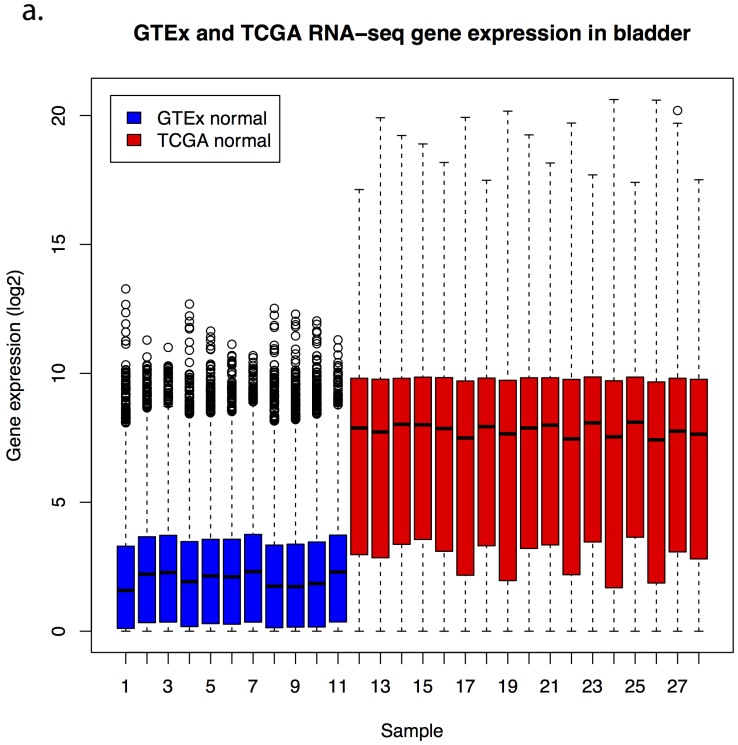
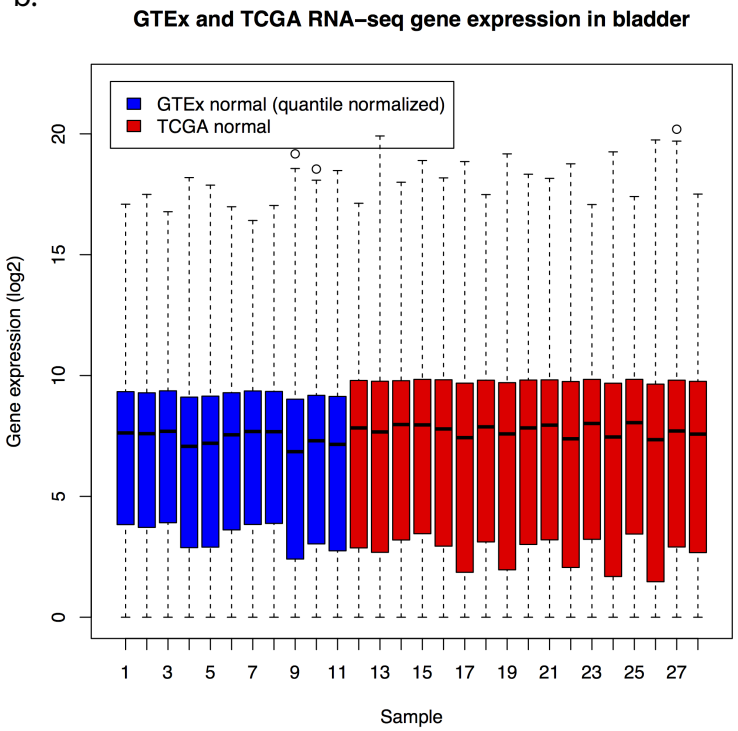
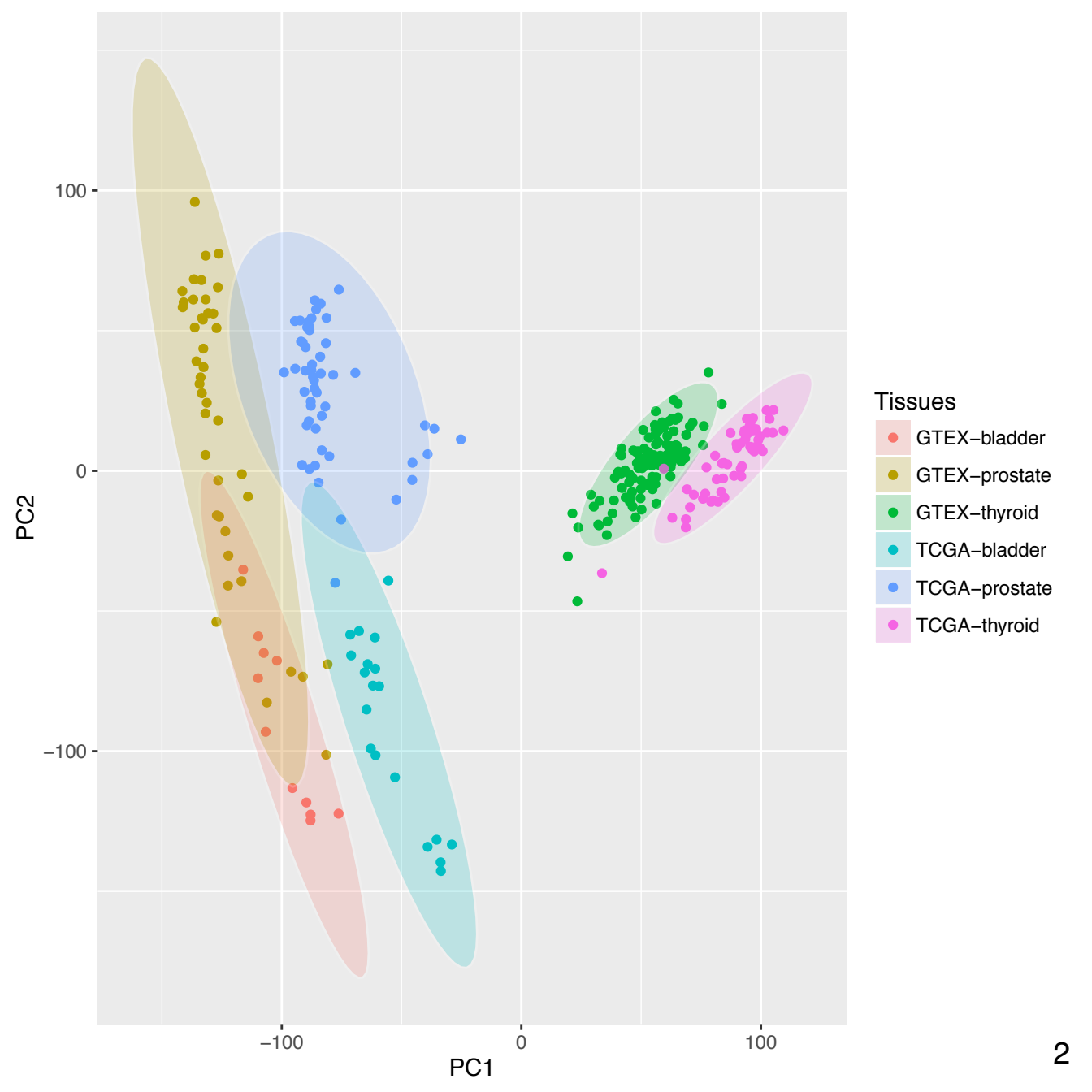
Supplementary information for *Unifying cancer and normal RNA sequencing data from different sources*

a.

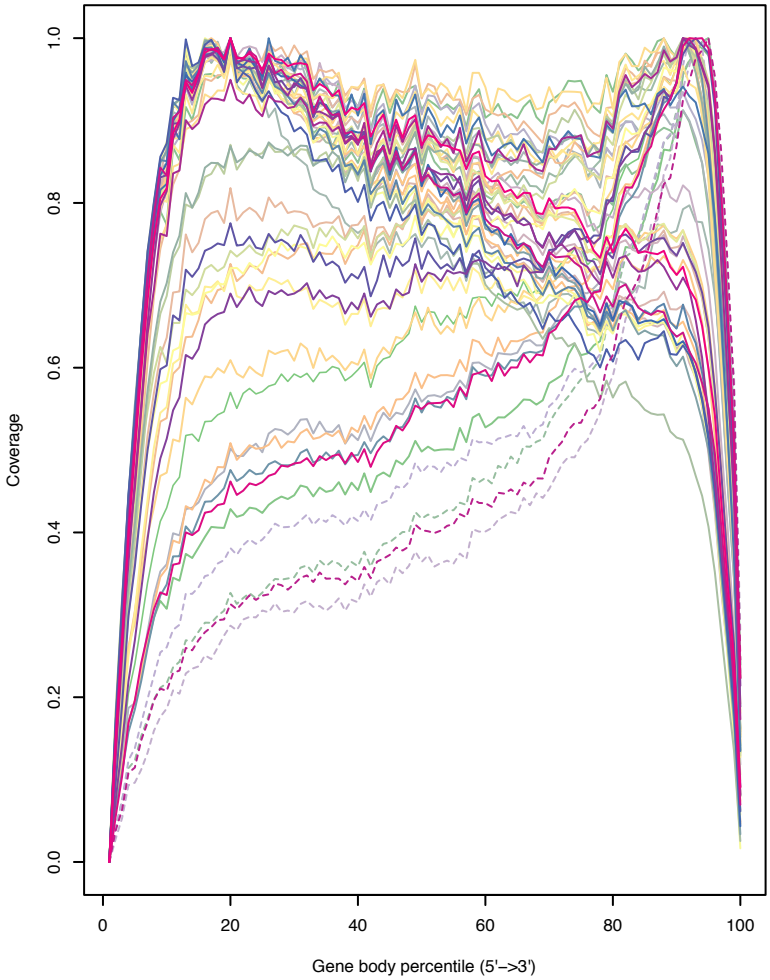

GTEx and TCGA RNA–seq gene expression in bladder
GTEx normal
TCGA normal
Gene expression (log2)
0
5
10
15
20
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Sample

b.

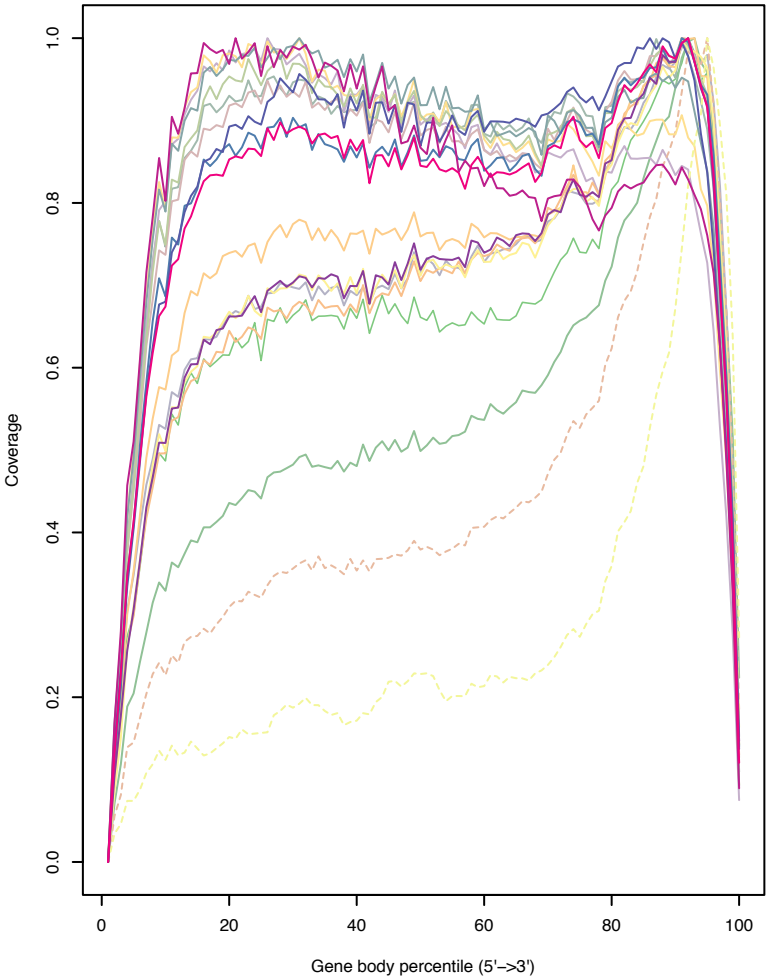

GTEx and TCGA RNA–seq gene expression in bladder
GTEx normal (quantile normalized)
TCGA normal
Gene expression (log2)
0
5
10
15
20
1
3
5
7
9
11
13
15
17
19
21
23
25
27
Sample


100
0
−100
PC2
−100
0
100
PC1
Tissues
GTEX−bladder
GTEX−prostate
GTEX−thyroid
TCGA−bladder
TCGA−prostate
TCGA−thyroid

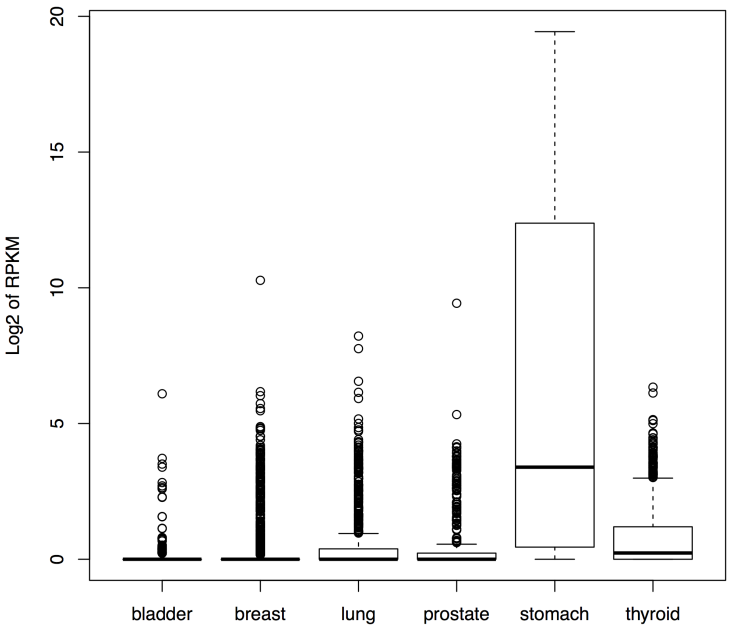
a.


Coverage
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
Gene body percentile (5'–>3')

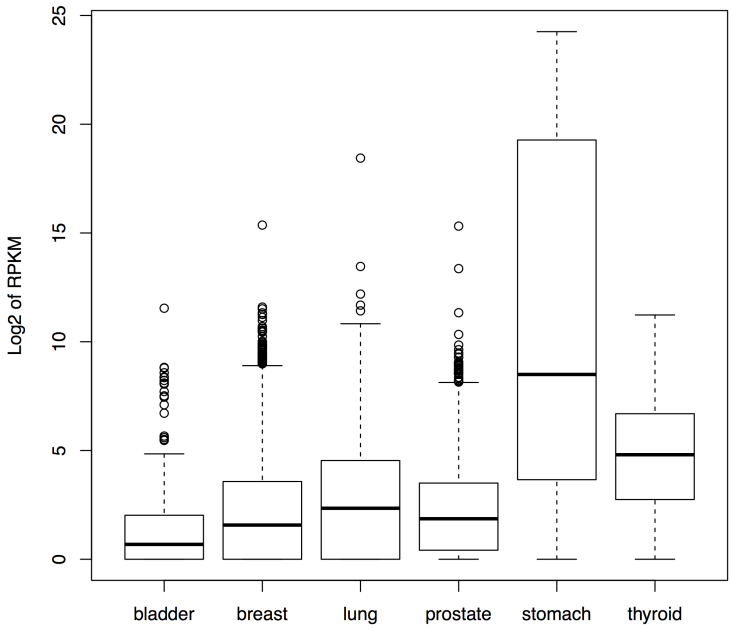
b.

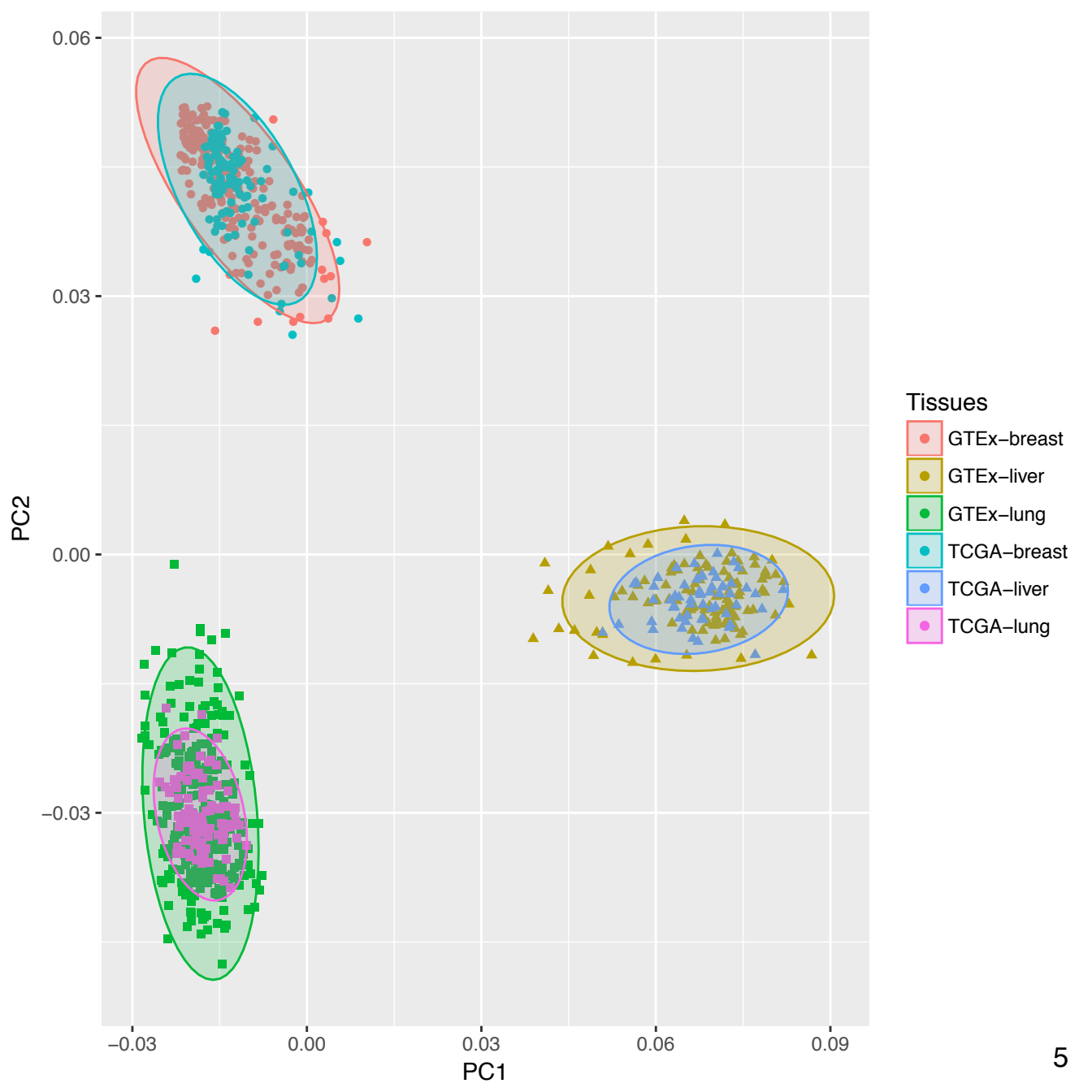

Coverage
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
Gene body percentile (5'–>3')

a.


PGA3 expression by FeatureCounts
Log2 of RPKM
0
5
10
15
20
bladder
breast
lung
prostate
stomach
thyroid

b.


PGA3 expression by RSEM
Log2 of RPKM
0
5
10
15
20
25
bladder
breast
lung
prostate
stomach
thyroid


0.06
0.03
0.00
−0.03
PC2
−0.03
0.00
0.03
0.06
0.09
PC1
Tissues
GTEx−breast
GTEx−liver
GTEx−lung
TCGA−breast
TCGA−liver
TCGA−lung

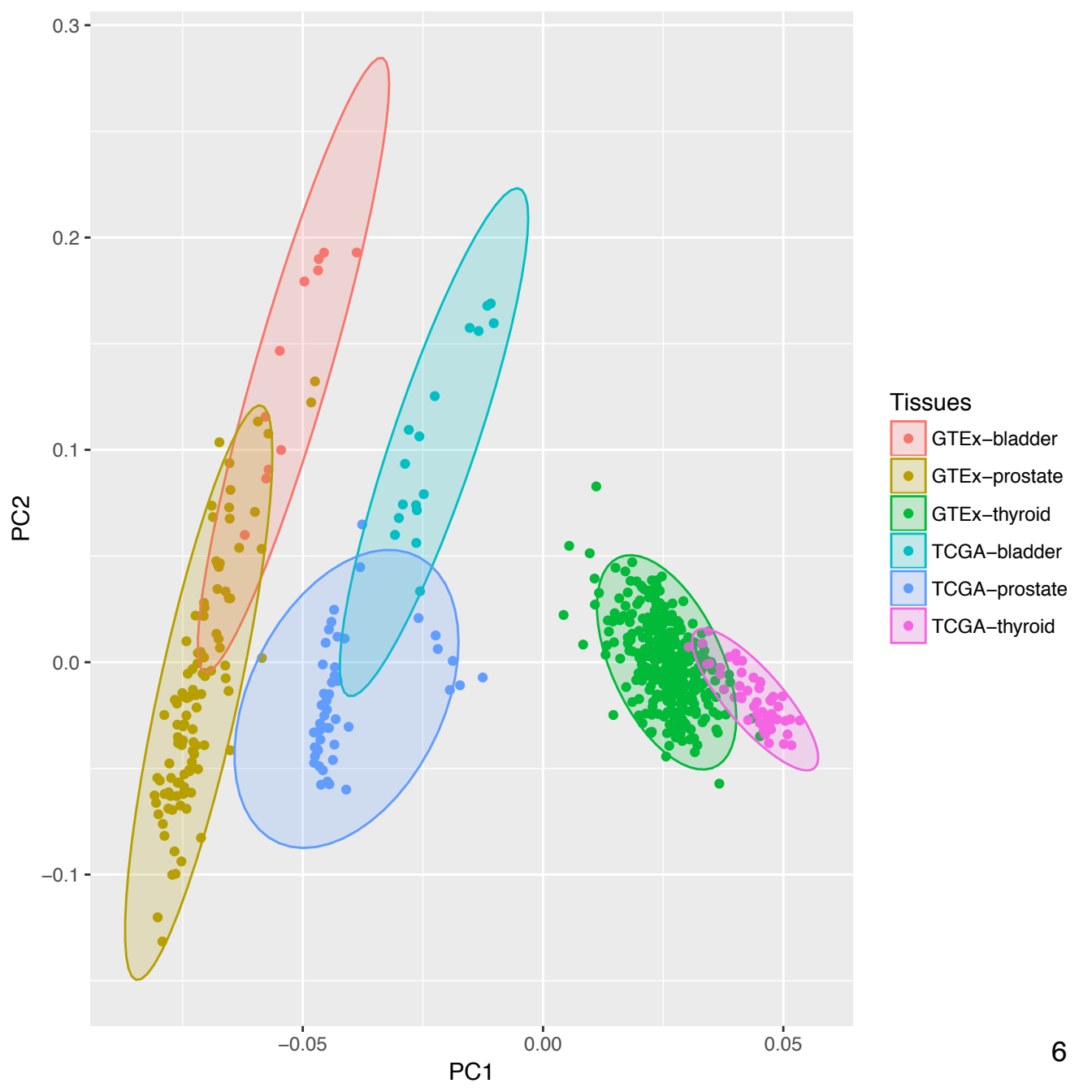

0.3
0.2
0.1
0.0
−0.1
PC2
−0.05
0.00
0.05
PC1
Tissues
GTEx−bladder
GTEx−prostate
GTEx−thyroid
TCGA−bladder
TCGA−prostate
TCGA−thyroid

Table S1. Samples with no or insufficient numbers of normal samples available in TCGA or GTEx.

GTEx tissue / TCGA cancer type	GTEx	TCGA normal	TCGA tumor	Total
adipose / sarc	621	2	259	882
blood / laml	456	0	0	456
none / chol	0	9	36	45
none / dlbc		0	48	48
adrenal gland / acc	159	0	79	238
adrenal gland / pcpg		3	179	182
brain / gbm	1403	0	156	1559
brain / lgg		0	516	516
ovary / ov	108	0	294	402
pancreatic / paad	197	4	178	379
skin / skcm	974	1	103	1078
small intestine / none	104	0	0	104
testis / tgct	203	0	150	353
none / thym		2	120	122
none / meso		0	87	87
none / uvm		0	80	80
Total	**4225**	**21**	**2285**	**6531**

Table S2. Parameters of ComBat for: (a) bladder, and (b) lung. TCGA lung cancer has two subtypes: lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). LUSC was designated in the same batch as LUAD.

(a) Parameters of ComBat for bladder

	GTEx bladder	TCGA BLCA normal	TCGA BLCA tumor
Batch	1	2	2
Variable of interest	normal	normal	tumor

(b) Parameters of ComBat for lung

	GTEx lung	TCGA LUAD normal	TCGA LUAD tumor	TCGA LUSC normal	TCGA LUSC tumor
Batch	1	2	2	2	2
Variable of interest	normal	normal	tumor	normal	tumor