| Method | | CIBERSORT proportion predict | | | RPC proportion predict | | | QP proportion predict | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | HCC | control | preg | HCC | control | preg | HCC | control | preg |
| Truth | HCC | 18 | 9 | 0 | 17 | 10 | 0 | 14 | 13 | 0 |
| | control | 7 | 25 | 0 | 7 | 25 | 0 | 3 | 29 | 0 |
| | preg | 0 | 1 | 16 | 0 | 1 | 16 | 0 | 0 | 17 |

Supplementary Table S1: Classification confusion matrices based on real data using different reference-based methods. HCC: hepatocellular carcinoma patients. control: healthy, unpregnant control people. preg: healthy pregnant women. Cibersort: using tissue proportions solved from Cibersort for prediction. RPC: using tissue proportions solved from Robust Partial Correlations for prediction. QP: using tissue proportions solved from Quadratic Programming procedure for prediction. Cibersort predicted accuracy: 0.78. RPC predicted accuracy: 0.76. QP predicted accuracy: 0.79.

Supplementary Figure S1: Boxplot of classification accuracies for NMF and QP under different noise level and sample size. NMF: Non-negative matrix factorization (NMF) approach. QP: using tissue proportions solved from Quadratic Programming procedure for prediction. *N* represent the total sample size used in simulation. 3-fold cross validation was conducted. A, low noise level; B, medium noise level; C, high noise level.

Supplementary Figure S2: Boxplot of classification accuracies for multiple methods in simulations. QP: using tissue proportions solved from Quadratic Programming procedure for prediction. OLS: using tissue proportions directly solved from Ordinary Least Square without constraint in QP. QP high var: using a 'high-noise reference' as the reference in the QP step to solve tissue proportions, then use the solved proportions for prediction. QP random: using a 'randomly shuffled reference' as the reference in the QP step to solve tissue proportions, then use the solved proportions for prediction. A, low noise level; B, medium noise level; C, high noise level.

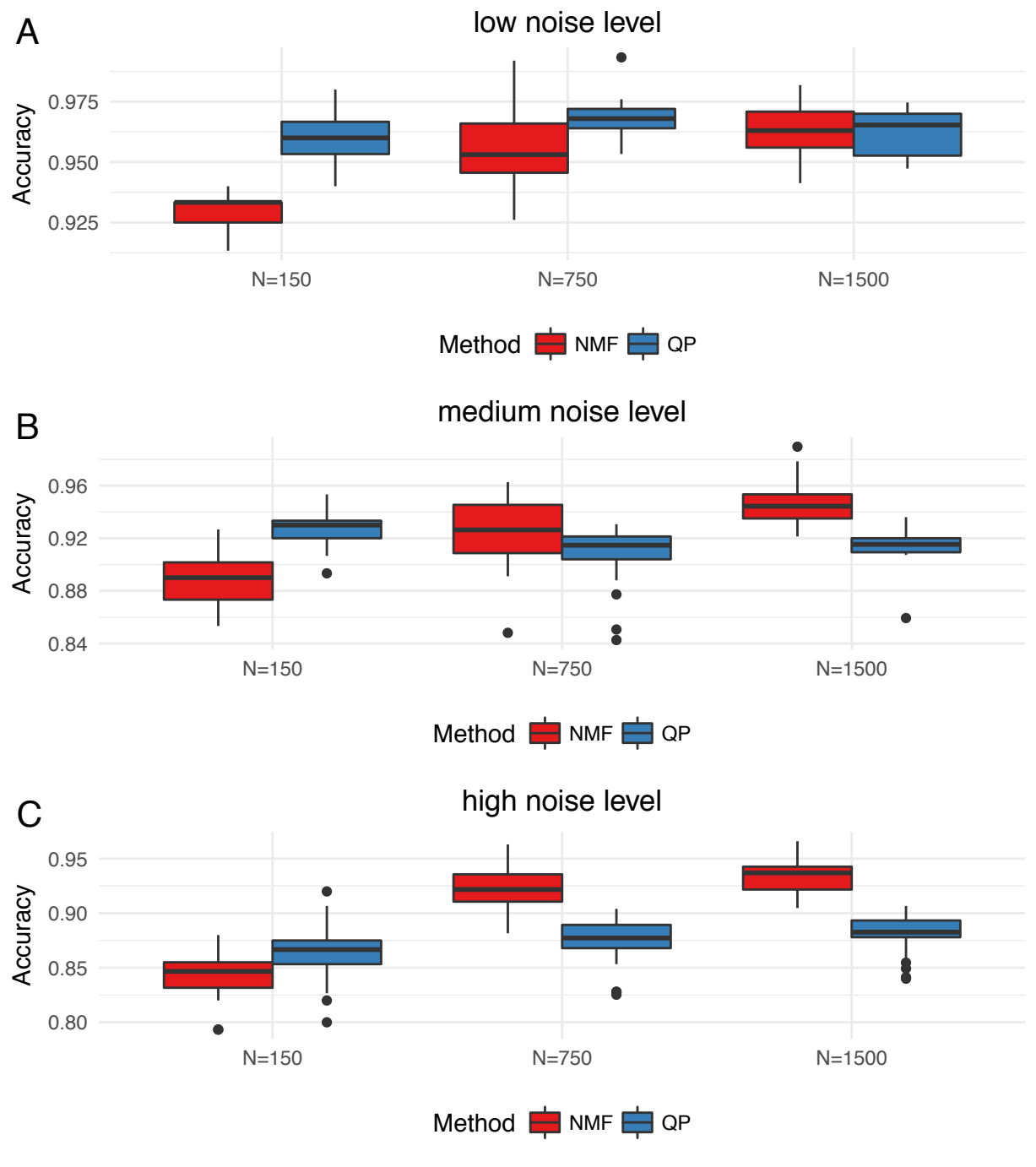
Supplementary Figure S3: Boxplot of classification accuracies for multiple reference-based methods in simulations. QP: using tissue proportions solved from Quadratic Programming procedure for prediction. RPC: using tissue proportions solved from Robust Partial Correlations for prediction. CBS: using tissue proportions solved from Cibersort for prediction. True prop: using true tissue proportions for prediction. A, low noise level; B, medium noise level; C, high noise level.

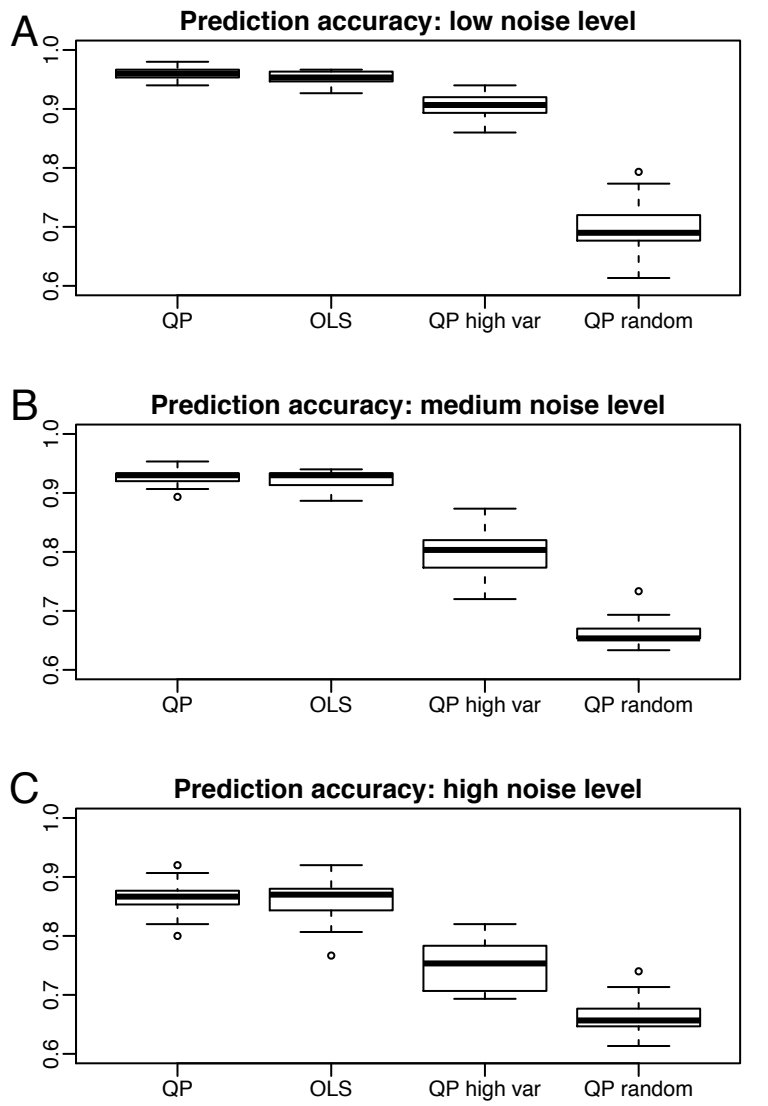
Supplementary Figure S4: Scatterplots of NMF estimated reference methylation levels versus true reference methylation levels in all 14 tissues in simulation. Spearman's correlation is shown in each panel.

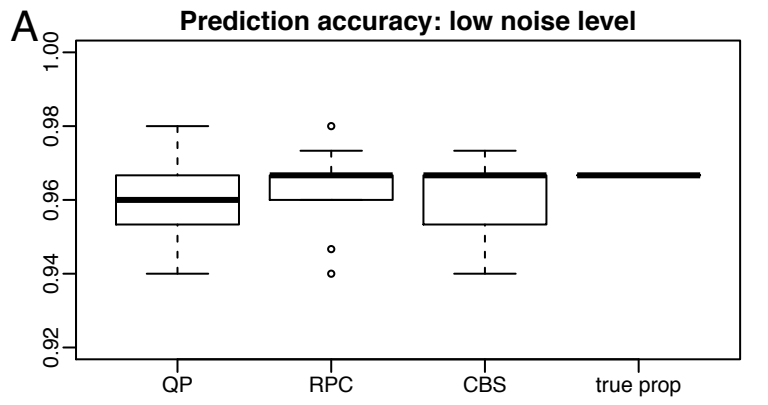
Supplementary Figure S5: Scatterplots of NMF estimated tissue proportions versus true tissue proportions in all 14 tissues in simulation. Spearman's correlation is shown in each panel.

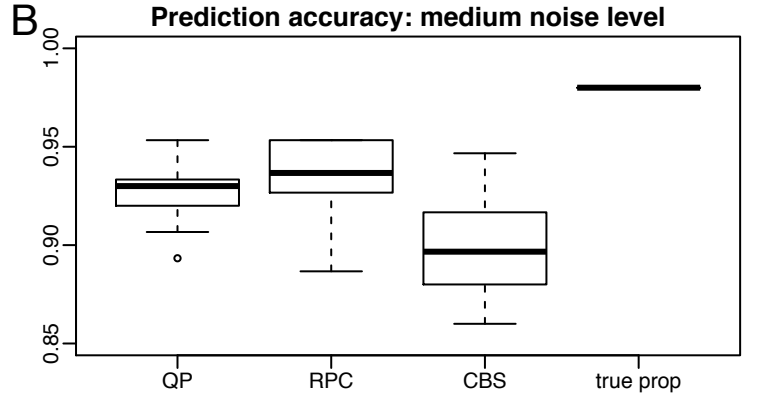
Supplementary Figure S6: Boxplot of real data solved tissue proportions for all 14 tissues, respectively, among 3 groups. One panel for each tissue.

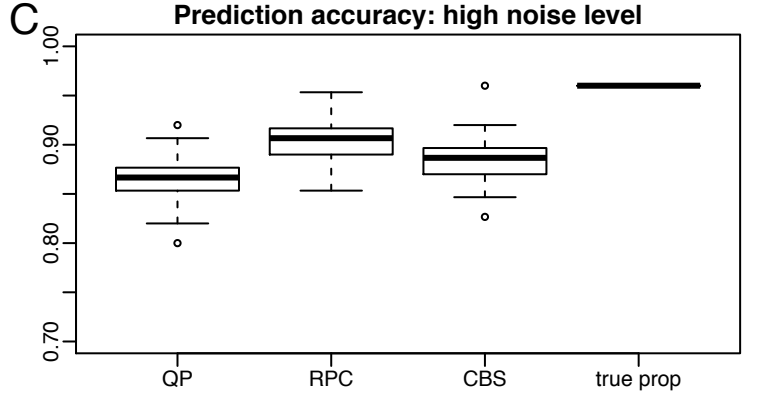
Supplementary Figure S7: Scatterplots of NMF estimated reference methylation levels versus true reference methylation levels in all 14 tissues in real data from Sun K et al. PNAS 2015. Spearman's correlation is shown in each panel.


A
low noise level
Accuracy
0.975
0.950
0.925
N=150
N=750
N=1500
Method
NMF
QP
B
medium noise level
Accuracy
0.96
0.92
0.88
0.84
N=150
N=750
N=1500
Method
NMF
QP
C
high noise level
Accuracy
0.95
0.90
0.85
0.80
N=150
N=750
N=1500
Method
NMF
QP

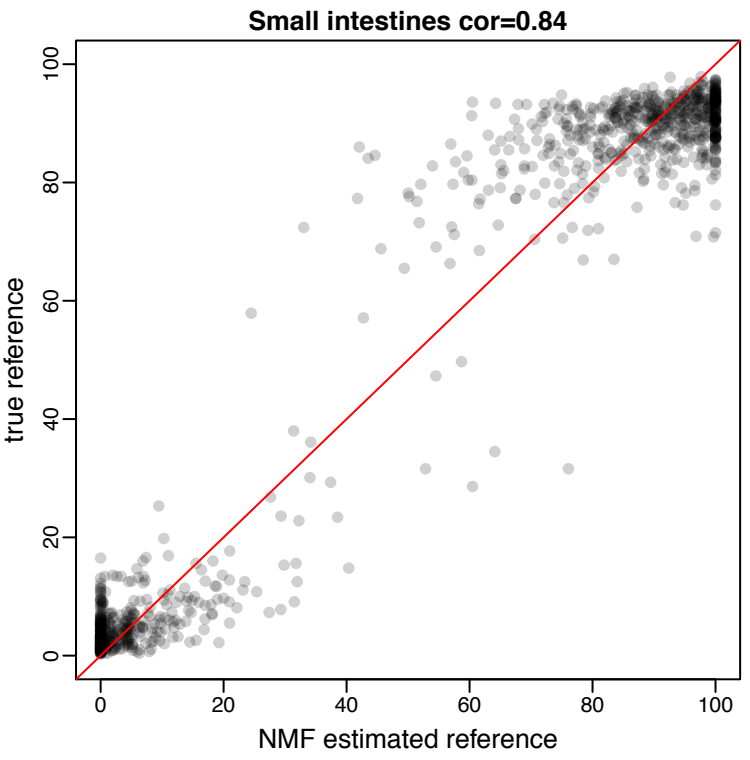

A
Prediction accuracy: low noise level
1.0
0.9
0.8
0.7
0.6
QP
OLS
QP high var
QP random
B
Prediction accuracy: medium noise level
1.0
0.9
0.8
0.7
0.6
QP
OLS
QP high var
QP random
C
Prediction accuracy: high noise level
1.0
0.9
0.8
0.7
0.6
QP
OLS
QP high var
QP random


A
Prediction accuracy: low noise level
1.00
0.98
0.96
0.94
0.92
QP
RPC
CBS
true prop

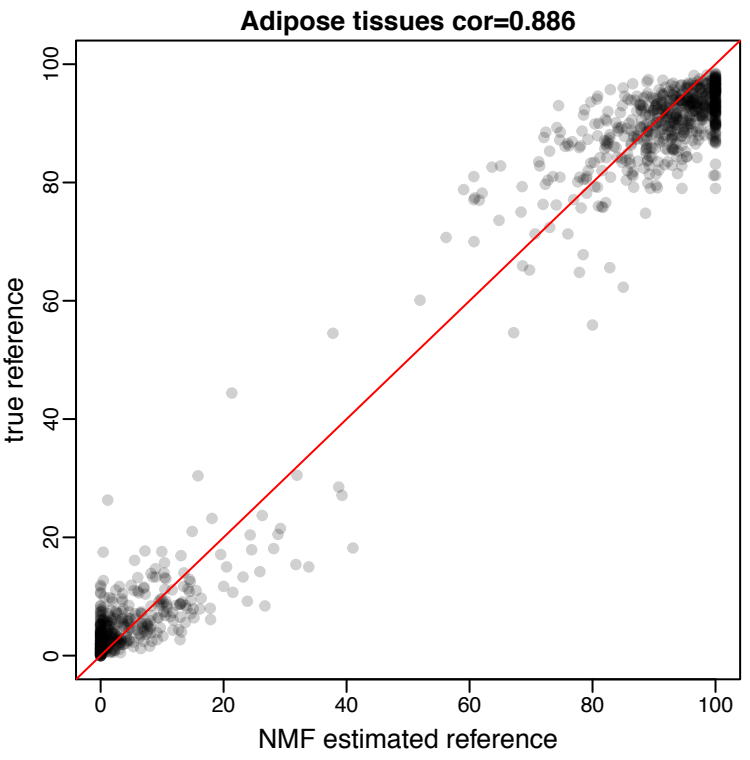

B
Prediction accuracy: medium noise level
1.00
0.95
0.90
0.85
QP
RPC
CBS
true prop


C
Prediction accuracy: high noise level
1.00
0.90
0.80
0.70
QP
RPC
CBS
true prop

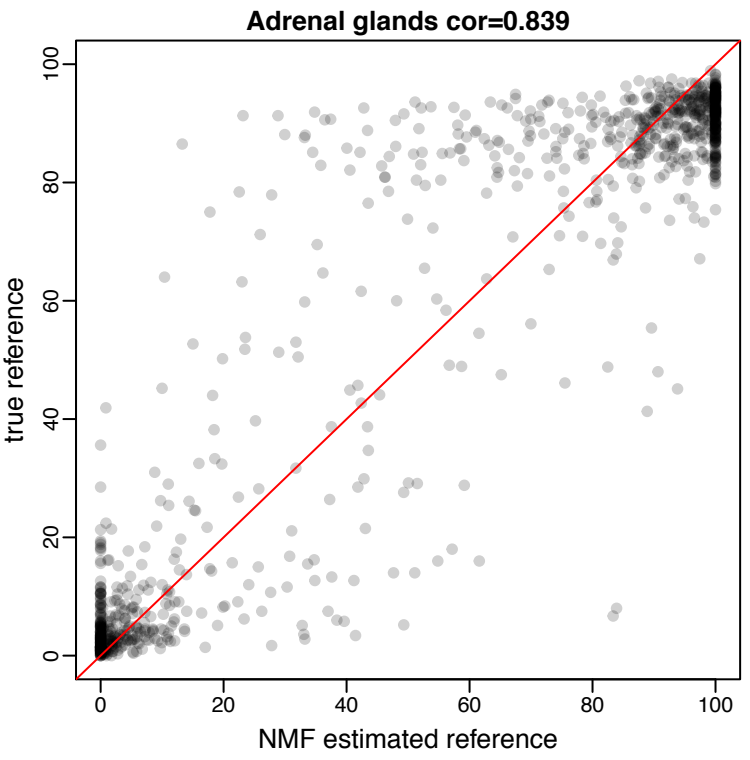
**Small intestines cor=0.84**



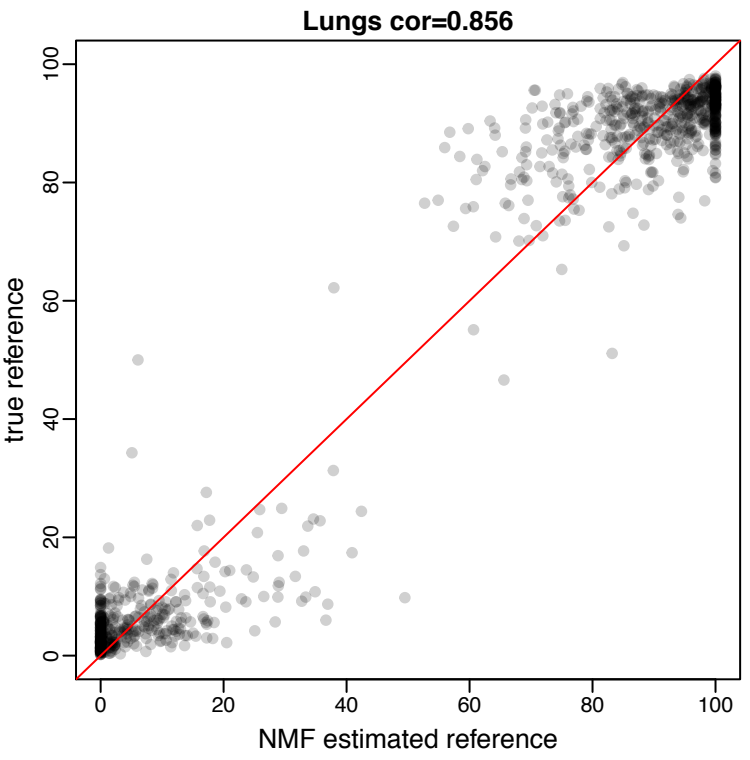

**Adipose tissues cor=0.886**



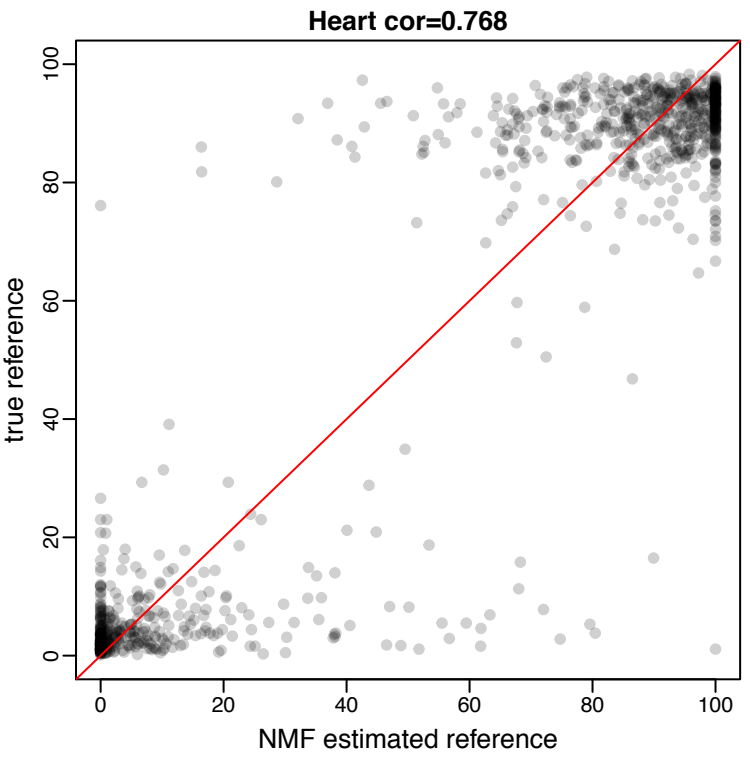

**Adrenal glands cor=0.839**



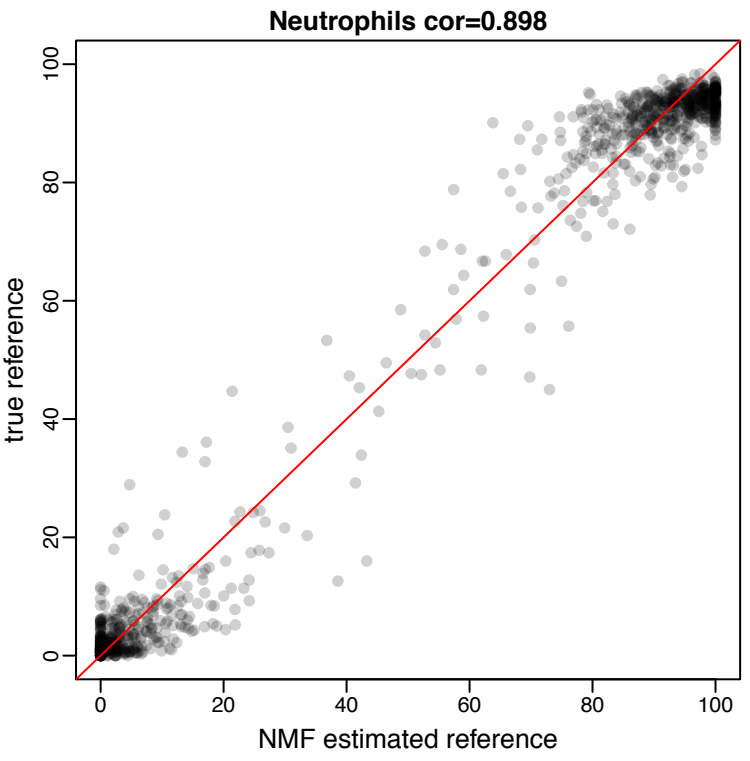

**Lungs cor=0.856**



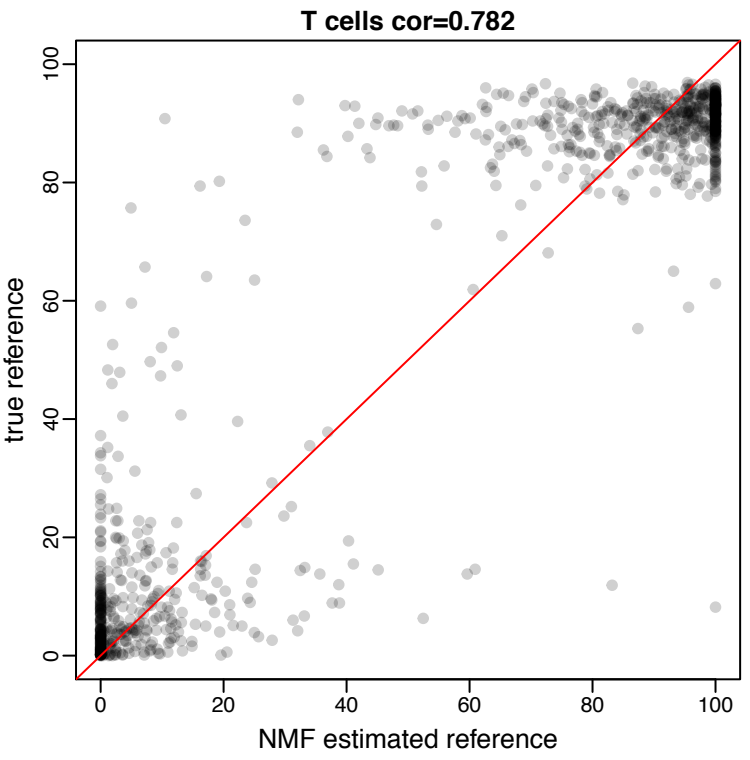

**Heart cor=0.768**



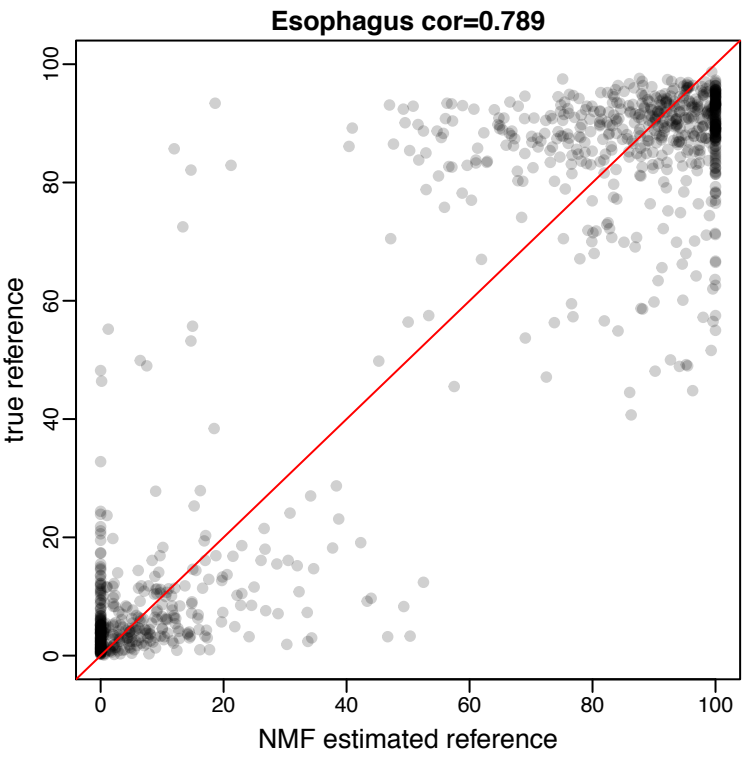

**Neutrophils cor=0.898**



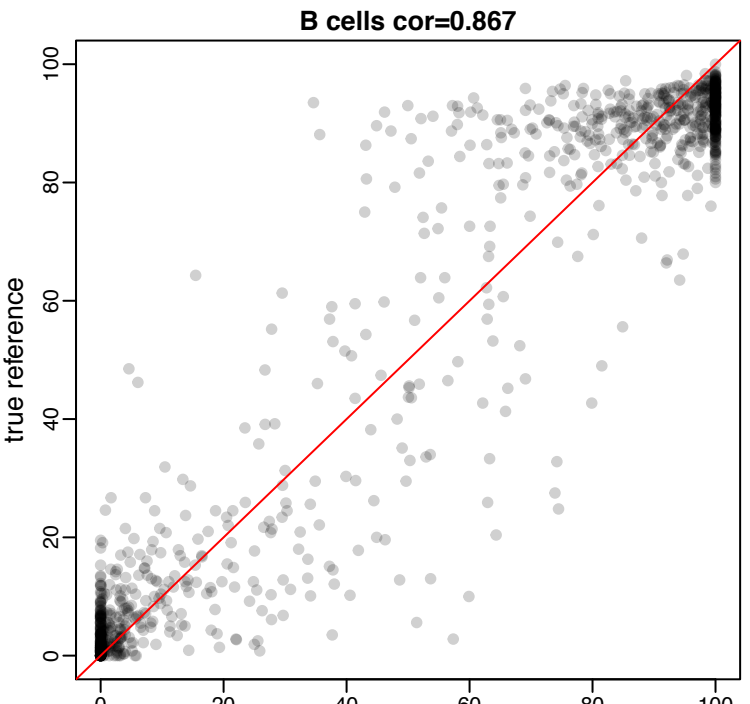

**T cells cor=0.782**



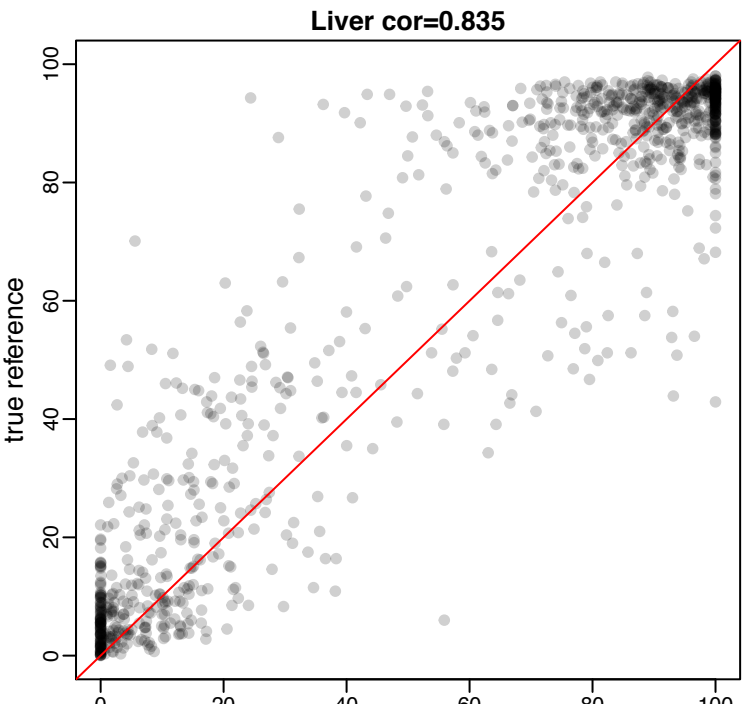

**Esophagus cor=0.789**



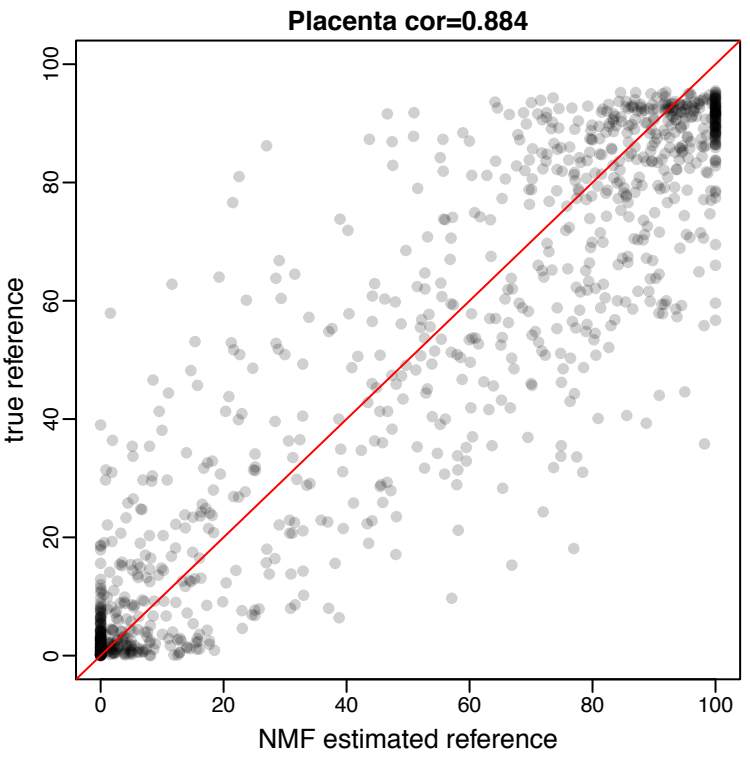

**B cells cor=0.867**



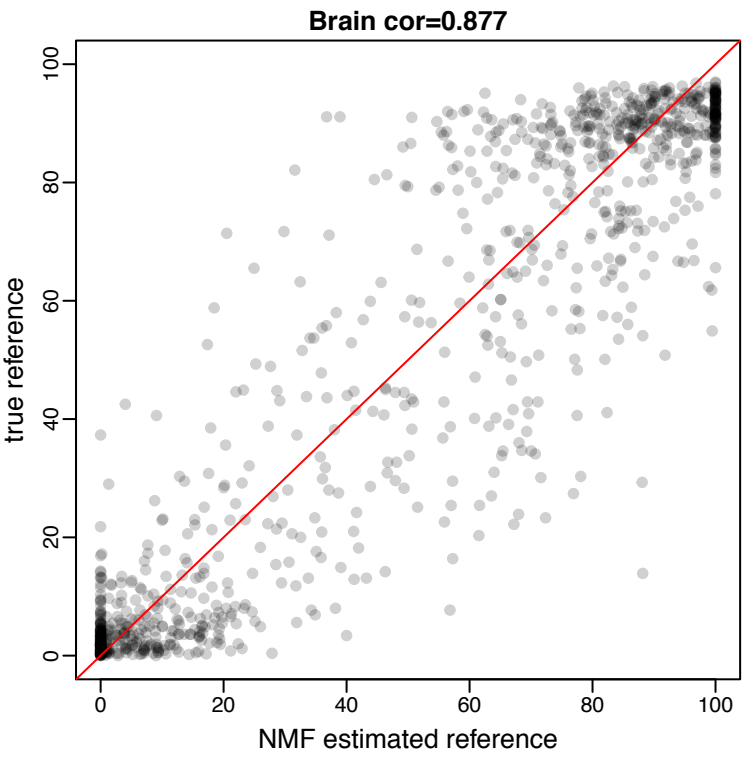

**Liver cor=0.835**



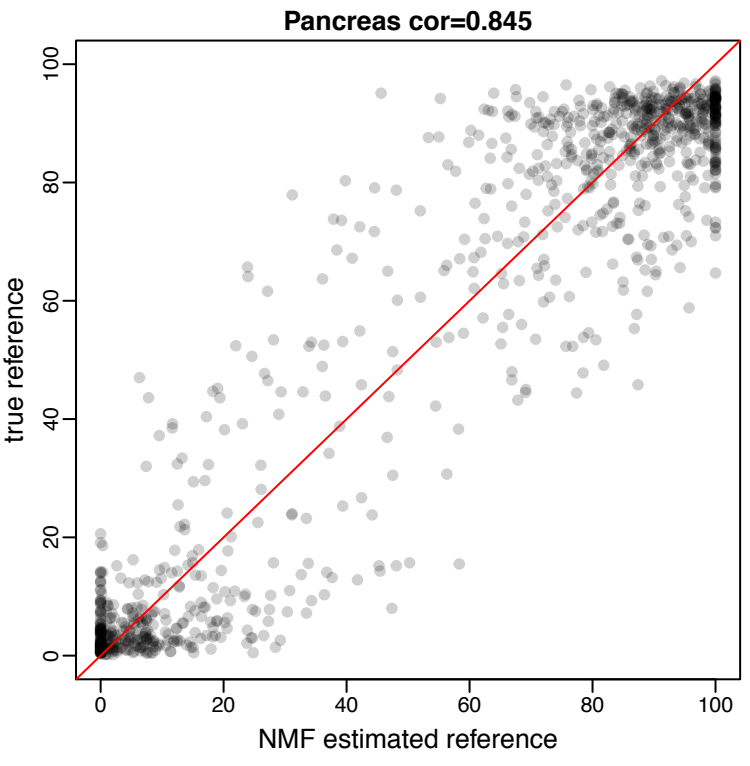

**Placenta cor=0.884**



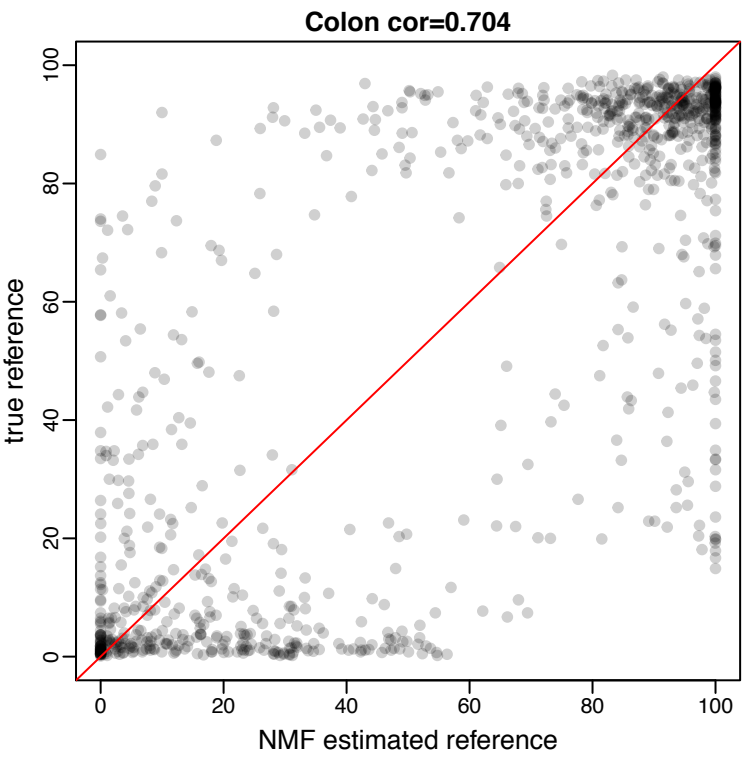

**Brain cor=0.877**



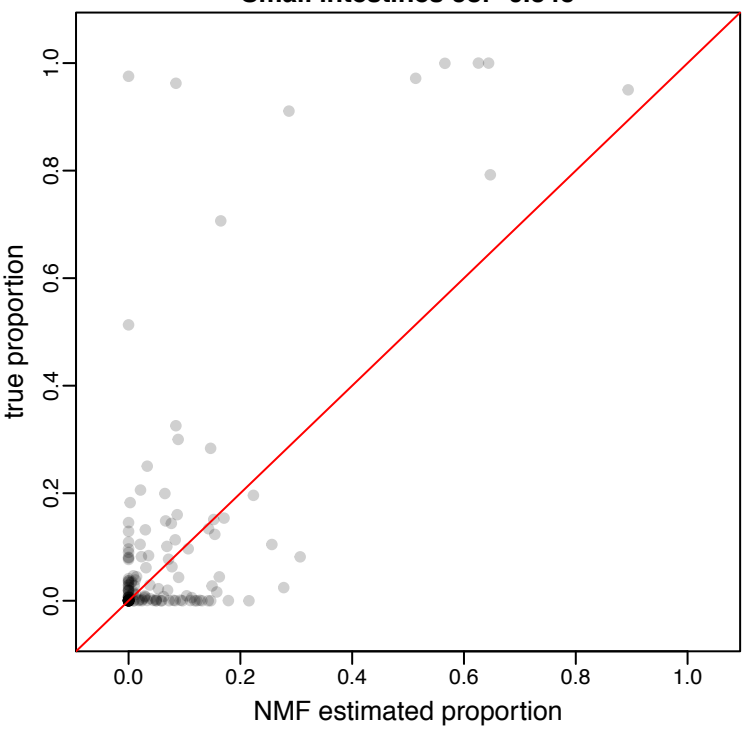

**Pancreas cor=0.845**



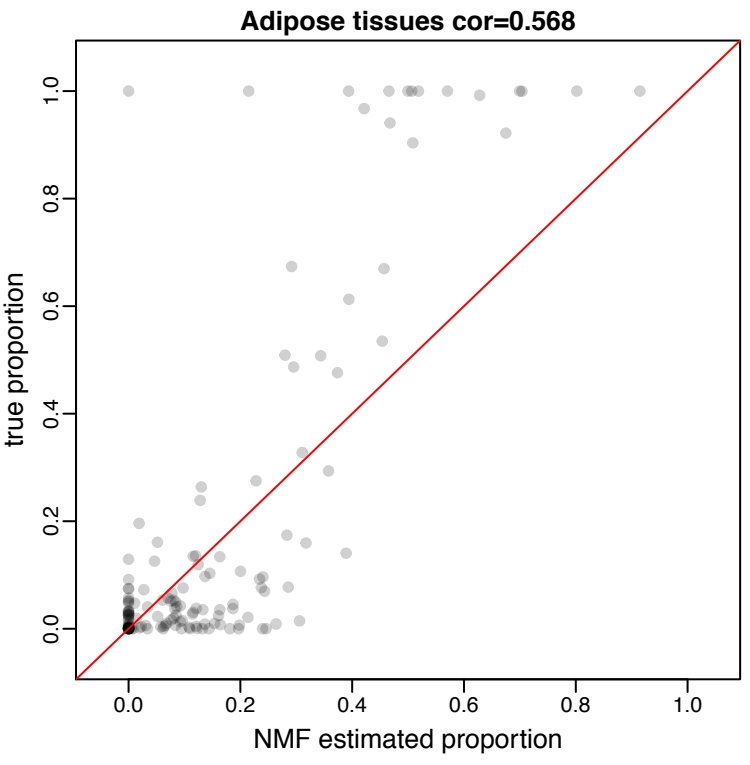

**Colon cor=0.704**



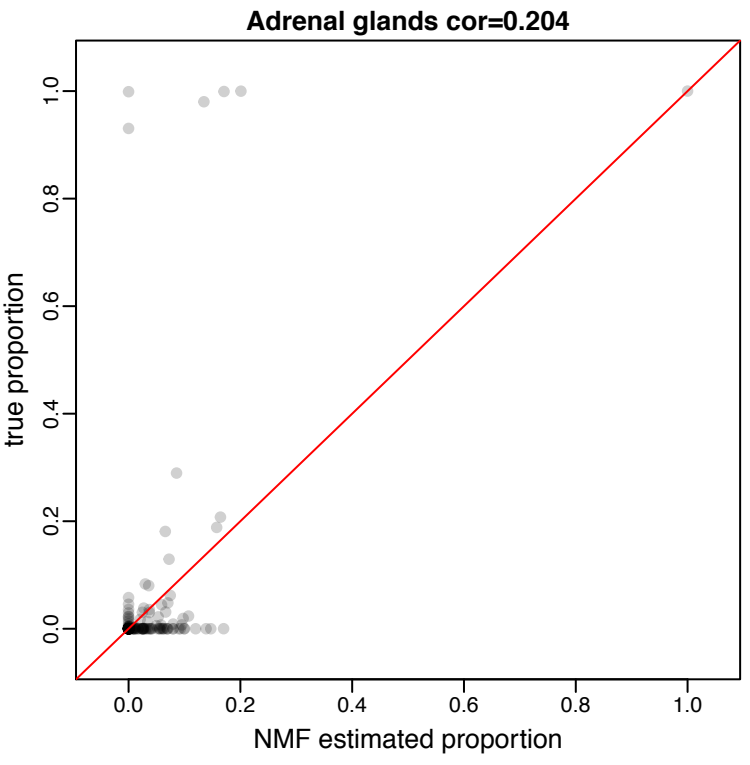
**Small intestines cor=0.345**



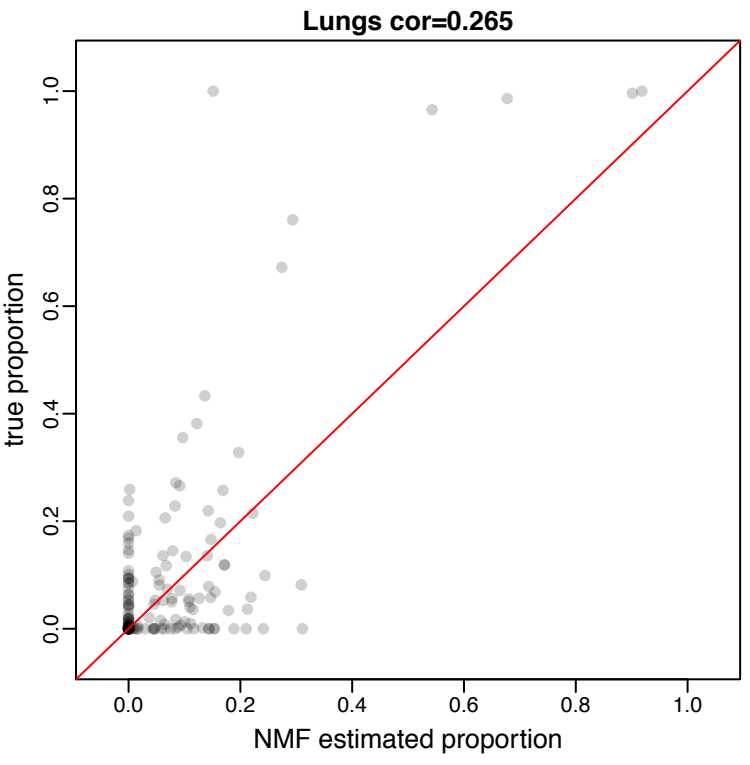

**Adipose tissues cor=0.568**



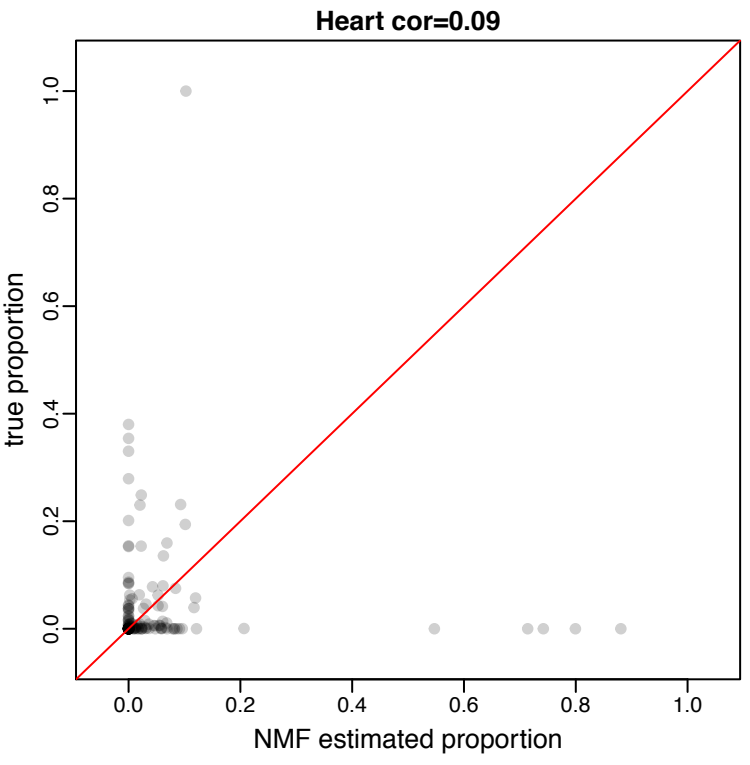

**Adrenal glands cor=0.204**



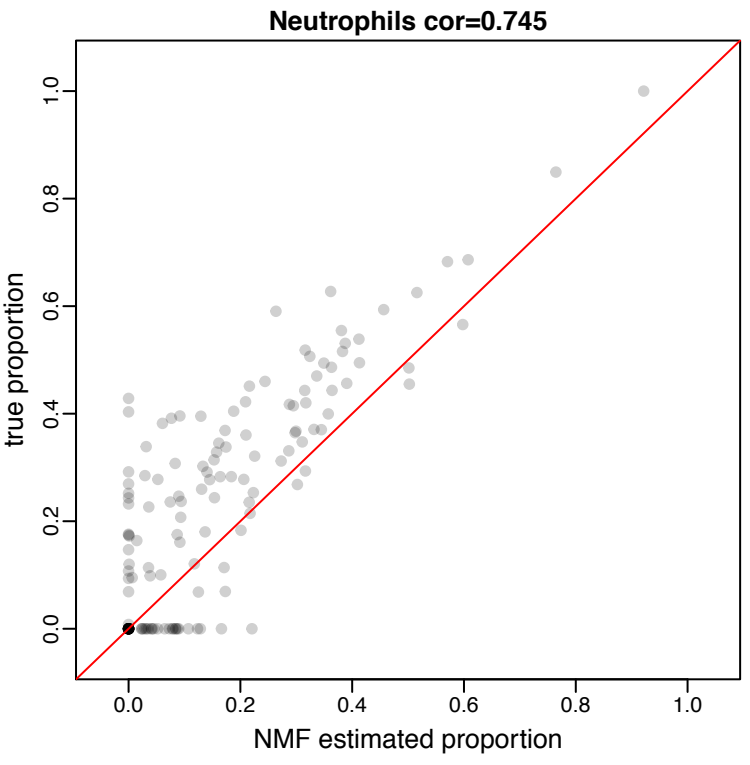

**Lungs cor=0.265**



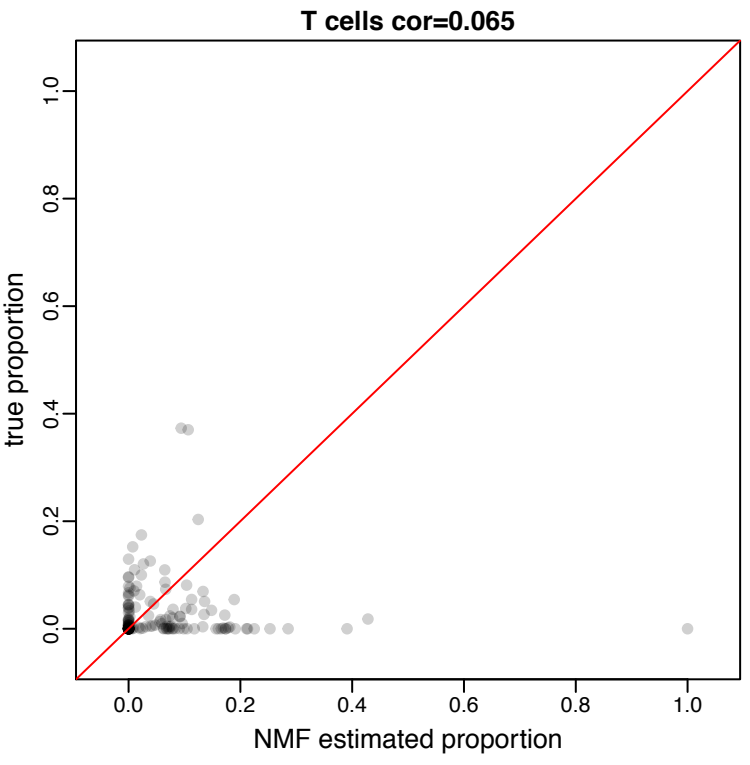

**Heart cor=0.09**



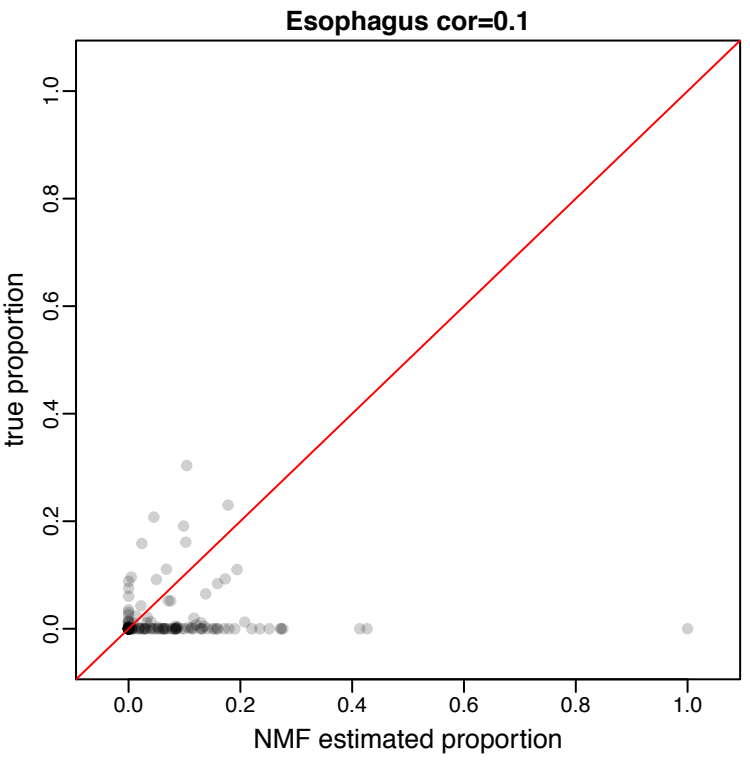

**Neutrophils cor=0.745**



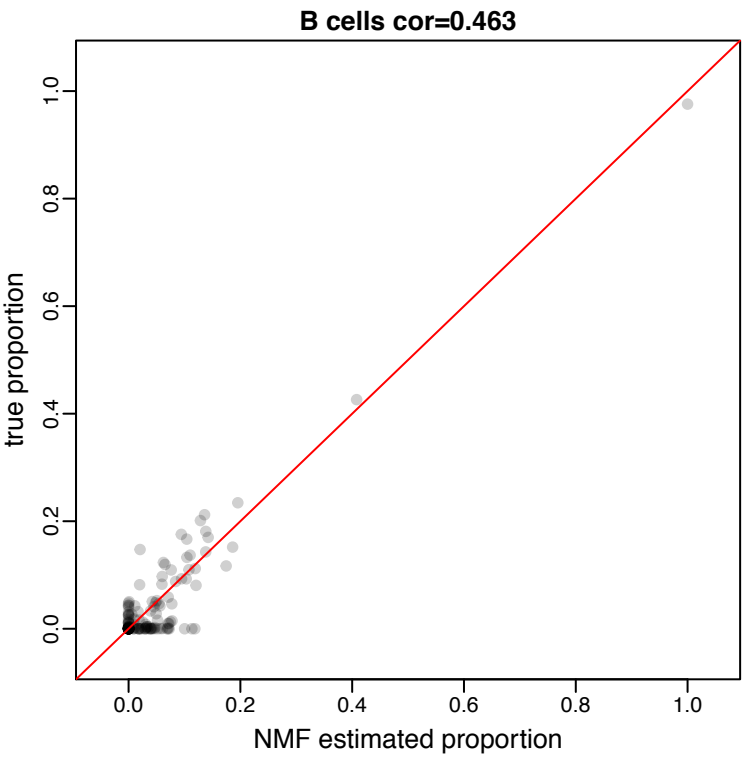

**T cells cor=0.065**



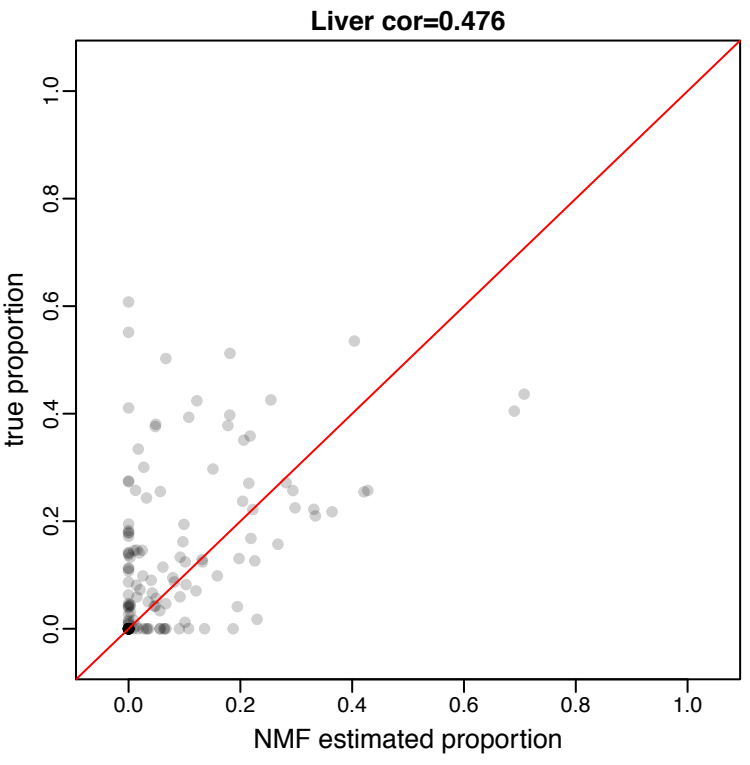

**Esophagus cor=0.1**



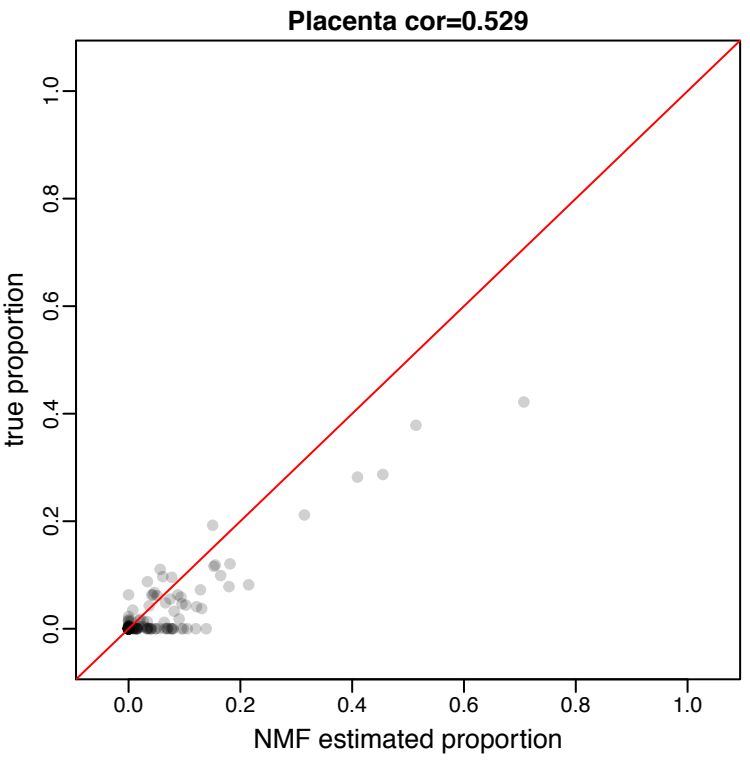

**B cells cor=0.463**



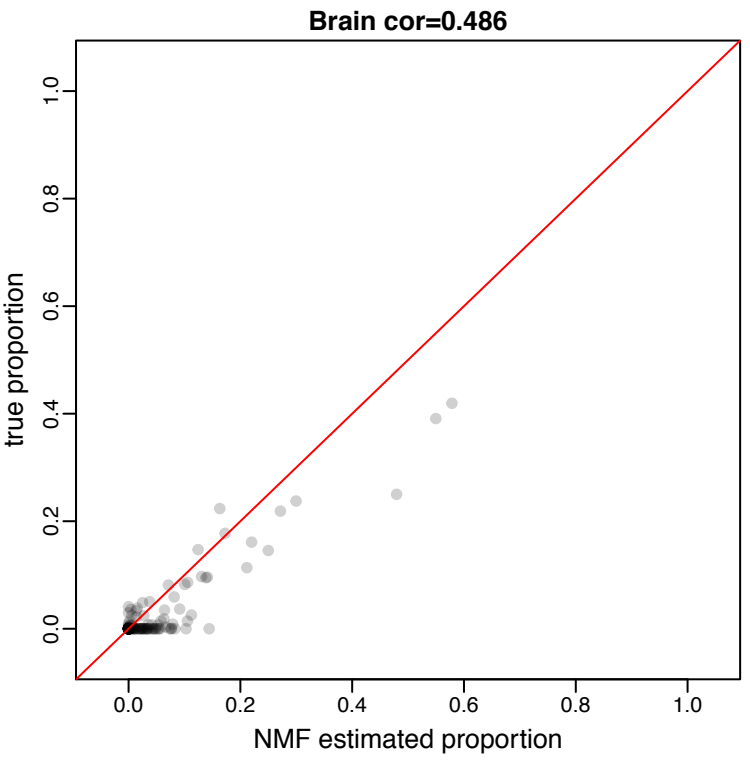

**Liver cor=0.476**



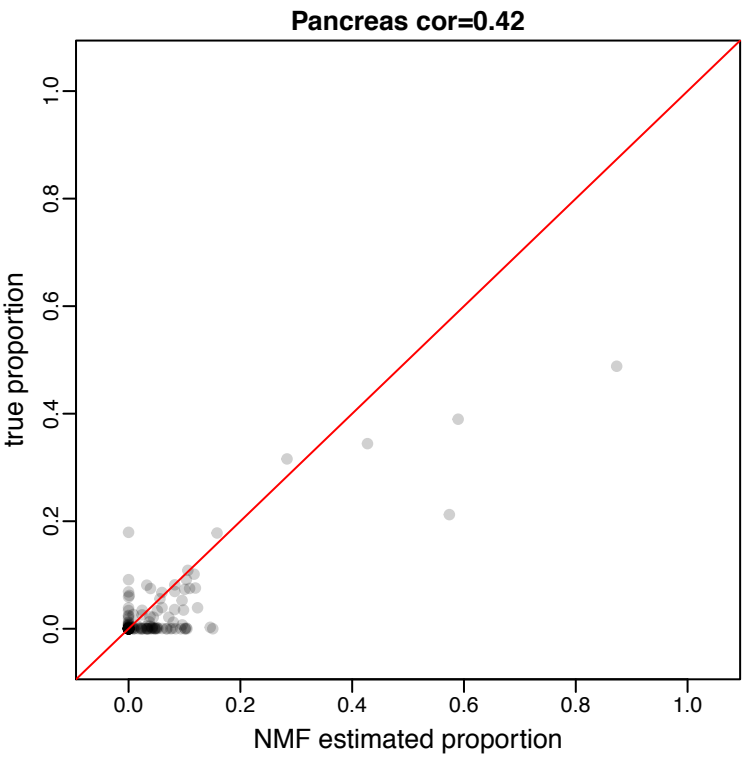

**Placenta cor=0.529**



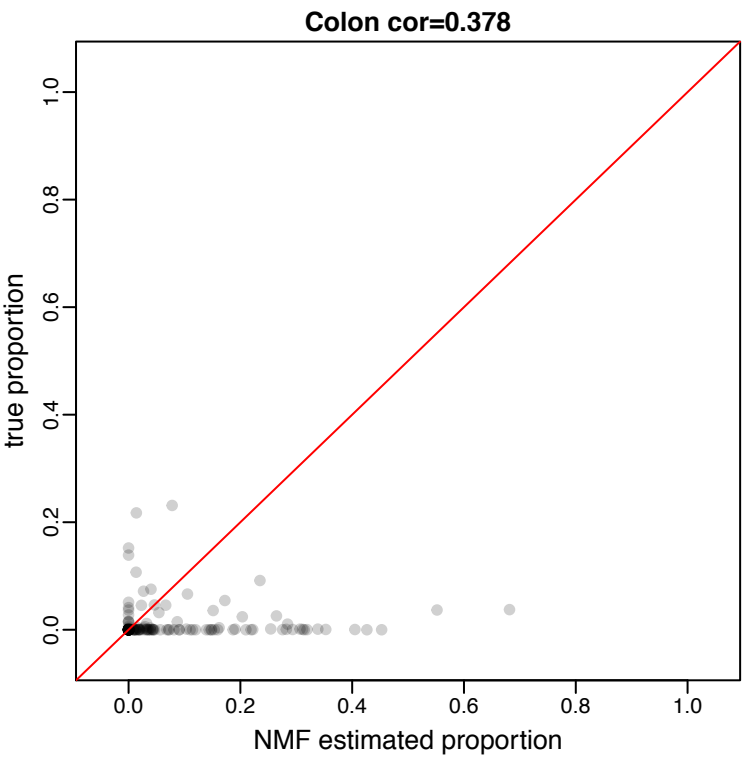

**Brain cor=0.486**

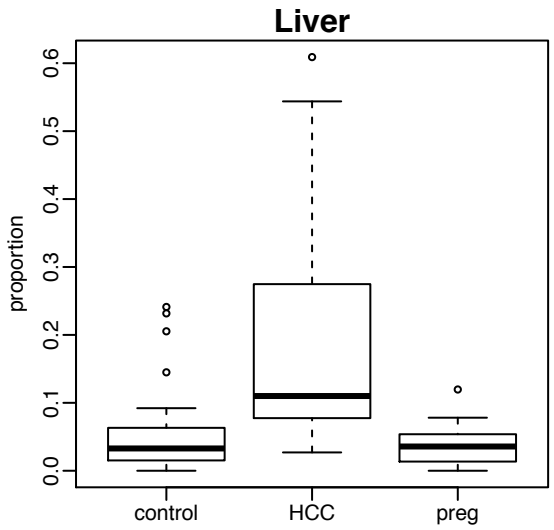


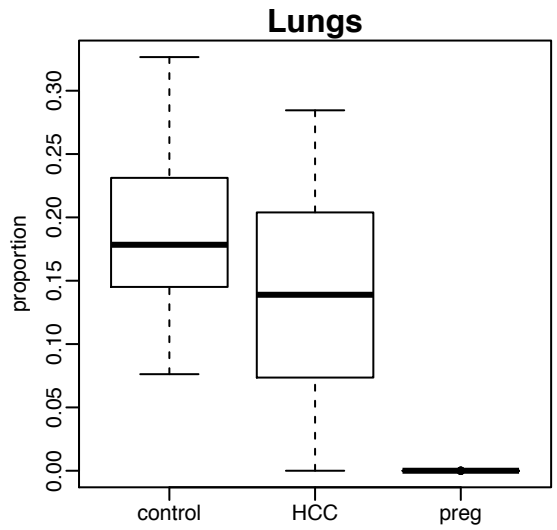

**Pancreas cor=0.42**

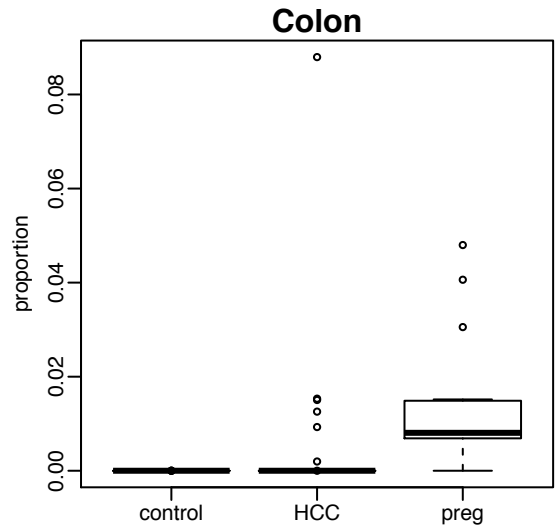


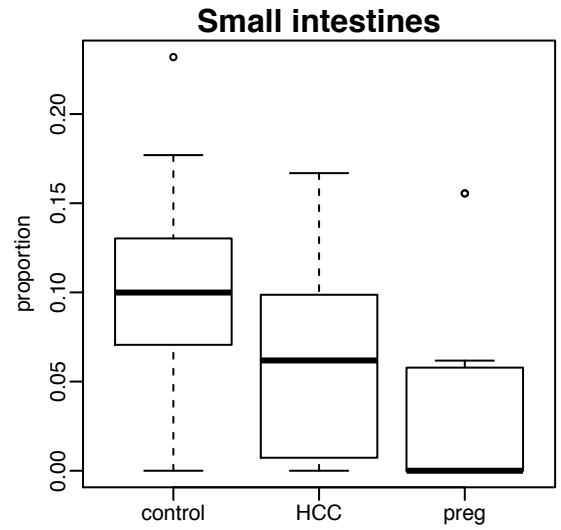

**Colon cor=0.378**

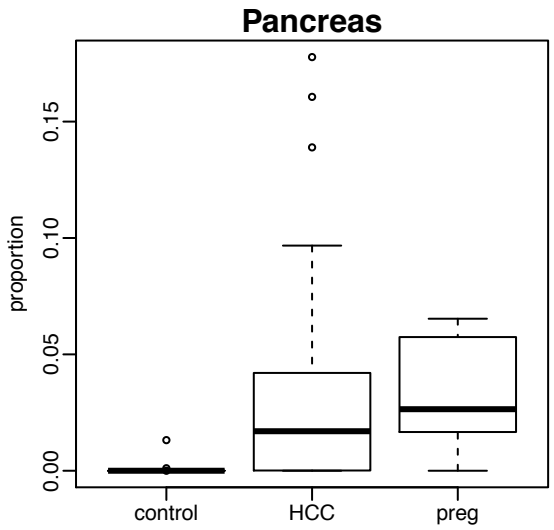


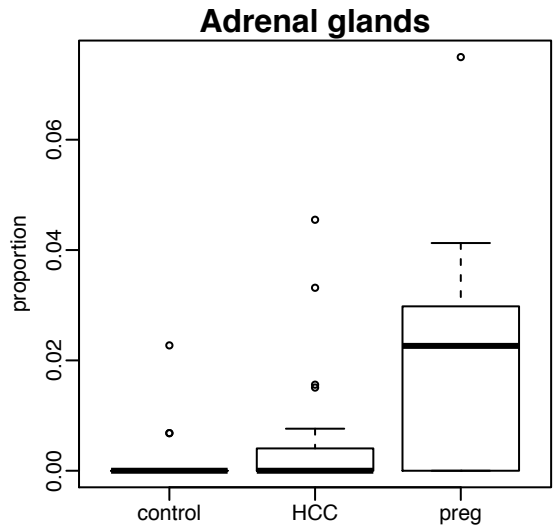

Liver
proportion
0.0
0.1
0.2
0.3
0.4
0.5
0.6
control
HCC
preg

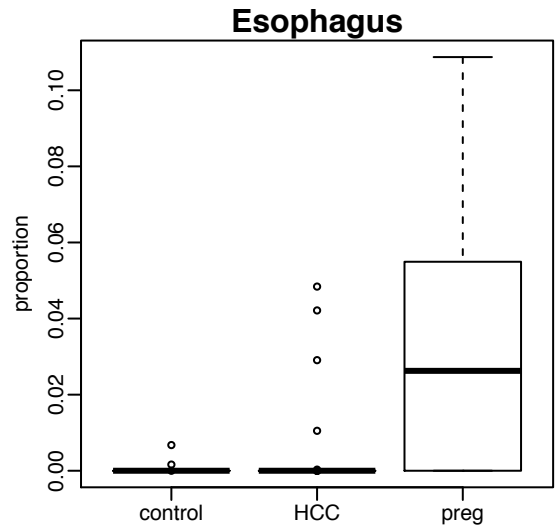

Lungs
proportion
0.00
0.05
0.10
0.15
0.20
0.25
0.30
control
HCC
preg

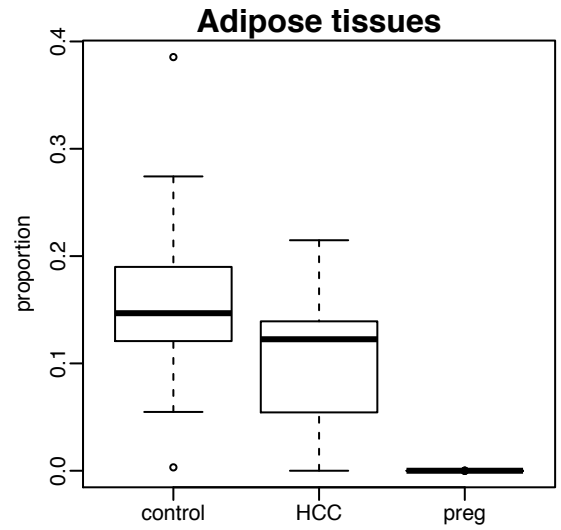

Colon
proportion
0.00
0.02
0.04
0.06
0.08
control
HCC
preg

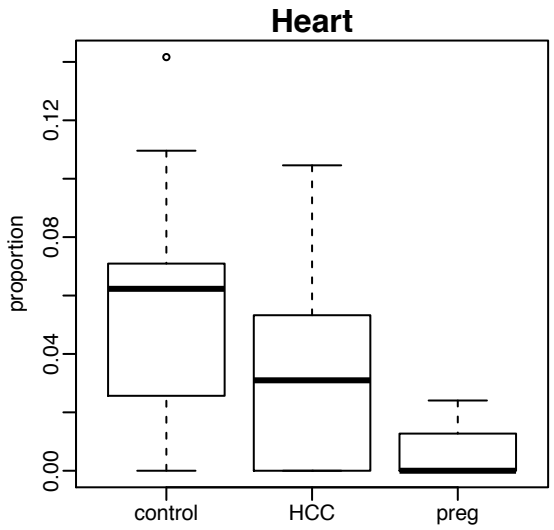

Small intestines
proportion
0.00
0.05
0.10
0.15
0.20
control
HCC
preg

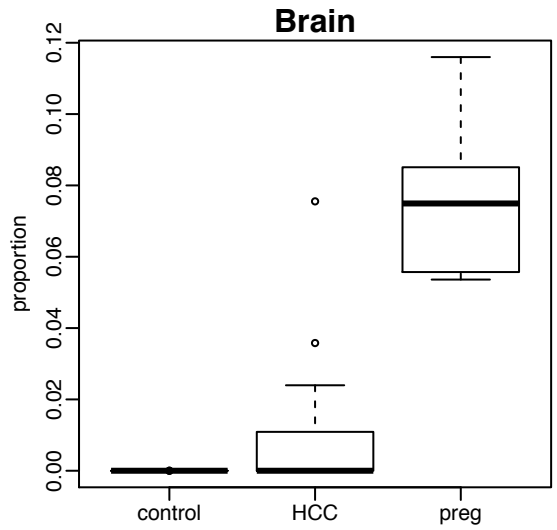

Pancreas
proportion
0.00
0.05
0.10
0.15
control
HCC
preg

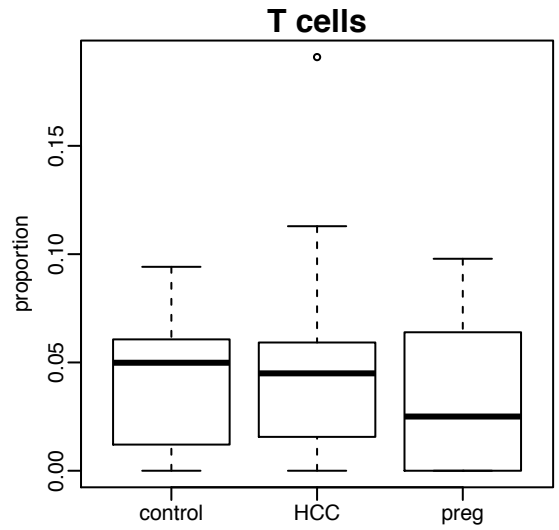

Adrenal glands
proportion
0.00
0.02
0.04
0.06
control
HCC
preg

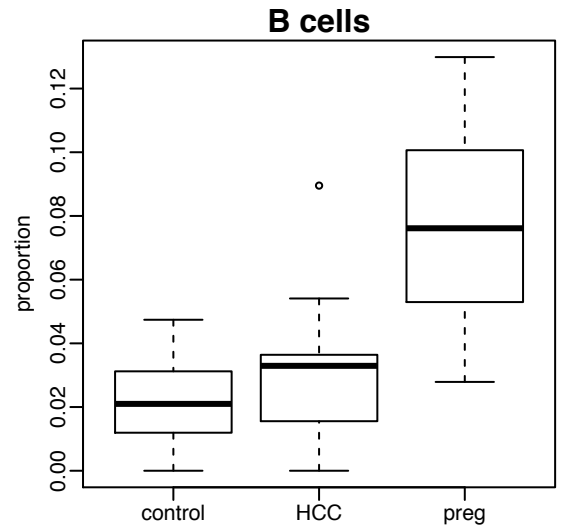

Esophagus
proportion
0.00
0.02
0.04
0.06
0.08
0.10
control
HCC
preg

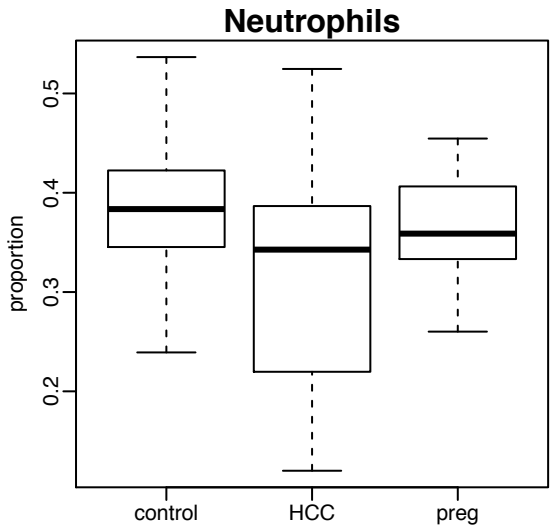

Adipose tissues
proportion
0.0
0.1
0.2
0.3
0.4
control
HCC
preg

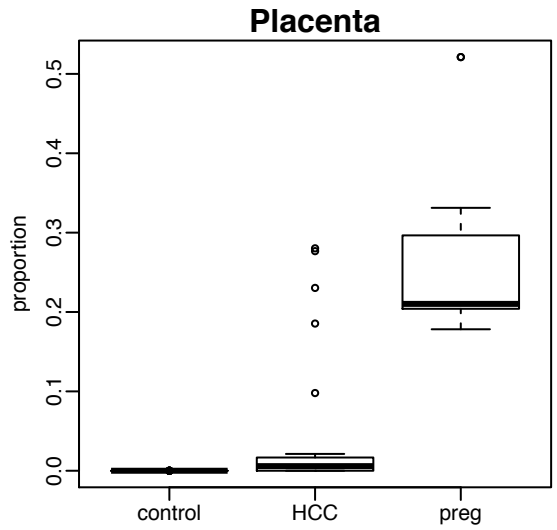

Heart
proportion
0.00
0.04
0.08
0.12
control
HCC
preg


Brain
proportion
0.00
0.02
0.04
0.06
0.08
0.10
0.12
control
HCC
preg

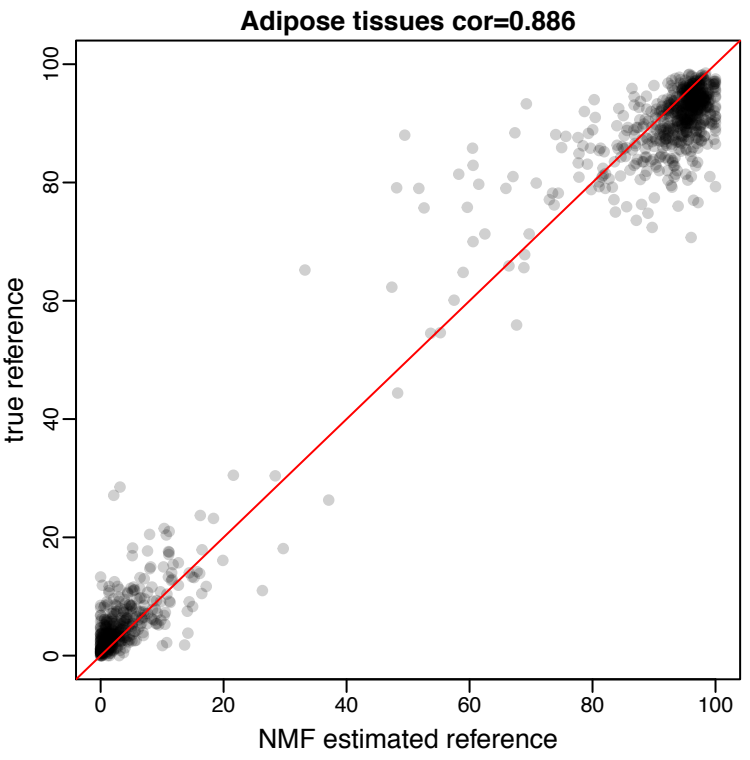

T cells
proportion
0.00
0.05
0.10
0.15
control
HCC
preg


B cells
proportion
0.00
0.02
0.04
0.06
0.08
0.10
0.12
control
HCC
preg

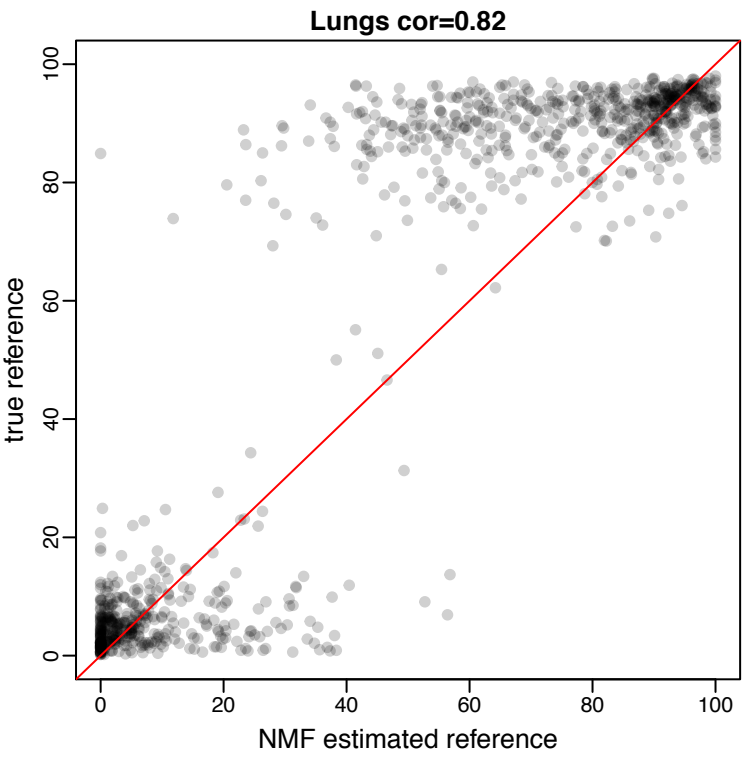

Neutrophils
proportion
0.2
0.3
0.4
0.5
control
HCC
preg


Placenta
proportion
0.0
0.1
0.2
0.3
0.4
0.5
control
HCC
preg

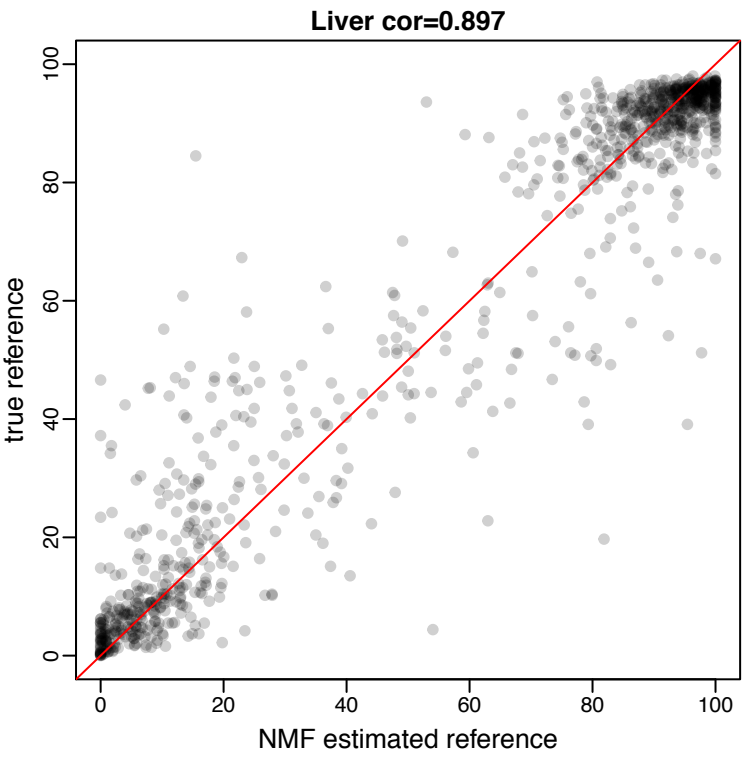
**Adipose tissues cor=0.886**



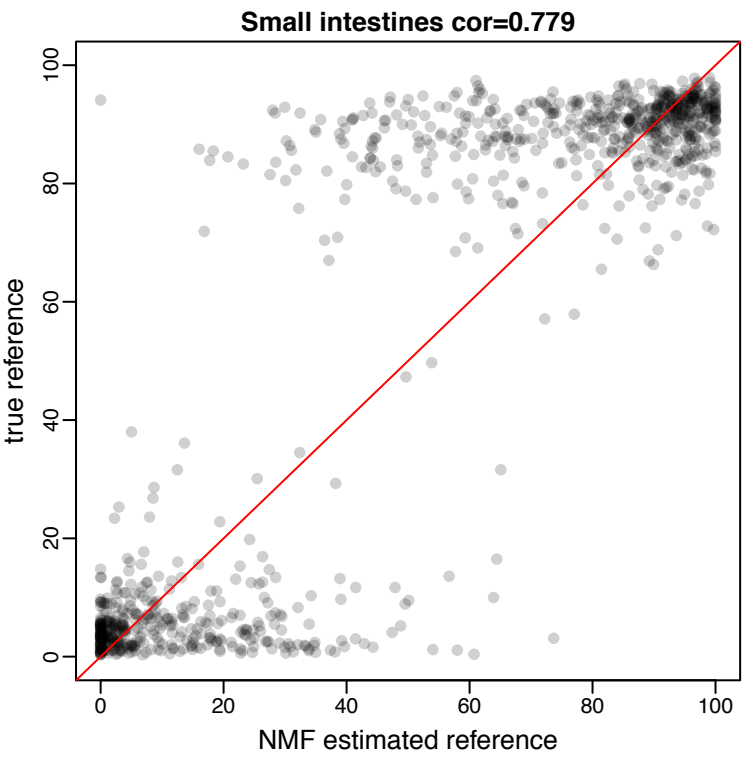

**Lungs cor=0.82**



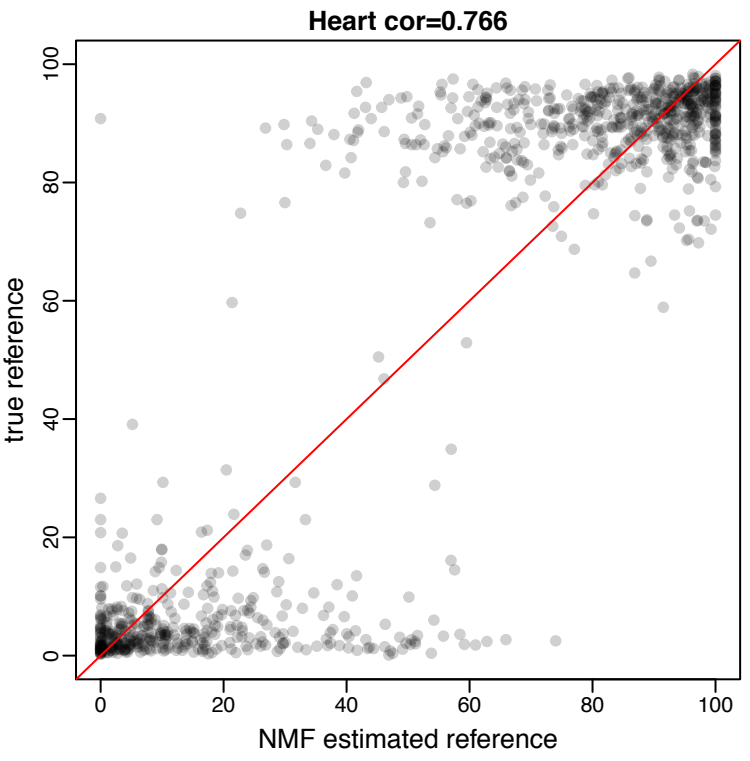

**Liver cor=0.897**



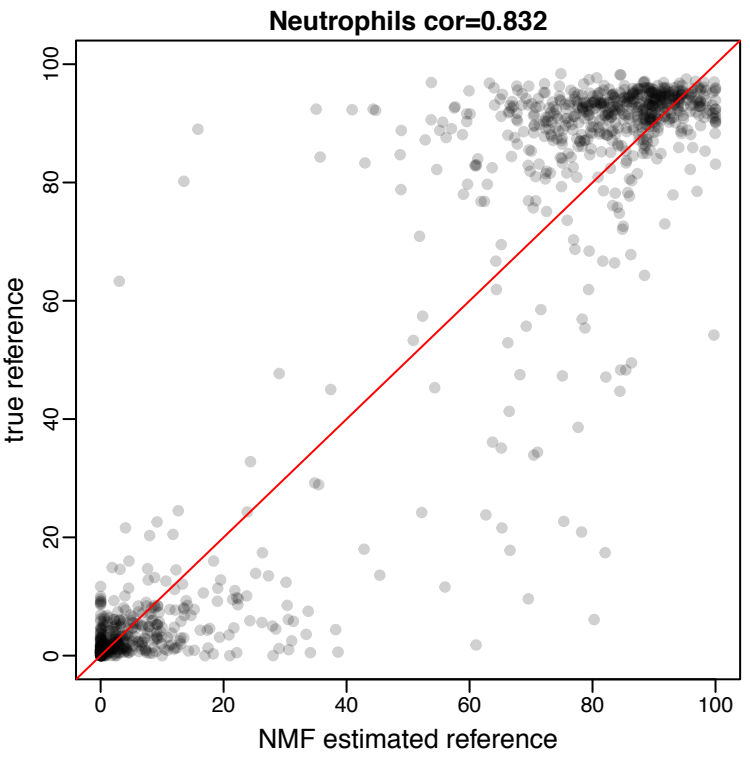

**Small intestines cor=0.779**



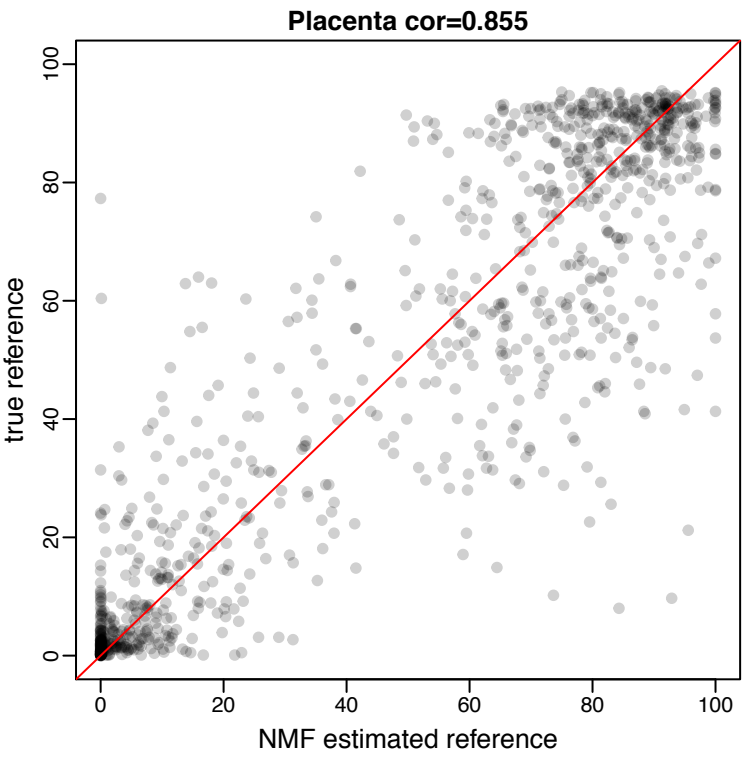

**Heart cor=0.766**



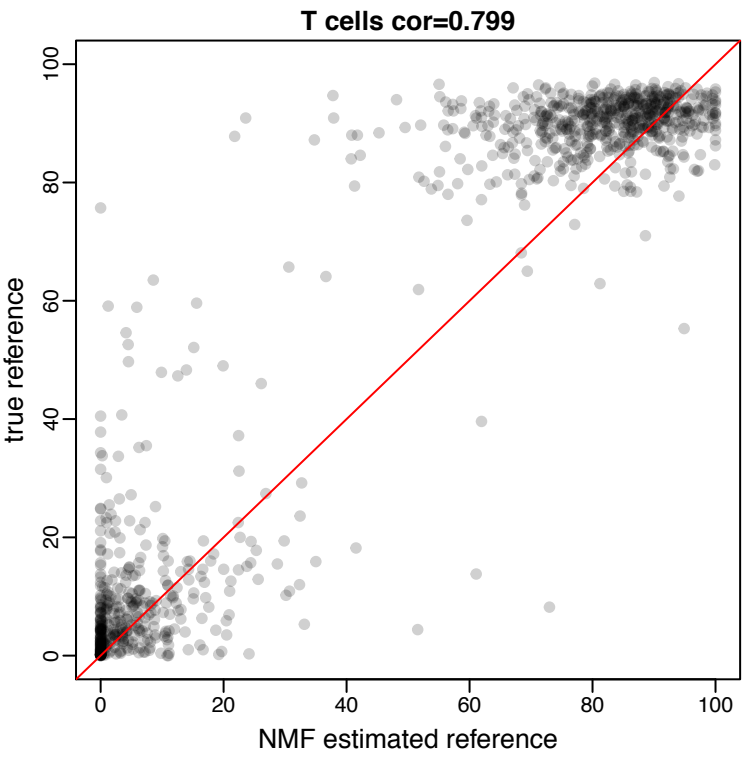

**Neutrophils cor=0.832**



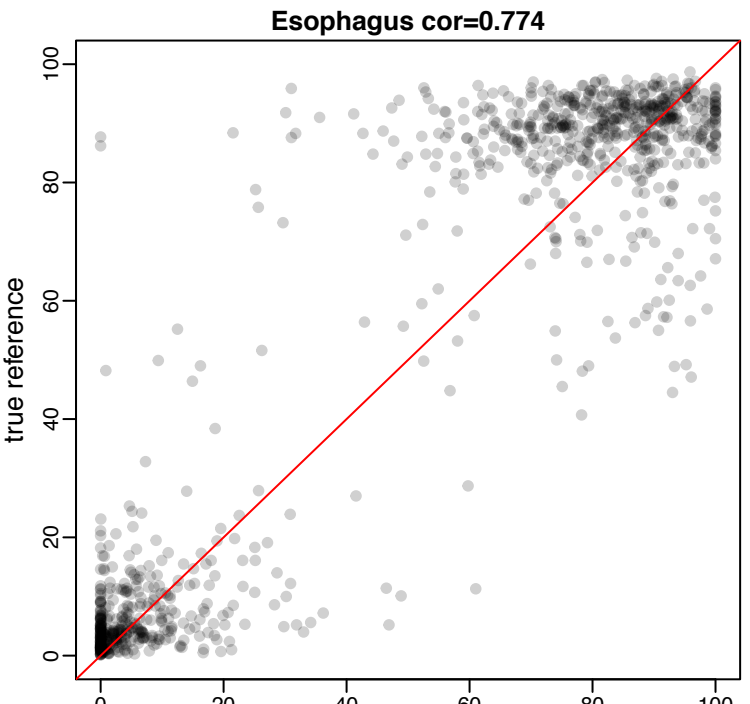

**Placenta cor=0.855**



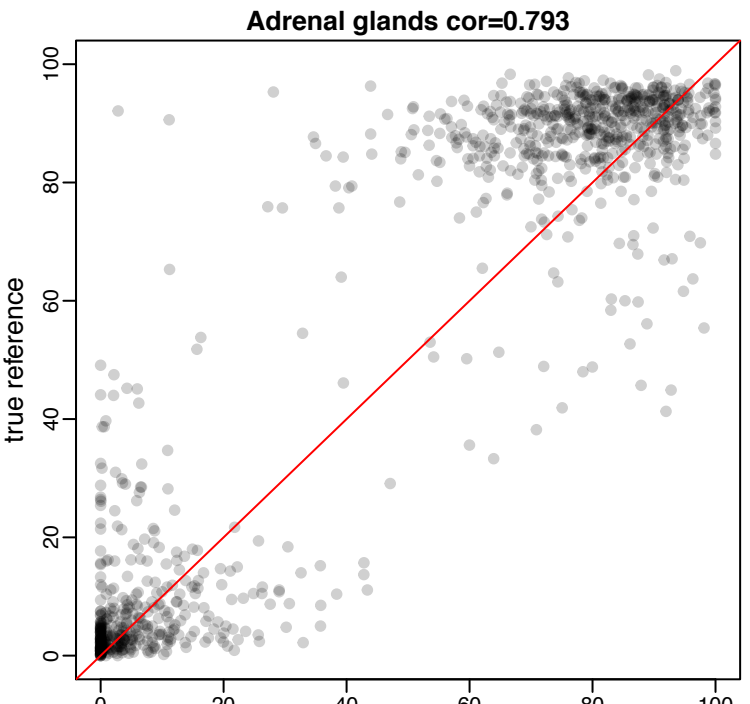

**T cells cor=0.799**



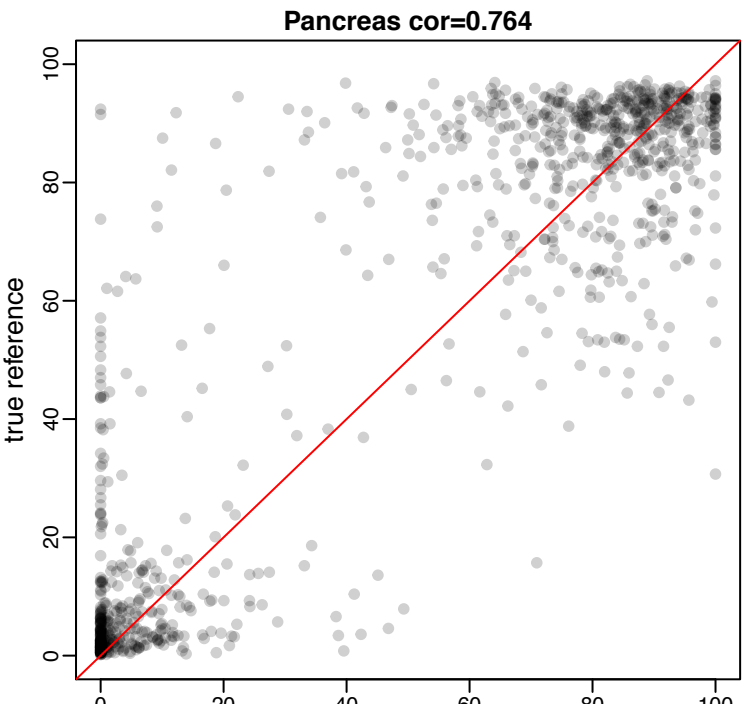

**Esophagus cor=0.774**



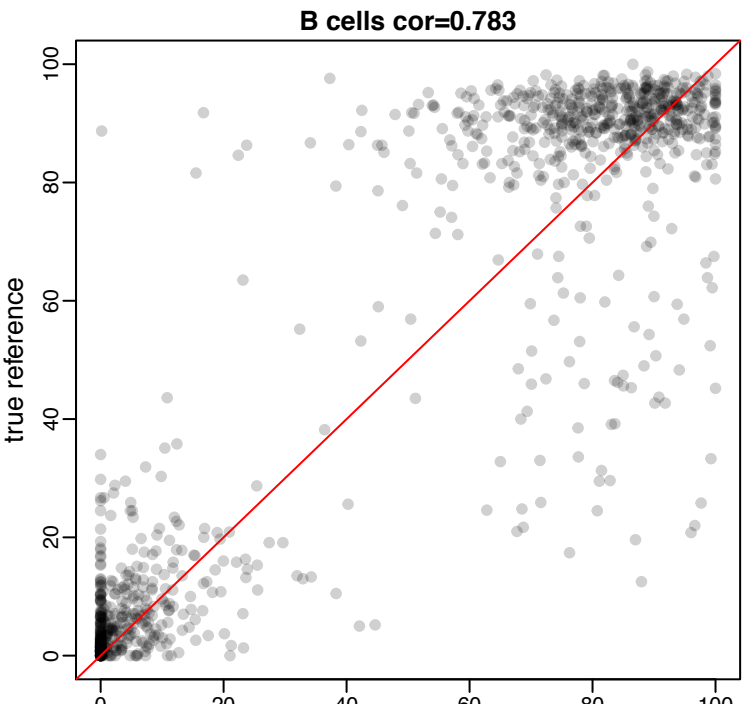

**Adrenal glands cor=0.793**



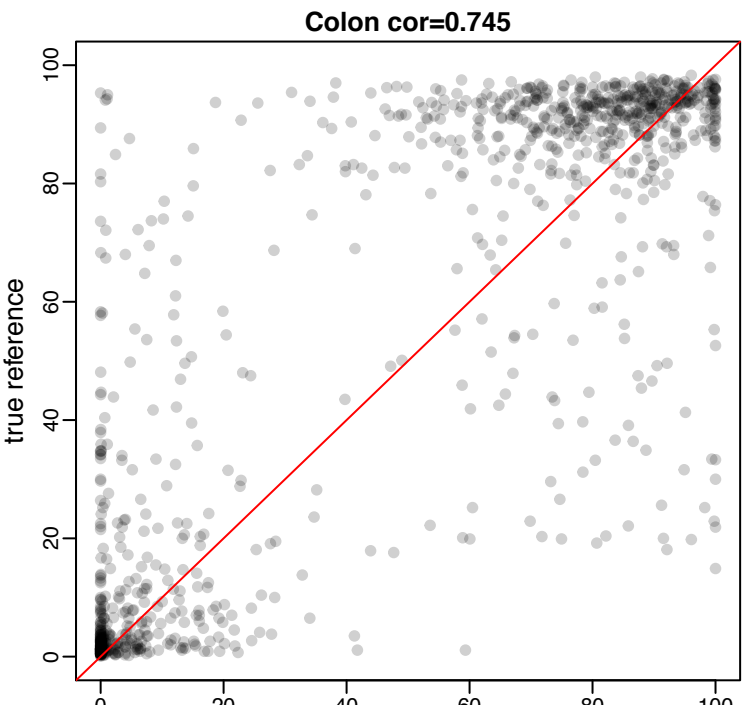

**Pancreas cor=0.764**

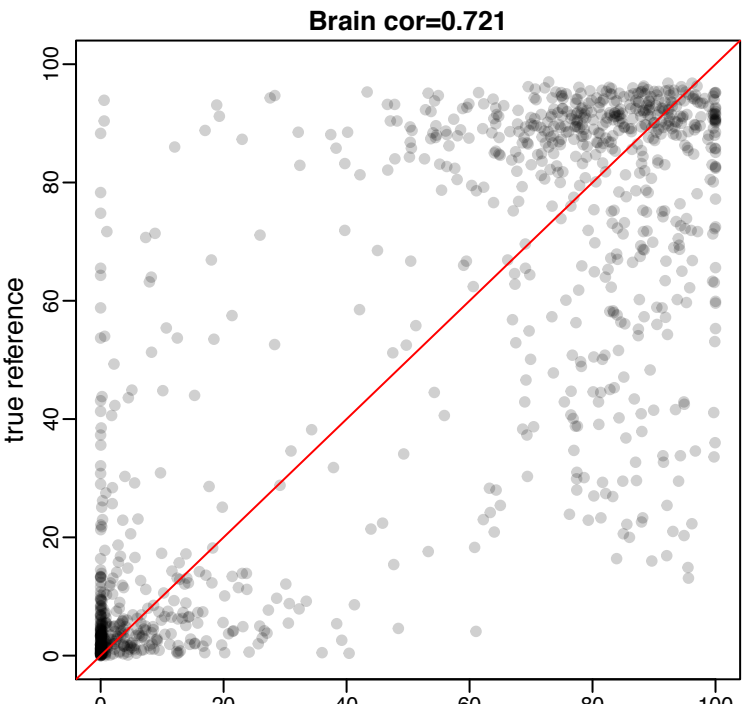



**B cells cor=0.783**




**Colon cor=0.745**




**Brain cor=0.721**