

# Workflow for high-throughput drug sensitivity screen

Junyan Lu  
December 2017  
Cambridge, UK

# Cancer is a heterogeneous disease

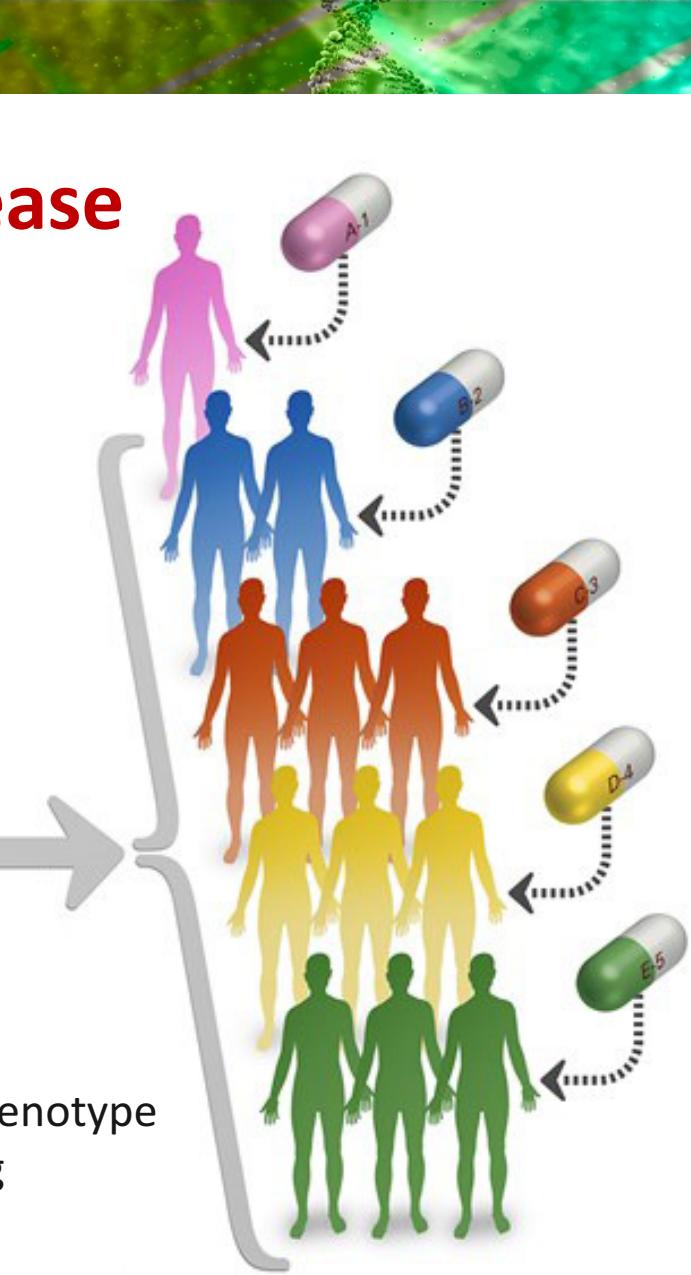


Patient cohort



- Drug sensitivity phenotype
- Molecular profiling

**Key step: drug sensitivity screen  
(cell line or patient sample based)**



Personalized therapy

## Application of drug sensitivity screen in personalized therapy

### LETTER

doi:10.1038/nature11003

The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity

### ARTICLE

doi:10.1038/nature11005

Systematic identification of genomic markers of drug sensitivity in cancer cells

### Cell

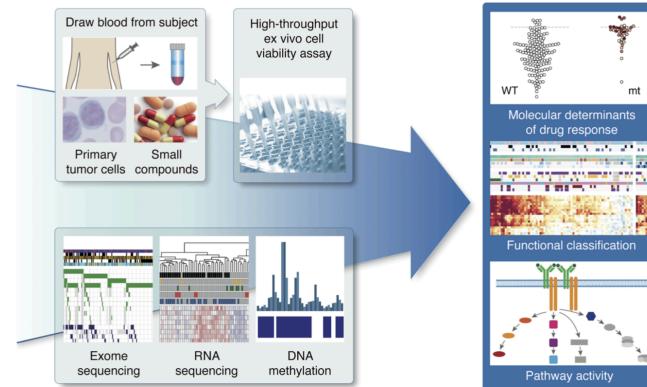
Resource

A Landscape of Pharmacogenomic Interactions in Cancer

RESEARCH ARTICLE

The Journal of Clinical Investigation

### Drug-perturbation-based stratification of blood cancer



(Dietrich, Oleś and Lu et al., 2017)

# Application of drug sensitivity screen in personalized therapy

## LETTER

The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity

doi:10.1038/nature11003

## ARTICLE

Systematic identification of genomic markers of drug sensitivity in cancer cells

doi:10.1038/nature11005

## Cell

A Landscape of Pharmacogenomic Interactions in Cancer

Resource

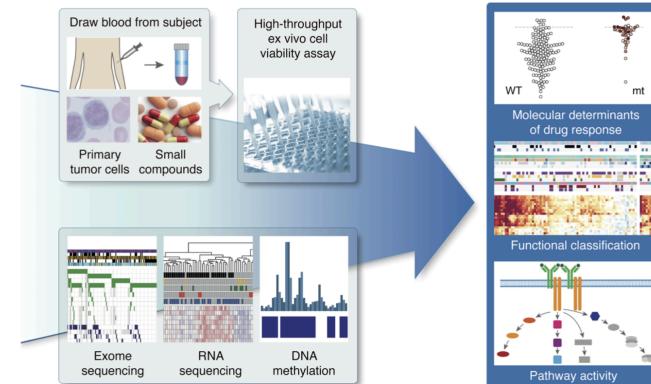
## Challenges:

- Reproducibility
- Data sharing and visualization

## RESEARCH ARTICLE

The Journal of Clinical Investigation

### Drug-perturbation-based stratification of blood cancer



(Dietrich, Oleś and Lu et al., 2017)

## ANALYSIS

doi:10.1038/nature12831

Inconsistency in large pharmacogenomic studies

Benjamin Haibe-Kains<sup>1,2</sup>, Nehme El & John Quackenbush<sup>3,8\*</sup>

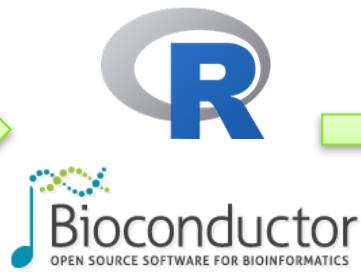
Reproducible pharmacogenomic profiling of cancer cell line panels

Peter M. Haverty<sup>1\*</sup>, Eva Lin<sup>2\*</sup>, Jenille Tan<sup>2</sup>, Yihong Yu<sup>2</sup>, Billy Lam<sup>2</sup>, Steve Lianoglou<sup>1</sup>, Richard M. Neve<sup>2</sup>, Scott Martin<sup>2</sup>, Jeff Settleman<sup>2</sup>, Robert L. Yauch<sup>2</sup> & Richard Bourgon<sup>1</sup>

## Workflow for high-throughput drug sensitivity screen



**AIM:** An end-to-end chain of tools for high-throughput drug sensitivity screen analysis , including data import, metadata management, quality assessment, visualization and sharing.

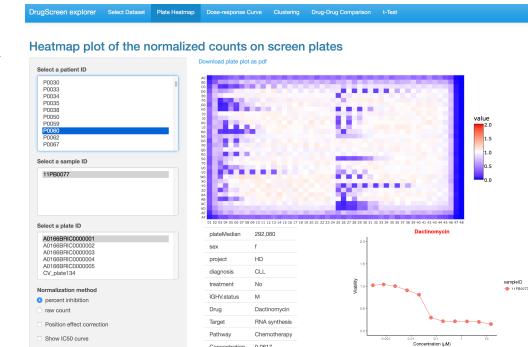


Raw data

Raw data processing and quality control for CPS1000  
Junyan Lu  
1/28/2016 (Updated on 10 November 2017)

Contents

- 1. Load packages
- 2. Import patient file
- 2.1. Function for reading raw counts on a plate
- 2.2. Read in the whole experiment
- 2.3. Fix the layout for P0001 (111P0003)
- 2.4. Add patient background information
- 2.5. Update the samples IDs according to the patient or autobox annotations
- 2.6. Check the raw count distribution of control and drug wells on the plate
- 2.7. Check the raw count distribution of control and drug wells on the entire array
- 2.8. Check the raw count value distribution of all samples
- 2.9. Fix the cell spotting issue
- 3. Data processing
- 3.1. Normalize raw counts values by internal negative control (OMSC) wells
- 3.1.1. For raw counts
- 3.1.2. For raw ATP counts (batch adjusted and no spotting effect)
- 3.2. Estimate the position effect by linear 2D surface fit
- 3.2.1. For raw ATP counts
- 3.3. Summarize the drug effect by averaging viability of concentrations 1-5
- 4. Quality assessment
- 4.1. Plot the count value on each plate
- 4.2. Plot the dose-response curves for each drug
- 4.3. Estimate the IC50 values
- 4.3.1. Well positions vs ATP counts of negative controls
- 4.3.2. Well positions vs viability of the lowest concentration
- 5. Annotate samples
- 5.1. Process patient background information
- 5.2. Add disease state information
- 5.3. Add drug information
- 6. Save data
- 7. Process a excel table
- 8. Estimate batch effect
- 9. Estimate the correlation between screen quality and basal ATP count or cell count



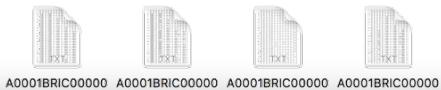
Quality report

Interactive visualization

R package: <https://github.com/lujunyan1118/DrugScreenExplorer> (in development)

# European Bioconductor Meeting 2017 sample annotation file (csv file)

## Step 1. Data import



`createWellInput()`

	A	B	C
1	wellID	name	concentration
2	A001	DMSO	0,3
3	A002	DMSO	0,3
4	A003	A-1210477	10
5	A004	Afatinib	10
6	A005	A-1210477	2
7	A006	Afatinib	2
8	A007	A-1210477	0,4
9	A008	Afatinib	0,4
10	A009	A-1210477	0,08

plate layout file  
(csv file)

Raw measurement  
(txt format)

`readScreen()`

`createPlateInput()`

	A	B	C	D
1	fileName	batch	sampleID	patientID
2	CTGLuminescence_20161124_	1	11PB0010	P0010
3	CTGLuminescence_20161124_	1	12PB0265	P0007
4	CTGLuminescence_20161124_	1	13PB0658	P0010
5	CTGLuminescence_20161124_	1	14PB0491	P0007
6	CTGLuminescence_20161124_	1	15PB0023	P0010
7	CTGLuminescence_20161124_	1	16PB0066	P0010
8	CTGLuminescence_20161125_	2	11PB0021	P0019
9	CTGLuminescence_20161125_	2	11PB0023	P0021

sample annotation file  
(csv file)

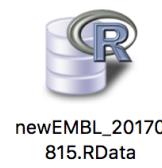
well	val	batch	patientID	sampleID	Drug	Concentration	Target
<chr>	<dbl>	<fctr>	<chr>	<chr>	<chr>	<dbl>	<chr>
A01	1083480	1	P0010	11PB0010	DMSO	3.0e-01	CONTROL
A02	856480	1	P0010	11PB0010	DMSO	3.0e-01	CONTROL
A03	1014920	1	P0010	11PB0010	A-1210477	1.0e+01	MCL1
A04	21000	1	P0010	11PB0010	Afatinib	1.0e+01	EGFR/ERBB2
A05	1151520	1	P0010	11PB0010	A-1210477	2.0e+00	MCL1
A06	557760	1	P0010	11PB0010	Afatinib	2.0e+00	EGFR/ERBB2
A07	1264000	1	P0010	11PB0010	A-1210477	4.0e-01	MCL1
A08	974880	1	P0010	11PB0010	Afatinib	4.0e-01	EGFR/ERBB2
A09	1265000	1	P0010	11PB0010	A-1210477	8.0e-02	MCL1
A10	1191200	1	P0010	11PB0010	Afatinib	8.0e-02	EGFR/ERBB2



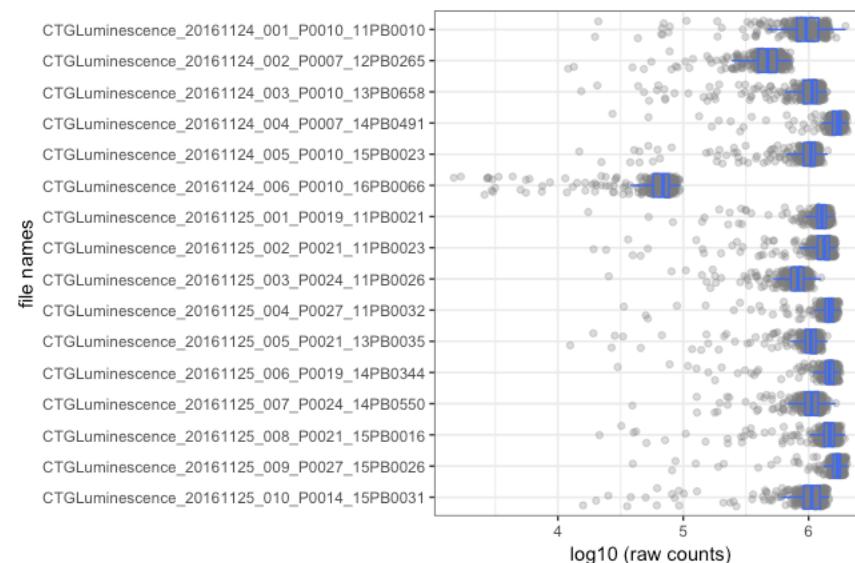
CPS1000\_170801.RData newEMBL\_20170 815.RData

Tidy table or  
SummarizedExperiment

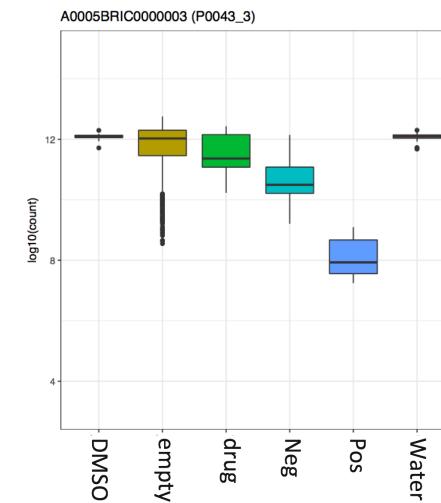
## **Step 2. Quality assessment**



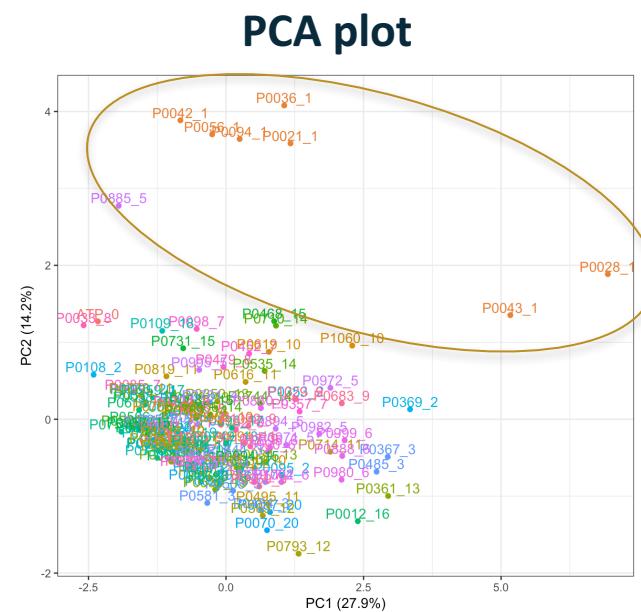
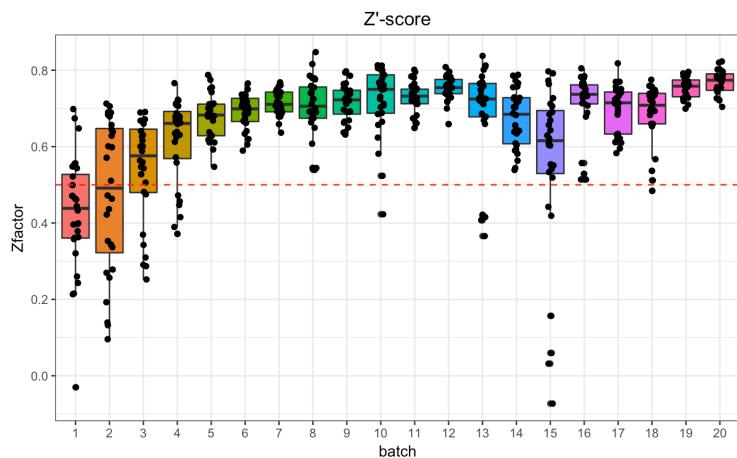
## raw signal distribution



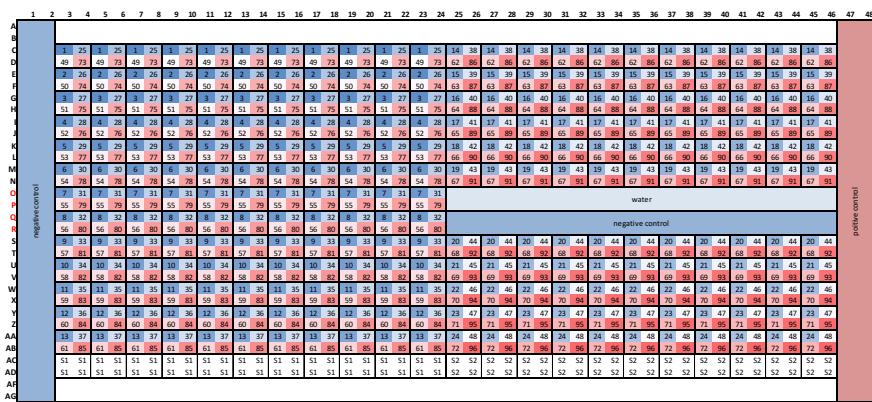
## per-plate signal distribution



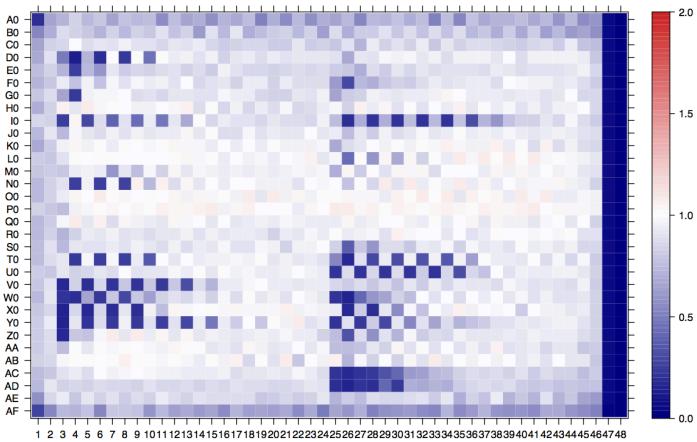
$$Z' - \text{factor} = 1 - 3(\delta_{\text{h,c}} + \delta_{\text{l,c}})/\mu_{\text{h,c}} + \mu_{\text{l,c}}$$



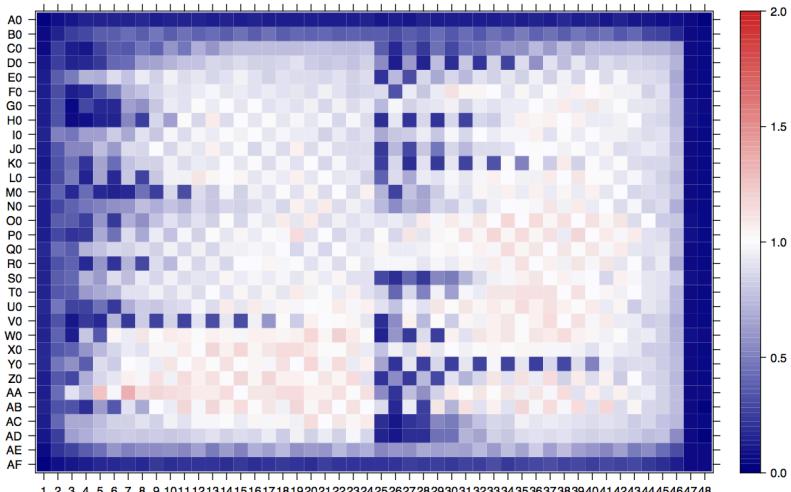
## Step 2. Quality assessment



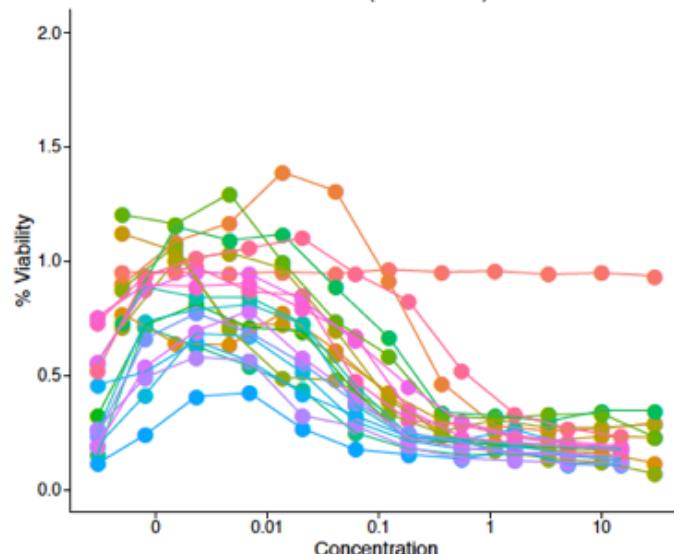
## Plate layout



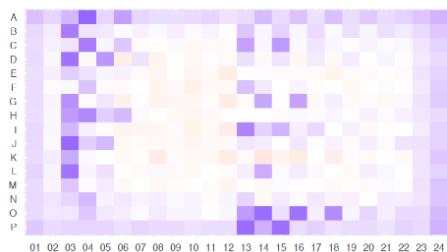
## ideal results



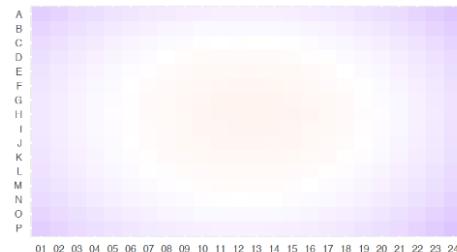
## edge effect



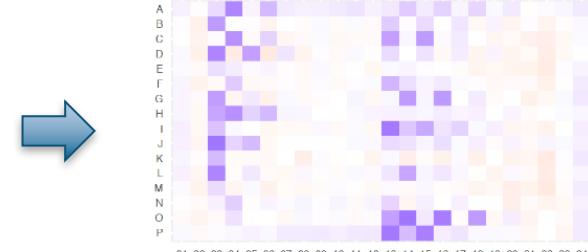
## Step 3. Post-processing



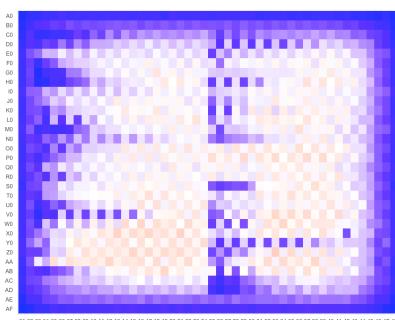
before correction(384-well plate)



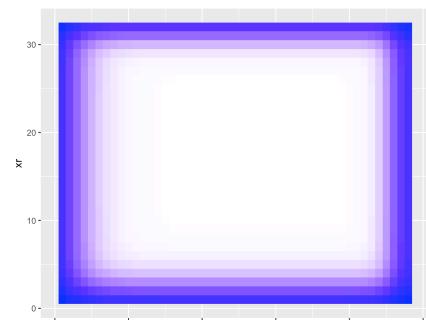
2D-local regression (loess)



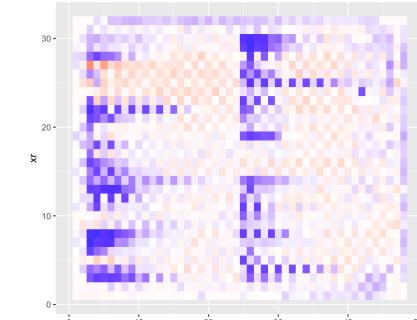
After correction



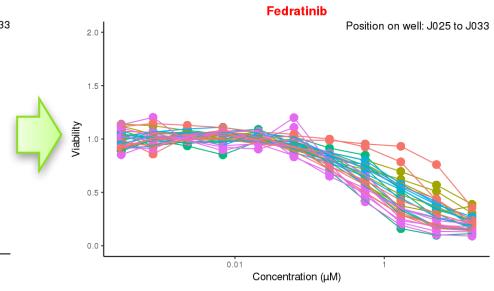
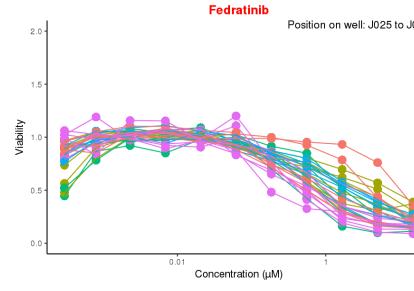
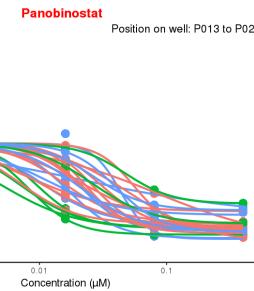
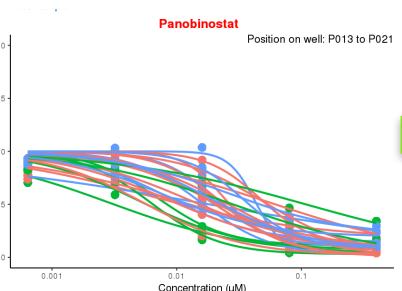
before correction(1536-well plate)



2D-sigmoid model



After correction



## Step 4. Summarization

- Percent inhibition at each concentration (normalized by negative or positive controls)
- IC50 curves and AUC (by "drc" package)
- Average over all (or selected) concentrations

`normalizePlate {DrugScreenExplorer}`

R Documentation

### Per-plate normalization

#### Description

This function calculates the percent inhibition values for each plate based on the control wells on the same plate.

#### Usage

```
normalizePlate(screenData, method = "negatives", discardLayer = 0)
```

## Step 5. Sharing and visualization

Automatic and customizable R markdown report and Shiny app



CPS1000\_170801.  
RData



makeReport()

makeShiny()  
*(in progress)*

```
# A tibble: 280,704 x 7
  well    val sampleID     Drug normVal normVal.cor normVal_auc
  <chr>  <dbl> <chr>      <chr>   <dbl>      <dbl>        <dbl>
1 A01 1083480 11PB0010 DMSO 1.10663071  1.0334577    NA
2 A02 856480 11PB0010 DMSO 0.87478041  0.7811308    NA
3 A03 1014920 11PB0010 A-1210477 1.03660579  0.9244394  1.2104588
4 A04 21000 11PB0010 Afatinib 0.02144871 -0.1072903  0.8047637
5 A05 1151520 11PB0010 A-1210477 1.17612453  1.0327358  1.2104588
6 A06 557760 11PB0010 Afatinib 0.56967766  0.4135397  0.8047637
7 A07 1264000 11PB0010 A-1210477 1.29100788  1.1239989  1.2104588
8 A08 974880 11PB0010 Afatinib 0.99571026  0.8196860  0.8047637
9 A09 1265000 11PB0010 A-1210477 1.29202925  1.1088242  1.2104588
10 A10 1191200 11PB0010 Afatinib 1.21665237  1.0280807  0.8047637
# ... with 280,694 more rows
```

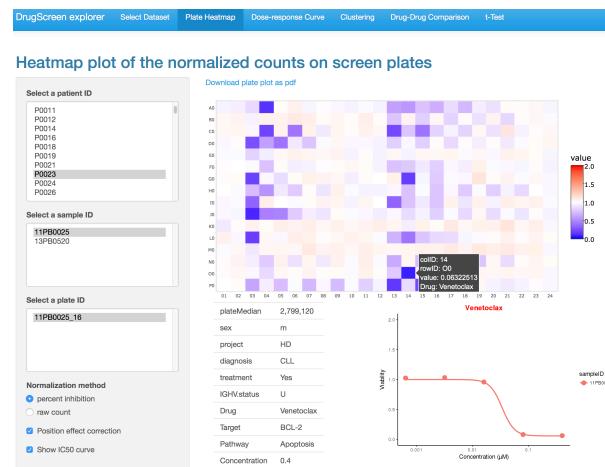
Report for my drug screening project

Junyan Lu

2 December 2017

Contents

- [1 Dataset summary](#)
  - [1.1 Basic information](#)
    - [1.1.1 Screen format](#)
    - [1.1.2 Plate layout](#)
  - [1.2 Raw signal distribution on each plate](#)
- [2 Quality assessment](#)
  - [2.1 Variance vs rank of mean](#)
  - [2.2 Edge effect](#)
    - [2.2.1 Plot intensity of edge effect](#)
    - [2.2.2 PCA plot](#)
    - [2.2.3 MDS](#)



# European Bioconductor Meeting 2017

<http://mozi.embl.de/drugScreens/>

DrugScreen explorer Select Dataset Plate Heatmap Dose-response Curve Clustering Drug-Drug Comparison t-Test

## Heatmap plot of the normalized counts on screen plates

Select a patient ID

- P0011
- P0012
- P0014
- P0016
- P0018
- P0019
- P0021
- P0023**
- P0024
- P0026

Select a sample ID

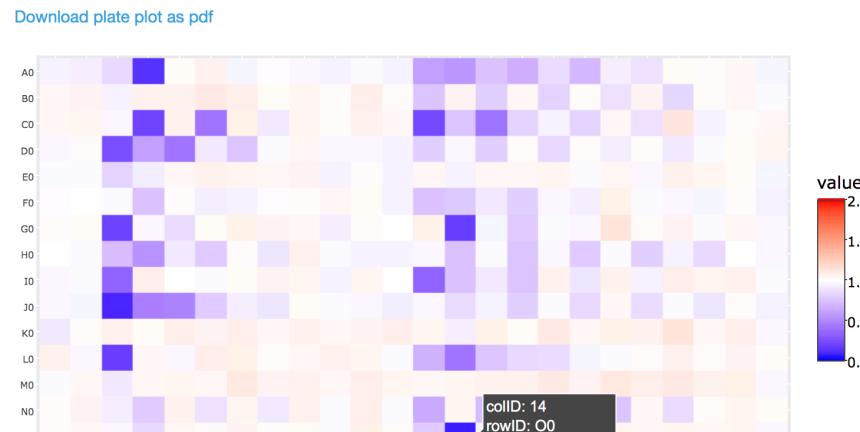
- 11PB0025**
- 13PB0520

Select a plate ID

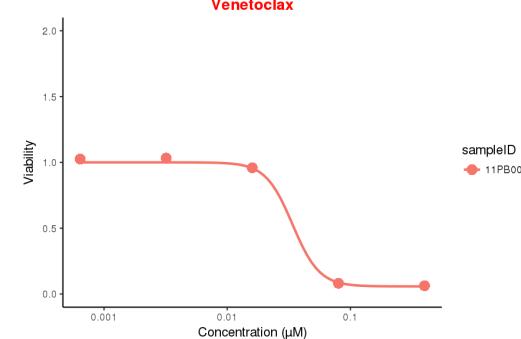
- 11PB0025\_16**

Normalization method

- percent inhibition
- raw count
- Position effect correction
- Show IC50 curve



plateMedian	2,799,120
sex	m
project	HD
diagnosis	CLL
treatment	Yes
IGHV.status	U
Drug	Venetoclax
Target	BCL-2
Pathway	Apoptosis
Concentration	0.4



## Quality check

[DrugScreen explorer](#)[Select Dataset](#)[Plate Heatmap](#)[Dose-response Curve](#)[Clustering](#)[Drug-Drug Comparison](#)[t-Test](#)

## Dose-response curves for specified patient samples

**Drug selection**

Single drug  
 Drug group  
 All drugs

**Select a drug**

Venetoclax

**Grouped by**

diagnosis

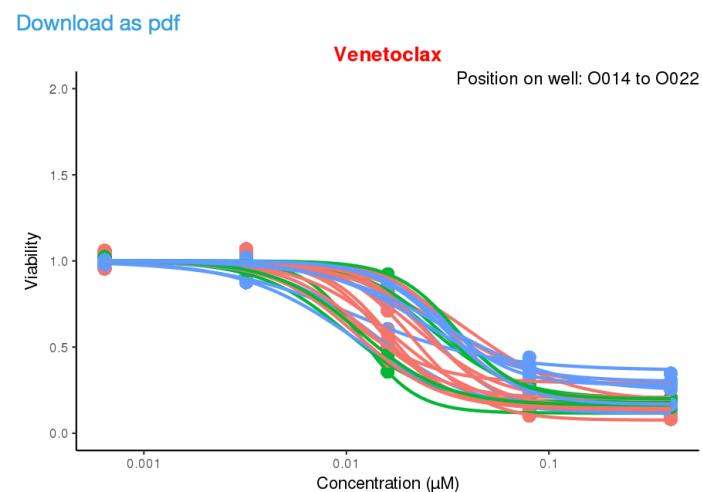
**Groups included**

CLL  
 MCL  
 AML

**Colored by**

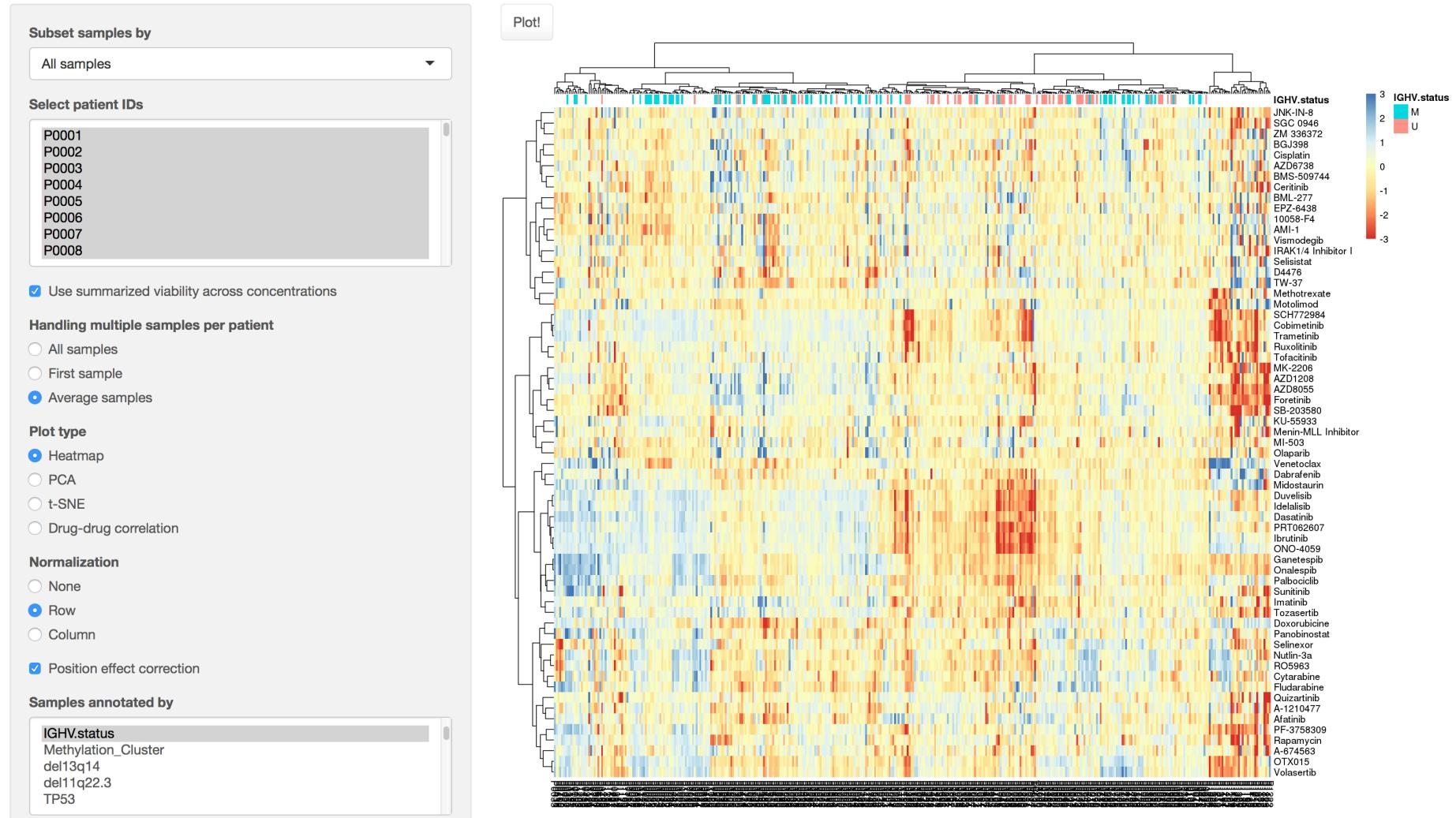
batch

Position effect correction  
 Show IC50 curve (maybe slow)  
 Remove NAs



## Dose-response analysis

## Clustering samples or drugs based on drug response



# Clustering

DrugScreen explorer

Select Dataset

Plate Heatmap

Dose-response Curve

Clustering

Drug-Drug Comparison

t-Test

## Associations between drug responses and genetic background

**Subset samples by**

All samples

**Search drug by name**

Nut

**Select a drug**

Nutlin-3a

Use summarized viability across concentrations

**Handling multiple samples per patient**

Average samples

First sample

Position effect correction

**Select a t-test method**

Equal variance

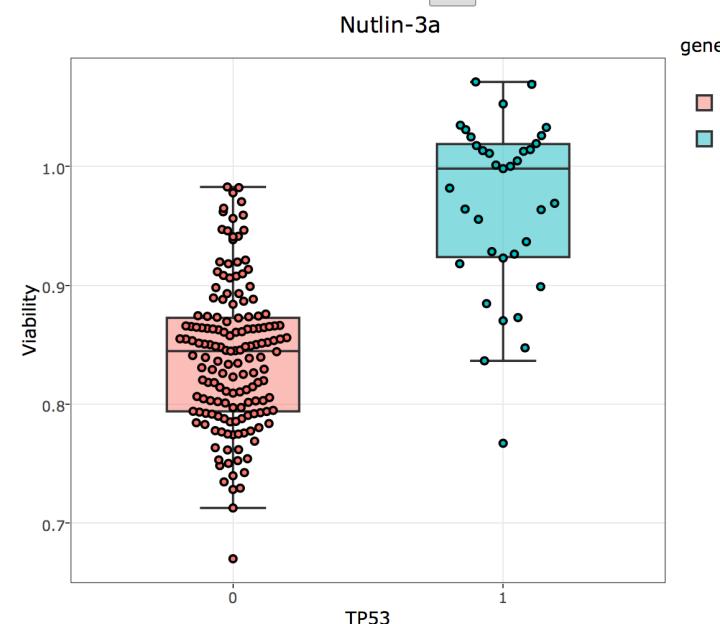
Unequal variance

Consider IGHV status

	p.raw	diffMean	p.adj
TP53	0.00	0.13	0.00
del17p13	0.00	0.09	0.00
Chromothripsis	0.00	0.09	0.05
VWF	0.01	0.14	0.08
trisomy12	0.01	-0.04	0.08

Showing 1 to 5 of 61 entries

Previous 1 2 3 4 5 ... 13 Next



## Hypothesis testing

# European Bioconductor Meeting 2017

**7 commits** **1 branch** **0 releases** **1 contributor**

Branch: master [New pull request](#) [Create new file](#) [Upload files](#) [Find file](#) [Clone or download](#)

lujunyan1118 update readme file Latest commit 1446d76 24 minutes ago

	Create new repository	37 minutes ago
R	Create new repository	37 minutes ago
inst	Create new repository	37 minutes ago
man	Create new repository	37 minutes ago
report	Create new repository	37 minutes ago
shiny	Create new repository	37 minutes ago
vignettes	Create new repository	37 minutes ago
DESCRIPTION	Create new repository	37 minutes ago
DrugScreenExplorer.Rproj	Create new repository	37 minutes ago
NAMESPACE	Create new repository	37 minutes ago
README.md	update readme file	24 minutes ago
index.html	add new front page	28 minutes ago

**README.md**

## DrugScreenExplorer

DrugScreenExplorer is an R package designed for streamlined processing, quality control and visualising of high-throughput drug sensitivity screen data.

Read more at: <https://lujunyan1118.github.io/DrugScreenExplorer/>

## Introduction to DrugScreenExplorer

**Junyan Lu**

**3 December 2017**

### Contents

- [1 Getting started](#)
- [2 Data import](#)
  - [2.1 Introduction of input file types](#)
  - [2.2 Functions for helping create annotation files](#)
    - [2.2.1 createPlateInput\(\)](#)
    - [2.2.2 createWellInput\(\)](#)
  - [2.3 Read in the whole experiment](#)
- [3 Quality control](#)
  - [3.1 Raw count distribution](#)
  - [3.2 Raw count distribution for each well type](#)
  - [3.3 Plate heatmap plot](#)
  - [3.4 Automatic screen quality report](#)
- [4 Adjusting incubation effect](#)
- [5 Interactive data exploration using Shiny app \(In development\)](#)

### 1 Getting started

DrugScreenExplorer is an R package designed for streamlined processing, quality control and visualizing of high-throughput drug sensitivity screen data.

```
library(DrugScreenExplorer)
library(tidyverse)
library(gridExtra)
library(tools)
```

### 2 Data import

#### 2.1 Introduction of input file types

- For creating the complete drug screen datasets, three types of data are needed:
1. the raw screening data in text format, which stores the values out from a plate reader;
  2. the well annotation or plate layout, which stores the metadata for all the wells on a screening plate. The most important annotations are drug names and concentrations;
  3. the plate annotation, which stores the metadata for each plate. For example, the sample information, if each plate represents one sample.

The examples for each type of the input files can be found in the test data affiliated with this package:

#### Raw screening data

```
system.file("testData/rawData", package = "DrugScreenExplorer")
```

<https://github.com/lujunyan1118/DrugScreenExplorer>

## Acknowledgement



UniversitätsKlinikum Heidelberg



**SOUND**

European Commission

Thorsten Zenz

Michelle Libério

Sebastian Scheinost

Sascha Dietrich

Sophie Rabe



Chemical Biology Core  
Facility

Kerstin Putzker

Joe Lewis

Wolfgang Huber Group



**Thank you for your attention!**