolution as a function of fold change difference from their measured abundance in pure tissue. Triplicate measurements of pure tissue were averaged and compared to estimated expression profiles deconvolved from mixture data. Shown are (**a**) total (**b**) overestimated and (**c**) underestimated probes as a function of fold-change. The fraction of misestimated probes is higher in lung, the tissue subset whose frequency is across all samples never goes beyond 25%, and is lowest in brain, the tissue subset whose variance between samples is highest.

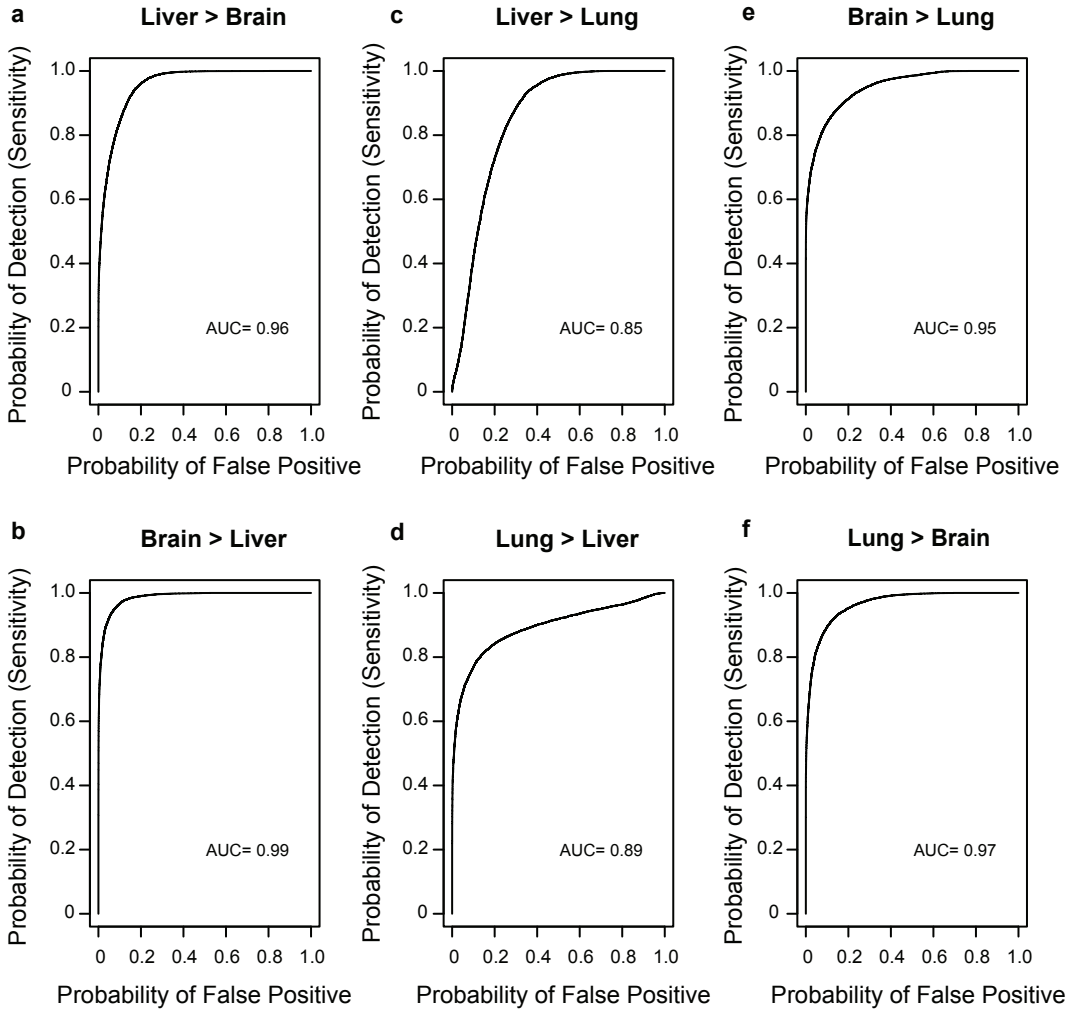
Supplementary Figure 3



Off diagonal probe effects are reproducible and are not due to biological inter-tissue interactions but rather to technical considerations.Shown are density plot comparisons of reconstituted versus measured sample mixtures of brain and liver tissues, for a single mixture 75% liver, 25% brain. Once when the mixing was performed at the (**a**) tissue level post homogenization, and once when performed at (**b**) the cRNA level. Though unlikely, inter-tissue biological interactions due to causes such as poor homogenization cannot be ruled out in tissue level mixtures, but should not occur at cRNA-level mixtures. Their reproducibility (over 50% overlap from one experiment to another) suggest that they are likely due to technical considerations such as probe cross-hybridization.

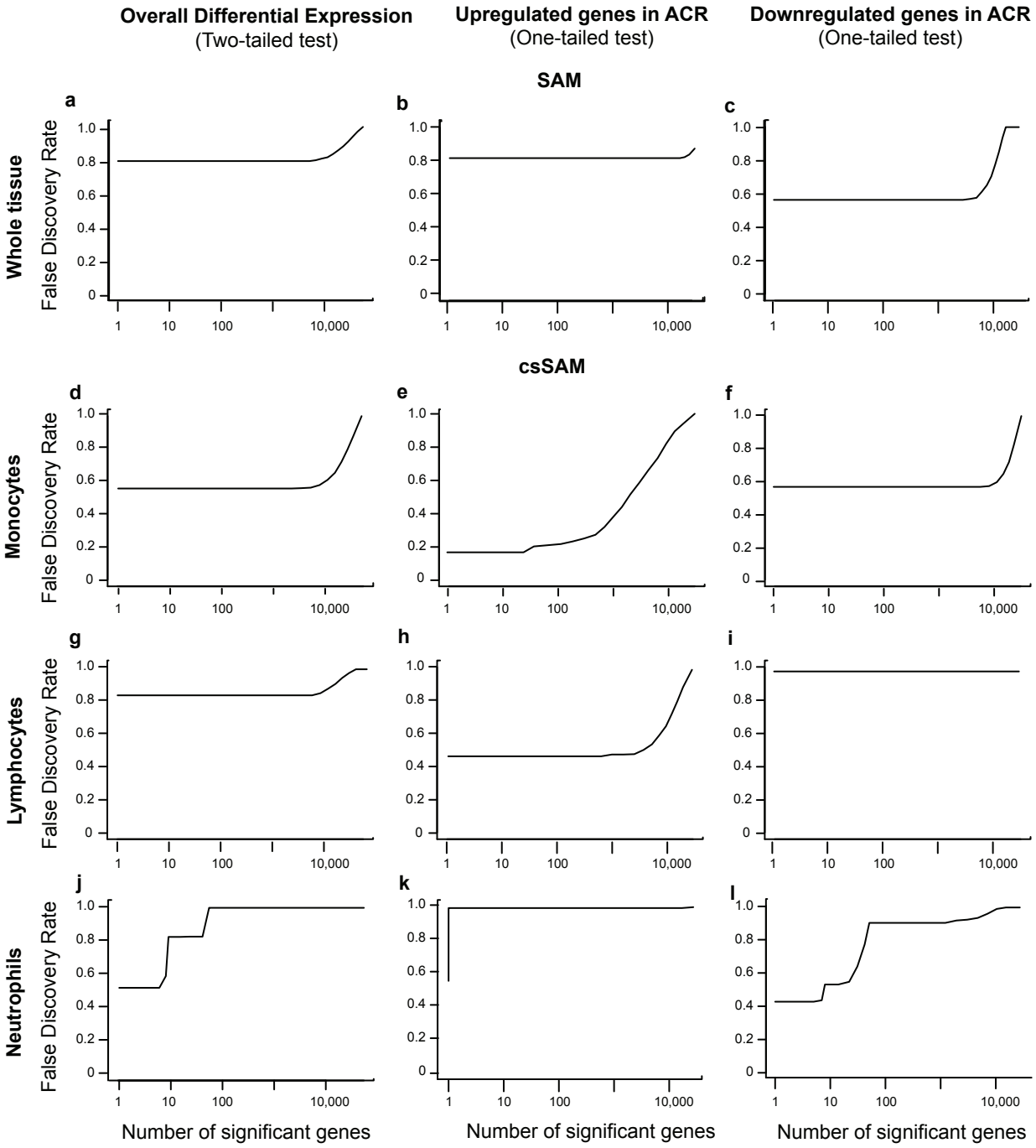
Nature Methods: doi: 10.1038/nmeth.1439

Supplementary Figure 4



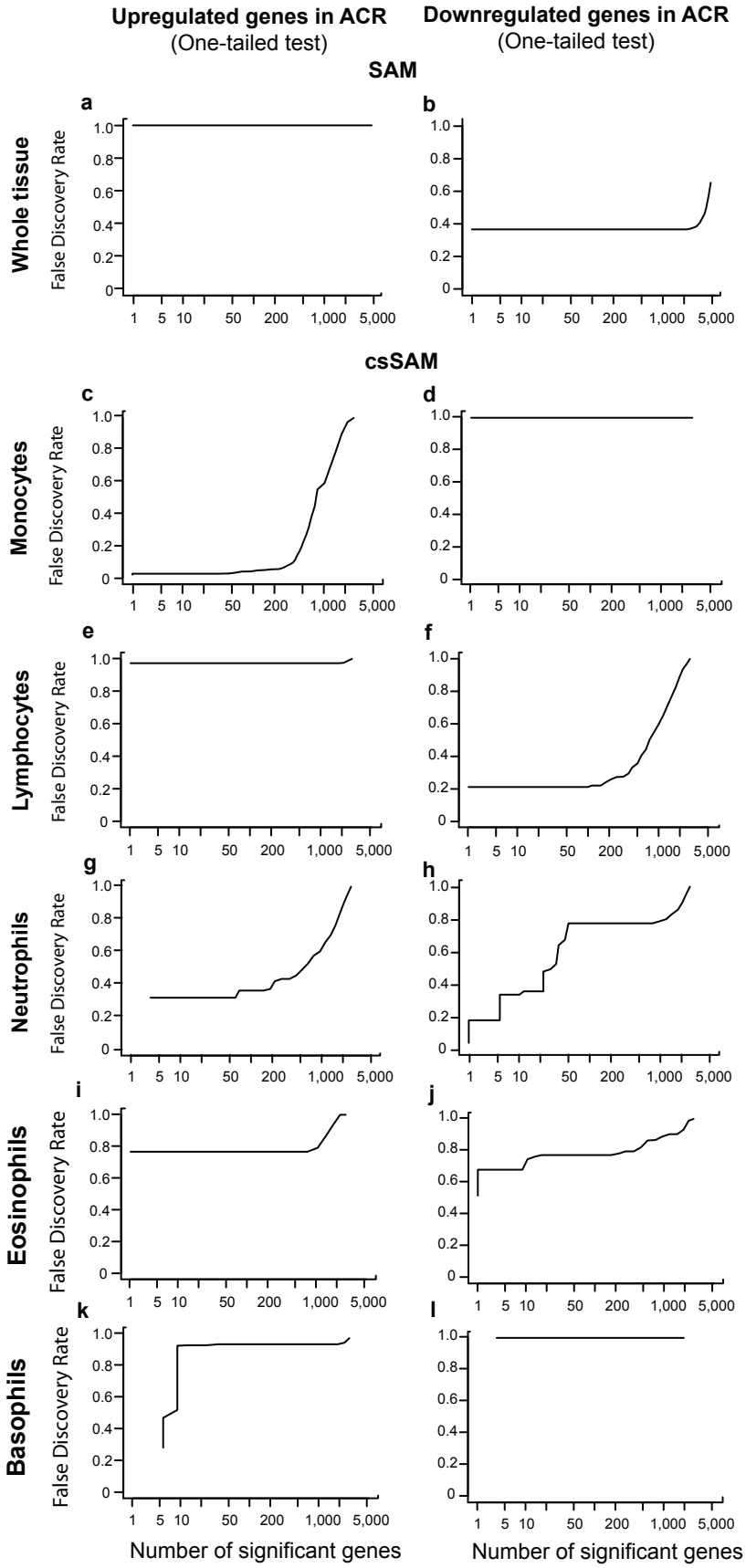
Differential expression analysis by statistical deconvolution is specific and sensitive. ROC curves of the ability to identify differentially expressed probes between two tissues by statistical deconvolution from mixed tissue samples. (**a-f**) For each comparison between tissues, a separate plot is given for detection of those probes upregulated in tissue 1 vs. 2 and vice versa.

Supplementary Figure 5



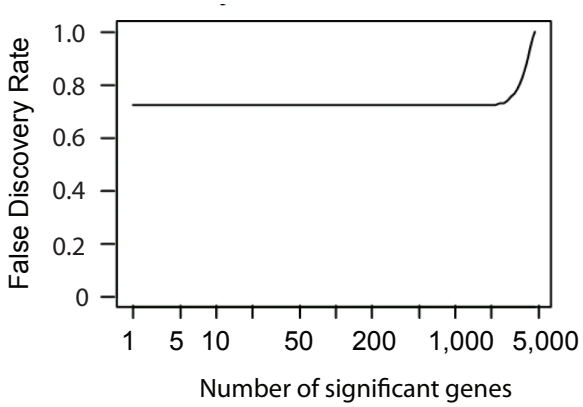
csSAM reveals cell-type specific differential expression undetectable at heterogeneous tissue level from whole genome data. Differential expression analysis between whole blood of acute rejection and stable kidney transplant patients was performed over the entire genome across all samples by a traditional microarray analysis technique (SAM) using both a (a) two-tailed and (b,c) one-tailed tests, but did not yield any genes differentially expressed at a reasonable FDR of 0.3. Deconvolution of cell-type subset expression profiles for each of the patients groups and their comparison via csSAM lowers FDRs but (e) identifies differentially expressed in only in a one-way test and only in monocytes. Varying the gene set size under consideration, in an unsupervised manner, can achieve much lower FDRs for several cell types.

Supplementary Figure 6



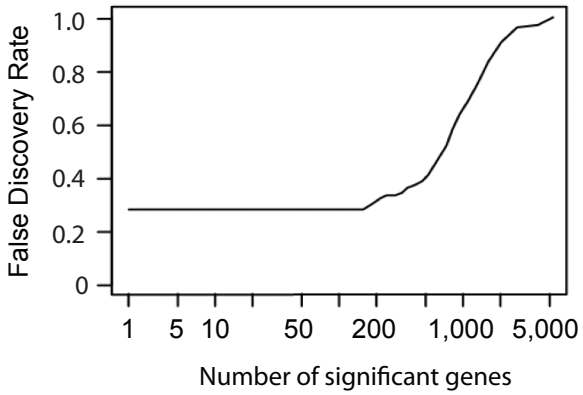
Restricting csSAM analysis to reduce the number of multiple hypothesis tested and separately assaying for up and down regulation yields improved performance of csSAM. Differential expression analysis between whole blood of acute rejection and stable kidney transplant patients was performed using one-tailed tests, over the top 5,000 most varying genes across all samples, by a traditional analysis technique (SAM) (**a,b**) but did not yield any genes differentially expressed at a reasonable FDR of 0.3. Yet, deconvolution of cell-type subset expression profiles for each of the patients groups and their comparison in a one-tailed test via csSAM for either upregulation or downregulation reveals an even larger number of differentially expressed genes (**c-h**) in monocytes, lymphocytes and neutrophils, masked by the variation in cell-type subset frequency variation in the original data. No gene expression differences are detectable in the basophils and eosinophils (**i-l**) of acute rejection and stable kidney transplant patients.

Supplementary Figure 7



Differential expression analysis of adjusted cell-type subset frequency invariant whole blood samples. Deconvolution of group specific cell-type expression enables sophisticated differential expression analysis tests. The adjustment test removes all the heterogeneity in the samples that is due to cell frequency variations, while maintaining sample individuality. Its application for analysis of differences between acute rejection and stable kidney transplant patients did not yield any differentially expressed genes at reasonable FDR.

Supplementary Figure 8



Differential expression analysis across all cell-type subset expression profiles. Deconvolution of group specific cell-type expression enables sophisticated differential expression analysis tests. The Omnibus Cell Test considers the total difference in expression of a gene between two groups by summing the observed squared differences across all of the deconvolved cell specific expression profiles. This strips the biological and technical variation between individual samples that is incorporated into the adjustment test and considers the expression of only those cells of interest. Application of the Omnibus Cell Test to our dataset, shows a reduction in FDR, down to 0.3, with more than a 100 genes are differentially expressed between the two patient groups.

Nature Methods: doi: 10.1038/nmeth.1439

Supplementary Table 1

Experimental design for rat microarray experiment. cRNA from brain, liver and lung tissues derived from a single rat were homogenaized, extracted and mixed in 13 different proportions three of which are each of the tissues in isolate (100% lung, 100% brain and 100% liver). The 10 other mixtures include RNA from each of the three tissues at varying proportions. Note: Mixture 4 samples was not used in the analysis, due to the high frequency of lung tissue

Exp #	FileName	Tissue Type	Liver	Brain	Lung	Mix Num.	Conc
1	NUID-0000-0094-1609.CEL	Pure	100	0	0	Mix 1_1	0.472
2	NUID-0000-0094-1620.CEL	Pure	100	0	0	Mix 1_2	0.472
3	NUID-0000-0094-1631.CEL	Pure	100	0	0	Mix 1_3	0.472
4	NUID-0000-0094-1642.CEL	Pure	0	100	0	Mix 2_1	0.473
5	NUID-0000-0094-1646.CEL	Pure	0	100	0	Mix 2_2	0.473
6	NUID-0000-0094-1647.CEL	Pure	0	100	0	Mix 2_3	0.473
7	NUID-0000-0094-1648.CEL	Pure	0	0	100	Mix 3_1	0.508
8	NUID-0000-0094-1649.CEL	Pure	0	0	100	Mix 3_2	0.508
9	NUID-0000-0094-1650.CEL	Pure	0	0	100	Mix 3_3	0.508
10	NUID-0000-0094-1610.CEL	Mixed	5	25	70	Mix 4_1	0.497
11	NUID-0000-0094-1611.CEL	Mixed	5	25	70	Mix 4_2	0.497
12	NUID-0000-0094-1612.CEL	Mixed	5	25	70	Mix 4_3	0.497
13	NUID-0000-0094-1613.CEL	Mixed	70	5	25	Mix 5_1	0.481
14	NUID-0000-0094-1614.CEL	Mixed	70	5	25	Mix 5_2	0.481
15	NUID-0000-0094-1615.CEL	Mixed	70	5	25	Mix 5_3	0.481
16	NUID-0000-0094-1616.CEL	Mixed	25	70	5	Mix 6_1	0.474
17	NUID-0000-0094-1617.CEL	Mixed	25	70	5	Mix 6_2	0.474
18	NUID-0000-0094-1618.CEL	Mixed	25	70	5	Mix 6_3	0.474
19	NUID-0000-0094-1619.CEL	Mixed	70	25	5	Mix 7_1	0.474
20	NUID-0000-0094-1621.CEL	Mixed	70	25	5	Mix 7_2	0.474
21	NUID-0000-0094-1622.CEL	Mixed	70	25	5	Mix 7_3	0.474
22	NUID-0000-0094-1623.CEL	Mixed	45	45	10	Mix 8_1	0.476
23	NUID-0000-0094-1624.CEL	Mixed	45	45	10	Mix 8_2	0.476
24	NUID-0000-0094-1625.CEL	Mixed	45	45	10	Mix 8_3	0.476
25	NUID-0000-0094-1626.CEL	Mixed	55	20	25	Mix 9_1	0.481
26	NUID-0000-0094-1627.CEL	Mixed	55	20	25	Mix 9_2	0.481
27	NUID-0000-0094-1628.CEL	Mixed	55	20	25	Mix 9_3	0.481
28	NUID-0000-0094-1629.CEL	Mixed	50	30	20	Mix 10_1	0.48
29	NUID-0000-0094-1630.CEL	Mixed	50	30	20	Mix 10_2	0.48
30	NUID-0000-0094-1632.CEL	Mixed	50	30	20	Mix 10_3	0.48
31	NUID-0000-0094-1633.CEL	Mixed	55	30	15	Mix 11_1	0.478
32	NUID-0000-0094-1634.CEL	Mixed	55	30	15	Mix 11_2	0.478
33	NUID-0000-0094-1635.CEL	Mixed	55	30	15	Mix 11_3	0.478
34	NUID-0000-0094-1636.CEL	Mixed	50	40	10	Mix 12_1	0.476
35	NUID-0000-0094-1637.CEL	Mixed	50	40	10	Mix 12_2	0.476
36	NUID-0000-0094-1638.CEL	Mixed	50	40	10	Mix 12_3	0.476
37	NUID-0000-0094-1639.CEL	Mixed	60	35	5	Mix 13_1	0.474
38	NUID-0000-0094-1640.CEL	Mixed	60	35	5	Mix 13_2	0.474
39	NUID-0000-0094-1641.CEL	Mixed	60	35	5	Mix 13_3	0.474
40	NUID-0000-0094-1643.CEL	Mixed	65	34	1	Mix 14_1	0.473
41	NUID-0000-0094-1644.CEL	Mixed	65	34	1	Mix 14_2	0.473
42	NUID-0000-0094-1645.CEL	Mixed	65	34	1	Mix 14_3	0.473

Supplementary Table 2

Off diagonal probes are due to technical reasons in experiment and not to statistical deconvolution. Overlap in off-diagonal probes between measured samples, deconvolution estimates of tissue expression and reconstituted mixtures. Shown are the number of off-diagonal probes that are shared between any two comparisons and their total percentage from the two experiments compared. Listed on the diagonal is the total number of off-diagonal probes for that experiment.

	Liver	Brain	Lung	Mix 5	Mix 6	Mix 7	Mix 8	Mix 9	Mix 10	Mix 11	Mix 12	Mix 13	Mix 14
Liver	1708	937 (55%,81%)	1291 (76%,58%)	1307 (77%,64%)	1554 (91%,71%)	1482 (87%,65%)	1431 (84%,65%)	1323 (77%,65%)	1246 (73%,61%)	1157 (68%,56%)	1026 (60%,46%)	862 (50%,63%)	1080 (63%,25%)
Brain		1159	973 (84%,44%)	1127 (97%,55%)	1045 (90%,47%)	1086 (94%,48%)	1061 (92%,48%)	1095 (94%,54%)	1032 (89%,50%)	963 (83%,46%)	676 (58%,30%)	1044 (90%,76%)	678 (58%,16%)
Lung			2229	1834 (82%,90%)	1860 (83%,84%)	1928 (86%,85%)	1982 (89%,90%)	1867 (84%,91%)	1972 (88%,96%)	1985 (89%,96%)	1887 (85%,84%)	1086 (49%,79%)	837 (38%,20%)
Mix 5				2048	1901 (93%,86%)	1980 (97%,87%)	1969 (96%,89%)	1958 (96%,96%)	1854 (91%,91%)	1767 (86%,85%)	1500 (73%,67%)	1246 (61%,90%)	914 (45%,21%)
Mix 6					2204	2115 (96%,93%)	2055 (93%,93%)	1922 (87%,94%)	1837 (83%,90%)	1748 (79%,84%)	1543 (70%,69%)	1165 (53%,85%)	1104 (50%,26%)
Mix 7						2263	2130 (94%,96%)	1987 (88%,97%)	1914 (85%,94%)	1828 (81%,88%)	1597 (71%,71%)	1221 (54%,89%)	1086 (48%,25%)
Mix 8							2213	1992 (90%,97%)	1958 (88%,96%)	1883 (85%,91%)	1655 (75%,74%)	1188 (54%,86%)	1004 (45%,24%)
Mix 9								2044	1886 (92%,92%)	1799 (88%,87%)	1536 (75%,68%)	1210 (59%,88%)	904 (44%,21%)
Mix 10									2046	1934 (95%,93%)	1676 (82%,75%)	1143 (56%,83%)	801 (39%,19%)
Mix 11										2075	1761 (85%,78%)	1084 (52%,79%)	729 (35%,17%)
Mix 12											2246	791 (35%,57%)	584 (26%,14%)
Mix 13												1378	734 (53%,17%)
Mix 14													4270

Supplementary Table 3

Cell-type subset frequency and abundance for each kidney transplant patient

Sample ID	Patient Group	Neutrophils %	Lymphocytes %	Monocytes %	Eosinophils %	Basophils %
B145	ACR	38	47.5	12.4	1.4	0.8
B196	ACR	70.7	17.1	11.8	0.1	0.3
B212	ACR	62.1	20.3	14.6	2.5	0.5
B246	ACR	90.5	2	4.8	2.4	0.3
B250	ACR	65.8	25	7.3	1.4	0.5
B286	ACR	39.2	37.1	14.1	9.2	0.4
B343	ACR	89.4	6.3	3.7	0.1	0.5
B355	ACR	91.3	1.6	4.6	2.5	0
B369	ACR	55.4	31.3	8.4	4.6	0.3
B386	ACR	14.5	61.5	21.9	1.3	0.8
B389	ACR	81	17.9	0.4	0.6	0.1
B565	ACR	53.8	37.2	8.4	0.4	0.2
B656	ACR	59.9	32.6	5.8	1.5	0.3
B726	ACR	77	18.5	4.1	0.2	0.2
B740	ACR	47.8	39.7	6.6	4.7	1.2
B022	Stable	76.4	19.4	3.6	0.4	0.2
B315	Stable	58.5	28	9.6	3.3	0.6
B342	Stable	60.5	27.8	8.3	2.8	0.6
B354	Stable	43.6	44	5.8	6.5	0
B401	Stable	63.4	23.1	11.6	1.6	0.3
B453	Stable	50.1	30.6	18.1	1.2	0
B514	Stable	54.8	15.2	27.8	1.6	0.6
B554	Stable	47.7	37.8	12.1	1.6	0.8
B737	Stable	30.4	52	9.4	7.7	0.5

Supplementary Note 1:

Determination of source and downstream affects of off-diagonal probes in rat experiment:

The primary purpose of our pre-designed mixture experiment was to determine the relationship and magnitude of the gene expression in mixed samples and the actual gene expression of the constituting cell-subsets. To date, applications of statistical deconvolution to infer cell-specific gene expression have assumed it to be linear[1-3]. However, there are many places where sequence dependent biases may be introduced experimental protocols and reported normalized expression values are relatively distant from actual transcript abundance in a sample, particularly in fluorescent based techniques such as microarrays. Thus, an assumption of linearity may not necessarily hold and if so, the implications for accurate cell-specific gene expression and more importantly, sensitive and specific differential expression analysis, are unknown.

We performed the pre-designed mixture ratio experiments a total of four times. The initial experiment (henceforth V1) comprised of two tissues only, liver and brain (data not shown). These were derived from a single rat and mixed at the tissue homogenate level in different proportions, 100% liver - 0% brain, 75% liver - 25% brain, 50% liver - 50% brain, 25% liver -75% brain, and finally 0% liver 100% brain, 3 technical replicates each. Snap frozen rat liver and brain was kept frozen while cutting it into pieces. The same protocol was followed for these biospecimens as discussed in the Methods section under the heading “Microarray analysis of rat brain, liver and lung”.

Comparison of the experimentally measured samples to the reconstituted ones in the V1 experiment showed overall high correlations (r=0.99,0.99 and 0.97 for 75% liver-25% brain, 50% liver-50% brain and 25% liver-75% brain, respectively) with only a small number of probes, of both mid- and high-expression level showing significant deviation from the diagonal. We identified off-diagonal probes, namely those whose change between simulated and measured mixtures was higher than two-fold. These constituted between 2.6 and 3.8% of the probes across the three mixtures and were more highly expressed in the measured mixtures than in the reconstituted samples from pure tissue. Importantly, they were highly overlapping with one another between mixtures though their abundance was affected by brain/liver ratio (**Supplementary Table 2**).

The probe deviation from the diagonal observed in our comparison of measured mixture to reconstituted samples in the V1 experiment could be due either to experimental design or error, technical artifacts, or inter-tissue effects. The two tissues in the V1 experiment were mixed at the tissue homogenate level. It is unexpected for inter-tissue biological interactions to occur between homogenized tissue mixtures. However, performing functional enrichment analysis of these off-diagonal probes (2-fold cutoff) using DAVID[4], showed that they were highly enriched for “extracellular localization” ($P < 10^{-23}$) and “response to stimuli” ($P < 10^{-10}$) functions, as well as for brain and liver tissue expression.

To identify whether this increased expression in mixtures is due to actual inter-tissue interaction as well as to rule out any experimental errors, we repeated the two-tissue mixture experiment both when the mixtures were performed at the tissue homogenate level (tissue-mixtures) as well as by extracting the RNA from each tissue separately and then mixing the cRNAs (cRNA-mixtures). In the cRNA-mixture case, no inter-tissue biological interaction is possible, whereas if the off-diagonal probes were due to an experimental error, than we would not expect to see them in either one of the new tissue and cRNA-mixtures.

Nature Methods: doi: 10.1038/nmeth.1439

Comparison of the new reconstituted samples from the tissue-mixture experiment to their corresponding measured mixture samples showed that a small fraction of probes were significantly more highly expressed in the measured mixtures than in the reconstituted samples, as was observed prior. These, "new" off-diagonal probes, highly overlapped with those probes identified as "off-diagonal" in the prior experiment. Analysis of the cRNA mixture samples showed similar results, again with a high fraction of the probes overlapping with the other experiments (**Supplementary Fig. 3**).

Taken together this suggests that the phenomenon of a fraction of probes significantly and consistently being more highly expressed in measured mixtures than in reconstituted samples, is unlikely to be due to experimental error nor to inter-tissue effects. More likely, this is due to a technical issue with microarray hybridization technology or normalization choices. The detection of off-diagonal probes at mid-level expression suggests that this is not due to probe saturation, but rather either to non-linear amplification of specific transcripts or to synergistic cross-hybridization of tissue specific transcripts in sample mixtures.

What are the effects of these probes on cell-type specific deconvolution? In all of these cases, comparison of measured isolated tissue to deconvolved estimates of the tissue-specific expression profile shows that these probes are overestimated by deconvolution. The variation in abundance and tissue-type proportion dependency of these probes between measured samples help buffer these effects to some extent from some of the deconvolved tissue specific expression profiles (For example, in the two tissue comparison, most of the off-diagonal probes are estimated as being highly associated with brain). Yet, fold change can be a very fragile, non-robust measure for downstream differential expression analysis, especially at low levels of expression [5] Using a t-statistic, our analysis of differences between estimated tissues compared to the gold-standard differences in measured tissues, shows very good ROC curves (**Supplementary Fig. 4**) for all tissue comparisons in all four of the pre-designed rat experiments. In agreement with what we observed in our analysis of deconvolution misestimated probes this suggests that many of the off-diagonal probes are, consistently and to a similar extent, misestimated in all tissues such that their net effect on differential expression analysis is negligible. Despite this, identification and exclusion of a core set of probes for which deconvolution does not yield accurate cell-type estimates, may be recommended.

1. R. O. Stuart, W. Wachsman, C. C. Berry et al., *Proc Natl Acad Sci U S A* **101** (2), 615 (2004).
2. M. Wang, S. R. Master, and L. A. Chodosh, *BMC Bioinformatics* **7**, 328 (2006).
3. H. Lahdesmaki, L. Shmulevich, V. Dunmire et al., *BMC Bioinformatics* **6**, 54 (2005).
4. G. Dennis, Jr., B. T. Sherman, D. A. Hosack et al., *Genome Biol* **4** (5), P3 (2003).
5. D. Witten and R. Tibshirani, A comparison of fold change and the t-statistic for microarray data analysis Available at http://www-stat.stanford.edu/~tibs/ftp/FCTComparison.pdf, (2007).

Nature Methods: doi: 10.1038/nmeth.1439

Supplementary Note 2:

Justification for treating adjusted data as real:

Consider the estimator $\hat{\theta}_j = \bar{c}_k^T r_{kj}$ where r_{kj} is the contrast between two deconvolved cell-specific expression profiles $r_{kj} = (\hat{h}^2_{kj} - \hat{h}^1_{kj})/\hat{s}e_{kj}$ and let $W_j = \hat{\theta}_j/\hat{\text{se}}(\theta_j)$ where the denominator is

$$\sqrt{\frac{n_1 \text{var}(\bar{w}_k^T \hat{h}^n_{kj}) + n_2 \text{var}(\bar{w}_k^T \hat{h}^w_{kj})}{(n1+n2)}}]$$

To show that it is appropriate to treat the adjusted data as real we prove that the T-test on the adjusted data, based on $\hat{X}_{ij}$ (adjusted expression, as defined in Methods), is numerically equivalent to the the Wald test statistic W_j.

Assume that the observations are ordered so that the first n_1 observations fall in group 1, and the remaining n_2 observations fall in group 2. Then $\hat{X}_j = W_j \hat{H}_j$ where $\hat{H}_j = (W_j^T W_j)^{-1} W_j^T X_j$.

Let $\mathbf{a}$ be a non-negative vector of length k with $\sum a_i = 1$. One obvious choice for $\mathbf{a}$ is the average of the rows of W.

The Wald test for the null hypothesis $\theta = \mathbf{a}^T(H_2 - H_1) = 0$ equals $\hat{\theta}/[\widehat{\text{var}}(\hat{\theta})]^{1/2}$ where

$$\begin{aligned} \hat{\theta} &= \mathbf{a}^T(\hat{H}_2 - \hat{H}_1) \\ \widehat{\text{var}}(\hat{\theta}) &= \mathbf{a}^T[(W_2^T W_2)^{-1} + (W_1^T W_1)^{-1}]\mathbf{a} \cdot \mathbf{v} \end{aligned}$$

Here $\mathbf{v} = (n_1 \hat{s}^2_{1j} + n_2 \hat{s}^2_{2j})/(n_1 + n_2)$, and $\hat{s}^2_{1j} = \sum_{i=1}^{n_1} (x_{ij} - \hat{x}_{ij})^2/(n_1 - k)$ and similarly for $\hat{s}^2_{2j}$.

For the T-test on the adjusted data we let

$$\tilde{X} = (\mathbf{1}_1 \mathbf{a}^T \hat{H}_1, \mathbf{1}_2 \mathbf{a}^T \hat{H}_2) + \left(R_1 P_{n_1 k}, R_2 P_{n_2 k}\right)$$

where $(R_1, R_2) = (X_1 - \hat{X}_1, X_2 - \hat{X}_2)$ and $(P_{n_1 k}, P_{n_2 k}) = (\sqrt{\frac{n_1 - k}{n_1 - 1}}, \sqrt{\frac{n_2 - k}{n_2 - 1}})$.

The numerator of the standard two-sample T-statistic is $\mathbf{1}_2^T \mathbf{1}_2 \mathbf{a}^T \hat{H}_2/n_2 - \mathbf{1}_1^T \mathbf{1}_1 \mathbf{a}^T \hat{H}_1/n_1 = \hat{\theta}$, where $\mathbf{1}_1, \mathbf{1}_2$ are vectors of n_1 and n_2 ones respectively, and we have used the fact that $\mathbf{1}^T R_j = 0$. The denominator is $\tilde{\mathbf{v}}^{1/2}$ where

$$\tilde{\mathbf{v}} = (1/n_1 + 1/n_2)(n_1 \tilde{s}^2_{1j} + n_2 \tilde{s}^2_{2j})/(n_1 + n_2),$$

$\tilde{s}^2_{1j} = \sum_{i=1}^{n_1} (x_{ij} - \tilde{x}_{ij})^2/(n_1 - k) = \hat{s}^2_{1j}$ and similarly for $\tilde{s}^2_{2j}$.

Thus $\tilde{\mathbf{v}}$ is proportional to $\hat{\mathbf{v}}$ and hence the T-statistic is equivalent to the Wald test statistic based on $\hat{\theta}$. The proportionality constant is

$$(1/n_1 + 1/n_2)/\mathbf{a}^T[(W_2^T W_2)^{-1} + (W_1^T W_1)^{-1}]\mathbf{a}$$

Nature Methods: doi: 10.1038/nmeth.1439

. This is typically close to 1, and does not play a role if the null distributions are estimated by permutations.

Nature Methods: doi: 10.1038/nmeth.1439