

Introduction to Bioinformatics

Introduction to NGS/Genomics Technologies (Microarray)

Credits to **Thomas Girke** of UCR and Reiner Sprang, 2021.

A preview of the past few weeks

Profiles are lists of quantified molecular features

A preview of the past few weeks

Profiles are lists of quantified molecular features

There are profiles of ...

- RNA Transcripts (mRNA, miRNA, lncRNA, ...)
- Proteins (total expression, phosphorylation, ubiquitination ...)
- Metabolites (intra cellular, secreted, ...)
- Epigenetics (DNA methylation, histone methylation, histone acetylation)
- Transcription factor binding (ChIP)
- DNA copy number variation
- Microbiomes (16S rRNA, Metagenomes, ...)

A preview of the past few weeks

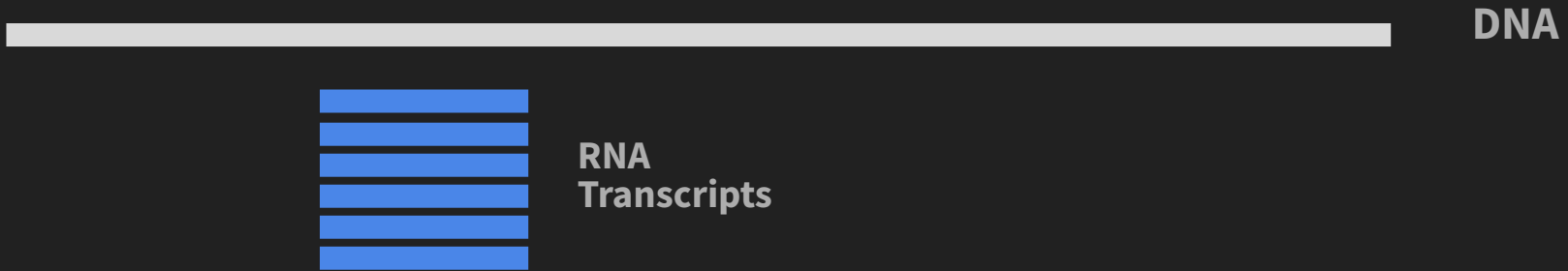
Profiles are lists of quantified molecular features

Profiles can be generated by different technologies

- RNA Transcripts (microarray, nanoString, RNAseq)
- Proteins (MassSpec, protein array)
- Metabolites (NMR, MassSpec,...)
- Epigenetics (ChIP-seq, bisulfate sequencing, ATAC-seq)
- Transcription factor binding (ChIP)
- DNA copy number variation (aCGH, NGS)
- Immune cell infiltration (FACS, imaging, proteomics)
- Microbiomes (arrays, 16S rRNA-seq)

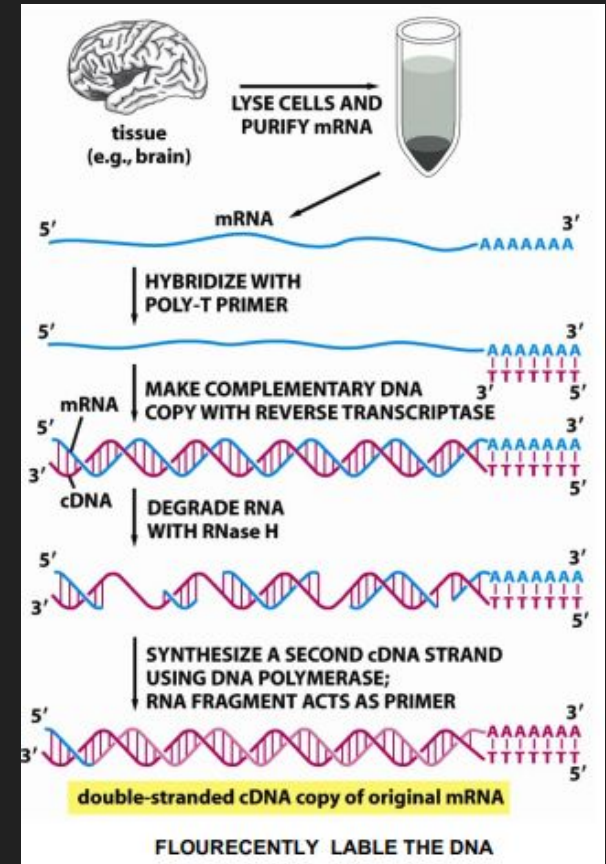
What is Microarray?

Microarrays measure mRNA abundances in a tissue or a cell culture



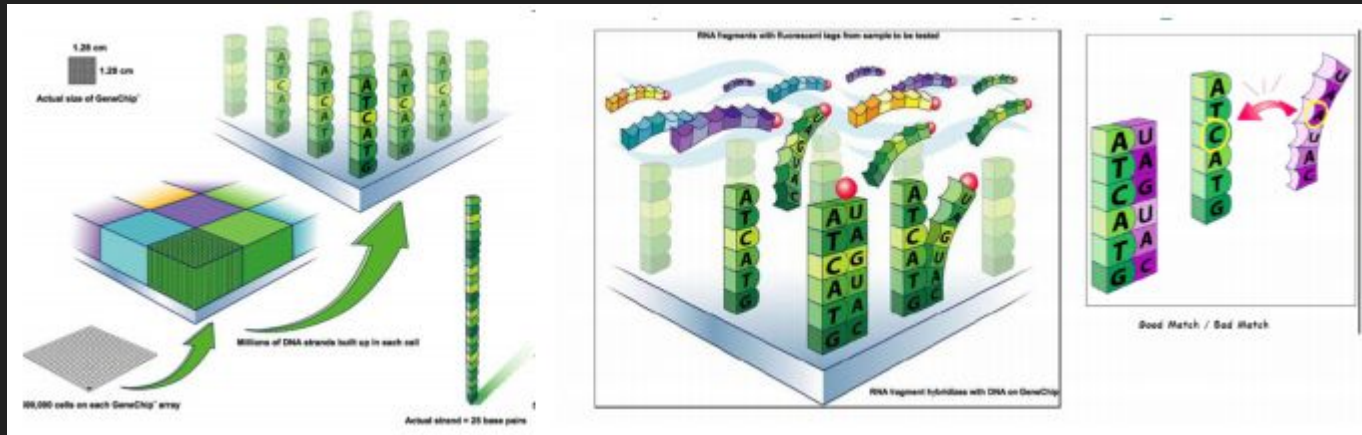
How do we measure it?

Purify mRNA and transform it to cDNA clones



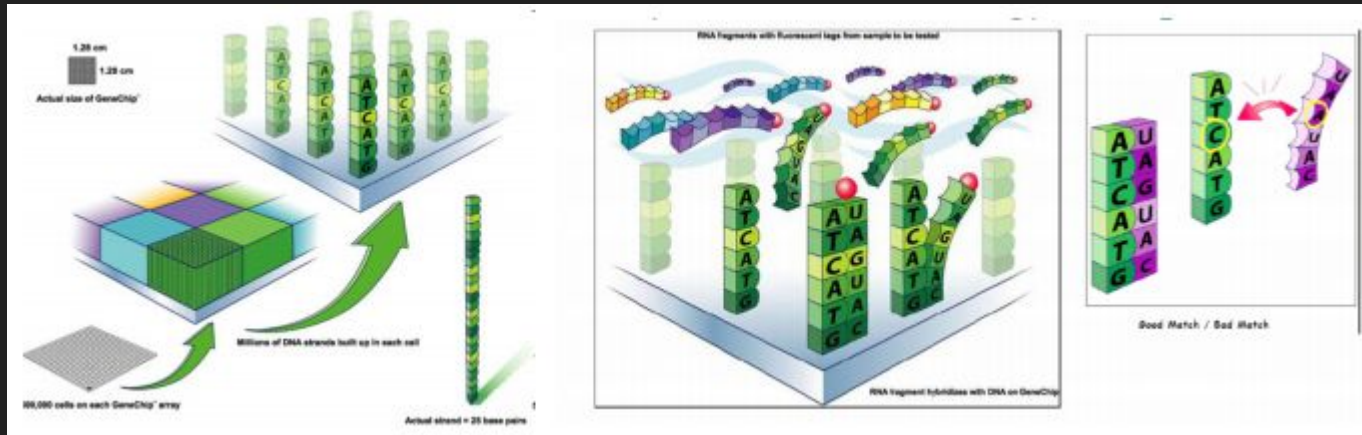
How do we measure it?

Transcriptome wide oligonucleotide libraries glued to a chip catch the cDNA in a hybridization reaction



How do we measure it?

Transcriptome wide oligonucleotide libraries glued to a chip catch the cDNA in a hybridization reaction

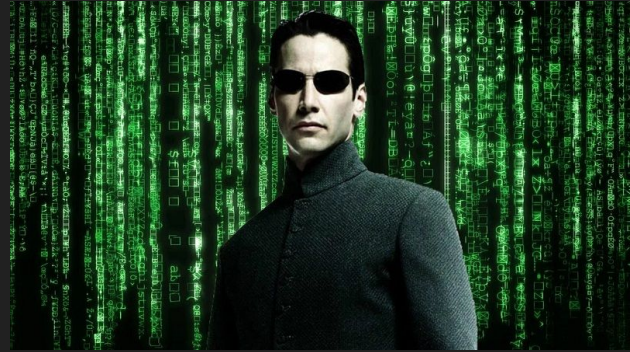
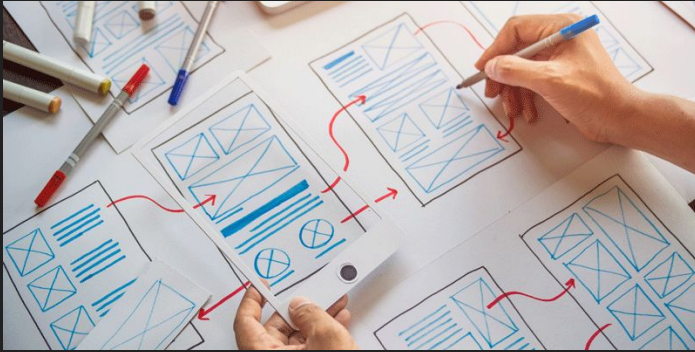


Design Matrix

This code right here is one of the most important part when we're performing differential gene exp

```
175 # this is the factor of interest
176 cellType <- factor(targets$Sample_Group)
177
178 targets
179
180 # use the above to create a design matrix
181 design <- model.matrix(~cellType, data=targets)
182 colnames(design) <- c(levels(cellType))
183
184 # fit the linear model
185 fit <- lmFit(mVals, design)
186
187 # create a contrast matrix for specific comparisons
188 contMatrix <- makeContrasts(GroupA-GroupB,
189                             levels=design)
```

What is Design Matrix?



Design Matrices

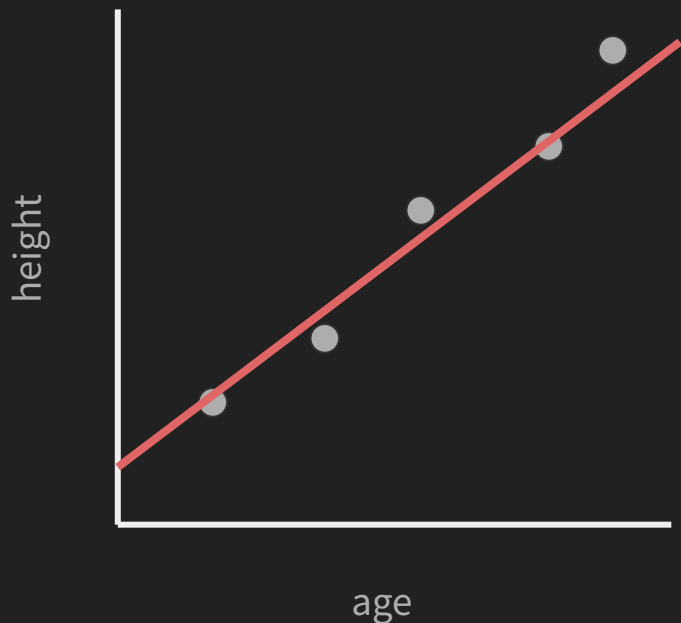
- In statistics and in particular in regression analysis, a **design matrix**, also known as **model matrix** or **regressor matrix** and often denoted by **X**, is a matrix of values of **explanatory variables** of a set of objects (Wikipedia).
- Used to define the form of a statistical model and to store observed values of the **explanatory variable(s)**.
- Used in the computation process to estimate **model parameters**.

Design Matrices

- In statistics and in particular in regression analysis, a **design matrix**, also known as **model matrix** or **regressor matrix** and often denoted by **X**, is a matrix of values of **explanatory variables** of a set of objects.
- Used to define the form of a statistical model and to store observed values of the **explanatory variable(s)**.
- Used in the computation process to estimate **model parameters**.

Simple Linear Model

CASE STUDY: How does human height change with age?



$$Y = mx + b$$

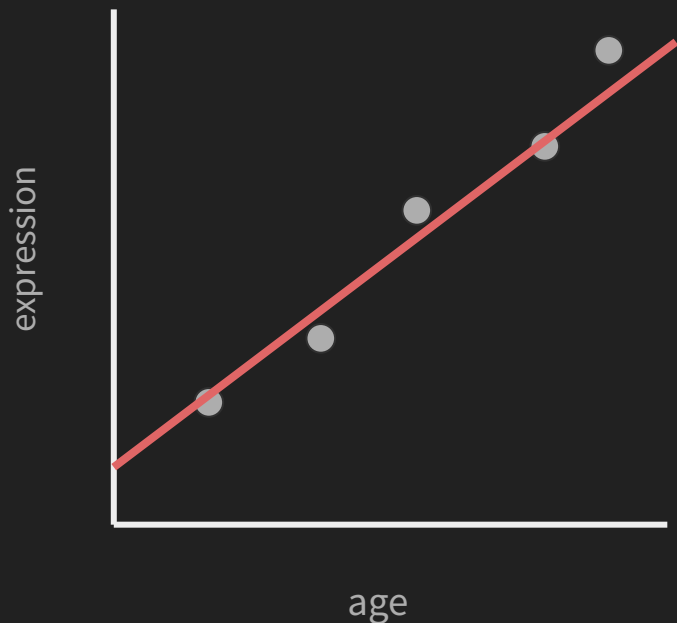
$$\text{height} = \beta_0 + \beta_1 \text{age}$$

Simple Linear Model

- There are two types of explanatory variable that we can use when we are performing differentially gene expression analysis or differentially methylated probe analysis, they are:
 - Covariates : Continuous variable, quantitative (age, weight, measurement from PCR)
 - Factors : Categorical variable associated with samples (disease status, genotype, cell type or treatment)

Simple Linear Model

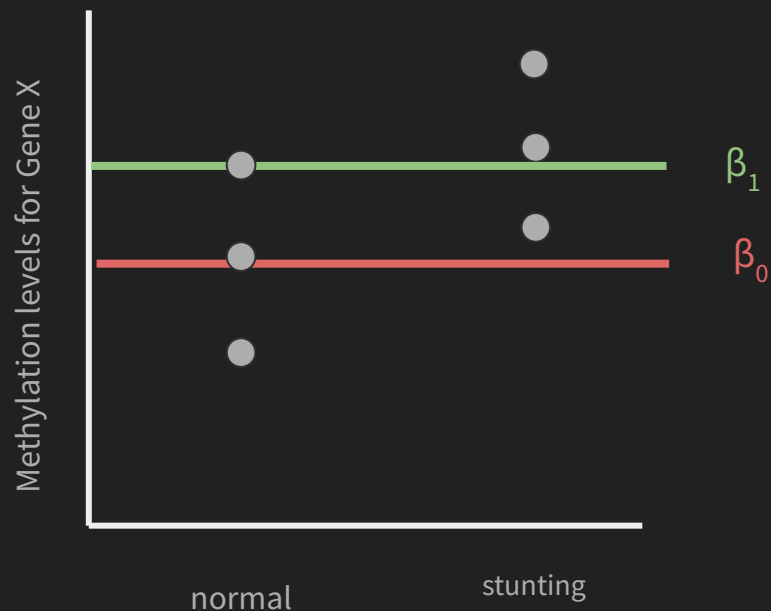
- For covariates, it is pretty straightforward, we can draw a straight line to the model that can describe the relationship



$$\text{Expression} = \beta_0 + \beta_1 \text{age}$$

Simple Linear Model

- For factors, there are two different types of model. The first one is the means model.

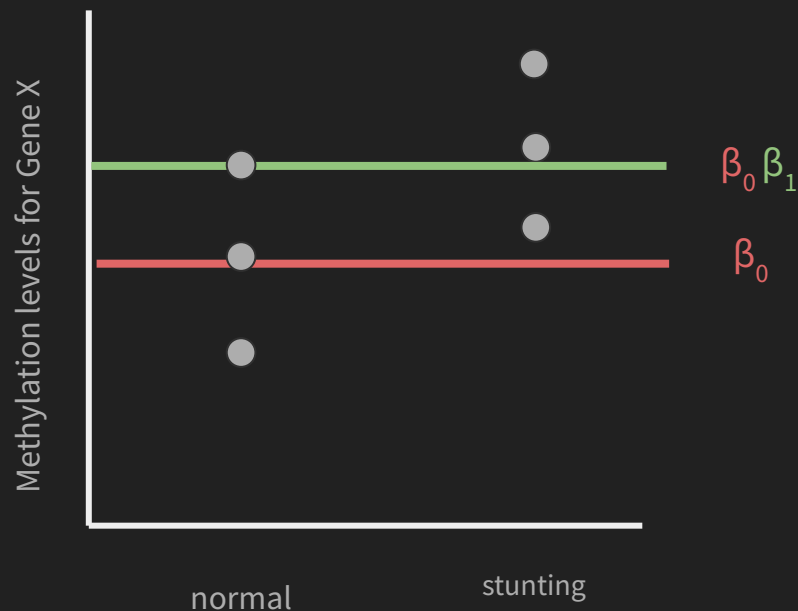


Means Model

Methylation levels = β_0 normal + β_1 stunting

Simple Linear Model

- For factors, there are two different types of model. The first one is the means model.



Means Reference Model

Methylation levels = $\beta_0 + \beta_1 \text{stunting}$