Integrated Assignment, Part 2 (Take-home)

**Exploratory Data Analysis of Biological Data using R**
May 21-22, 2015, Toronto, ON

This assignment is an optional continuation of the Integrated Assignment. Some of the topics may cover material from Day 2 of the workshop or may be slightly more advanced.

In addition to the datasets described in Part 1, the Encyclopedia also includes:

1. Mutation statuses of the cell lines for a gene panel (`mut`)
2. Pharmacological profiles of the cell lines in response to a panel of anti-cancer drugs (`pharm`)

These datasets are summarized in heatmaps below. Only a portion of the data is shown below.
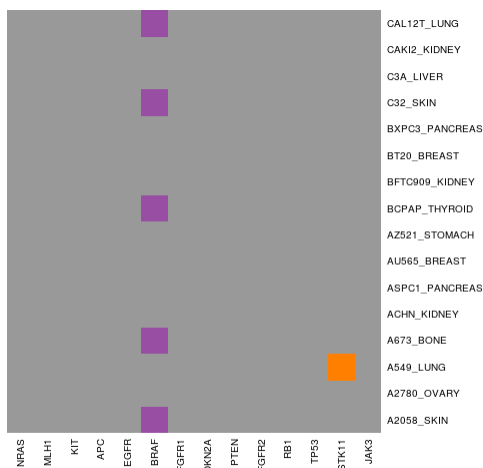
Mutation status of select genes.
Purple, missense. *Orange*, nonsense.
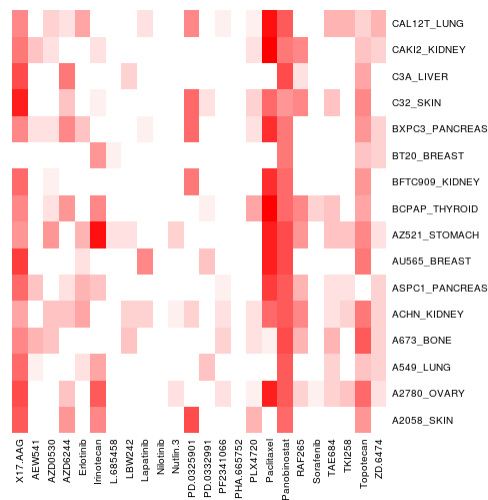
Sensitivity to anti-cancer drugs.
White, insensitive. *Red*, sensitive.

The pharmacological profiles were pre-analyzed in the CCLE project, as described in Barretina, 2012. Briefly, drugs were administered to cell cultures at 8 different concentrations and cell viability was measured and normalized to negative and positive controls (where 0% indicates viability is comparable to negative controls, and -100% indicate viability is comparable to positive controls). Sigmoid curves were fitted to the dose-response data. Activity area, maximum activity (A_{max}), effective concentration (EC50), and inhibitory concentration (IC50) were determined as illustrated in the below schematic. These four measures of dose-response assess the efficacy of a drug against cancer cell lines. A drug that is effective at inhibiting the growth of a cell line would have low IC50 and EC50, as well as high activity area and maximum activity in the dose-response curve against the given cell line.
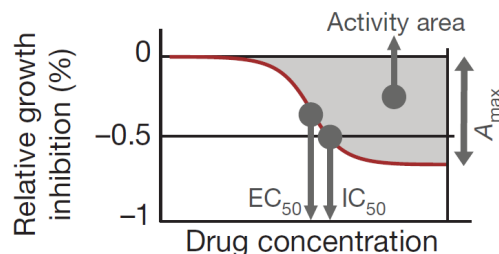
Instructions

(Continued from Part 1)

Section 8 Explore the gene expression again

RUNX1 is a transcription factor that regulates the differentiation of haematopoeitic stem cells into mature blood cells. It is involved in many different cancers. Further, it is known to physically interact with JUN. Given that physically interacting proteins are likely to be co-expressed, are RUNX1 and JUN expression correlated?

Follow the instructions in *Stats2015_IntegratedAssignment_Part2_Questions.R.*

Section 9 Explore the pharmacological profiles

Four response variables were measured in the pharmacological study (`ic50`, `ec50`, `act.area`, and `act.max`). To analyze them visually, we can create scatter plots with two variables at a time.

Such scatter plots show the *bivariate joint distribution* of two variables, while histograms show how a single variable is distributed (i.e. *univariate distribution*).

Follow the instructions in *Stats2015_IntegratedAssignment_Part2_Questions.R.*

Section 10 Create publication-quality plots

The ggplot2 library has become popular among R users.

Follow the instructions in *Stats2015_IntegratedAssignment_Part2_Questions.R* to create beautiful plots with ggplot2.