

Datasets that elucidate the role of the microenvironment in modulating cellular phenotypes

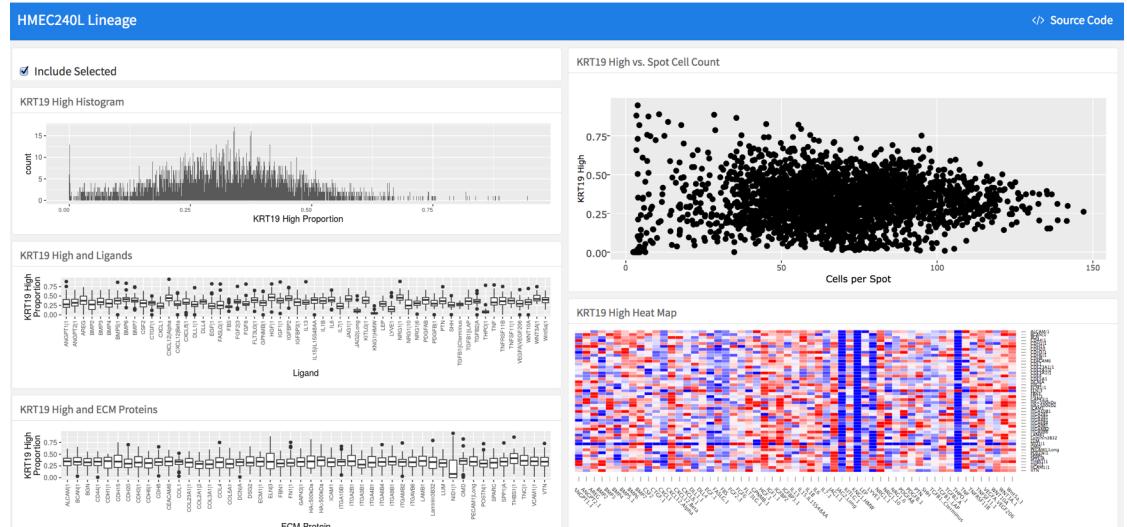
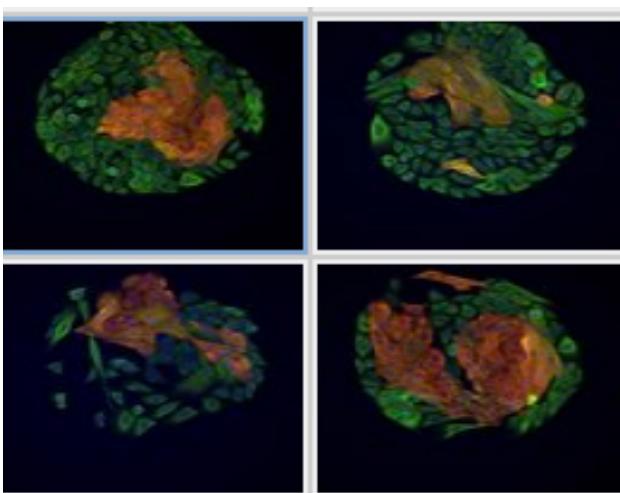
Mark Dane

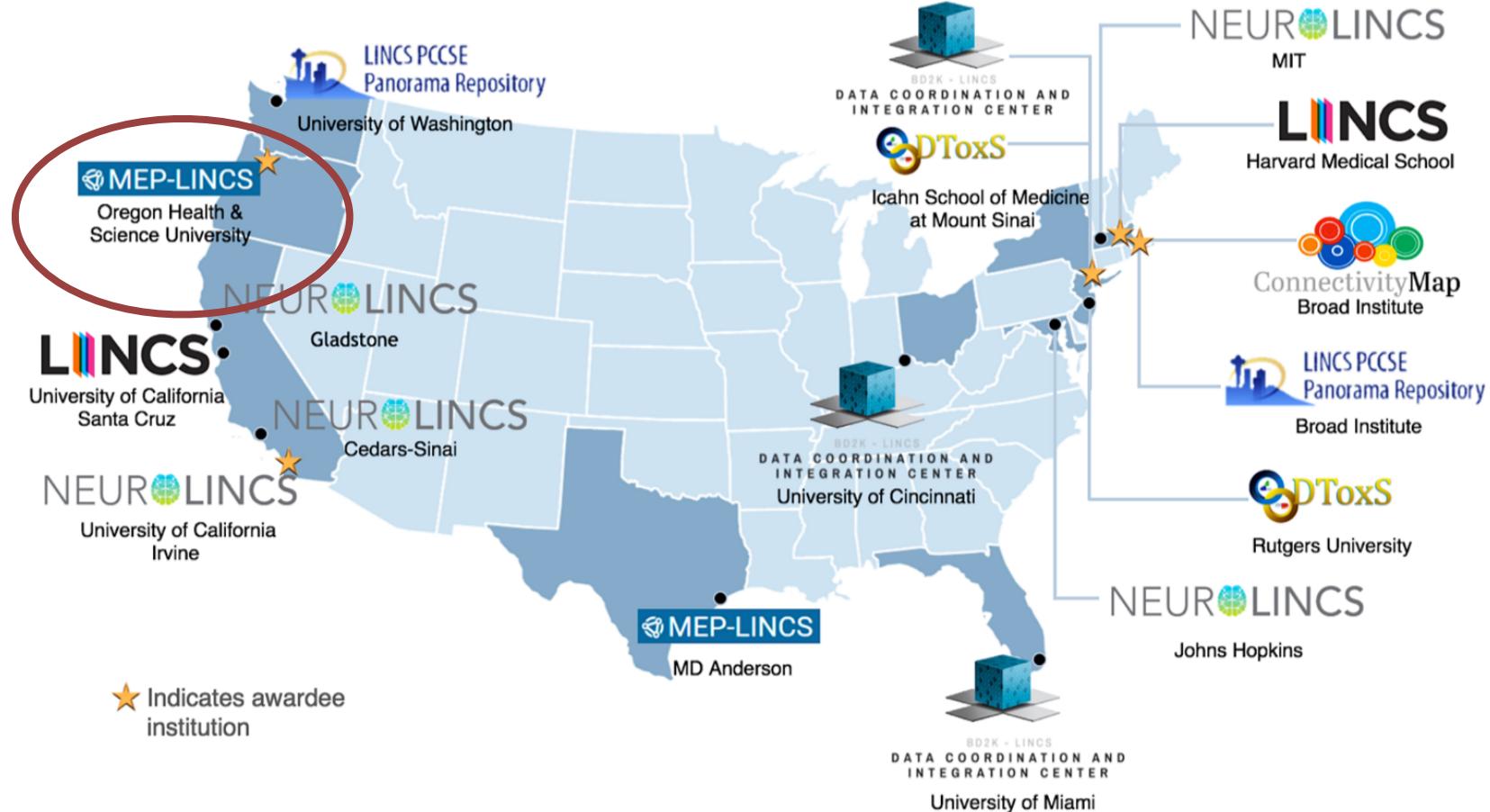
Oregon Health and Science University

Heiser Lab

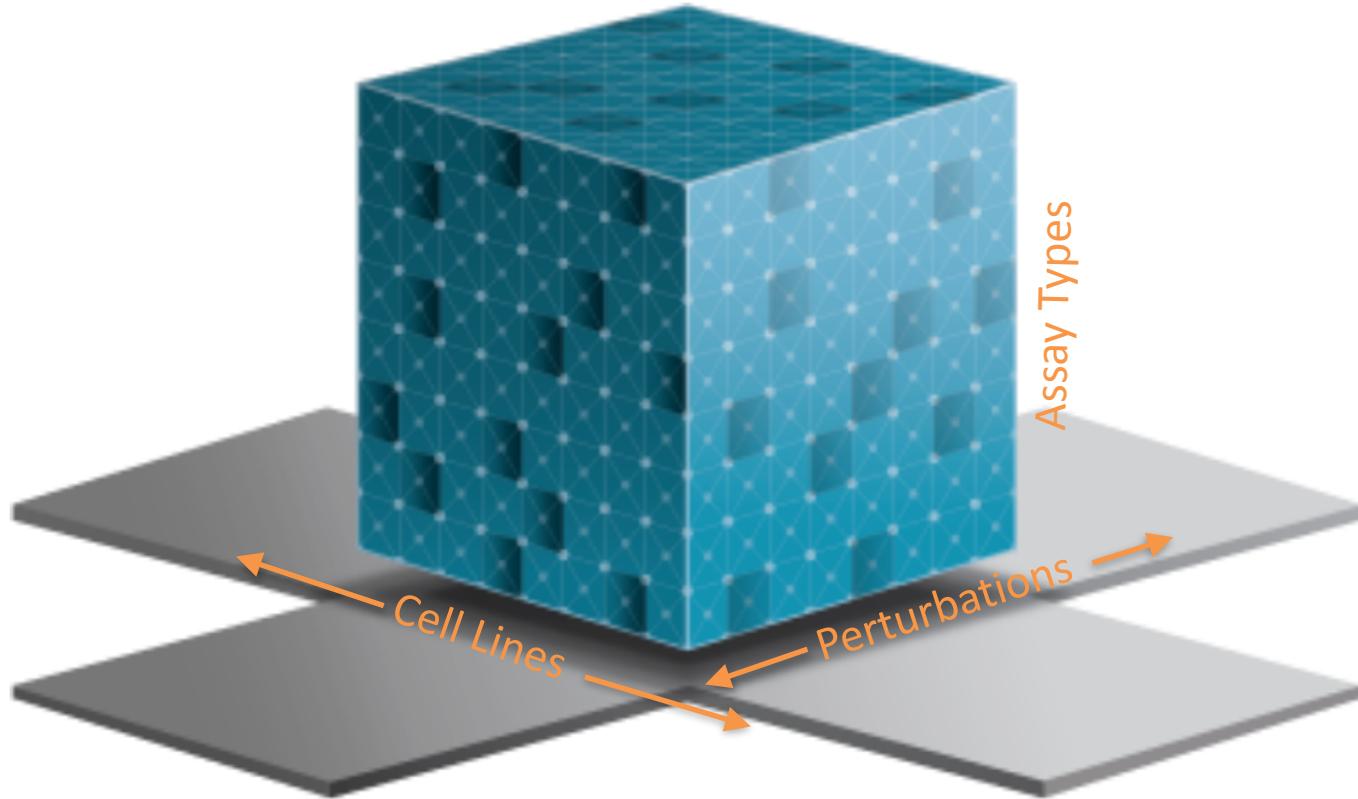
Cascadia R Conference

6/3/2017





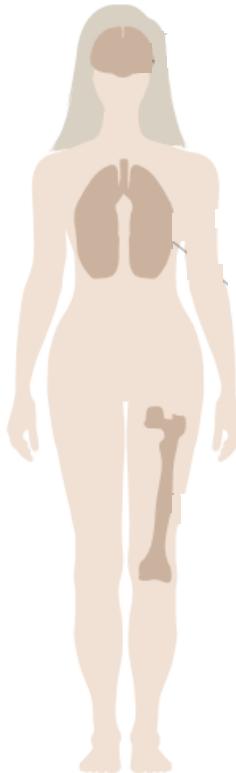
The **Library of Network-Based Cellular Signatures (LINCS)** program aims to create a network-based understanding of biology by cataloging changes in gene expression and other cellular processes that occur when cells are exposed to a variety of perturbing agents.



The **Library of Network-Based Cellular Signatures (LINCS)** program aims to create a network-based understanding of biology by cataloging changes in gene expression and other cellular processes that occur when cells are exposed to a variety of perturbing agents.

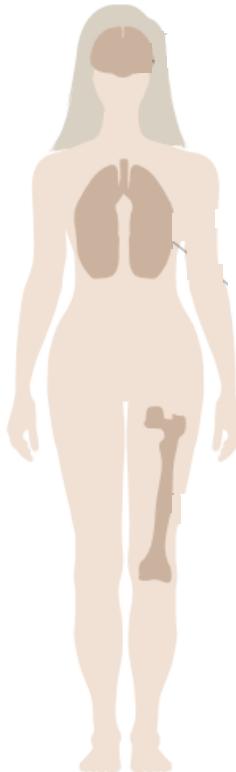
Microenvironments (ME) elicit varied responses

**Different
MEs in our
bodies**



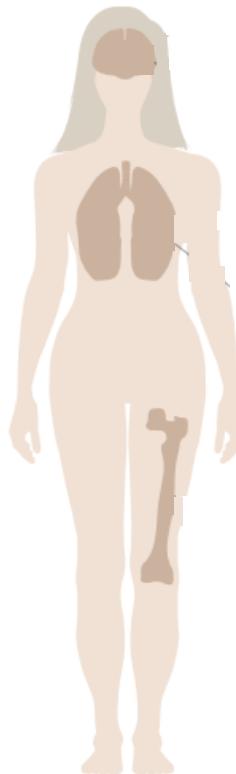
Microenvironments (ME) elicit varied responses

Different
MEs in our
bodies

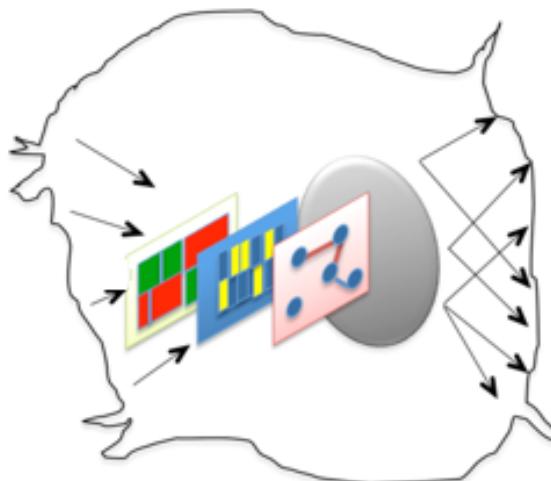


Microenvironments (ME) elicit varied responses

Different
MEs in our
bodies

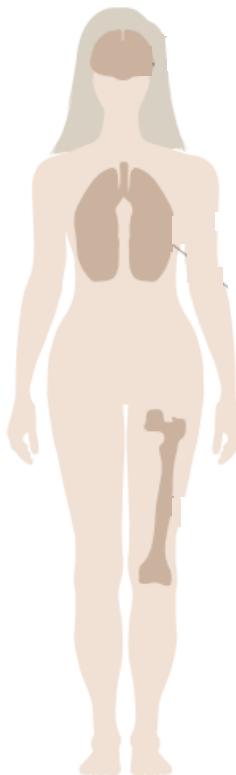


Cells integrate
ME signals

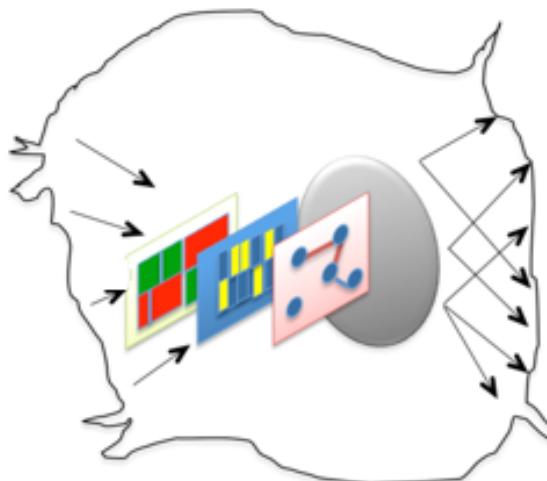


Microenvironments (ME) elicit varied responses

Different
MEs in our
bodies



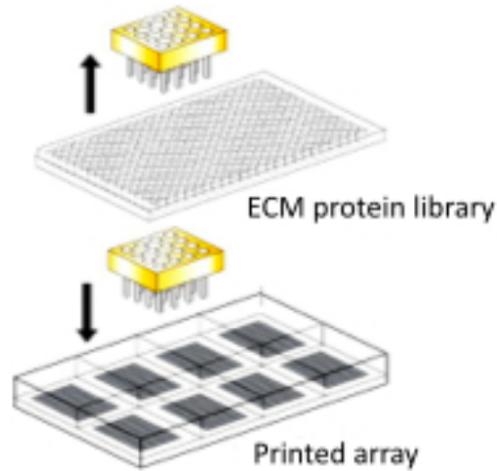
Cells integrate
ME signals



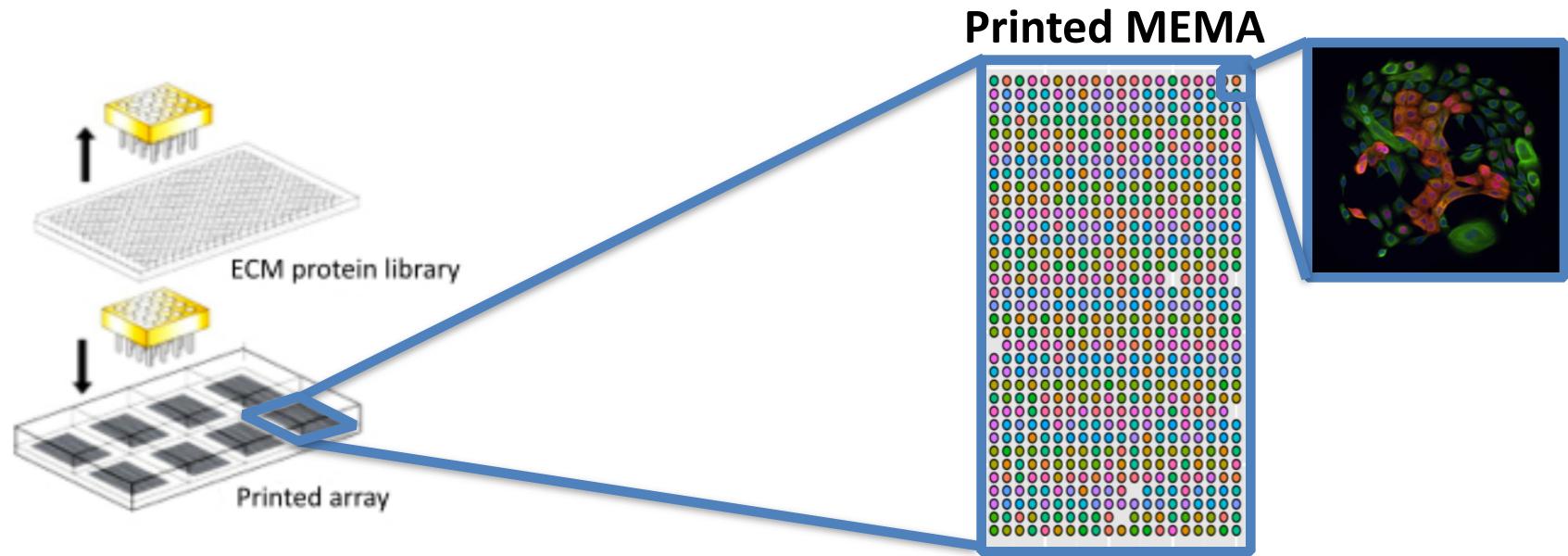
Phenotypes

- Proliferation
- Differentiation
- Metabolism
- Senescence

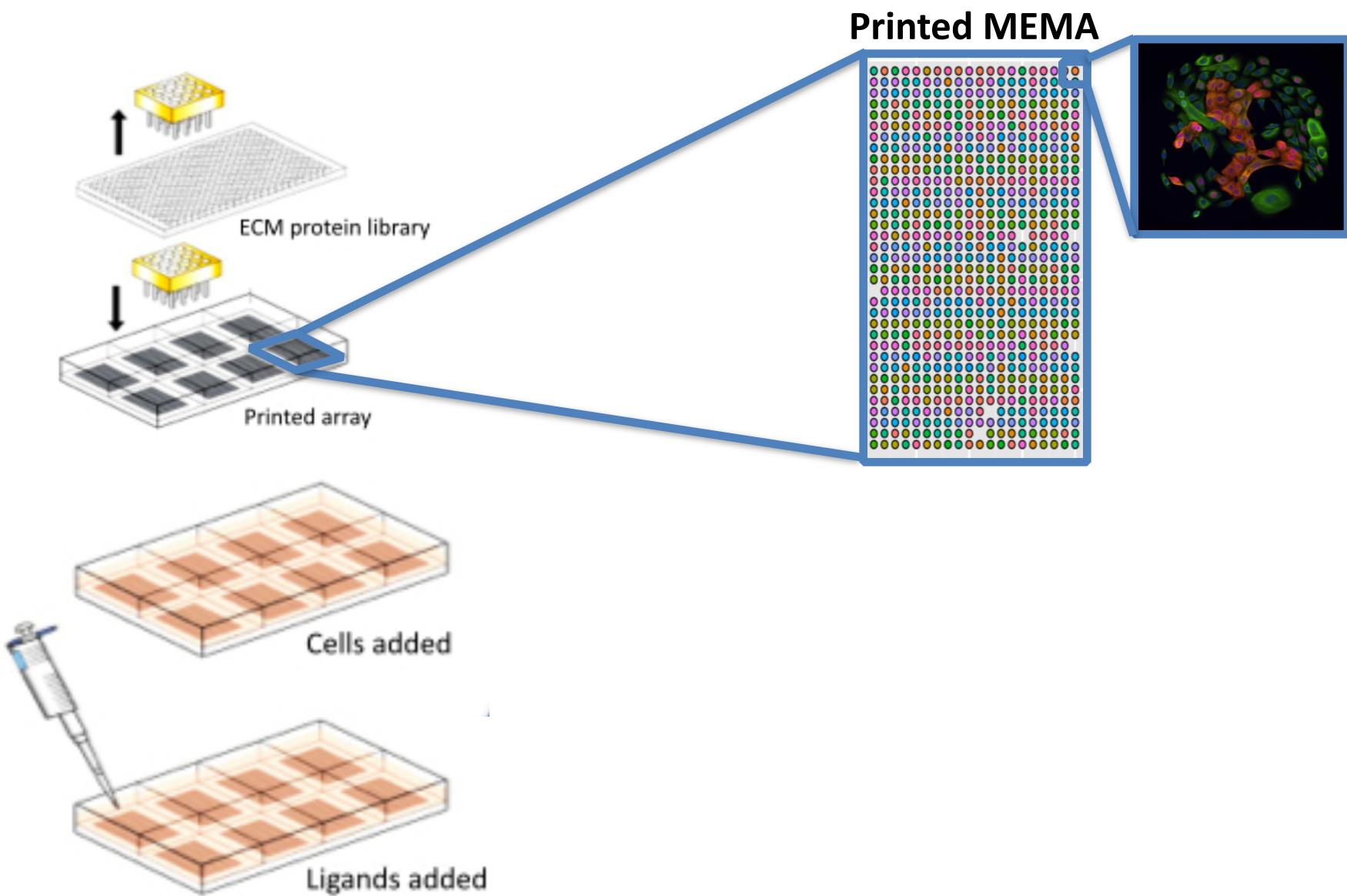
Cells are grown in ~2700 different microenvironments



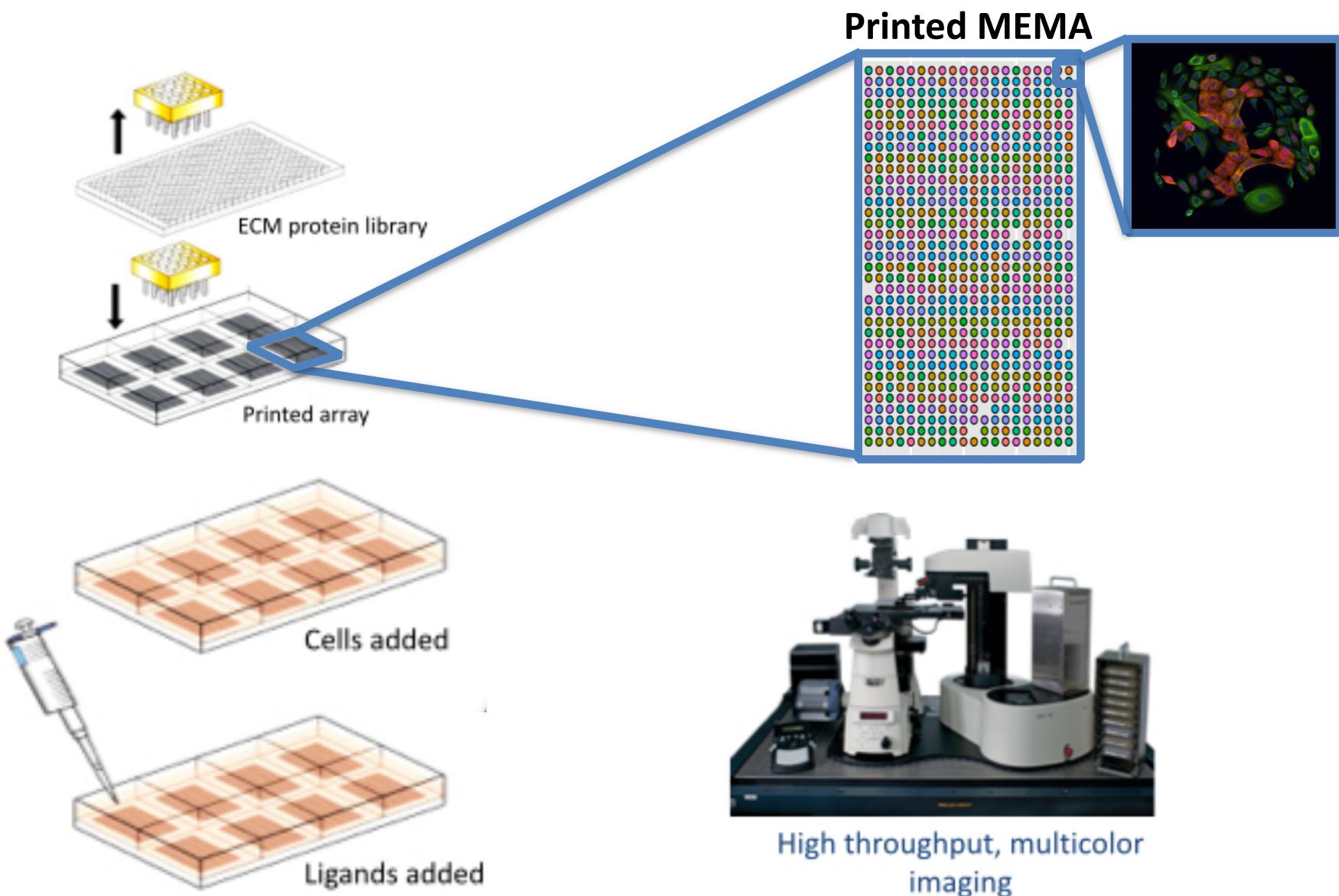
Cells are grown in ~2700 different microenvironments



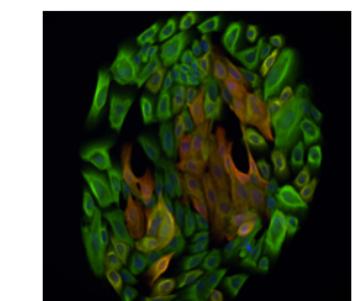
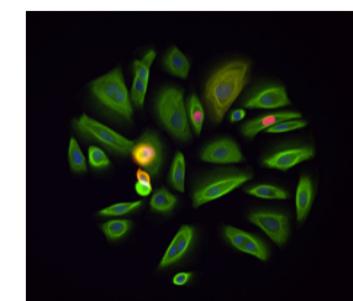
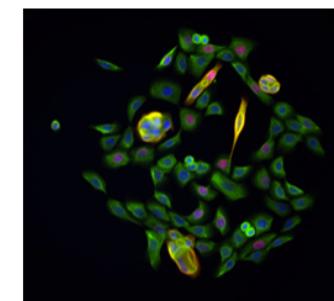
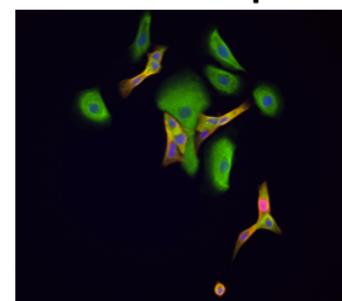
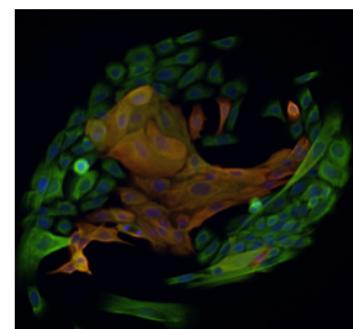
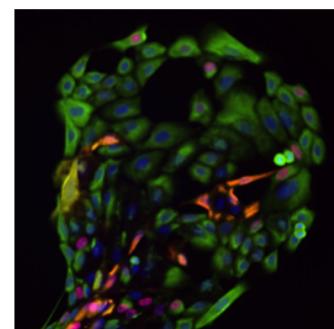
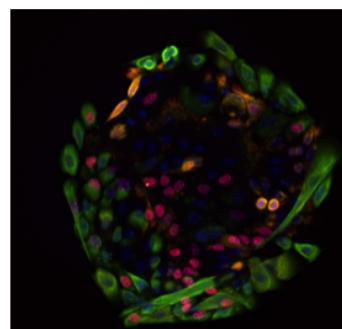
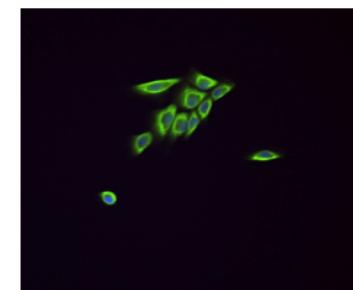
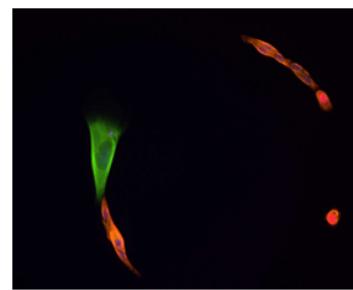
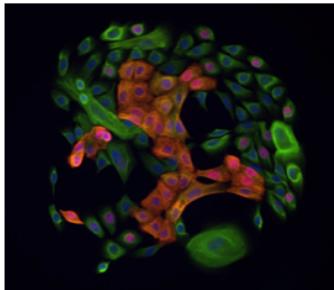
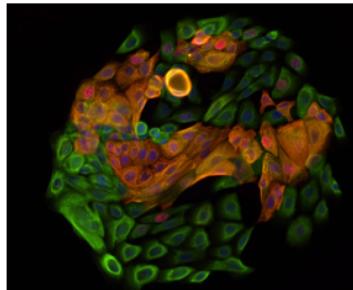
Cells are grown in ~2700 different microenvironments



Cells are grown in ~2700 different microenvironments



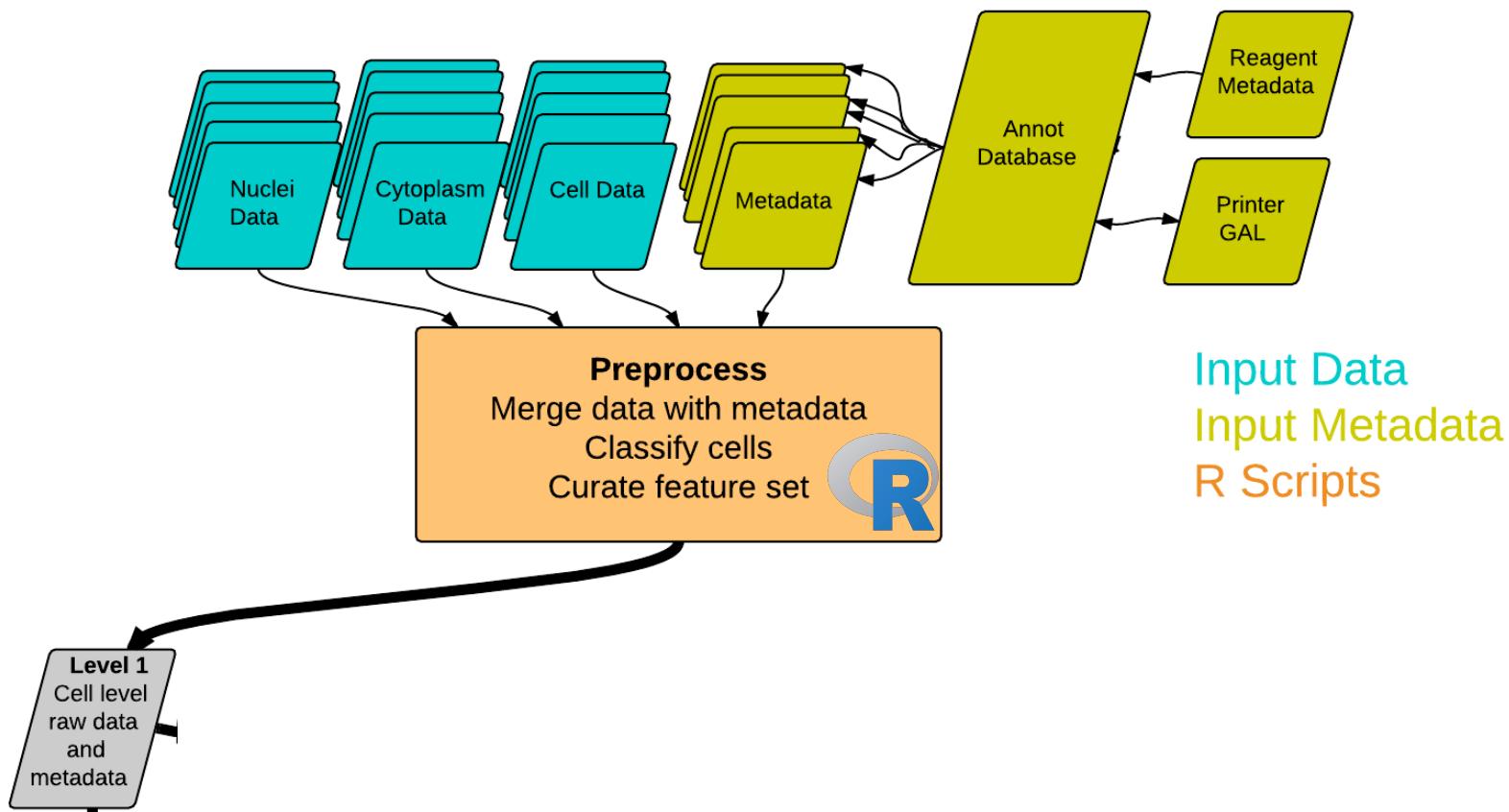
Human Mammary Epithelial Cells



- EdU (proliferation)
- DAPI (DNA)
- KRT5 (basal)
- KRT19 (luminal)

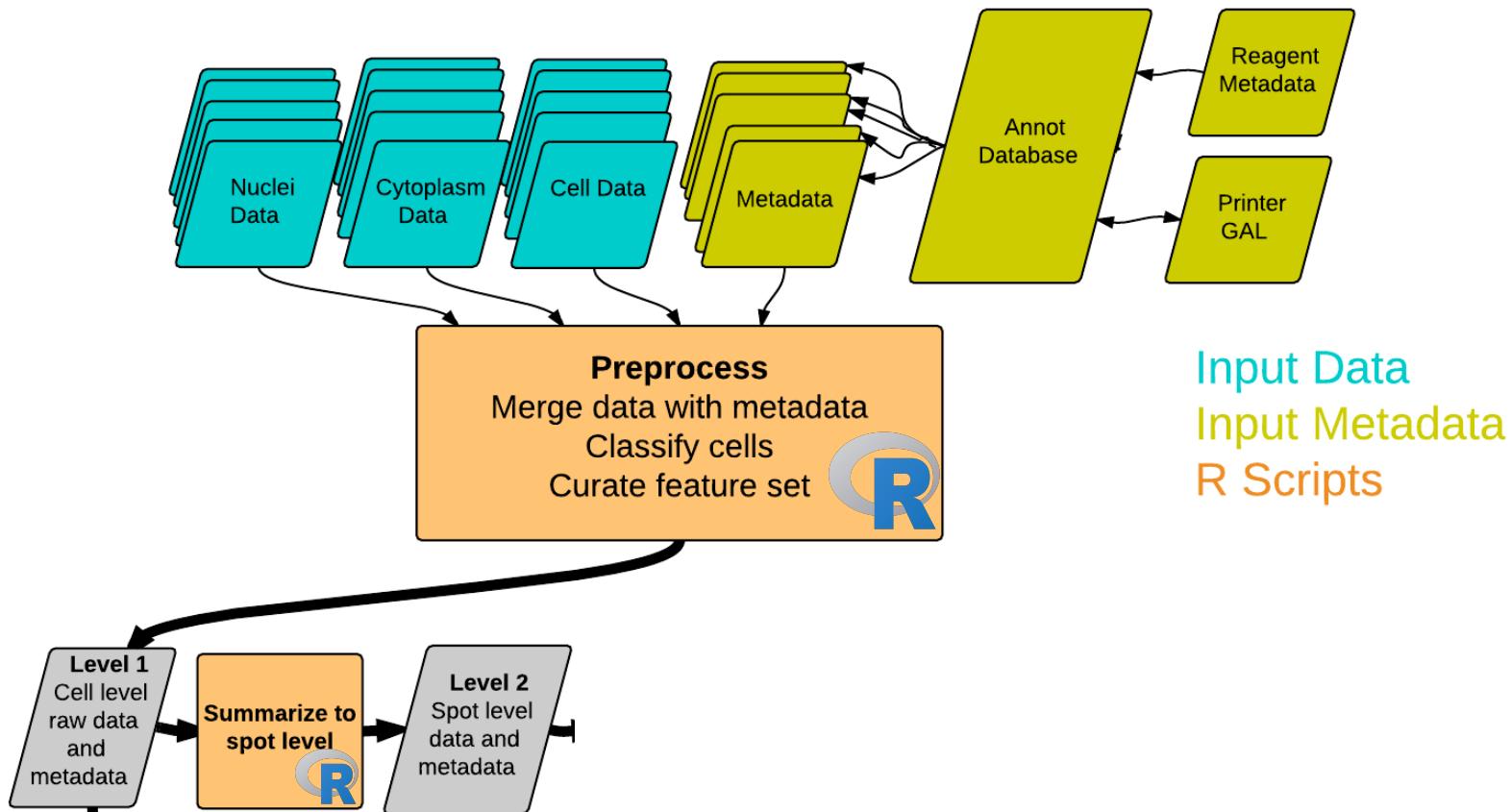
MEP-LINCS Computational Pipeline

5/31/2017



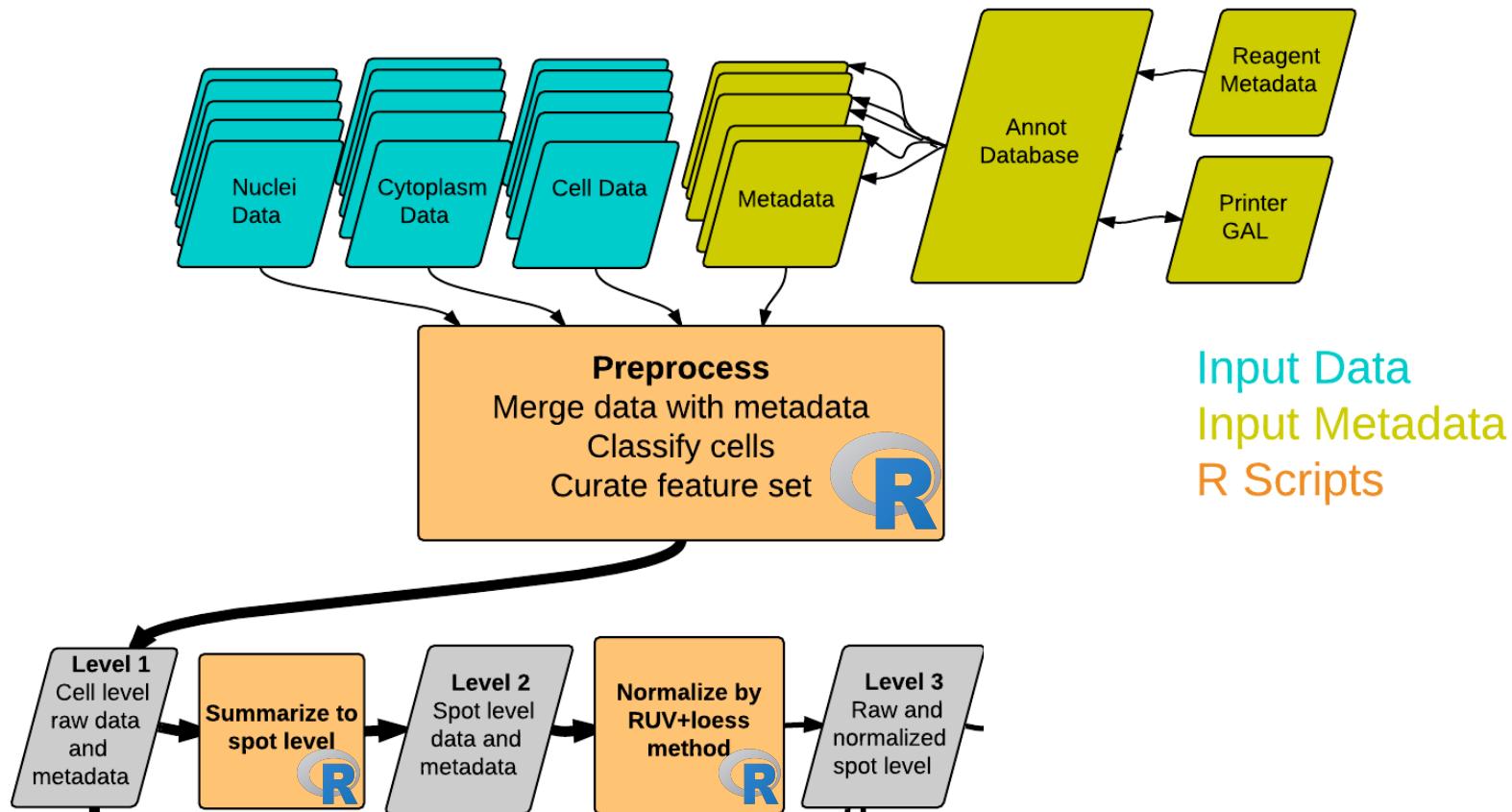
MEP-LINCS Computational Pipeline

5/31/2017



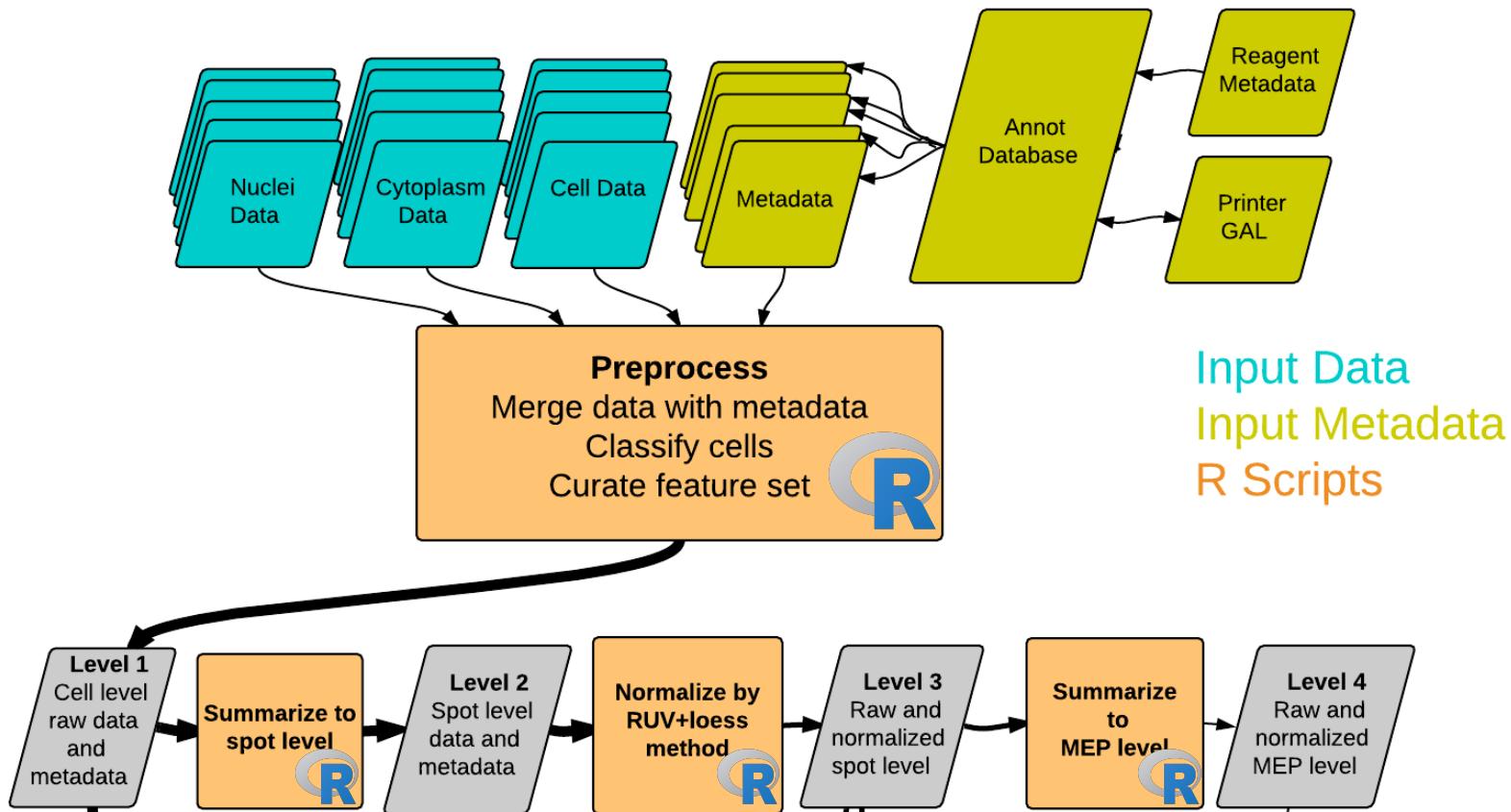
MEP-LINCS Computational Pipeline

5/31/2017



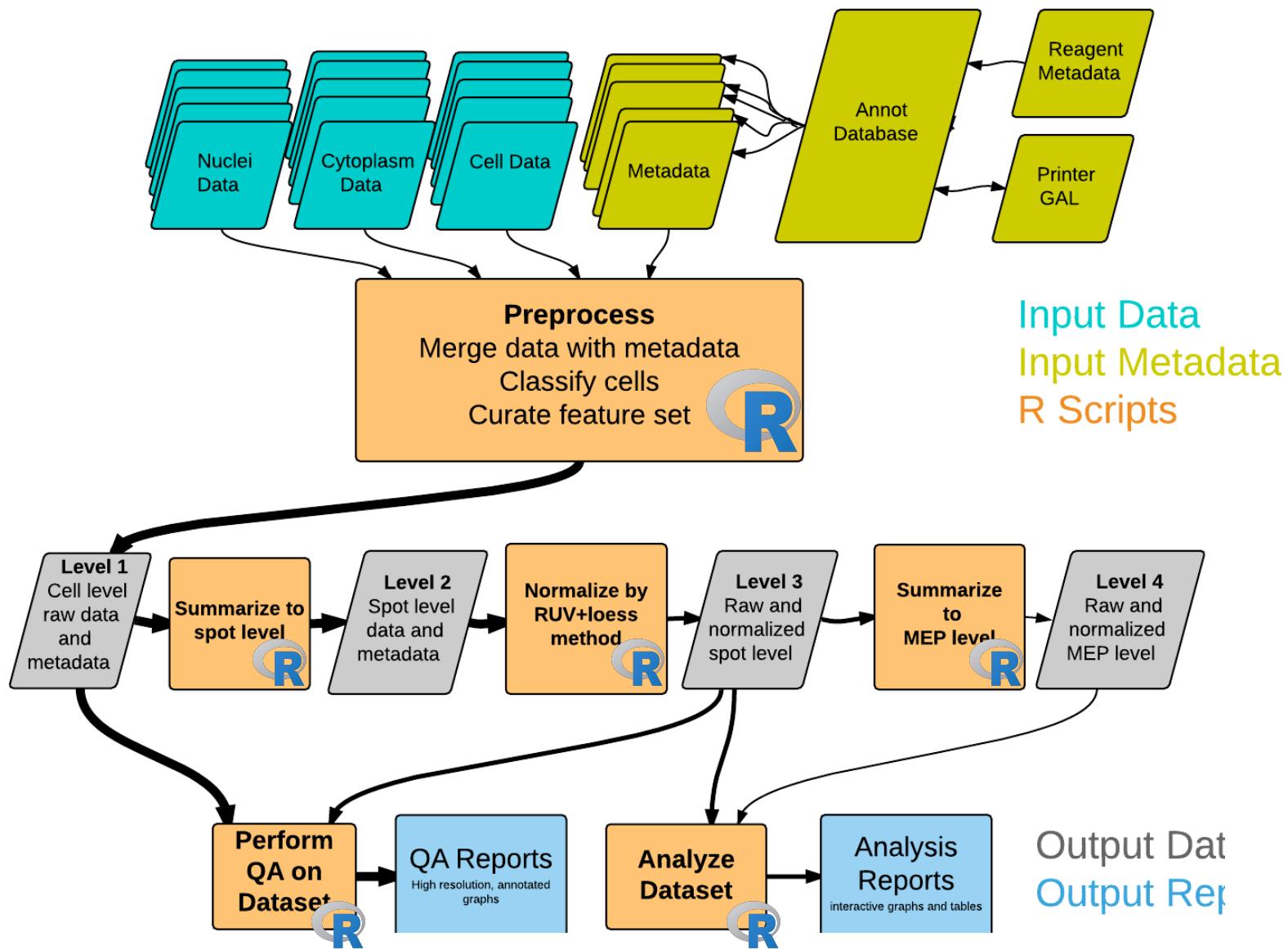
MEP-LINCS Computational Pipeline

5/31/2017



MEP-LINCS Computational Pipeline

5/31/2017



MEMA package and scripts

This repository Search Pull requests Issues Marketplace Gist

MEP-LINCS / MEMA

Code Issues 0 Pull requests 0 Projects 0 Wiki Settings Insights

R Package with preprocessing, QA and normalizing functions for MEP-LINCS data

Add topics

143 commits 5 branches 7 releases 2 contributors

Branch: master New pull request Create new file Upload files Find file Clone or download

markdane committed on GitHub Merge pull request #32 from MEP-LINCS/develop ... Latest commit 17f0eaa 16 days ago

R Fixed typos in KRT5KRT19 feature names 24 days ago

inst Updated vignette to match Synapse table id and fix title 16 days ago

man Converging vignette and data release pipelines 29 days ago

tests Deleted redundant columns from omero data 4 months ago

vignettes Updated vignette to match Synapse table id and fix title 16 days ago

.DS_Store Added support for normalizing replicate ligands besides FBS 8 months ago

.Rbuildignore Added functions developed outside of MEMA package 2 years ago

.gitignore add tar.gz to ignore 4 months ago

DESCRIPTION Updated vignette to match Synapse table id and fix title 16 days ago

LICENSE Renamed vignettes and reformatted 2 months ago

MEMA.Rproj Putting some functions into MEMA package 2 years ago

NAMESPACE Converging vignette and data release pipelines 29 days ago

README.md Update README.md 3 months ago

README.md

MEMA

The MEMA package includes functions to preprocess, QA, normalize and explore Microenvironment Microarray (MEMA) experiments. MEMAs are used in Oregon Health and Science's (OHSU) MEP-LINCS project. An overview of the project along with all data and protocols are available at <https://www.synapse.org/#ISynapse:syn2862345/wiki/72486>

Vignettes to create and explore datasets

Preprocess and QA MEMA Cell (level 1) and Spot (level 2) Data

Mark Dane
2017-05-17

Introduction

The MEMA package provides functions to preprocess, QA, analyze and explore level 1 and 2 data from the National Institute of Health's (NIH) LINCS project. MEP-LINCS uses immunofluorescent imaging to interrogate cellular responses to different microenvironments and drugs. A microenvironment is a combination of a spot of insoluble extracellular matrix proteins with media containing growth factors, cytokines, genes and drugs.

This vignette is based on the raw data from the MEP-LINCS experiment that uses Human Epithelial Mammary Cells (HEME) from a healthy 19 year old woman, cell line HMEC24L. The vignette loads one set of raw plate of raw data and its metadata from the Synapse website, preprocesses them for downstream analysis, summarizes them on the spot level and displays some Quality Assessment

plots.

Setup

We start with loading the necessary packages.

Preprocess and QA MEMA Level 3 and 4 Data

Mark Dane
2017-05-17

Introduction

The MEMA package provides functions to preprocess, QA, analyze and explore level 3 and 4 data from the National Institute of Health's (NIH) LINCS project. MEP-LINCS uses immunofluorescent imaging to interrogate cellular responses to different microenvironments and drugs. A microenvironment perturbation or MIP is a combination of a spot of insoluble extracellular matrix proteins with media containing growth factors, cytokines, genes and drugs.

This vignette is based on the summarized data from the MEP-LINCS experiment that uses Human Epithelial Mammary Cells (HEME) from a healthy 19 year old woman, cell line HMEC24L. This vignette loads one set of 8 well plates of 2 data from the Synapse website, normalizes them, summarizes them to the replicate level and displays some quality assessment and analysis figures.

Setup

We start with loading the necessary packages.

```
library(RHME)
library(httr)
library(jsonlite)
library(parallel)
library(stringr)
suppressPackageStartupMessages(library(dplyr))
library(ggplot2)
```

Synapse Login

The first step is to provide login credentials to Synapse from a configuration file previously stored in the default location `~/.synapseConfig`. An alternative is to provide a Synapse username and password as arguments in this command. More information about logging into and using Synapse is available at http://www.synapse.org/#!Synapse:getting_started#nmLoggingIntoSynapse.

suppressMessages(suppressLogin())

Download Metadata

Each plate in the dataset has a unique ID that is encoded in a barcode label attached to the plate. We provide each plate sequentially enabling sequential processing on smaller systems and parallel processing on larger systems.

We next define a query that uses annotations to locate the plate's metadata on Synapse. We run the query and use the results to download the metadata to a local file.

barcode <- "1B9805"

metadata <- sprintf("select id from syn7494872 WHERE DataFile>`Metadata` AND

Barcode='%"

barcode

Barcode%'")

metadataData <- jsonlite::fromJSON(metadata, verbose=FALSE)\$values

metadataFiles <- lapply(metadata\$labelId, synGet)

metadataFiles <- list(unlist(metadata\$getFileLocation(metadataFiles[[1]])))

Next we read and process the metadata.

metadata <- getMetadata(metadataFiles, useJsonMetadata = TRUE)

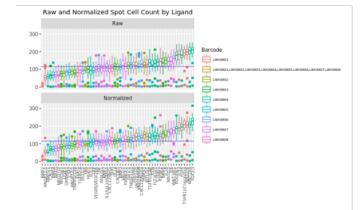
The metadata is stored in a database where each row is a spot in the plate. The metadata columns

are features that describe the microenvironment at each spot, imaging information, position labels

and other information that is useful for downstream analysis. There are 3000 spots and 66

features in the metadata. The head of some important columns are shown below.

head(metadata[, c(Barcode, Well, Spot, Ligand, EC0, Cellline, StainingSet)])



MEMA package and scripts

Optparse to handle command line arguments

Raw to Level 1

The PreprocessMEMACell.R script reads in one plate's data and metadata, merges them, gates some of the signals and writes out the results as Level 1 data in a tab-delimited file. The input data comes from an image segmentation pipeline. The metadata data comes from either OHSU's Annot (An!) database or from structured Excel files. The output can be written to the Synapse website or to a local server. Each row in the output file represents data from one cell in the experiment. The columns are either metadata about the cell, intensity and morphology measurements from the segmentation pipeline or derived features computed within the script. Help for the PreprocessMEMACell.R script is:

```
Usage: PreprocessMEMACell.R [options] BARCODE OUTPUTFILE

Options:
  -v, --verbose
    Print extra output

  -e, --excelMetadata
    Get metadata from Excel files instead of from the An! database

  -i PATH, --inputPath=PATH
    Path to local input data directory or Synapse ID for a File View.

  -r RAWDATAVERSION, --rawDataVersion=RAWDATAVERSION
    Raw data version from local server [default "v2"]

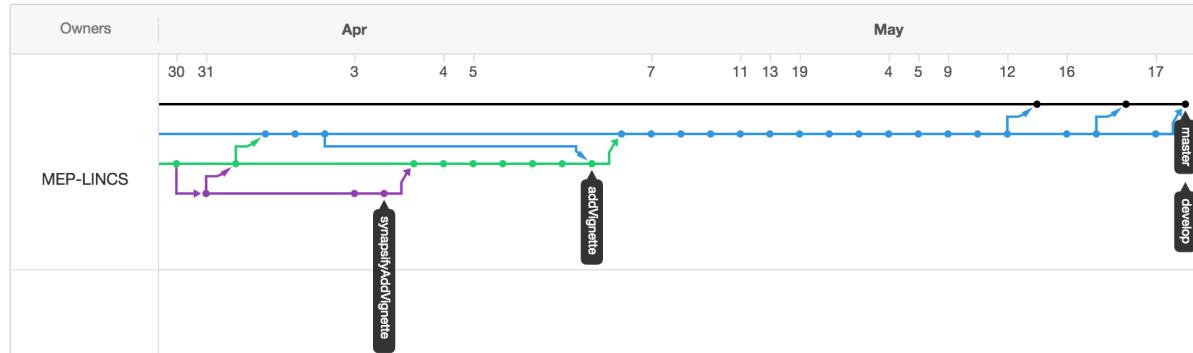
  --synapseStore=SYNAPSEID
    Store output file in Synapse directory (provide Synapse ID of Folder to store).

  -h, --help
    Show this help message and exit
```

The required BARCODE command line argument identifies the plate to be processed. Values will be similar to LI8X00641. The OUTPUTFILE argument is the full path on the local server for the output file. On many systems an output name that starts with "/tmp/" can be used if the output is to be stored on Synapse and not retained on the local server.

MEMA package and scripts

Collaborative development



Screenshot of the GitHub Issues page for the 'MEP-LINCS / MEP_Processing' repository. The page shows 7 open issues. The issues listed are:

- Add scriptlink to provenances
- get LINCS IDs as annotations
- check raw files from previous release
- turn this into a package
- determine if 'rawDataVersion' annotation is being used
- Update SOPs and link into raw data provenance
- Info about CellProfiler pipeline

Each issue includes a link to its details, the author, labels (e.g., enhancement, synapse), and a count of comments or reviews.

Synapse website for assay explanation and data access

synapse.org/MEP_LINCS

Microenvironment Perturbagen (MEP) LINCS ★

Synapse ID: syn2862345 DOI: doi:10.7303/syn2862345 Storage Location: Synapse Storage

Wiki Files Tables Discussion Docker Beta

Microenvironment Perturbagen (MEP) LINCS

- Assay Overview
- Data Description
- Data Releases
- Interactive Analysis
- MEP Data Explorer
- Cell Lines
- Metadata
- Protocols
- Quality Assessment
- External Data Resources

Edit Order

NIH LINCS PROGRAM THE UNIVERSITY OF TEXAS MDAnderson Cancer Center

[Assay Overview](#) [Data Description](#) [Download Data](#) [Explore Data](#) [Ask Questions](#)

New Data Release (16-May-2017)! - See [here](#) for details.

The MEP-LINCS center is one of six Data and Signature Generating Centers of the NIH Library of Integrated Network-based Cellular Signatures (LINCS) program.

The goal of MEP-LINCS is to develop computational strategy and generate data to elucidate how microenvironment signals affect cell intrinsic intracellular transcriptional- and protein-defined molecular networks to generate experimentally observable cellular phenotypes.

Data generated from the MEP-LINCS project is freely available to download. New to Synapse? Click below to register for free in order to download files ([More info on registering](#)).

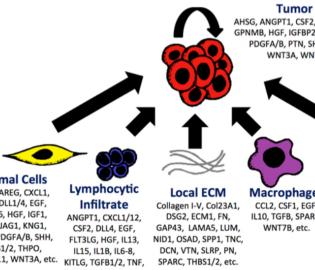
[Create a Free Synapse Account](#)

Have a question? Post it in the [Discussion Forum](#).

We will infer regulatory relationships by combining measurements of microenvironment perturbagen-induced changes in multiple cellular phenotypes, RNA expression, and regulatory protein expression in a core set of cell lines with measurements of responses of the same lines to chemical and genomic perturbagens made across the [LINCS Data and Signature Generating Centers](#).

Assay Overview

The biological behaviors of normal and diseased cells and their responses to therapeutic agents are strongly influenced by the regulatory signals they receive from the microenvironment (ME) in which they reside. These signals come from direct interactions with insoluble extracellular matrix and cellular proteins as well as soluble proteins, peptides, or glycoproteins.



Discussion Board

Topic

Welcome to MEP-LINCS: READ ME FIRST! 7 months ago

Are there microenvironments that alter the lineages in the human mammary epithelial cells (HMEC) cell lines? 4 months ago

How do I identify microenvironments that effect proliferation? 5 months ago

How do I download data from the current data release? 6 months ago

Synapse Discussions

The Synapse team is happy to announce the release of the new Discussion feature to help your team create a record of communications within your project. We will continue to improve and extend this feature area based on your feedback. Please let us know how you're using this feature and any problems or ideas for improvements by emailing us at discussion@synapse.org.

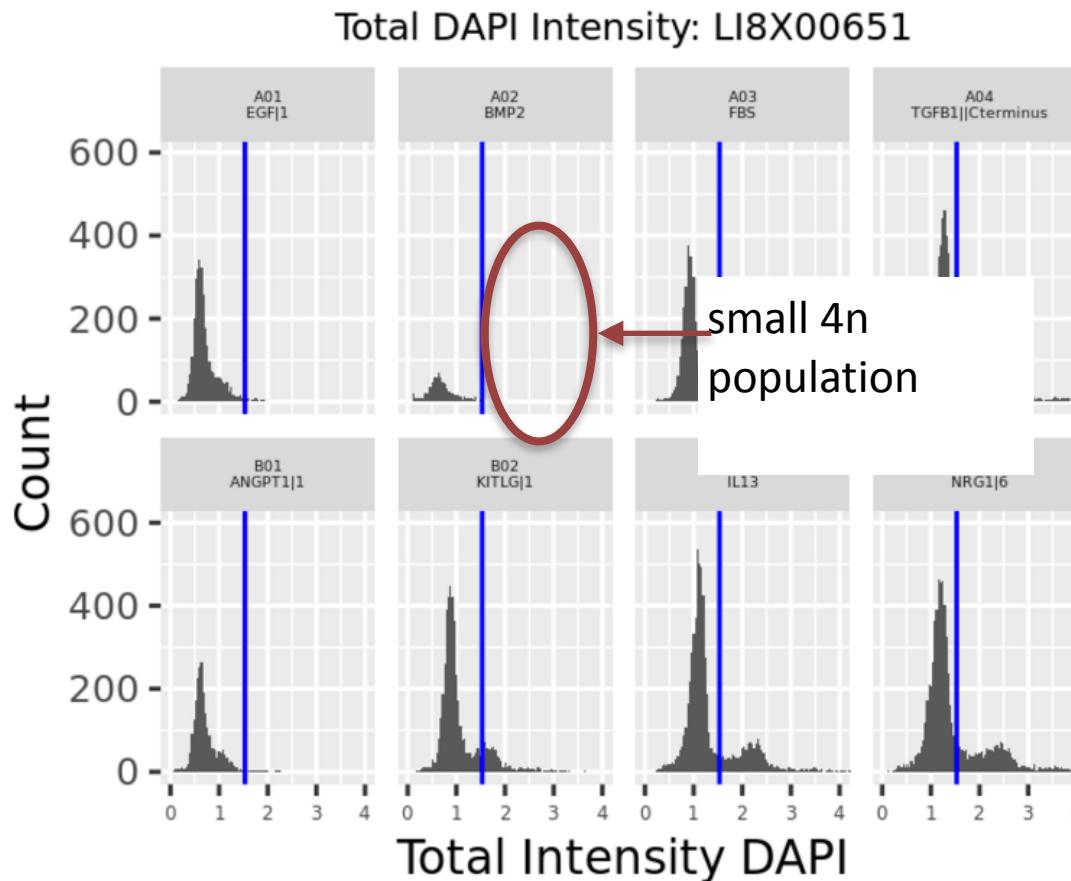
[Learn More](#)

Links to Data and Reports

Level 1 data consists of eight files per study (cell line + staining set).			
CellLine	StainingSet	1	2
HMEC12L_SS1		syn0518693 , syn0518699 , syn0518706 , syn0548921 , syn0548704 , syn0548707 , syn0518714 , syn0518721 , syn0518756 , syn0547010 , syn0548716 , syn0548772 , syn0518822 , syn0518825	syn0547111 , syn054719 , syn054723
HMEC12L_SS4		syn0518687 , syn0518690 , syn0518696 , syn0548371 , syn0548377 , syn0548484 , syn0548487 , syn0548497 , syn0518702 , syn0518709 , syn0518718 , syn0548850 , syn0548886 , syn0548893 , syn0548976 , syn0518793 , syn0518794	syn0548891 , syn0548992 , syn0548993
HMEC12L_SS3		syn0537446 , syn0537467 , syn0537490 , syn0549495 , syn0549591 , syn0549594 , syn0537518 , syn0537530 , syn0537567 , syn0549967 , syn0549984 , syn0549990 , syn0537783 , syn0537883 , syn0537986	syn0549997 , syn0549993 , syn0549997 , syn0549987 , syn0549984
HMEC24L_SS4		syn0537483 , syn0537577 , syn0537617 , syn0549085 , syn0549999 , syn0549108 , syn0537967 , syn0538017 , syn0538064 , syn0549119 , syn0549129 , syn0548406 , syn0538452	syn0549122 , syn0549123 , syn0547138
HMEC24L_SS2		syn0537144 , syn0537169 , syn0537176 , syn0549049 , syn0549051 , syn0549053 , syn0537200 , syn0537204 , syn0540061 , syn0540071 , syn0540073 , syn0537781 , syn0538044 , syn0538420 , syn0540989 , syn0549117 , syn0538321 , syn0539549 , syn0540495 , syn0549126 , syn0549128 , syn0537967 , syn0537972 , syn0538045	syn0549054 , syn0549126 , syn0549127 , syn0549134 , syn0549136
MCF10A_SS1		syn0537300 , syn0537304 , syn0537374 , syn0538074 , syn0538078 , syn0538148 , syn0540061 , syn0540073 , syn0540075 , syn0537781 , syn0538044 , syn0538420 , syn0540989 , syn0549117 , syn0538651 , syn0539182	syn0549052 , syn0549053 , syn0549054 , syn0549117 , syn0549118 , syn0549119 , syn0549120 , syn0549121 , syn0549122 , syn0549123 , syn0549124 , syn0549125 , syn0549126 , syn0549127
MCF10A_SS2		syn0537791 , syn0537957 , syn0537962 , syn0538045 , syn0538046 , syn0538126 , syn0549238 , syn0538277 , syn0538298	syn0549238 , syn0549244 , syn0549246 , syn0549248 , syn0549252 , syn0549253 , syn0549262
MCF10A_SS3		syn0537894 , syn0538045 , syn0538126 , syn0549241 , syn0549244 , syn0549246 , syn0549253	syn0549262 , syn0549263
MCF10A_SS4		syn0537891 , syn0538045 , syn0538126 , syn0549241 , syn0549244 , syn0549246 , syn0549253	syn0549262 , syn0549263

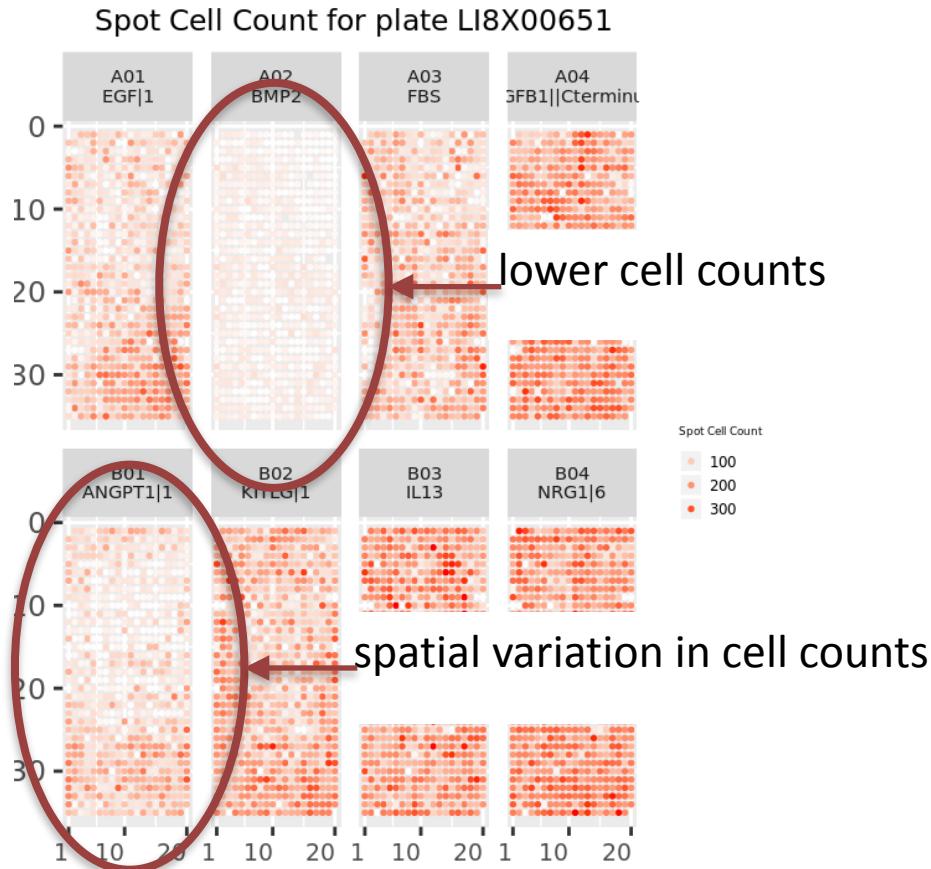
Cell and spot level quality assessment

- QA to look for technical and biological variations
- knitr-created html files with explanations and lightly-interactive figures



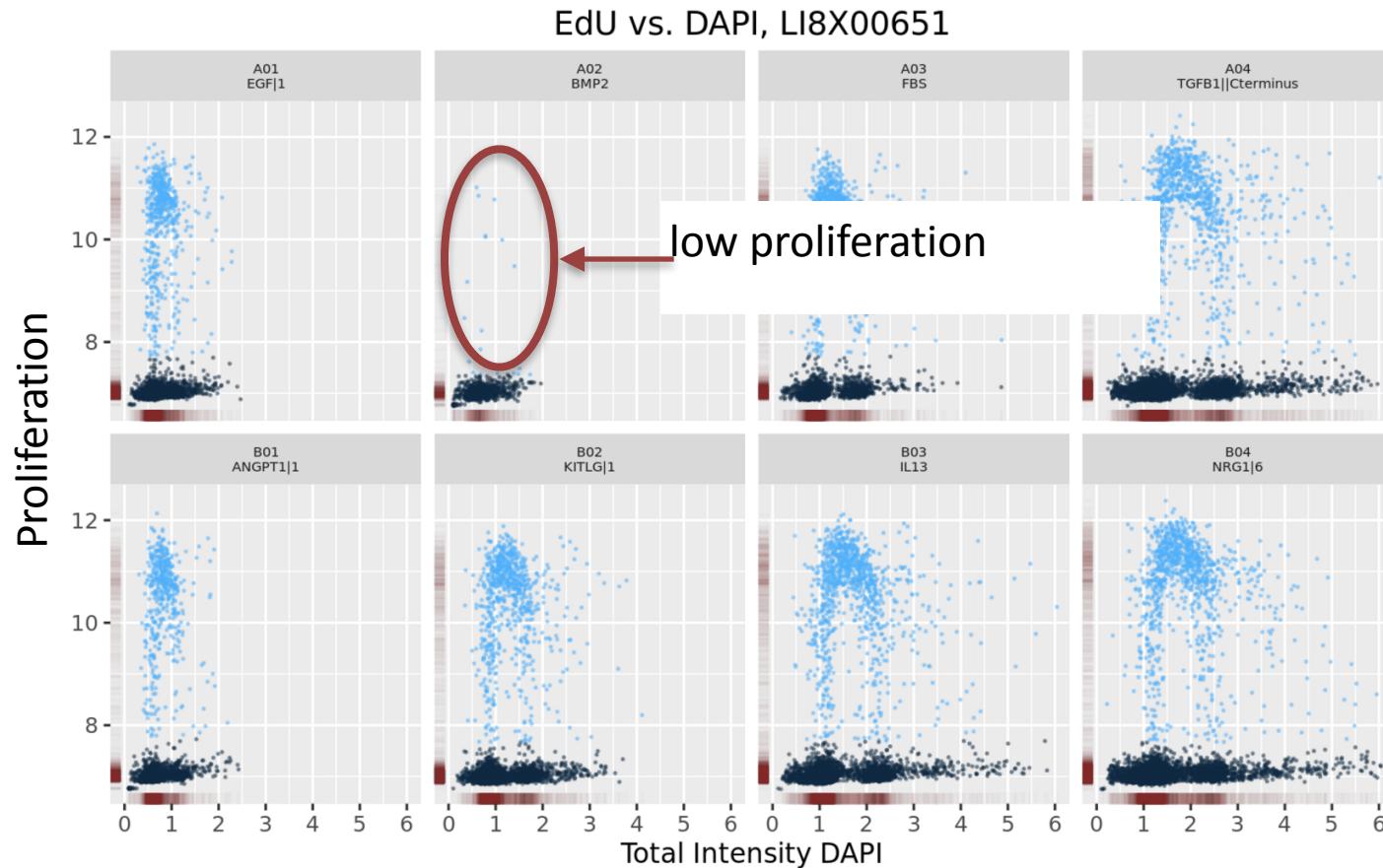
Cell and spot level quality assessment

- QA to look for technical and biological variations
- knitr-created html files with explanations and lightly-interactive figures

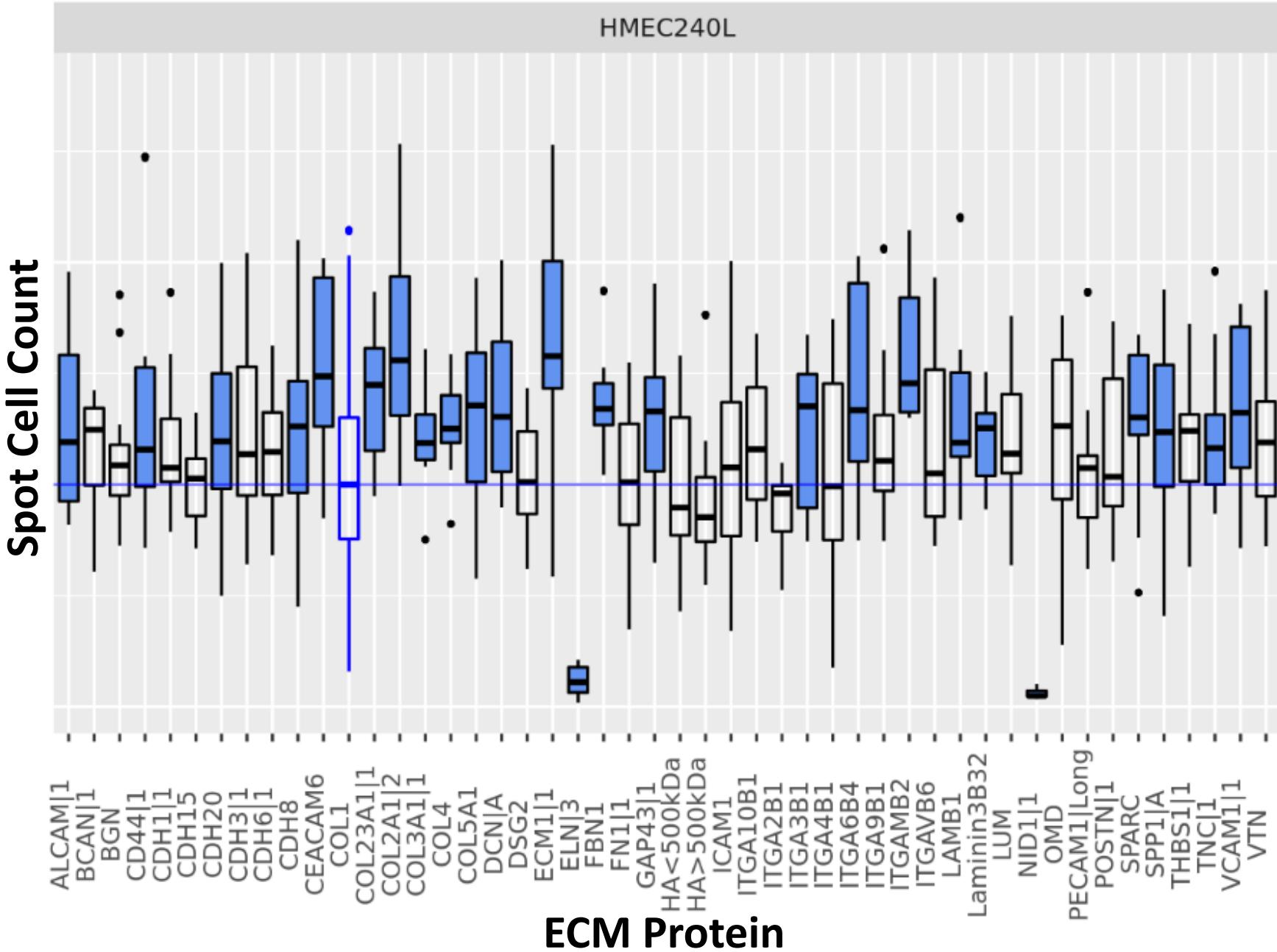


Cell and spot level quality assessment

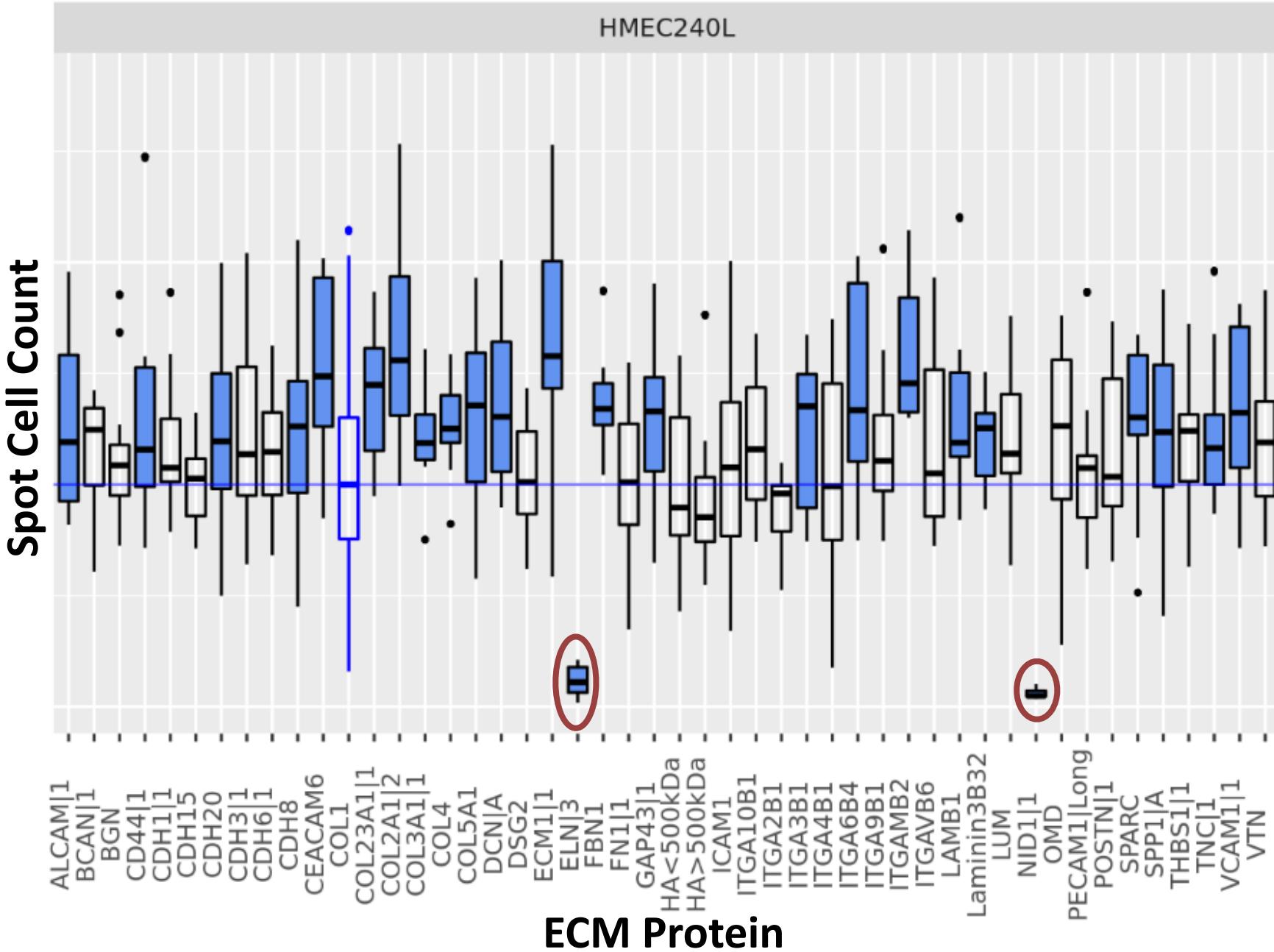
- QA to look for technical and biological variations
- knitr-created html files with explanations and lightly-interactive figures



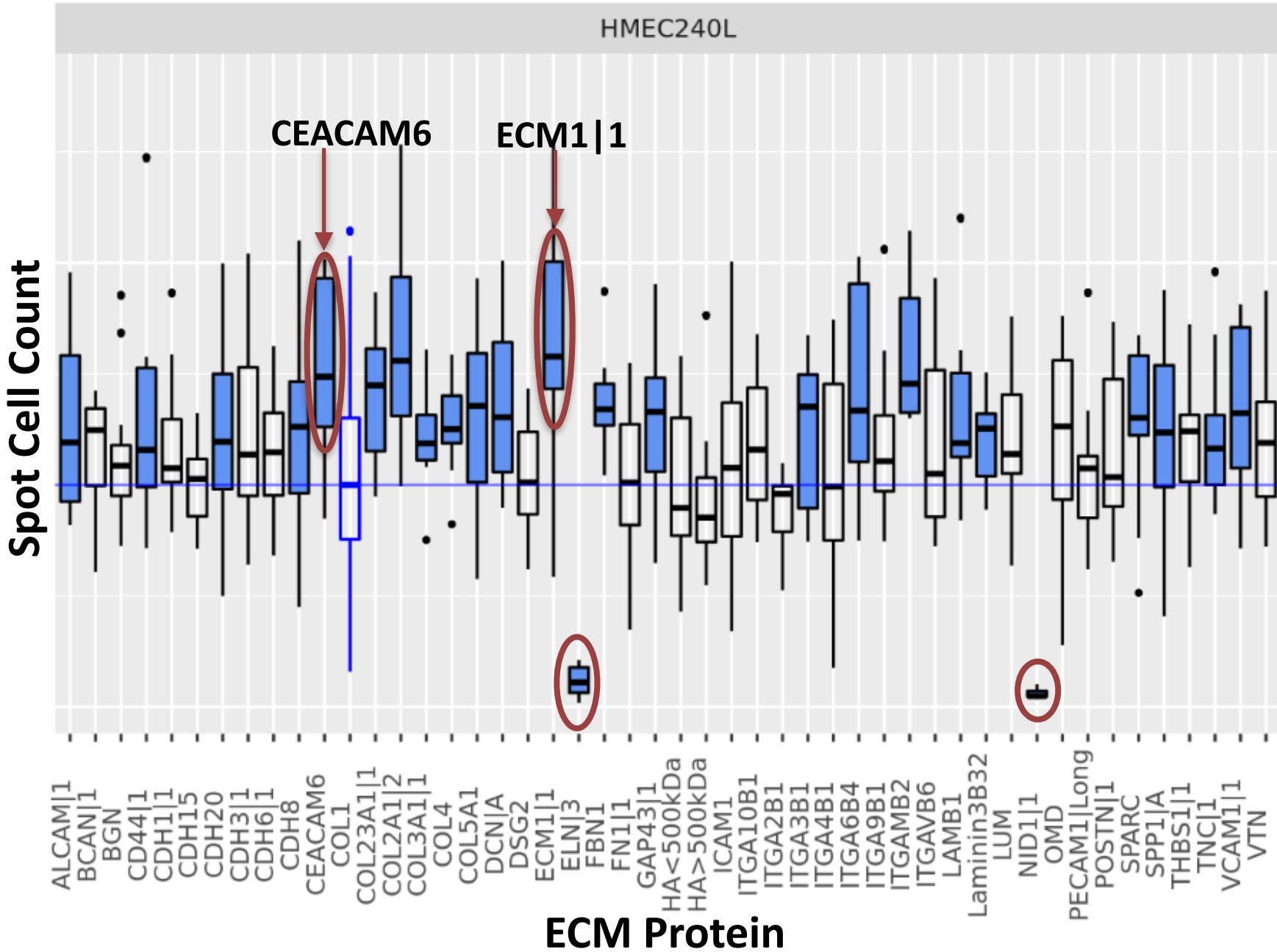
ECM proteins modulate cell growth and adhesion



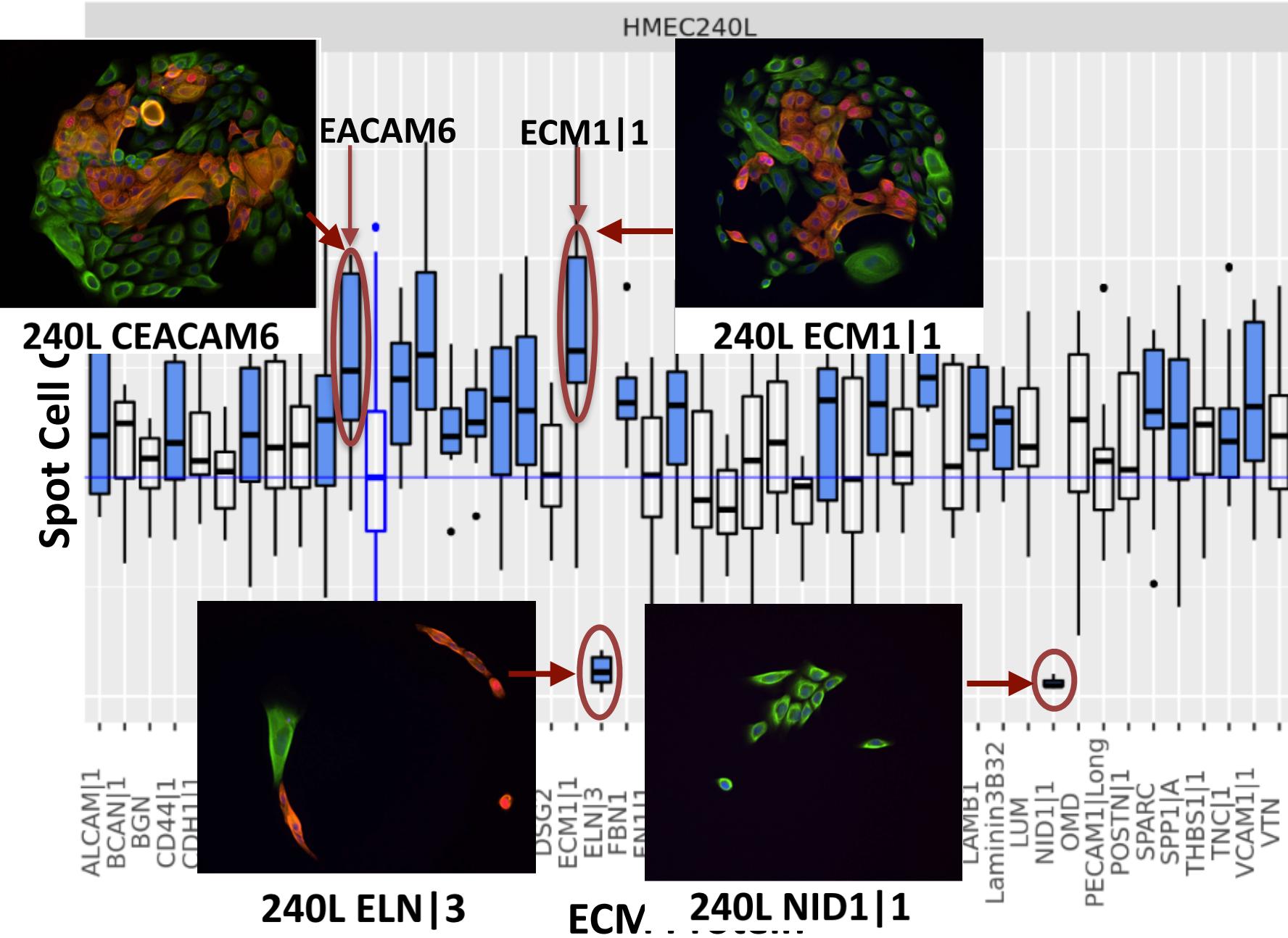
ECM proteins modulate cell growth and adhesion



ECM proteins modulate cell growth and adhesion



ECM proteins modulate cell growth and adhesion



Shiny apps for data exploration

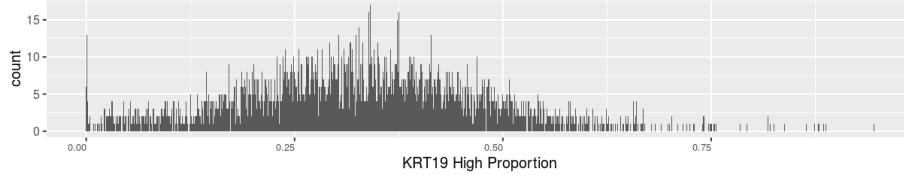
<https://dane.shinyapps.io/HMEC240LLineageDashboard/>

HMEC240L Lineage

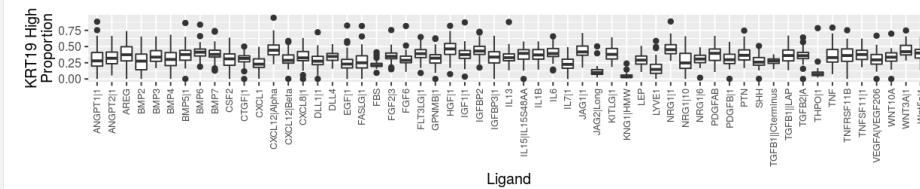
</> Source Code

Include Selected

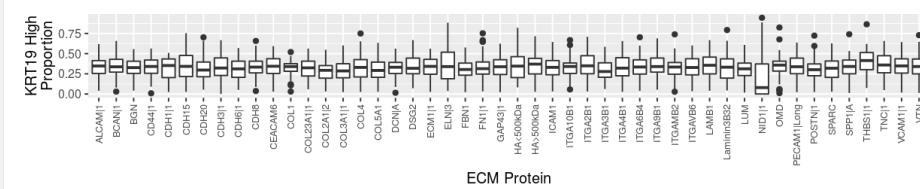
KRT19 High Histogram



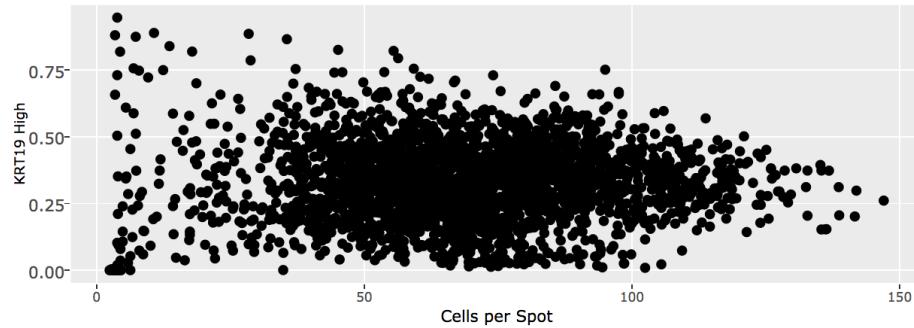
KRT19 High and Ligands



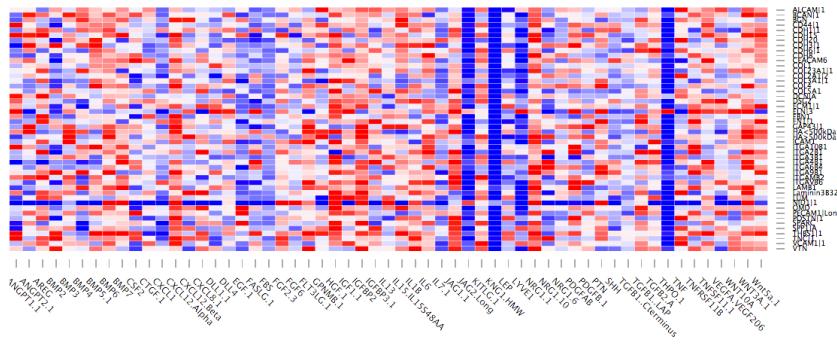
KRT19 High and ECM Proteins



KRT19 High vs. Spot Cell Count



KRT19 High Heat Map



Next Steps

- Apply machine learning directly to the images
- Gather proteomics (RPPA) and transcriptional data
- Explore the cell level data to explore heterogeneity within the cell populations
- Move protocols into 96 well plates
- Explore microenvironment modulation of drug responses

Microenvironment Perturbagen LINCS

www.synapse.org/MEP_LINCS

Overall PIs

Joe Gray

Gordon Mills

Data generation

James Korkola (PI)

Mark LaBarge

Gordon Mills

Sean Gross

Tiera Liby

Rebecca Smith

Wallace Thompson

Kaylyn Devlin

David Kilburn

Paul Spellman

Spencer Watson

Data analysis

Laura Heiser (PI)

Michel Nederlof

Damir Sudar

Adam Margolin

Elmar Bucher

Outreach

Laura Heiser (PI)

Larsson Omberg

Kenny Daily

Heidi Feiler





<https://github.com/MEP-LINCS>

synapse.org/MEP_LINCS

<https://omero.lincsclarion.org>

dane@ohsu.edu