

input data: more than one ExpressionSet with valid annotation link

ExpressionSet A

ExpressionSet B

extract chip platform

load current annotation, respectively

if not available, fetch from internet

Experiment study A

ID	sample x	sample y	sample z
probe x	5.1	5.2	5.005
probe y	8.3	9.1	8.4
probe z	12.8	13.1	12.9
...

raw
expression data

Experiment study B

ID	sample a	sample b	sample c
probe a	5.2	5.005	5.1
probe b	9.1	8.4	8.3
probe c	12.9	12.8	13.1
...

+

+

ID	symbol
probe x	XA1
probe y	YB2
probe z	ZC3
...	...

current
annotation

ID	symbol
probe a	XA1
probe b	YB2
probe c	ZC3
...	...

discard non-matching lines

collapse multiple values targeting
the same gene by given function

ID	sample x	sample y	sample z
XA1	5.1	5.2	5.005
YB2	8.3	9.1	8.4
ZC3	12.8	13.1	12.9
...

reannotated
expression data

ID	sample a	sample b	sample c
XA1	5.2	5.005	5.1
YB2	9.1	8.4	8.3
ZC3	12.9	12.8	13.1
...

discard non-matching lines

concatenate expression matrices

combined
expression matrix

ID	batch 1	batch 1	batch 1	batch 2	batch 2	batch 2
	sample x	sample y	sample z	sample a	sample b	sample c
XA1	5.1	5.2	5.005	5.2	5.005	5.1
YB2	8.3	9.1	8.4	9.1	8.4	8.3
ZC3	12.8	13.1	12.9	12.9	12.8	13.1
...

batch effect removal using empirical
bayes networks

modified
expression matrix

ID	batch 1	batch 1	batch 1	batch 2	batch 2	batch 2
	sample x	sample y	sample z	sample a	sample b	sample c
XA1	5.1	5.5	5.005	5.2	5.005	5.6
YB2	8.3	9.1	8.4	10.1	8.4	8.3
ZC3	10.0	13.1	12.9	12.9	12.8	13.1
...