Project 2: Cellular Response to Drug Perturbations

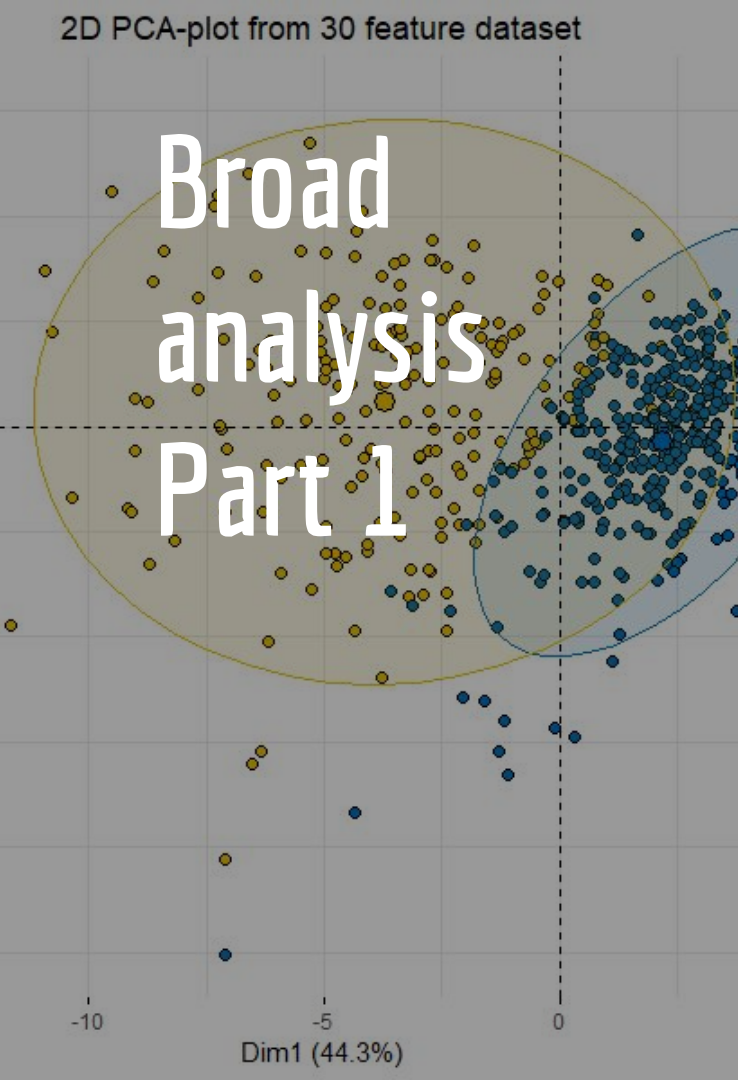
Analysis of the effects of Gemcitabine

Group 3

Sophie Zhu, Tabita Häßner, Shivani Nundoo, Diego Montero Solano

Available Data Sets

- **Gene expression** of 819 samples, 13299 entries
  - treated
  - untreated
- **Metadata**: cell, drug, dose, time and tissue
- **Somatic mutations** from cell lines
- **GI50** values of 61 cancer cell lines treated with 15 drugs
- **Gene copy-number** alterations
- **Basal gene expression** of the cell lines
- Annotation charts with **additional information** about the **cell lines** and **drugs**

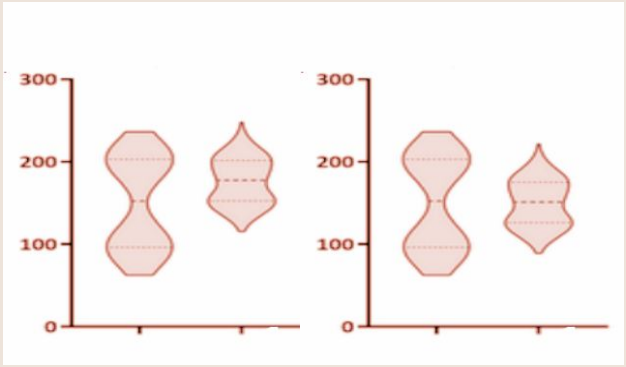


Broad analysis Part 1

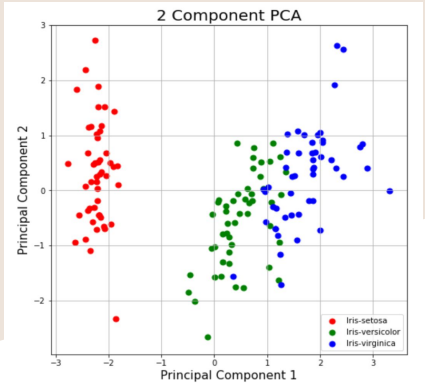
- ➔ Two paired samples test (treated and untreated) are available.
- ➔ Ensure that they are normalised.
- ➔ Reduce dimensions to eliminate data redundancy. (using PCA)

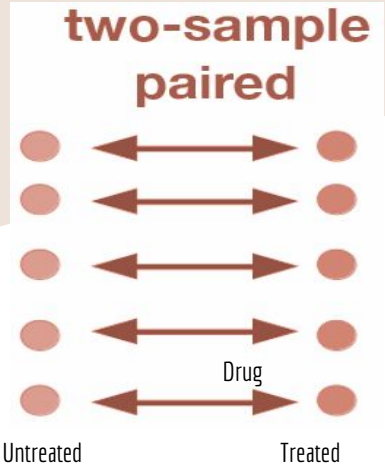
Is a batch effect or any relation between the sets of data identified?

Broad Analysis Part 1



Before Normalisation After Normalisation





Untreated Treated

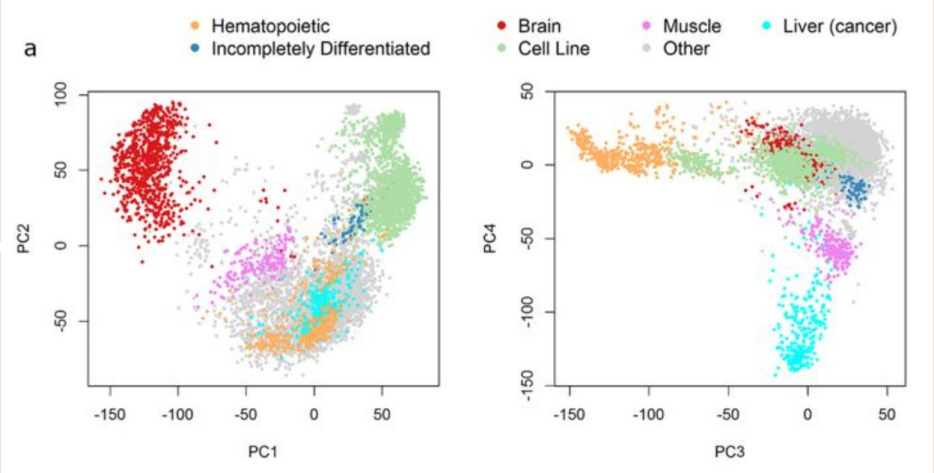
Normalisation

PCA associated with metadata

Paired two samples t-test

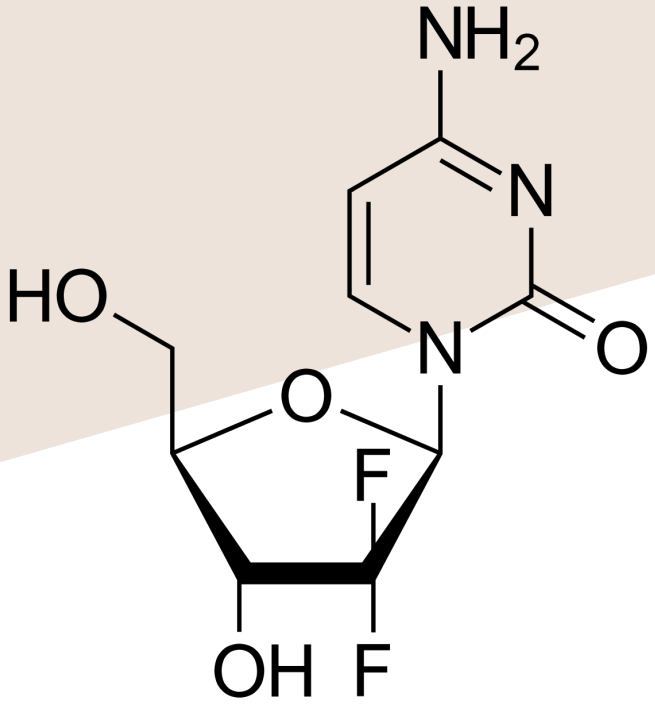
Broad analysis Part 2 - Influence of drug treatment

⟶ Fold Change
⟶ Heatmap (Sample = celllines, rows= mean of gene expression difference after drug treatment): Are there drugs with a similar pattern?
⟶ PCA of all genes (for example tissue associated?)



Gemcitabine

- Chemotherapy medication
- Faulty base
- Incorporated into DNA with normal nucleoside added next to it → base-excision repair cannot recognize it → irreparable error
- inhibition of further DNA synthesis
- Ultimately leads to cell death

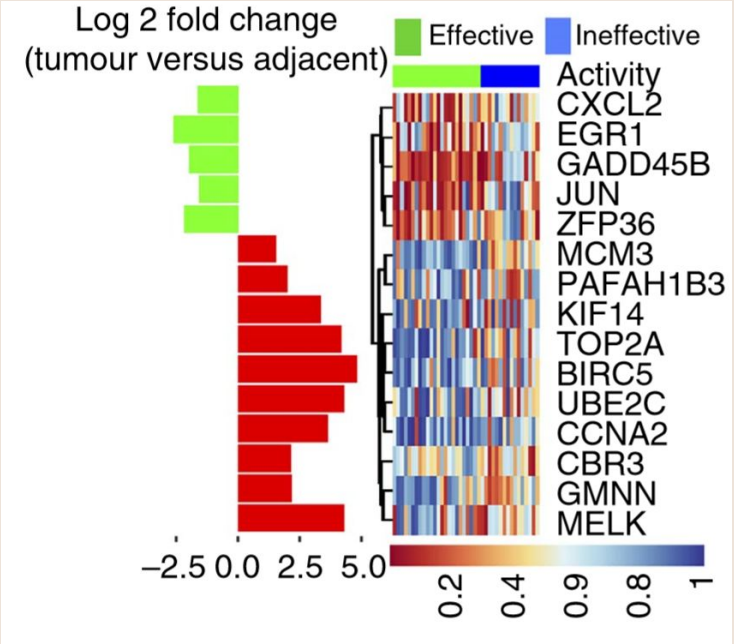


https://en.wikipedia.org/wiki/Gemcitabine

Finding drug response biomarkers

Comparison of gene expression

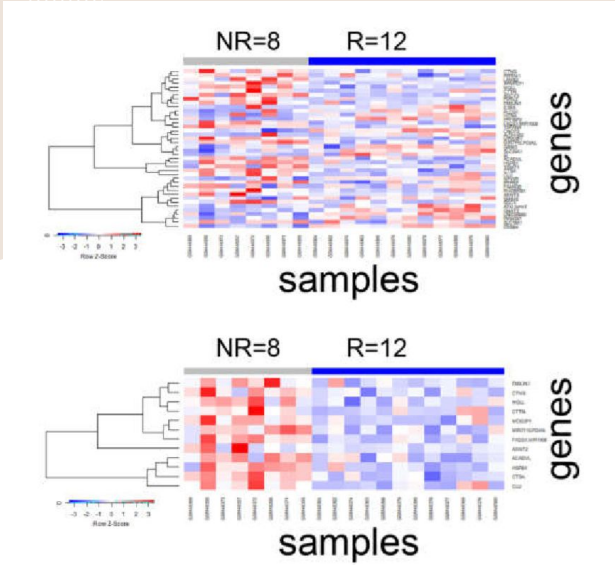
- Databases used: treated and untreated
- Look at difference between cell lines
- Filter out the 20 Genes with the biggest difference ⟶ Biomarkers
- Which genes are up or downregulated by Gemcitabine?



Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets (Chen *et al.*, 2017)

Does microsatellite instability (MSI) affect the impact of gemcitabine?

1. **High GI_{50} Values ⟶ rather MSI-H (High)?**
   ⟶ Violinplots
2. **MSI-H celllines ⟶ more mutations ?** (2 Violinplots: MSI-H, MSI-L/mutations)
3. Heatmap separated by MSI-H/L; colour⟶ Fold change of biomarkers



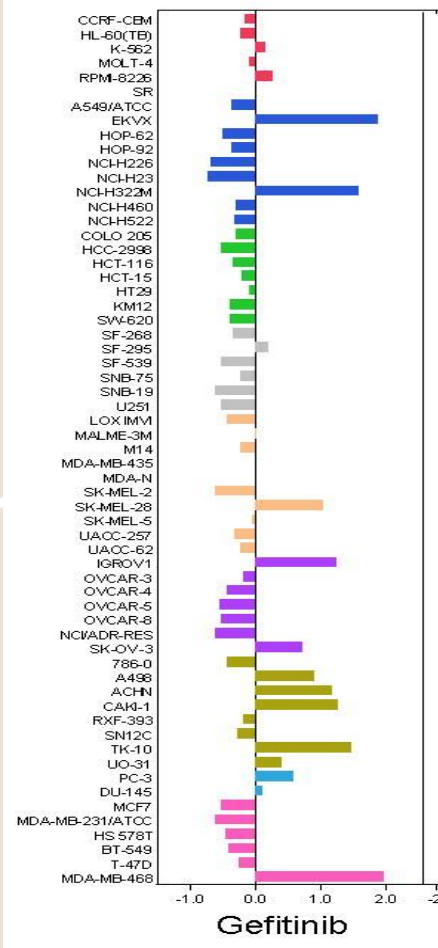
What is the effect of the somatic mutation type on gemcitabine sensitivity?

1. <u>Sub Question: Is the drug sensitivity for gemcitabine dependend on the cell line?</u>

- Using NegLogGI50.rds data
- Values are -log10 transformes -› higher values indicate higher sensitivity
- mean graph plot of GI50 values for Gemcitabine
- coding according to tissue of origin

2. <u>Sub Question: Is there a relation between mutation type and cell line?</u>

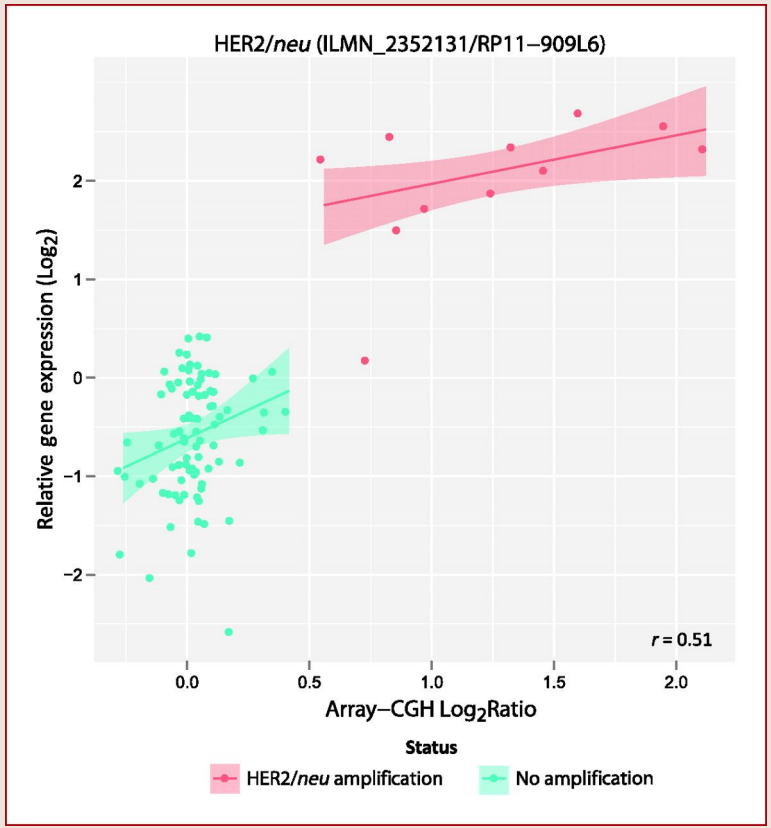
- Using CCLE_mutations.rds data
- several variables as criteria for mutation type eg: Hugo symbol , codon_change, protein_change
- create a matrix which counts of the same type of mutation
- **Chi square test** to investigate relationship between variables and different cell lines



Holbeck, Susan L et al. "Analysis of Food and Drug Administration-approved anticancer agents in the NCI60 panel of human tumor cell lines." *Molecular cancer therapeutics* vol. 9,5 (2010): 1451-60. doi:10.1158/1535-7163.MCT-10-0106

Can the drug sensitivity be predicted based on gene expression and gene copy-number?

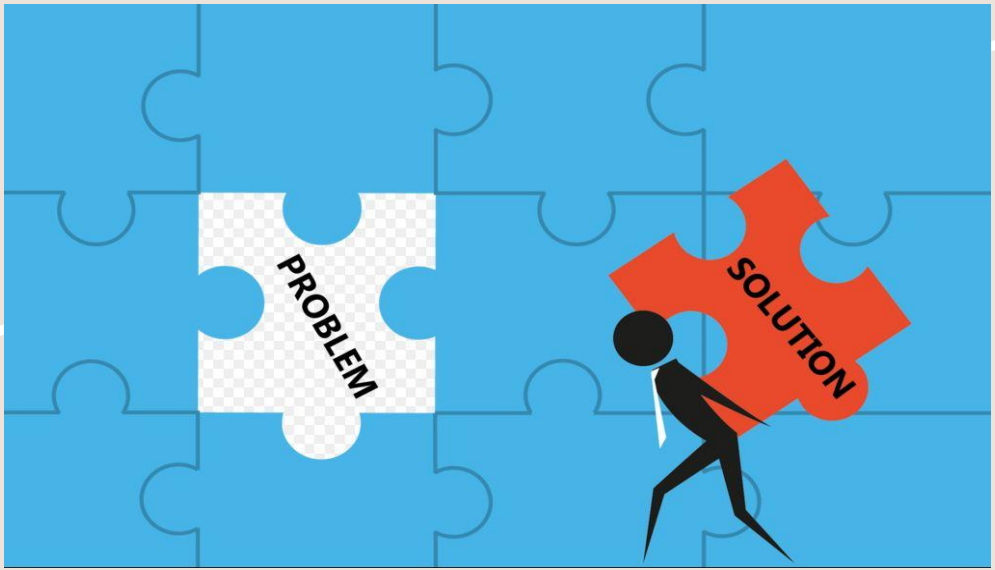
- Check correlation between gene expression and gene copy number
- Predict drug sensitivity using **linear regression**
  - Two separate LR if correlation is low
  - Only one LR if correlation is high



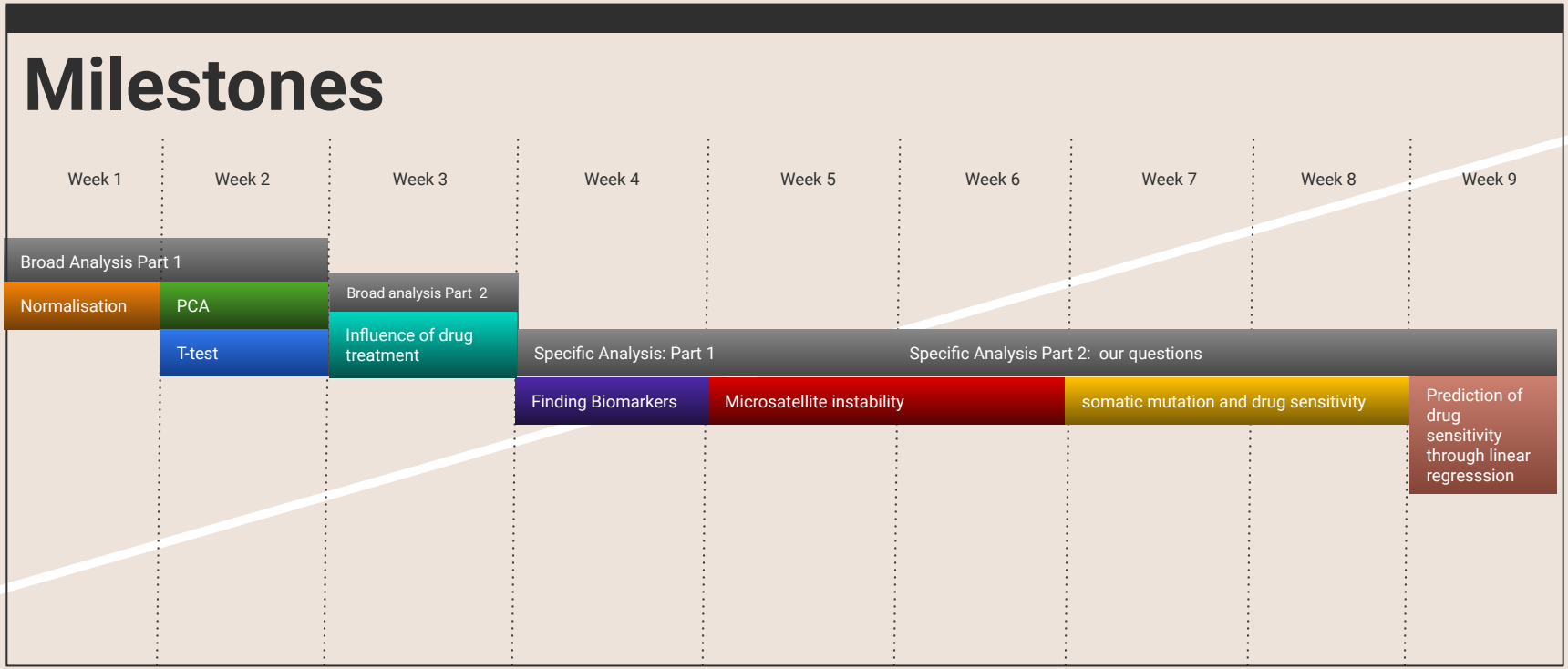
Clinical implications of Gene Dosage and Gene Expression Patterns in Diploid Breast Carcinoma (Parris *et al.*, 2010)

Specific analysis - Final step

- Conduct extra tests if the results give rise to any new interesting questions.
- Analyze results, compare with literature and solve any unforeseen problems that come up.



Timeline


Milestones
Week 1
Week 2
Week 3
Week 4
Week 5
Week 6
Week 7
Week 8
Week 9
Broad Analysis Part 1
Normalisation
PCA
T-test
Broad analysis Part 2
Influence of drug treatment
Specific Analysis: Part 1
Specific Analysis Part 2: our questions
Finding Biomarkers
Microsatellite instability
somatic mutation and drug sensitivity
Prediction of drug sensitivity through linear regresssion

Thank you for your attention