Paper Introduction:

Geeleher *et al. Genome Biology* (2016) 17:190
DOI 10.1186/s13059-016-1050-9

Genome Biology

METHOD Open Access

Cancer biomarker discovery is improved by accounting for variability in general levels of drug sensitivity in pre-clinical models

Paul Geeleher[1], Nancy J. Cox[2,3] and R. Stephanie Huang[1*]

Background

- Most biomarkers are initially identified through cell line drug sensitivity screening.
- Countless failures when biomarker predictions from pre-clinical data have been applied in the clinic.

- Three largest publicly available cell line pharmacogenomics studies:
  - GDSC: 138 compounds/drugs
  - CCLE: 24 compounds/drugs
  - CTRP: 481 compounds/drugs

MDR: Multi-drug resistance

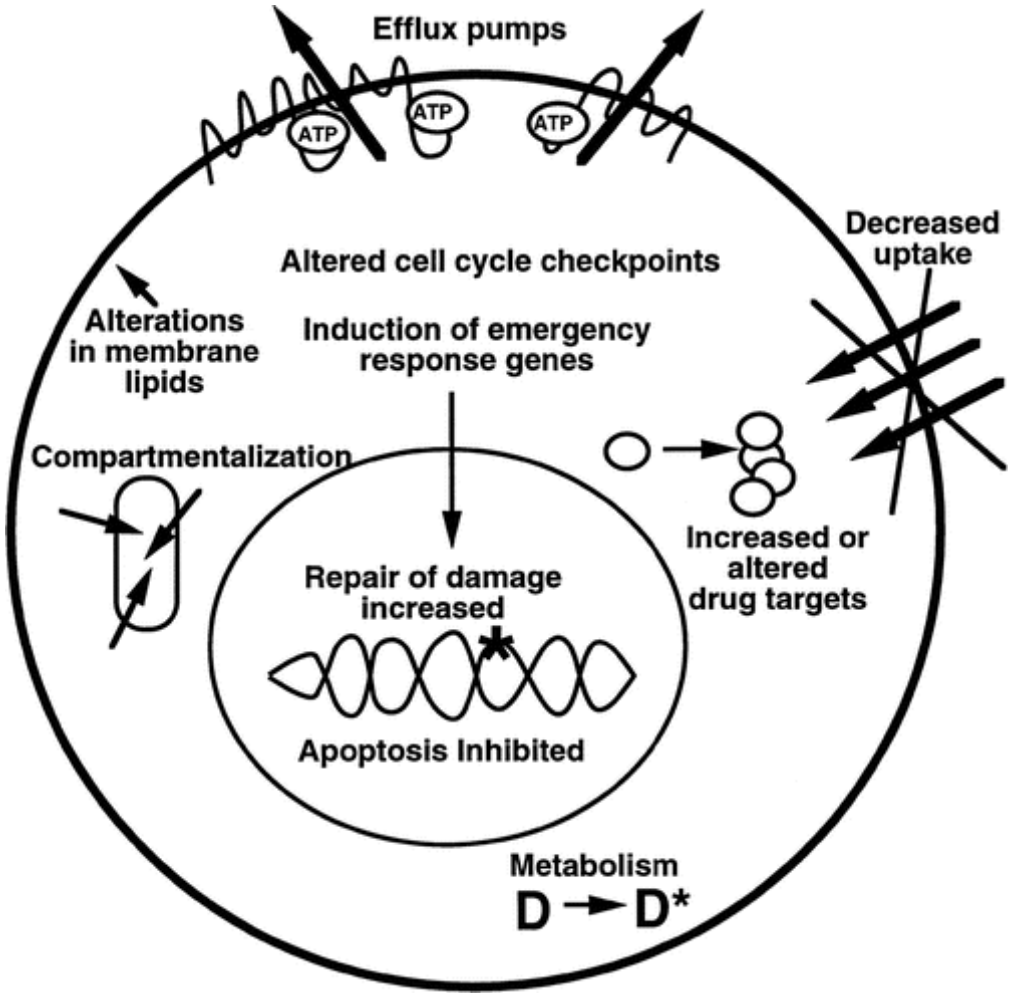
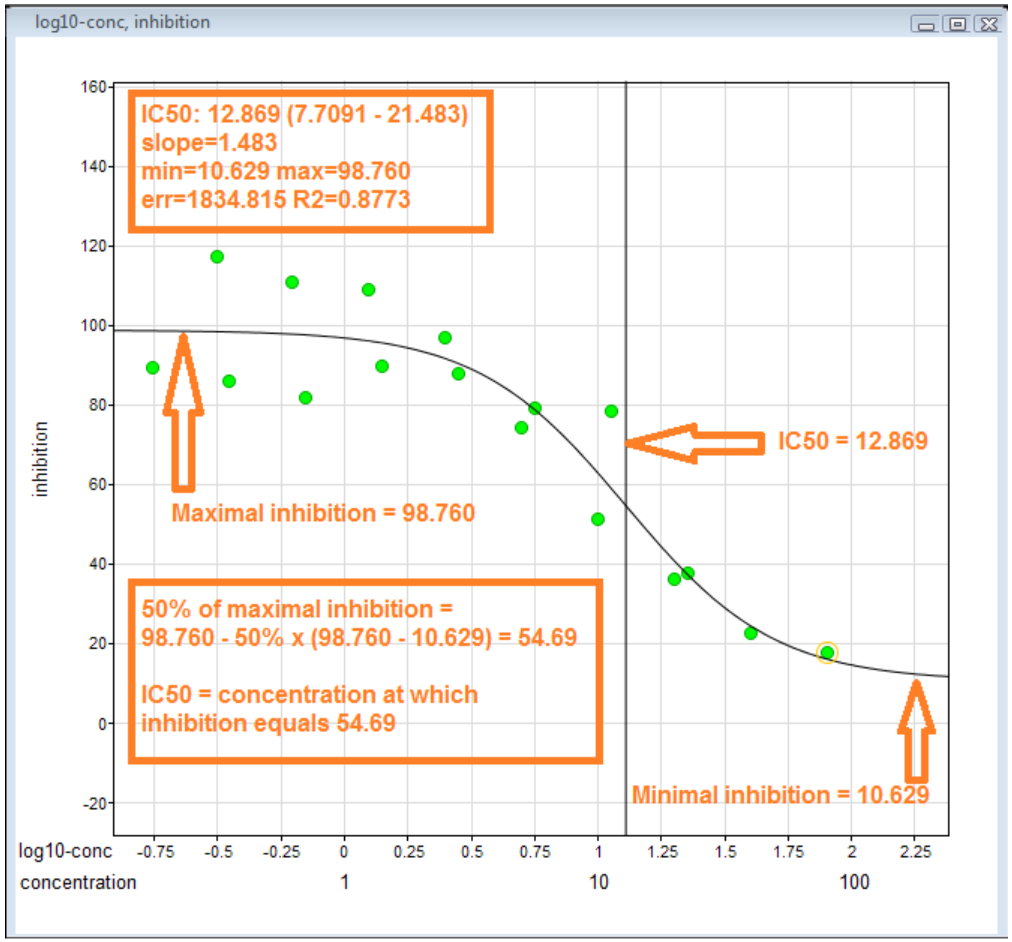


Figure: Ways of MDR. Figures from Gottesman, 2002.

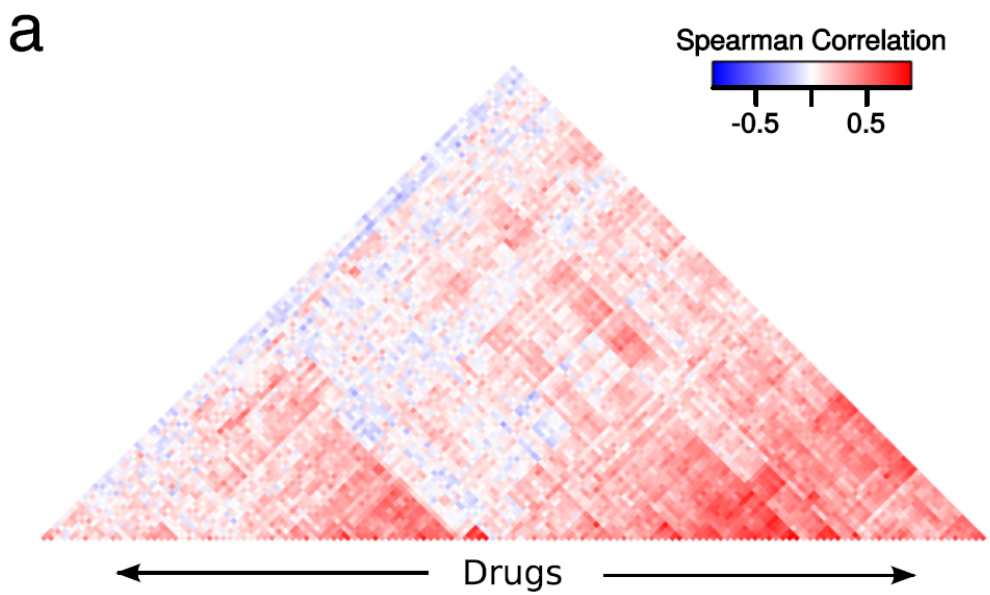
GLDS: General Level of Drug Sensitivity

IC_{50}: half maximal inhibitory concentration.

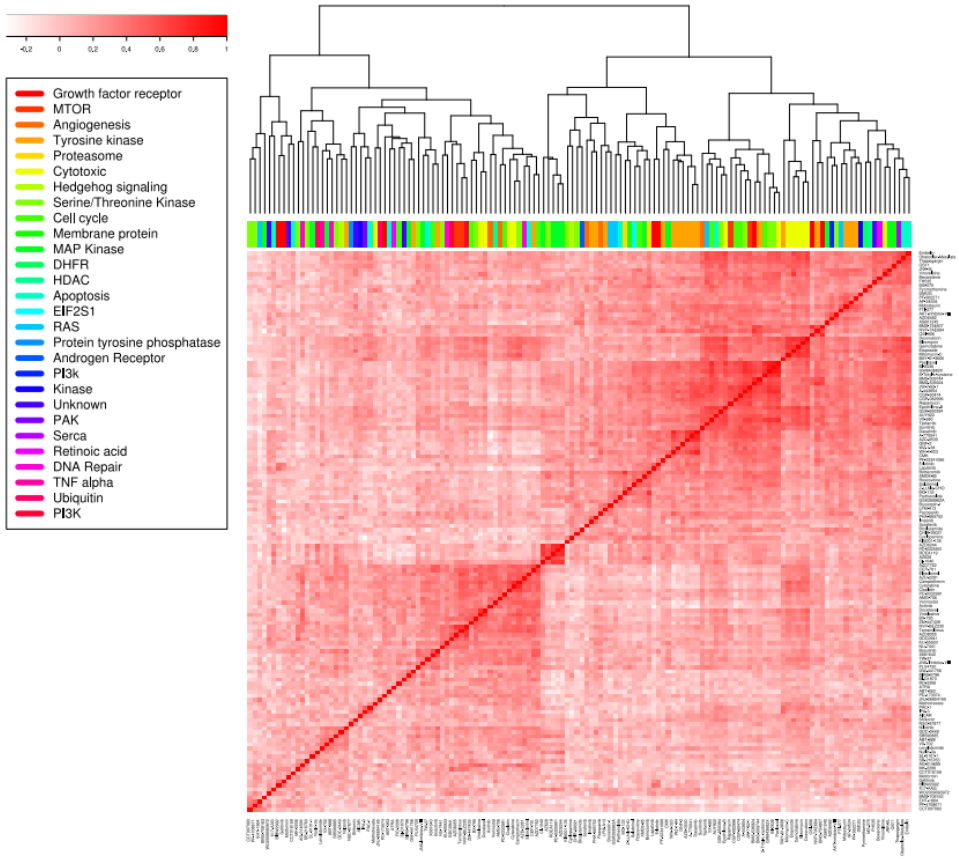


GLDS: General Level of Drug Sensitivity

(1). Pairwise correlation between IC_{50} values of all 138 drugs across all 714 cell lines in GDSC.



Similar classes of drugs don't clustering together strongly.



Strong correlations are not only observed between drugs within the same class, but also drigs with different mechanisms.

(2). Iterative matrix completion of IC_{50} values matrix $\boldsymbol{X}$.

- Initially, impute missing values by mean values of the same drug.
- Iteration:
  - Estimate PCs of $\boldsymbol{X}$;
  - For each cell line, using PCs of other cell lines to estimate the missing values. (lasso regression)

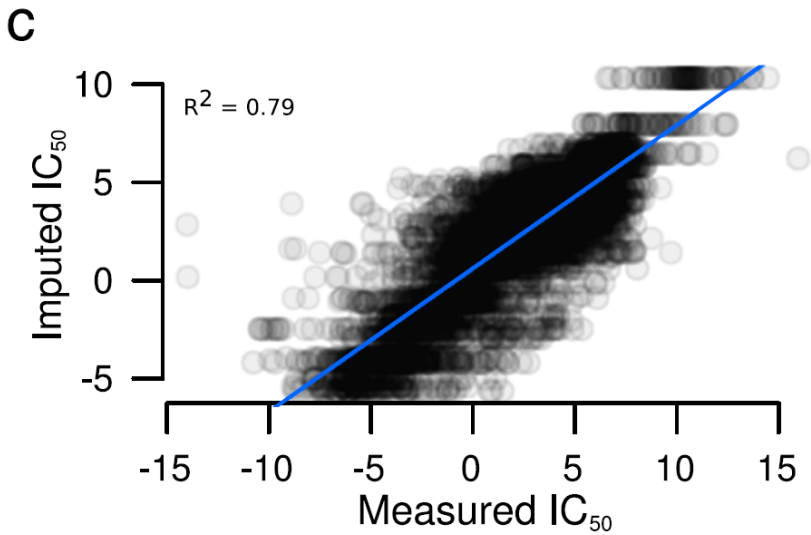


Figure: Imputed against measured IC_{50} values from 8-fold CV.

(3). Summarize the pattern of GLDS using SVD/PCA.

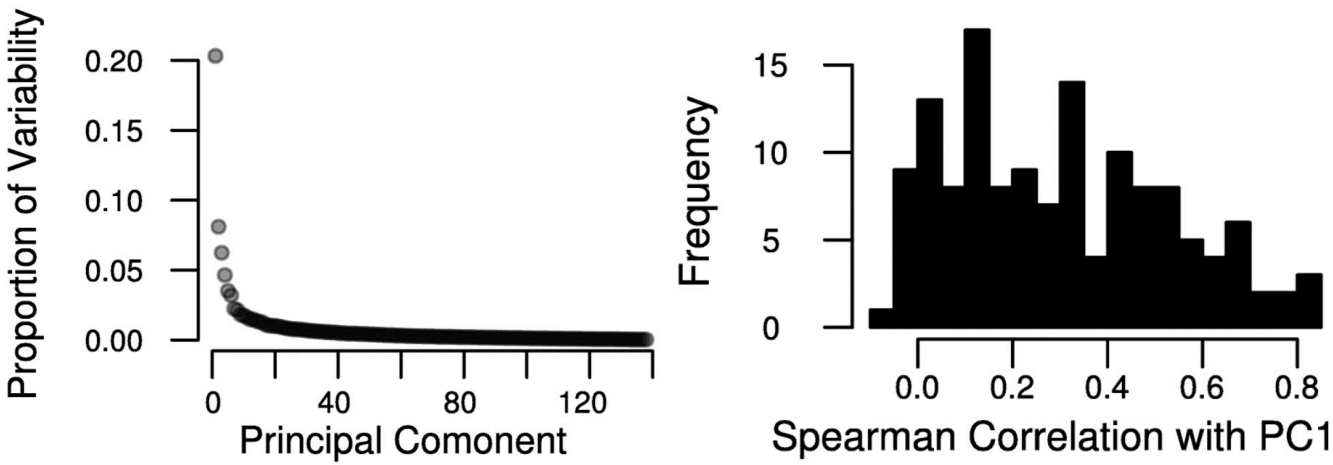


Figure: PCA of drug's IC_{50} values.

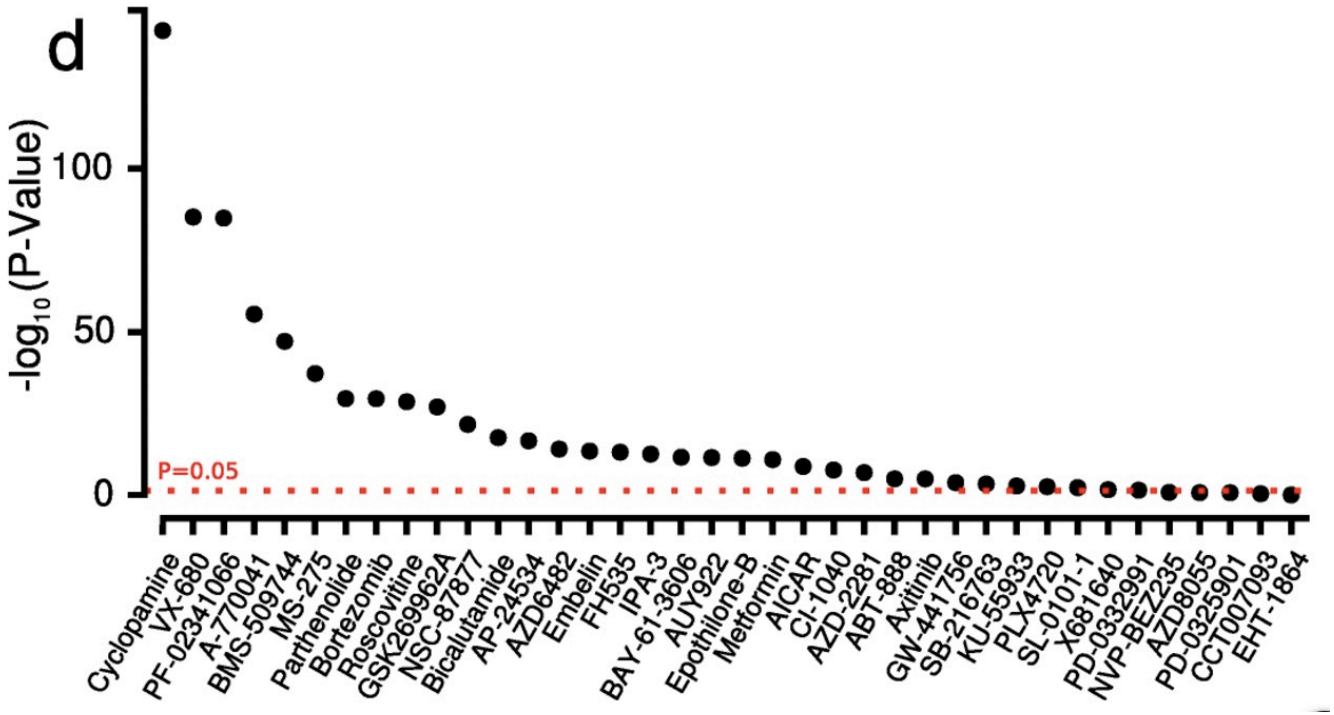


Figure: Use 38 drugs' PCs to predict other 100 drugs' IC_{50} values.

Cell lines tend to exhibit sensitivity or resistance to many drugs, regardless of canonical drug mechanisms.

(4). Biological drivers of GLDS in cell lines.

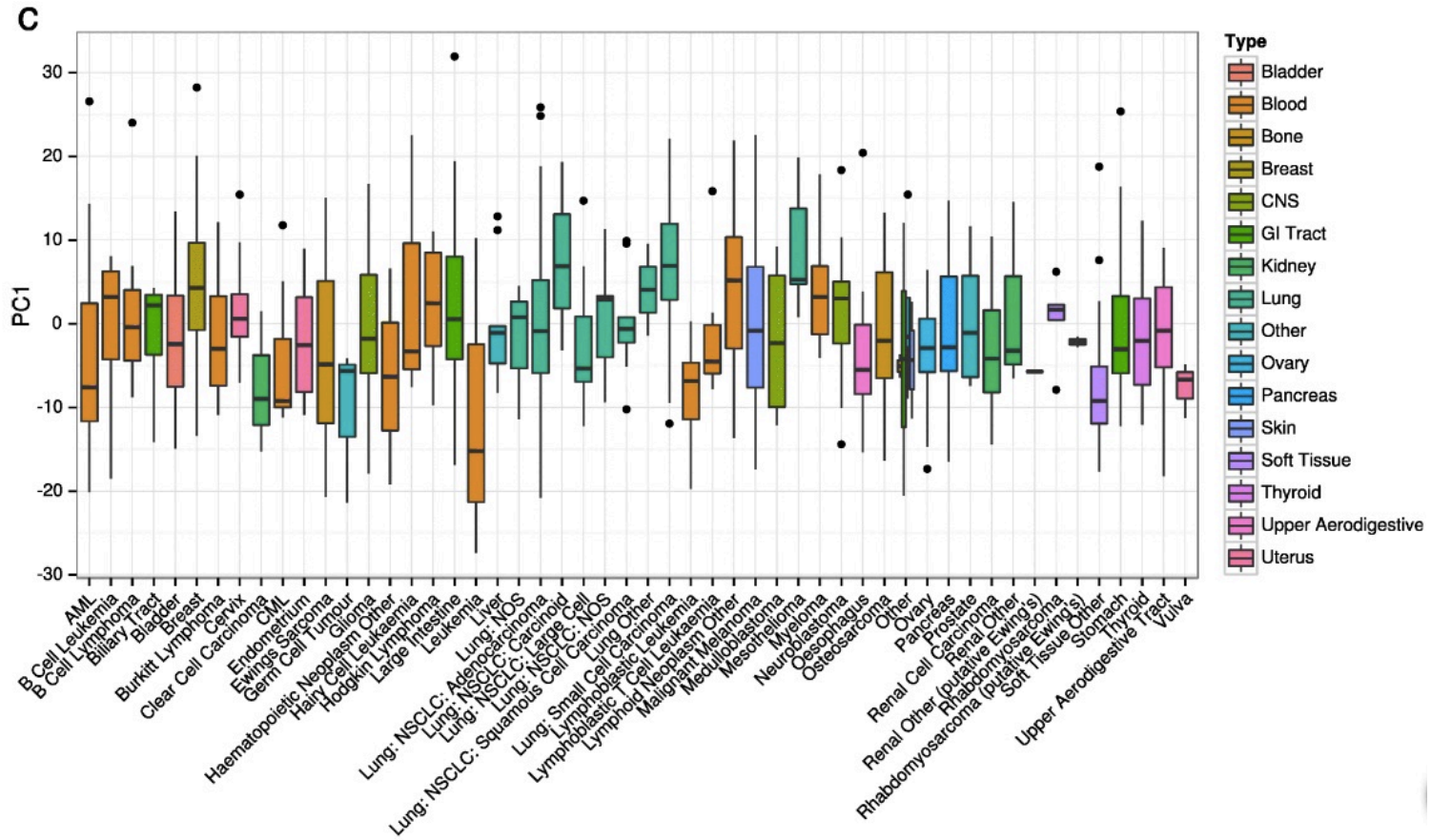
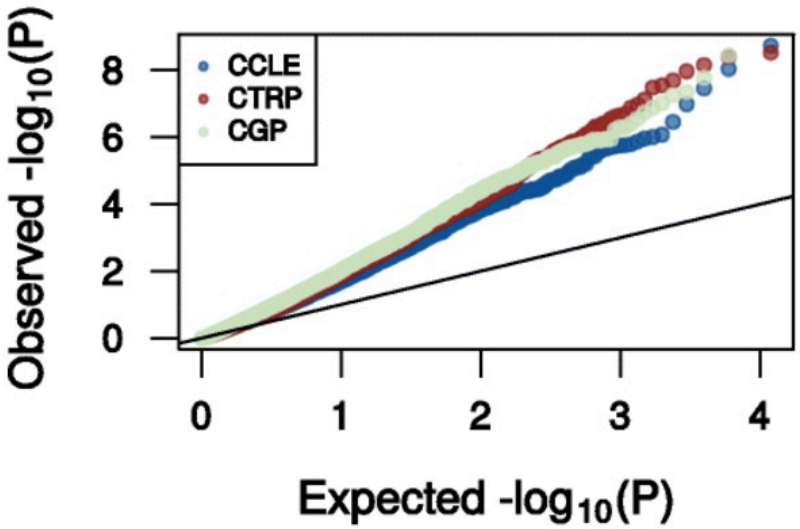


Figure: Boxplot of PC1 (estimated in all 714 cell lines) against tissue-of-origin in CGP.

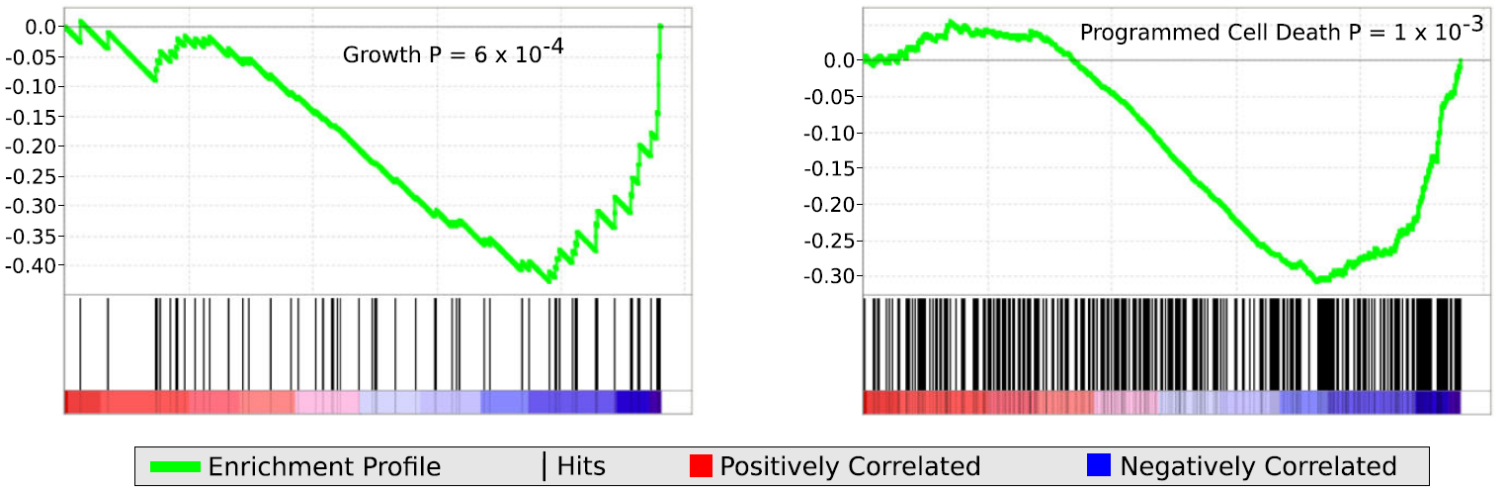
Explain only 8% of variance.

Detect genes associated with GLDS:



- 810, 4,680 and 4,457 were detected significantly associated with $PC1$ in GDSC, CCLE, CTRP, respectively.
- 185 genes were found in all 3 studies.

Gene Set Enrichment Analysis:



- $MDR1$ gene (efflux protein) is associated with GLDS in all 3 studies.
- All associated with cell cycle, growth and apoptosis in GDSC. In CTRP,
  - "Growth" ia associated with $PC2$.
  - "Regulation of apoptosis" is associated with $PC3$ and $PC5$.
  - "Lipid Transporter Activity" is associated with $PC1$.

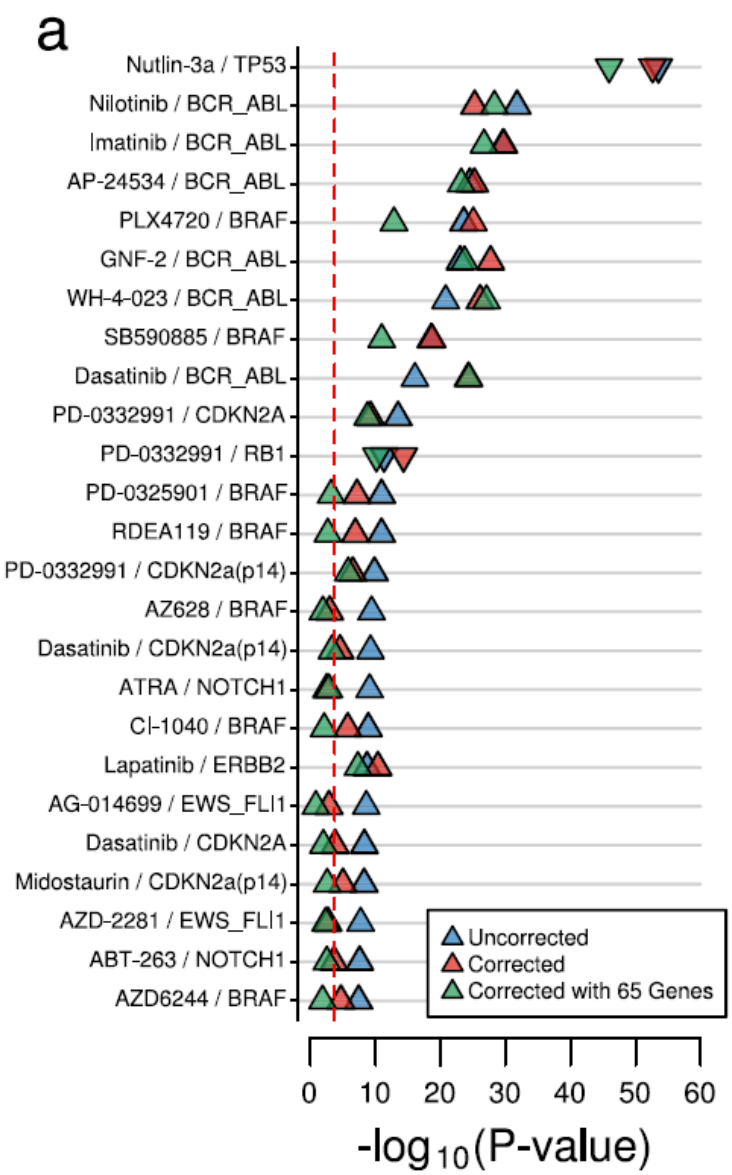
Prediction based on GLDS

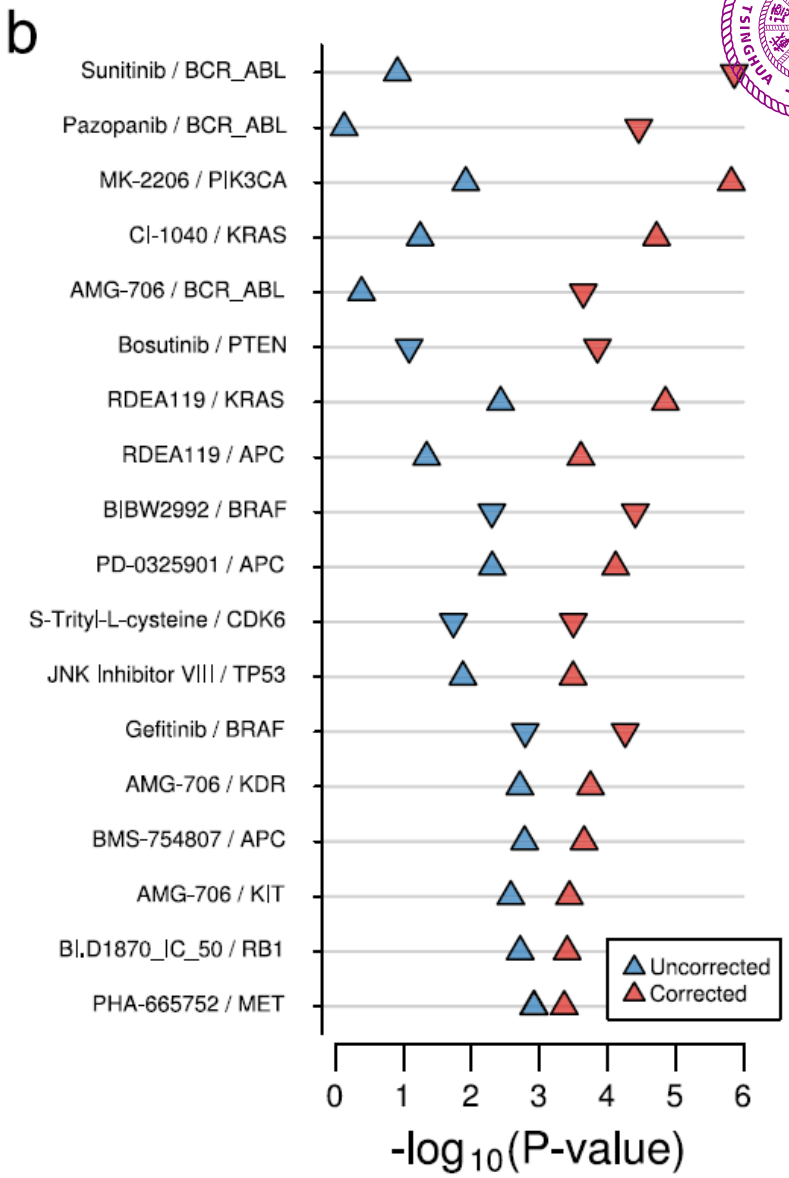
Why controlling for GLDS:

- Cancer drug biomarkers, are often subsequently tested on relapsed patients, who have undergone multiple rounds of chemotherapy and developed resistance to many drugs.
- New drugs are often tested in addition to existing standard-of-care multi-drug regimes.

For each drug,

- Select drugs as "negative controls": unrelated mechanism of action; not highly correlated.
- Use the fist 10 PCs of these drugs as covariates.
- Use linear model to test IC_{50} values against the mutation status of cancer genes.


a
Nutlin-3a / TP53
Nilotinib / BCR_ABL
Imatinib / BCR_ABL
AP-24534 / BCR_ABL
PLX4720 / BRAF
GNF-2 / BCR_ABL
WH-4-023 / BCR_ABL
SB590885 / BRAF
Dasatinib / BCR_ABL
PD-0332991 / CDKN2A
PD-0332991 / RB1
PD-0325901 / BRAF
RDEA119 / BRAF
PD-0332991 / CDKN2a(p14)
AZ628 / BRAF
Dasatinib / CDKN2a(p14)
ATRA / NOTCH1
Cl-1040 / BRAF
Lapatinib / ERBB2
AG-014699 / EWS_FLI1
Dasatinib / CDKN2A
Midostaurin / CDKN2a(p14)
AZD-2281 / EWS_FLI1
ABT-263 / NOTCH1
AZD6244 / BRAF
Uncorrected
Corrected
Corrected with 65 Genes
0
10
20
30
40
50
60
-log10(P-value)


b
Sunitinib / BCR_ABL
Pazopanib / BCR_ABL
MK-2206 / PIK3CA
Cl-1040 / KRAS
AMG-706 / BCR_ABL
Bosutinib / PTEN
RDEA119 / KRAS
RDEA119 / APC
BIBW2992 / BRAF
PD-0325901 / APC
S-Trityl-L-cysteine / CDK6
JNK Inhibitor VIII / TP53
Gefitinib / BRAF
AMG-706 / KDR
BMS-754807 / APC
AMG-706 / KIT
BI.D1870_IC_50 / RB1
PHA-665752 / MET
Uncorrected
Corrected
0
1
2
3
4
5
6
-log10(P-value)

- Of 25 significant associations before,
  - 9, P values improved. (supported by existing evidence).
  - 4, no longer significant ($FDR > 0.05$):
    PARP inhibitors - EWS-FLI1
- 18 new mutation-drug associations:
  - MK-2206 - PIK3CA mutation (phase II clinical studies)
  - CI-1040 sensitivity - KRAS mutation (in vitro and in vivo data)
  - Bosutinib - PTEN wild-type (mechanism documented)

- Reproducibility between large pharmacogenomics datasets:
  - 15 drugs and 63 sequenced cancer genes common in CCLE and GDSC
  - Improved from 47%(11 of 23 significant associations) to 62.5% (10 of 16)

GLDS estimated by expression

- Identify genes most associated with GLDS using a linear model.
- In GDSC, 65 genes are identified.
- Compare results of uncorrected, 65 genes corrected with GLDS corrected in other dataset:
  - $FDR < 0.05$, 62(uncorrected), 53(expression corrected) (62% and 68% of GLDS corrected)
  - $FDR < 0.25$, 760(uncorrected, only 232 GLDS-corrected) and 368(expression corrected, 201 GLDS-corrected)
- In TCGA breast cancer samples, 32 of 60 (sequenced of 65) genes were associated with alive/dead status.

Conclusion

- Identify GLDS as a novel phenomenon comfounding biomarker discovery.
- This bias may be found in all pre-clinical models: cell lines, mouse xenografts and in data derived directly from clinical studies.
- Develop methods to estimate and remove this confounder.
- Improve dramatically the clinical success rate of drug discovery.

Paper Introduction:

ARTICLES

Robust enumeration of cell subsets from tissue expression profiles

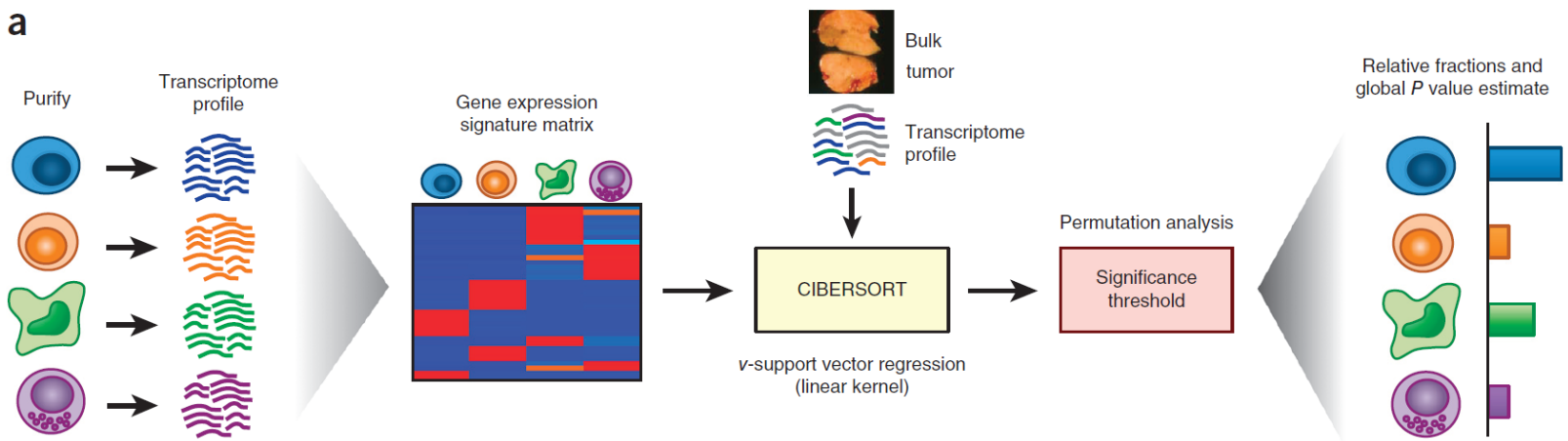
Aaron M Newman[1,2,10], Chih Long Liu[1,2,10], Michael R Green[2,3,9], Andrew J Gentles[3,4], Weiguo Feng[5], Yue Xu[6], Chuong D Hoang[6], Maximilian Diehn[1,5,7] & Ash A Alizadeh[1–3,7,8]

Introduction

- Levels of different cell types are associated with tumour growth, cancer progression and patient outcome.
- Can we predict fractions of multiple cell types in **gene expression profiles (GEP)**?
  - mixtures with unknown contents and noise (for example, solid tumour);
  - mixture of closely related cell types (for example, naïve and memory B cells).

CIBERSORT Methods

Cell-type Identifiction By Estimating Relative Subsets Of RNA Transcripts



- **Deconvolution model:**

$$\boldsymbol{m} = \boldsymbol{f} \times \boldsymbol{B}$$

  - $\boldsymbol{m}$: a mRNA mixture ($1 \times g$, suppose we use g genes.)
  - $\boldsymbol{f}$: fraction of cell -types ($1 \times k$, suppose we have k cell types.)
  - $\boldsymbol{B}$: signature matrix

Genes with expression profiles enriched in each cell type can be leveraged to impute unknown cell fractions from mixture profiles.

- **Signature matrix:**
  - Obtain purified or enriched cell populations
  - Detect significanlly differentially expressed genes between each cell population and all other populations($q < 0.3$)
  - Adaptively select genes by condition number:

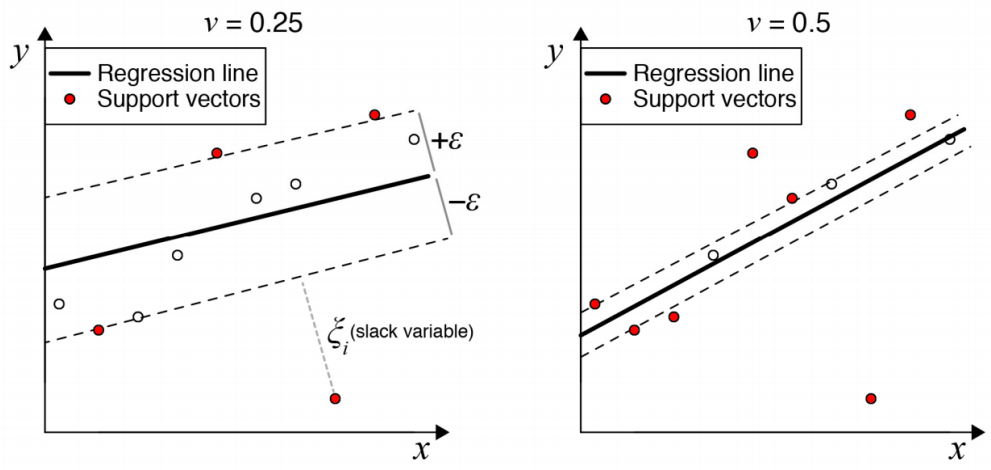
$$\kappa(\boldsymbol{B}) = \frac{\sigma_{\max}(\boldsymbol{B})}{\sigma_{\min}(\boldsymbol{B})}$$

    Use top G marker genes from each type to combine a matrix $\boldsymbol{B}$ and iterate G from 50 to 200, selecting the lowest condition number.
  - Normalized $\boldsymbol{B}$ to zero mean and unit variance.

Purified cell populations: (LM22)

LM22 Cells	Cell Type Description	Reference (PMID)	Authors	Cell Separation Method	Markers used	Purity
B cells	B cells naive	15789058	Abbas AR et al.	MACS® CD138 microbeads and CD19 microbeads	CD19+CD27- IgG/A-	Not stated
	B cells memory	15789058	Abbas AR et al.	MACS® CD138 microbeads and CD19 microbeads, then FACS	CD19+ CD27+	Not stated
PCs	Plasma cells	15789058	Abbas AR et al.	MACS® CD138 microbeads, then FACS	CD20+, CD138+ and CD19+	Not stated
CD8 T	T cells CD8	15789058	Abbas AR et al.	RosetteSep™ CD8+ T-cell enrichment cocktail, CD-8 subset	CD3, CD8, CD45RA	>90
CD4 T cells	T cells CD4 naive	16791882	Rasheed AU et al.	Ficoll, then MACS CD4+ T cell isolation kit	CD4+	>98%
	T cells CD4 memory resting	15789058	Abbas AR et al.	Ficoll, then FACS	$CD45RO^{high}$	Not stated
	T cells CD4 memory activated	15789058	Abbas AR et al.	Ficoll, then FACS, then activated by anti-CD3 (plate-bound) + anti-CD28 (soluble)	$CD45RO^{high}$; CD69, CD25 for activation	>90%
	T cells follicular helper	16791882	Rasheed AU et al.	Ficoll, then MACS CD4+ T cell isolation kit, then FACS	$CXCR5^{hi}$, $ICOS^{hi}$	>95%
	T cells regulatory (Tregs)	16702978	Ocklenburg F et al.	Ficoll-Hypaque, then MACS CD4+ T cell isolation kit, then FACS	CD4+ $CD25^{hi}$	>98%
Gamma delta T cells	T cells gamma delta	16339519	Chtanova T et al.	Ficoll, then FACS	Not stated	Not stated
NK cells	NK cells resting	15789058	Abbas AR et al.	RosetteSep™ NK-cell enrichment cocktail + CD2 Microbeads	CD56	Not stated
	NK cells activated	15789058	Abbas AR et al.	RosetteSep™ NK-cell enrichment cocktail + CD2 Microbeads + IL2 or IL15 for activation	CD56 + CD69	Not stated
Monocytes and Macrophages	Monocytes	15789058	Abbas AR et al.	MACS® CD14 Microbeads, monocyte subset	N/A	Not stated
	Macrophages M0	15789058	Abbas AR et al.	Differentiated from monocytes	None known; identified by morphology and phagocytic capability	Not stated
	Macrophages M1	17244792	Cho HJ et al.	Histopaque 1.077, then Miltenyi negative selection monocyte isolation kit and LS columns, then differentiated with 1% medium supplement nutridoma-HU + 100 nM M-CSF, then activated with 20 ng/ml IFN-g+ 100 ng/ml LPS	None known; identified by morphology and phagocytic capability	>97% (at monocyte stage)
	Macrophages M2	17244792	Cho HJ et al.	Histopaque 1.077, then Miltenyi negative selection monocyte isolation kit and LS columns, then differentiated with 1% medium supplement nutridoma-HU + 100 nM M-CSF, then activated with 20 ng/ml IFN-g+ 100 ng/ml LPS and 20 ng/ml IL-4	None known; identified by morphology and phagocytic capability	>97% (at monocyte stage)
Dendritic cells	Dendritic cells resting	15789058	Abbas AR et al.	Monocytes differentiated with 17 ng/ml IL4, and 67 ng/ml GMCSF	N/A	Not stated
	Dendritic cells activated	15789058	Abbas AR et al.	Monocytes differentiated with 17 ng/ml IL4, and 67 ng/ml GMCSF, then stimulated with 1 ug/ml LPS	N/A	Not stated
Mast cells	Mast cells resting	16339519	Chtanova T et al.	Ficoll of cord blood, then 100 ng/ml SCF + 10 ng/ml IL-10 + 5 ng/ml IL-6	N/A	95%
	Mast cells activated	16339519	Chtanova T et al.	Ficoll of cord blood, then 100 ng/ml SCF + 10 ng/ml IL-10 + 5 ng/ml IL-6 + IgE receptor activation	N/A	
Eos	Eosinophils	16339519	Chtanova T et al.	0.6% Dextran T500, then Percoll gradient (70%, 80%), then negative selection with MACS CD16 Microbeads	N/A	>97%
PMNs	Neutrophils	16339519, 15789058	Chtanova T et al., Abbas AR et al.	0.6% Dextran T500, then Percoll gradient (70%, 80%), then negative selection with MACS anti-CCR3 + anti-mouse IgG Microbeads	CD62L	>97%

- **Support vector regression:**



Fit a hyperplane $y = \langle \boldsymbol{w}, \boldsymbol{x} \rangle + b$, such that:

$$\min \quad \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$$

$$\text{s.t.} \quad \begin{cases} y_i - \langle \boldsymbol{w}, \boldsymbol{x_i} \rangle - b \leq \epsilon + \xi_i \\ -y_i + \langle \boldsymbol{w}, \boldsymbol{x_i} \rangle + b \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

- **Summary of methods:**

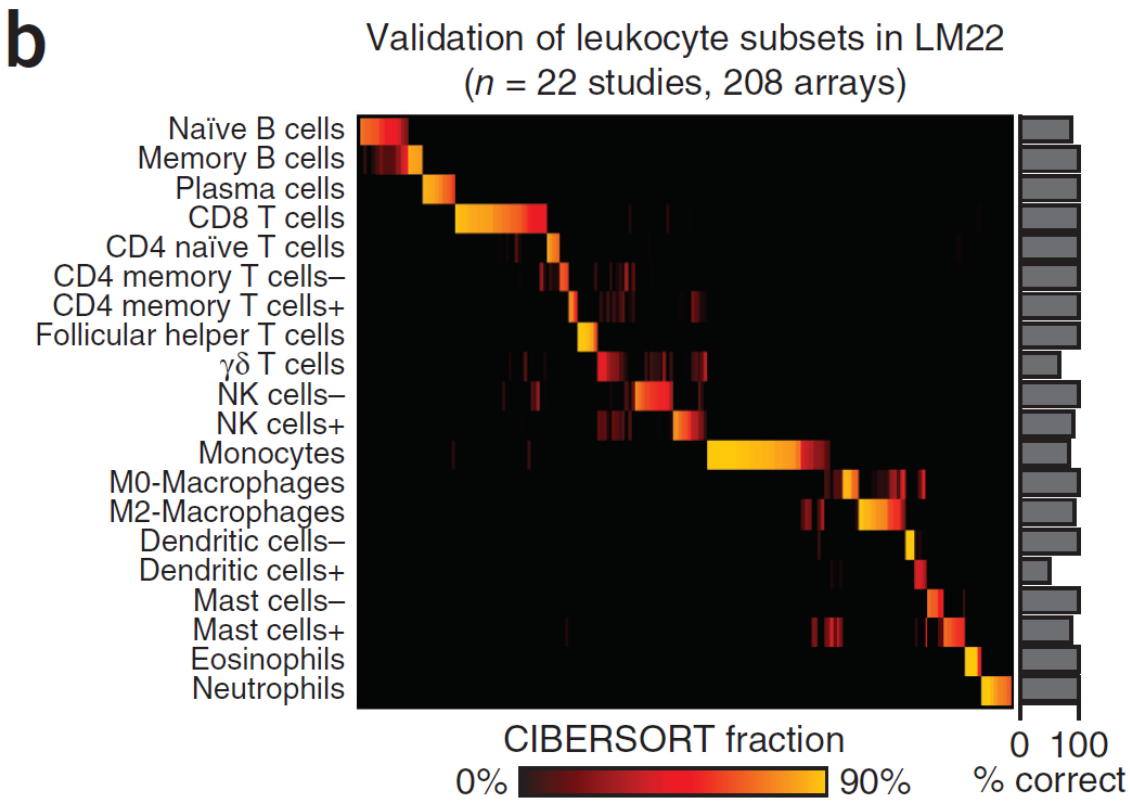
$$\boldsymbol{m} = \boldsymbol{f} \times \boldsymbol{B}$$

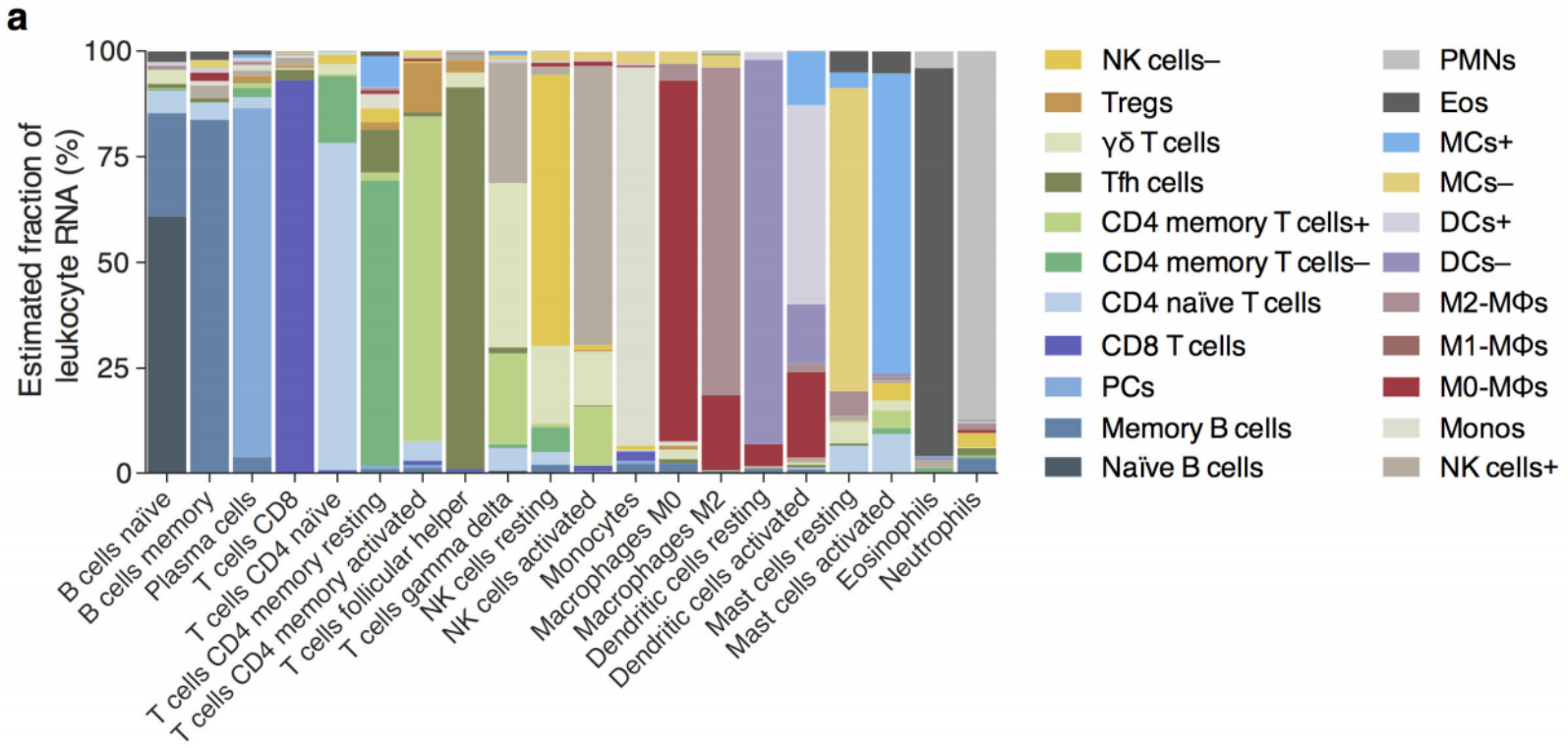
  - Construction of signature matrix $\boldsymbol{B}$
  - Using SVR to obatin $\hat{\boldsymbol{f}}$
  - $\boldsymbol{f}_i = \max\{\hat{\boldsymbol{f}}_i, 0\}$ and then normalized to 1.

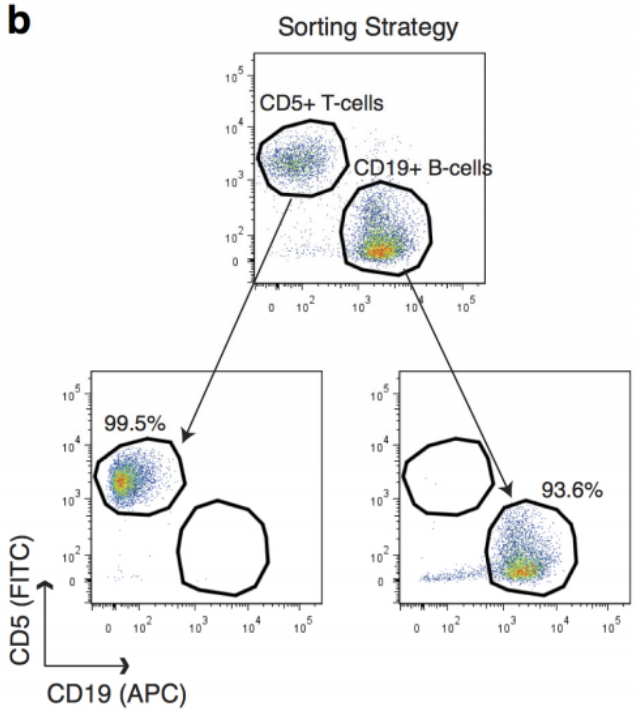
- Robustness to noise and overfitting owning to SVR and feature selection of genes from signature matrix.

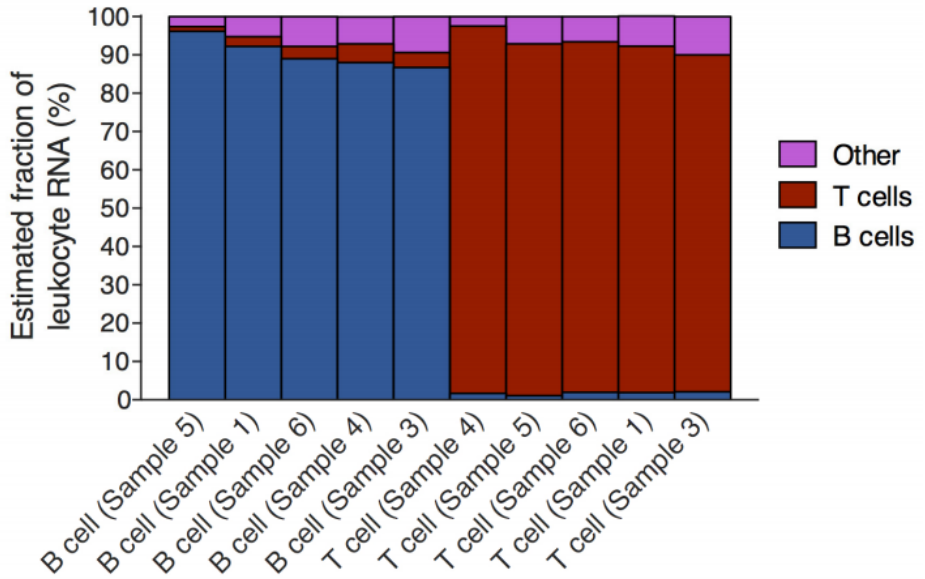
Results

- Perfomance in external datasets of variably purified leukocyte subsets:

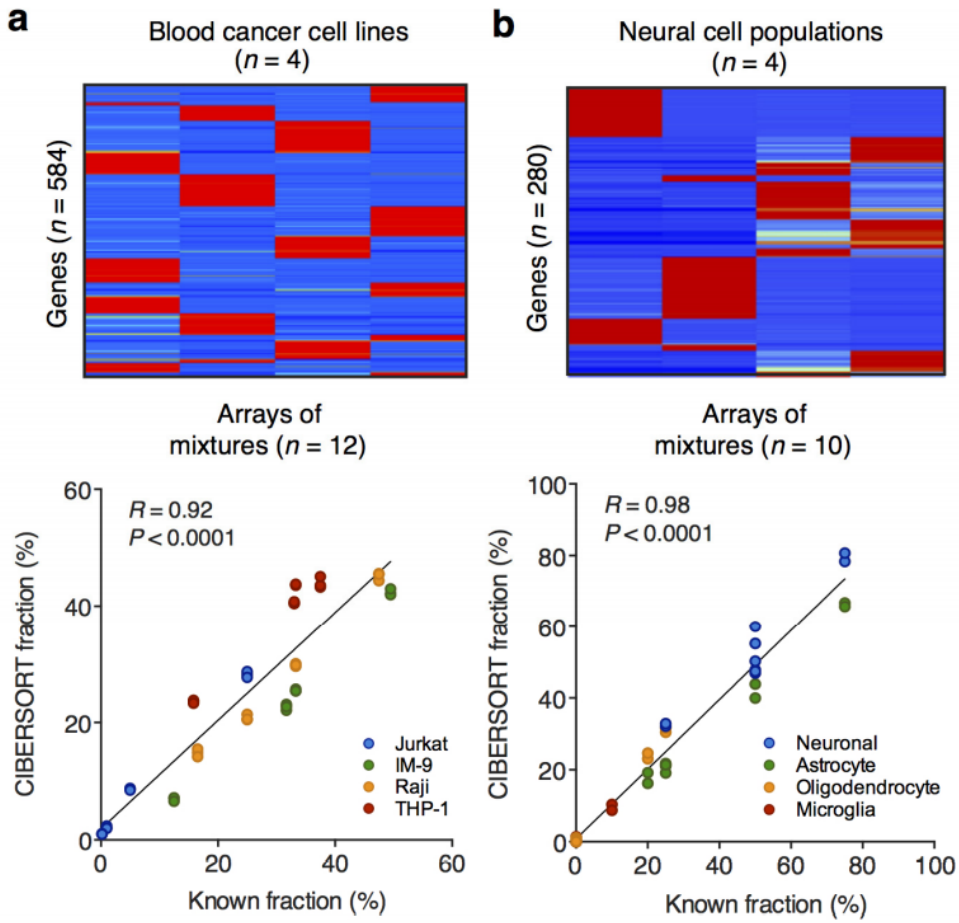



a
Estimated fraction of leukocyte RNA (%)
100
75
50
25
0
B cells naïve
B cells memory
Plasma cells
T cells CD8
T cells CD4 naïve
T cells CD4 memory resting
T cells CD4 memory activated
T cells follicular helper
T cells gamma delta
NK cells resting
NK cells activated
Monocytes
Macrophages M0
Macrophages M2
Dendritic cells resting
Dendritic cells activated
Mast cells resting
Mast cells activated
Eosinophils
Neutrophils
NK cells–
Tregs
γδ T cells
Tfh cells
CD4 memory T cells+
CD4 memory T cells–
CD4 naïve T cells
CD8 T cells
PCs
Memory B cells
Naïve B cells
PMNs
Eos
MCs+
MCs–
DCs+
DCs–
M2-MΦs
M1-MΦs
M0-MΦs
Monos
NK cells+

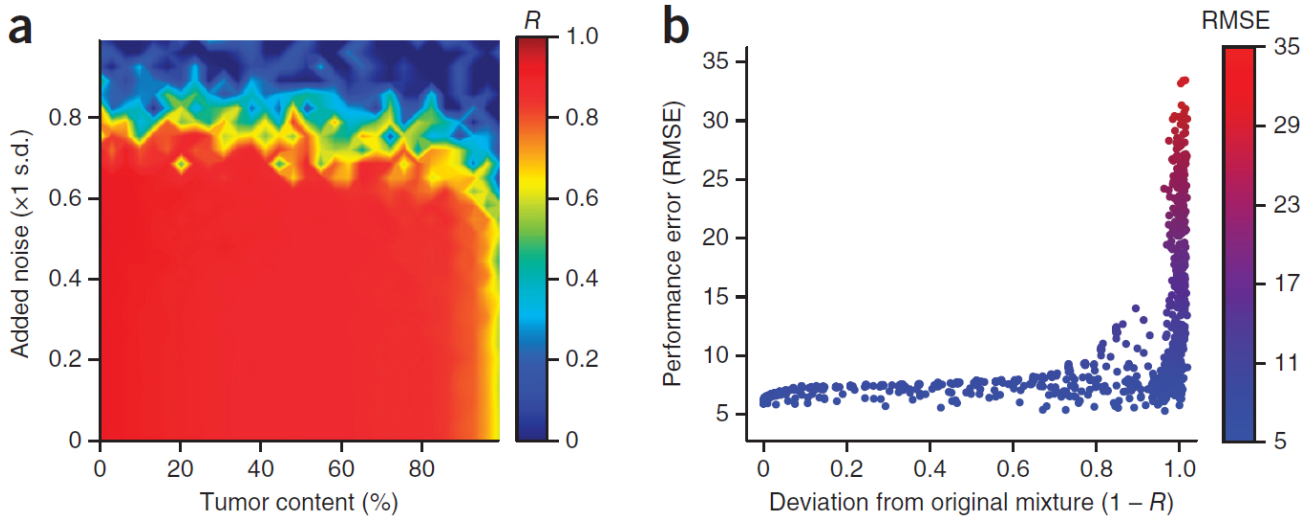

b
Sorting Strategy
CD5+ T-cells
CD19+ B-cells
99.5%
93.6%
CD5 (FITC)
CD19 (APC)


Estimated fraction of leukocyte RNA (%)
100
90
80
70
60
50
40
30
20
10
0
B cell (Sample 5)
B cell (Sample 1)
B cell (Sample 6)
B cell (Sample 4)
B cell (Sample 3)
T cell (Sample 4)
T cell (Sample 5)
T cell (Sample 6)
T cell (Sample 1)
T cell (Sample 3)
Other
T cells
B cells

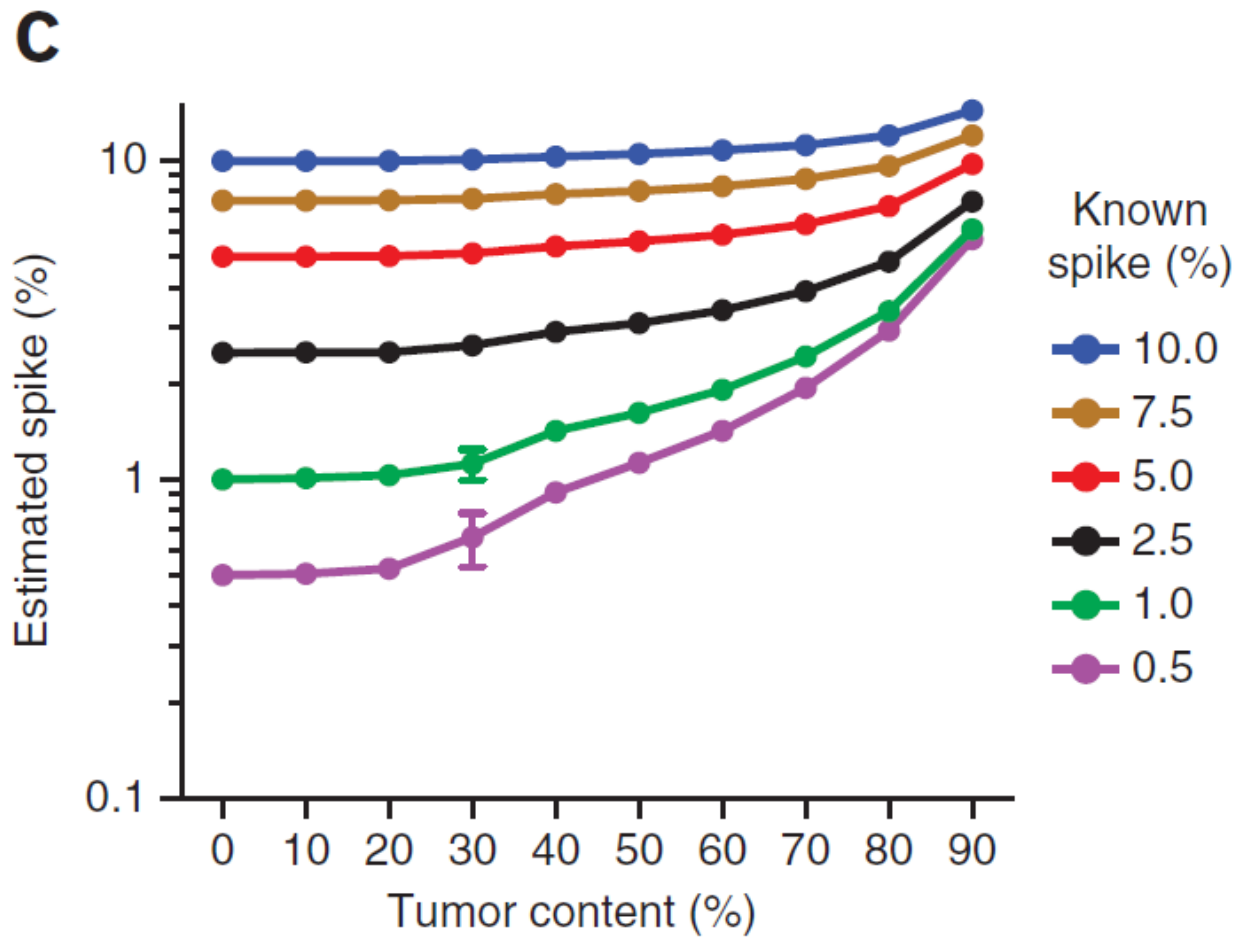
- Performance on well-defined mixtures:



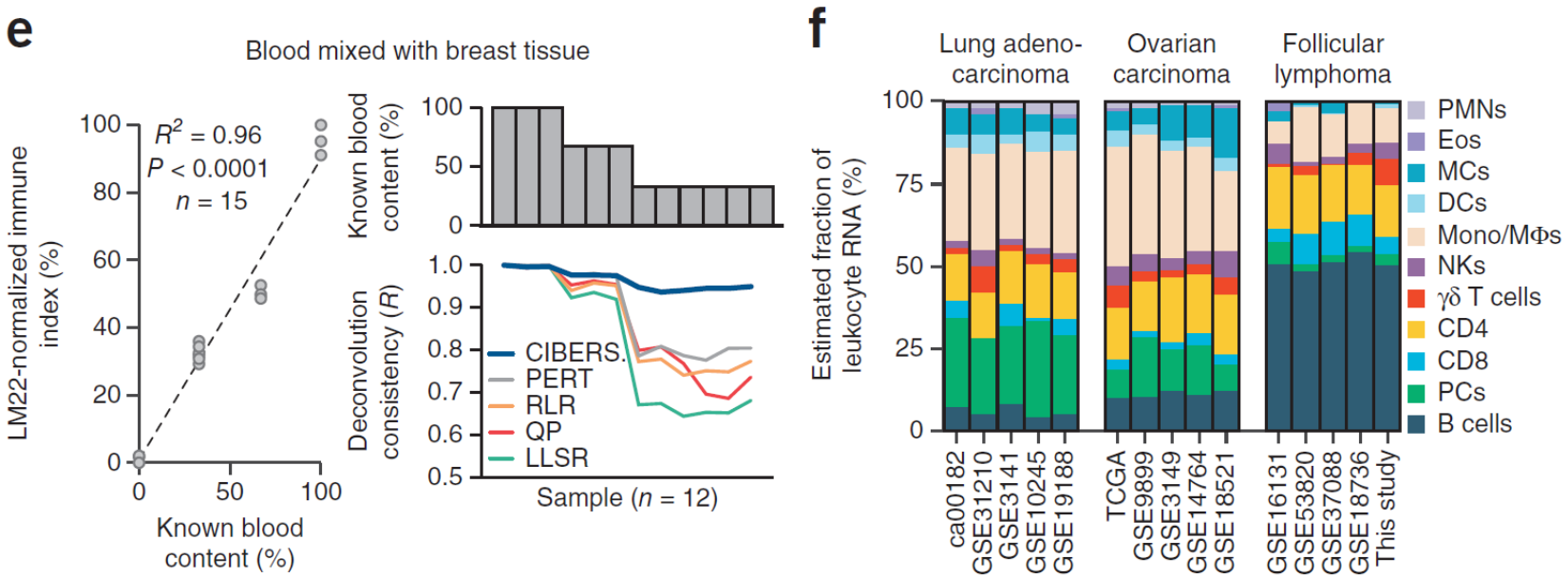
- Simulation of bulk tissues:
  - Tumor content (from 0% to $< 100\%$) (a colon cancer cell line)
  - Mixtures of 4 blood cell lines (simulate tomour with infiltrating leukocytes)
  - Add noise from log-Gaussian: $2^{\mathcal{N}(0, f \times \sigma)}$



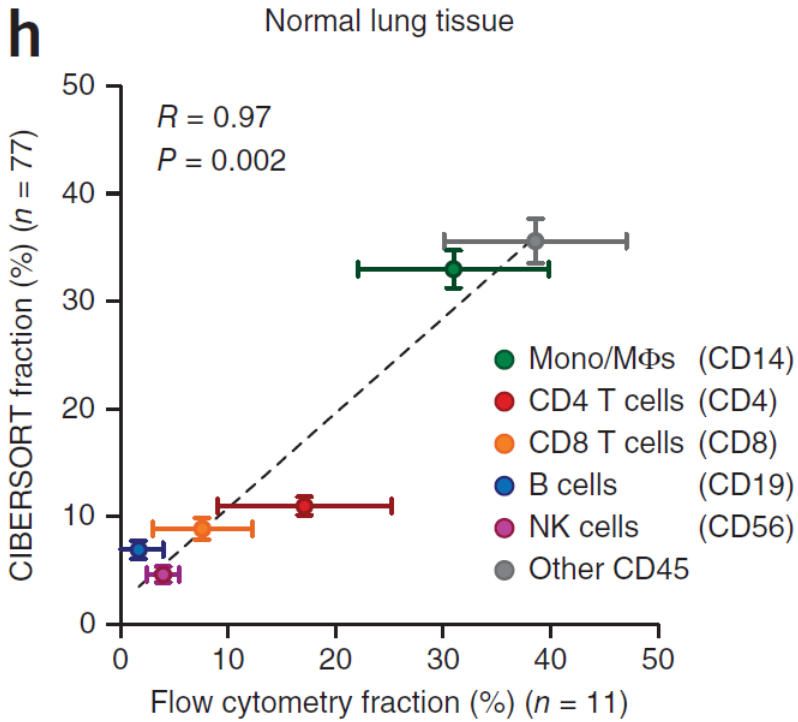
- Detection of rare cell type:

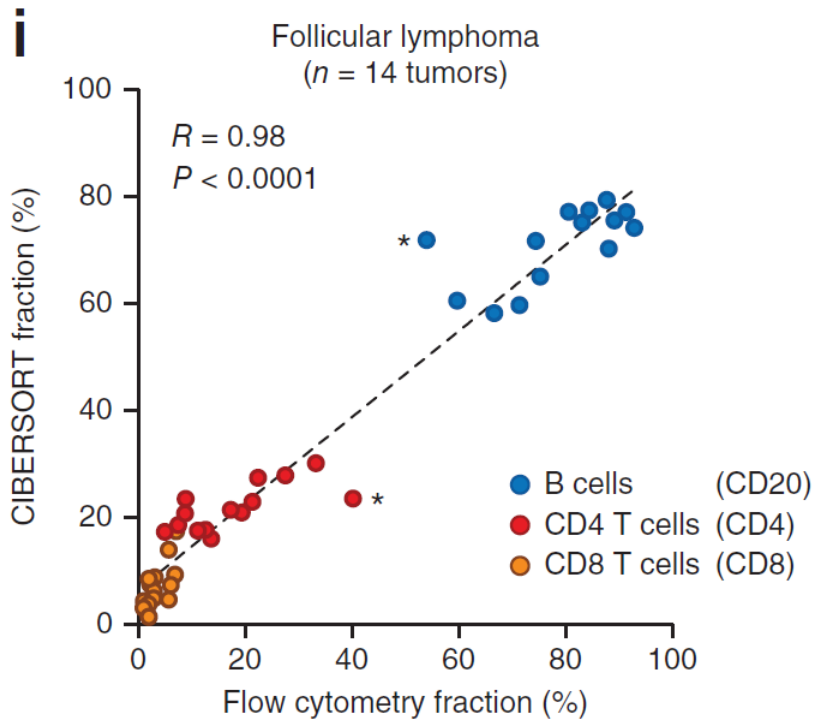


- Consistency on mixtures with unknown content or noise:



- Comparision to flow cytometry(ground-truth measurements of leukocyte content in solid tissues)





Conclusion

- Characterize cell heterogeneity using RNA mixtures from nearly any tissues.
- Fidelity of reference profiles, which could deviate in cells undergoing heterotypic interactions, phenotypic plasticity or disease-induced dysregulation.

Paper Introduction:

Perspective

Translating cancer 'omics' to improved outcomes

Emily A. Vucic,[1,2,6,7] Kelsie L. Thu,[1,6] Keith Robison,[3] Leszek A. Rybaczyk,[4] Raj Chari,[1,2,5] Carlos E. Alvarez,[4] and Wan L. Lam[1,2]

[1]British Columbia Cancer Research Centre, Vancouver V5Z 1L3, Canada; [2]Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver V6T 1Z4, Canada; [3]Warp Drive Bio, Cambridge, Massachusetts 02142, USA; [4]Center for Molecular and Human Genetics, The Research Institute at Nationwide Children's Hospital, Columbus, Ohio 43205, USA; [5]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA