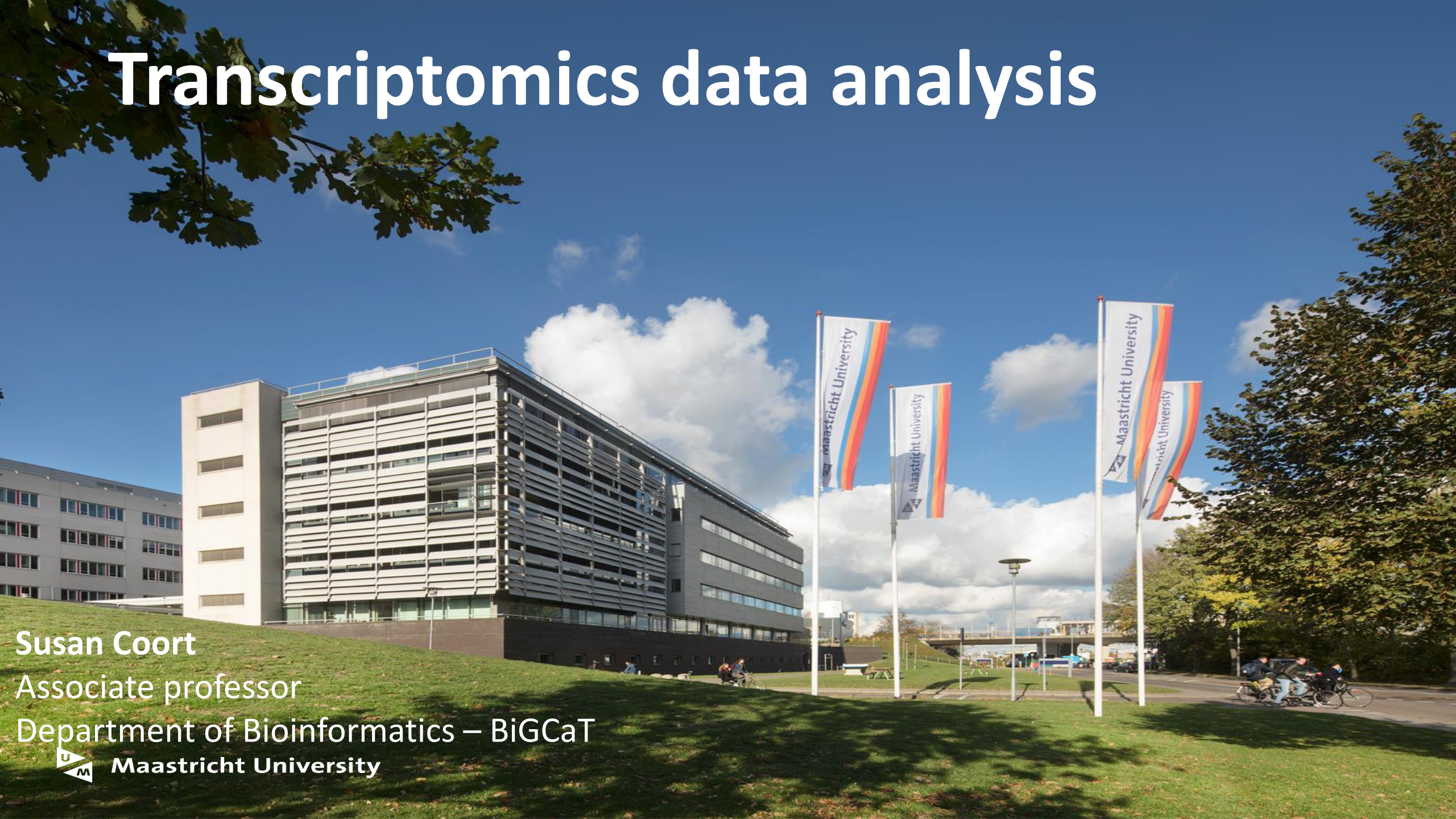
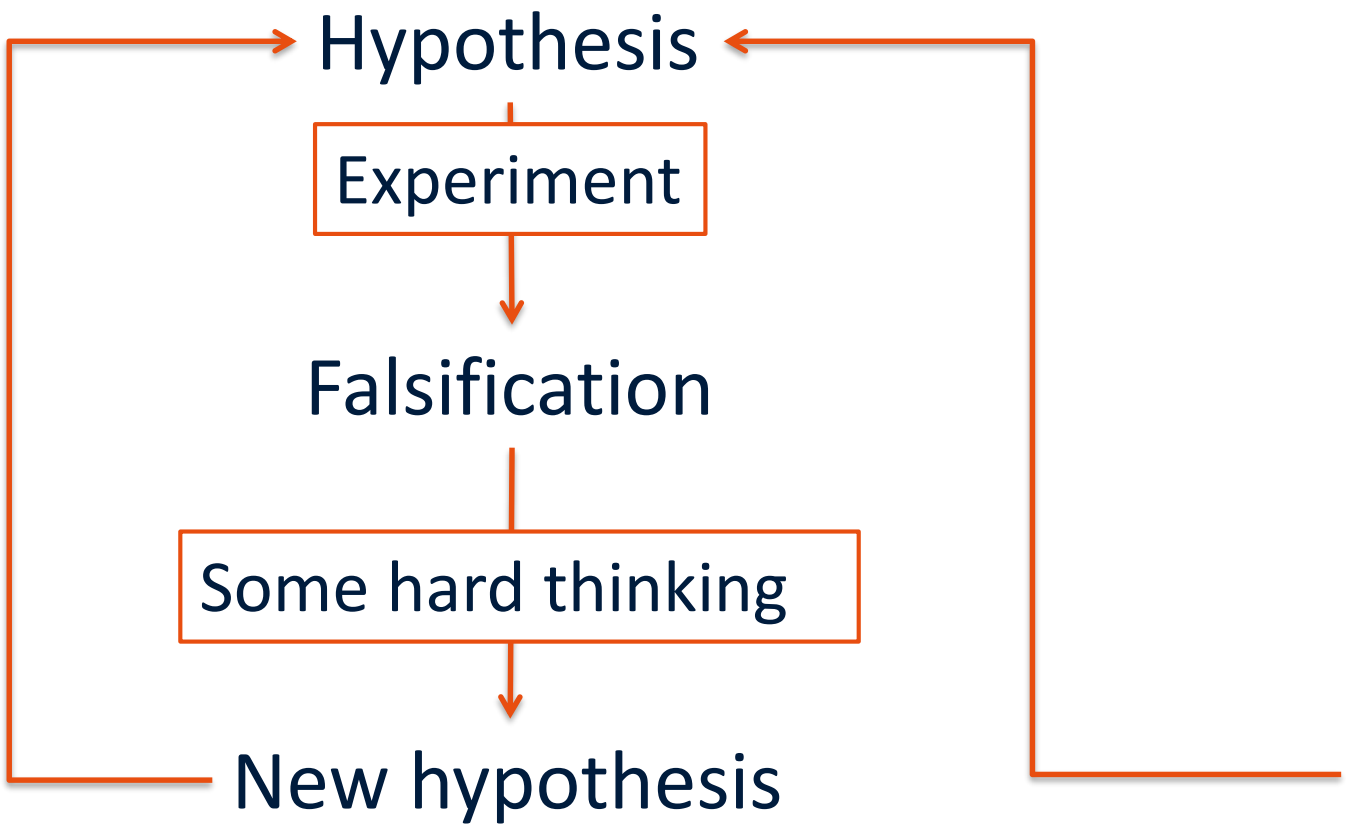
Transcriptomics data analysis



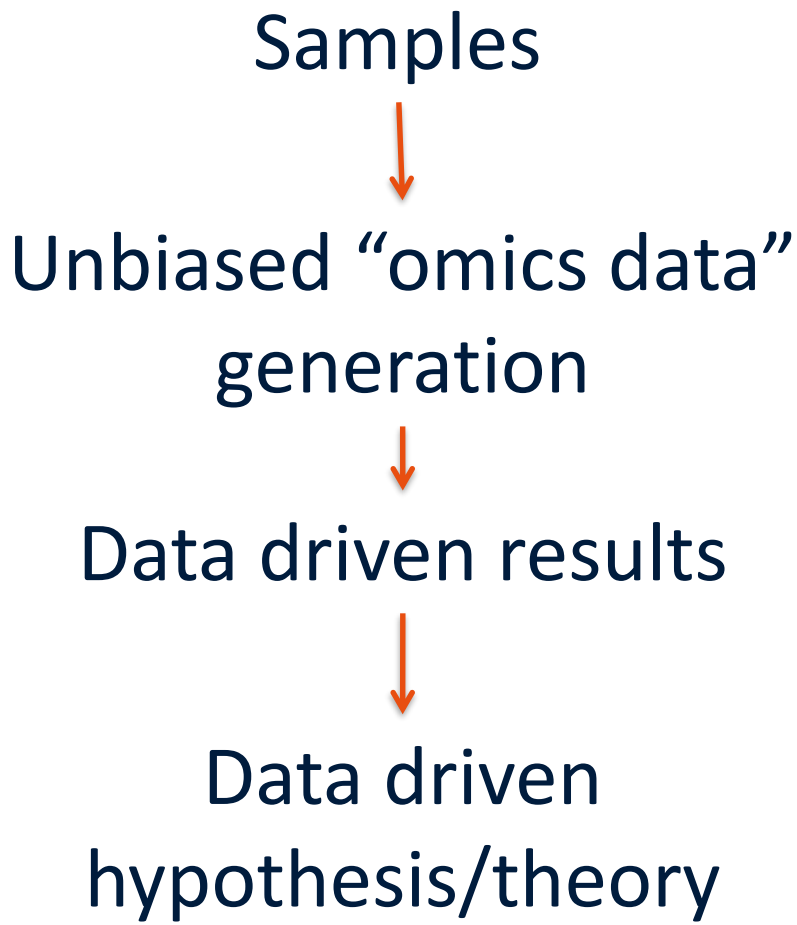
Susan Coort
Associate professor
Department of Bioinformatics – BiGCaT
Maastricht University

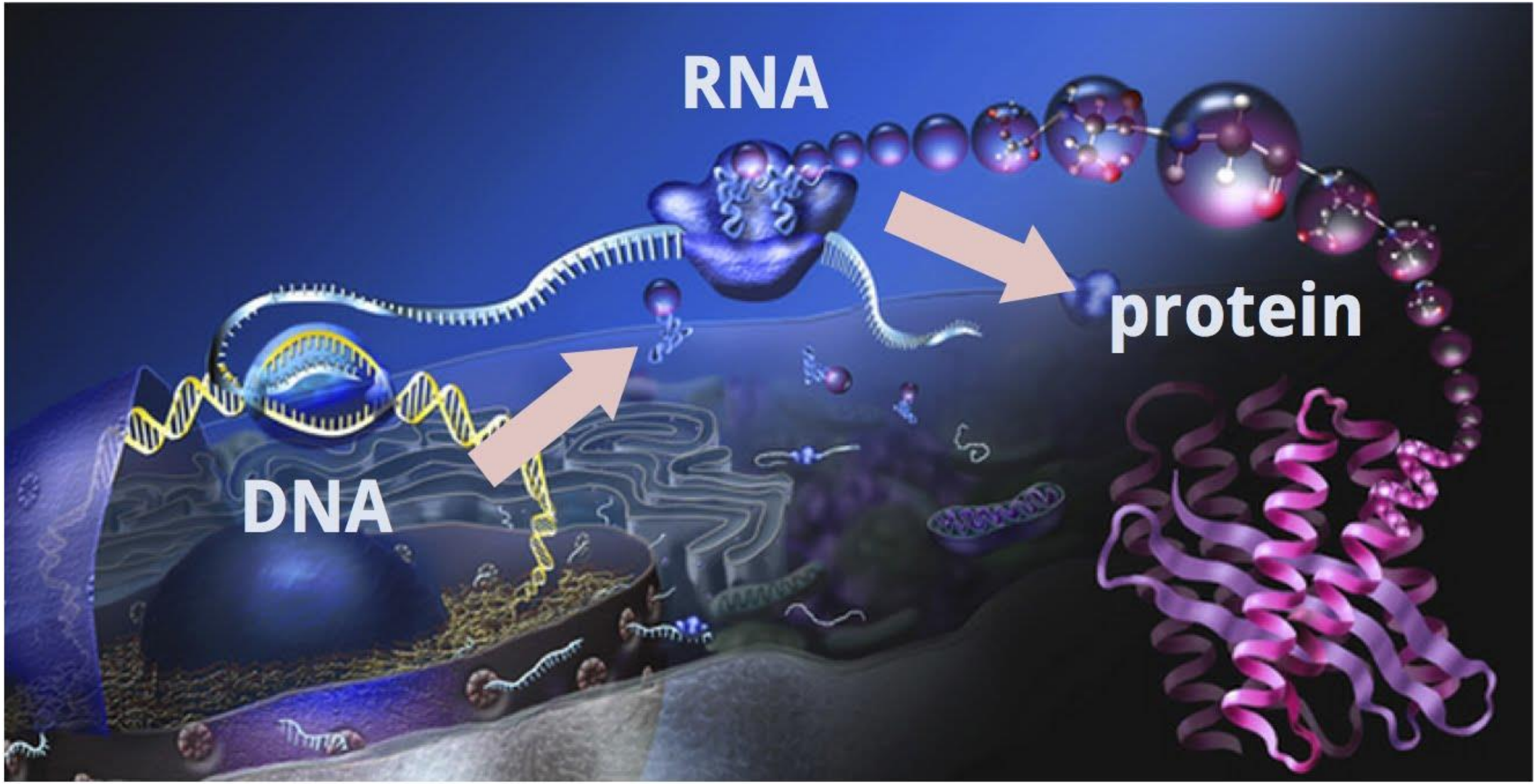
Hypothesis driven research vs. data driven research

- Hypothesis-driven research "reductionistic"



- Data driven research "holistic"



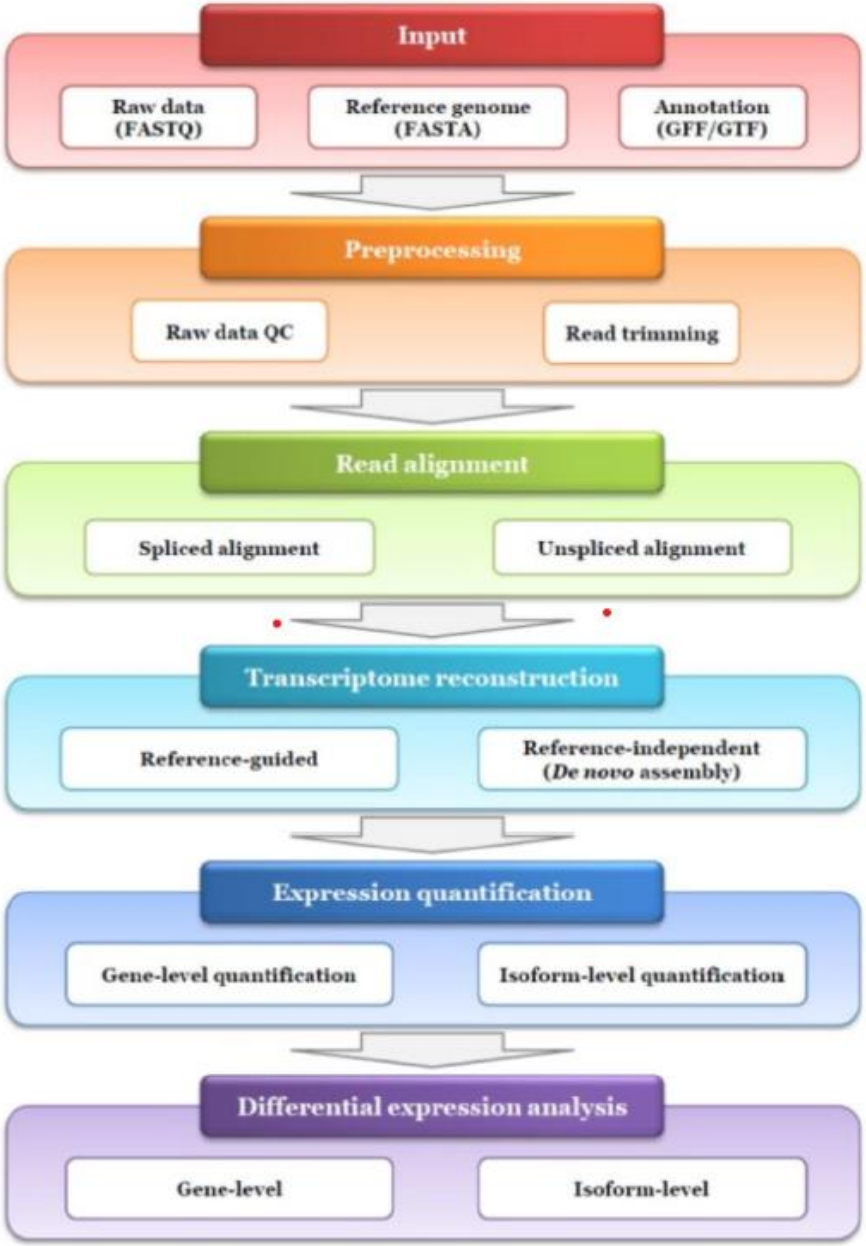

RNA
protein
DNA

RNA sequencing (in short)

- Extract RNA from cell or tissue samples
- Select mRNA by poly-A tail
- Translate to cDNA
- Quantify by PCR
- DNA sequencing

- Result: DNA reads in FASTQ format

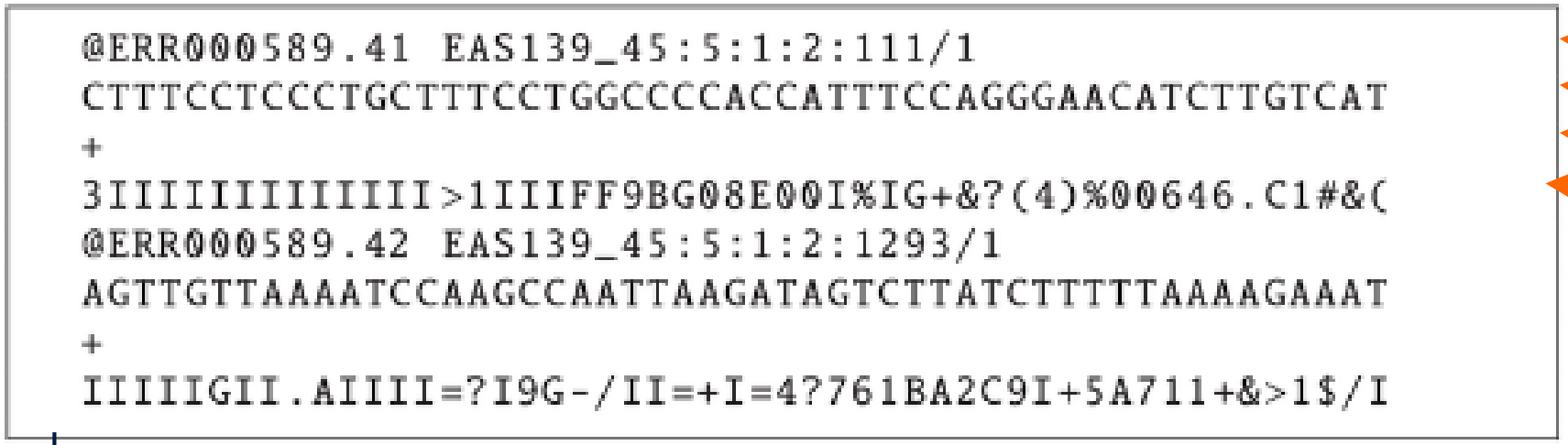
Workflow


Input
Raw data (FASTQ)
Reference genome (FASTA)
Annotation (GFF/GTF)
Preprocessing
Raw data QC
Read trimming
Read alignment
Spliced alignment
Unspliced alignment
Transcriptome reconstruction
Reference-guided
Reference-independent (De novo assembly)
Expression quantification
Gene-level quantification
Isoform-level quantification
Differential expression analysis
Gene-level
Isoform-level

Fastq files

Which read belongs to which gene?

- Contain the generated reads
- And quality scores



Many more lines (gigabytes) → okay? → computer-aided quality control (QC)

Maastricht University

Image source: Abuín Mosquera, José & Pichel, Juan & Pena, Tomas & Amigo, Jorge. (2016). PLOS ONE. 11. e0155461. 10.1371/journal.pone.0155461.

Genome mapping

- Map measured sequence to known sequence of transcripts

- Needs a reference genome
  - Always check which version has been used!
    (positions change over versions)

- Reads belonging to one gene will not map continuously
  - We only have exonic reads
  - Some reads may even map split to the genome!
  - Exon-junction spanning reads

Reads per gene = raw counts


Just with one button, you do all analysis! Right?
Look! Do you see this red button? I especially made it for you. Everything is just done with this button, so easy!
DESeq2
www.biocomicals.com

From raw counts to differentially expressed genes (DEGs)

- DESeq2 R package
- Original publication: Love, M.I., Huber, W., Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**:550. 10.1186/s13059-014-0550-8
- Improved version of DESeq
- Other options are: limma/voom, edgeR, DSS
- What does it do?
  - **Input: Raw counts** and **metadata** (sample description)
  - **Operation:** Uses a statistical model to calculate the difference in gene expression between two or more groups of samples
  - **Output:** Produces fold change (FC) and log2FC for **difference**, and p-value and adj. p-value (FDR) for **significance**

Advantages of DESeq2

- **Accurate correction for variability**: DESeq2 uses a statistical model to correct for variability in the sequencing data, leading to more accurate results.
- **Support for complex designs**: DESeq2 can handle complex experimental designs, such as multiple groups with multiple replicates.
- **Widely used and well-documented**: DESeq2 is a widely used tool in the bioinformatics community, with a large user base and a well-documented user manual.

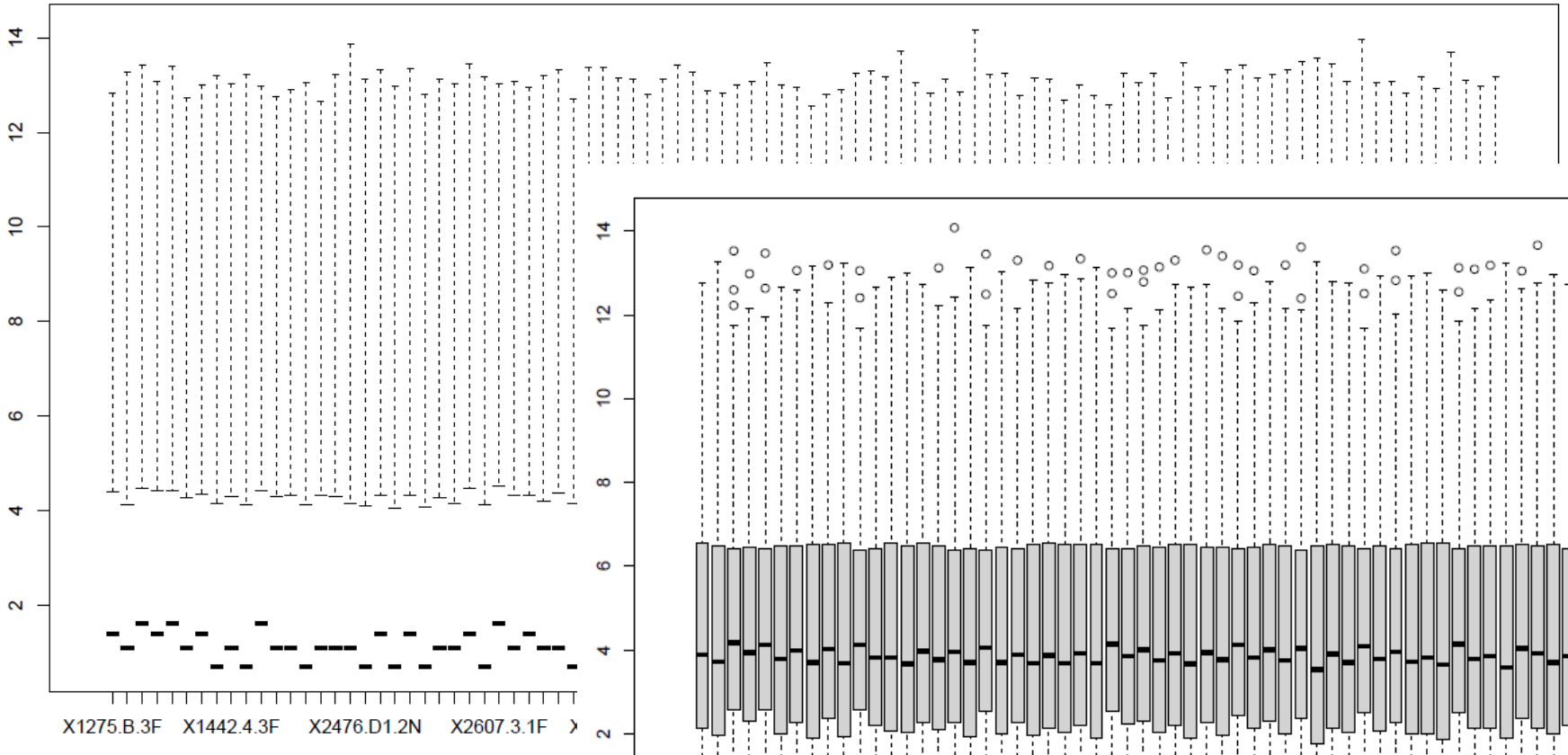
DESeq2 workflow – in short

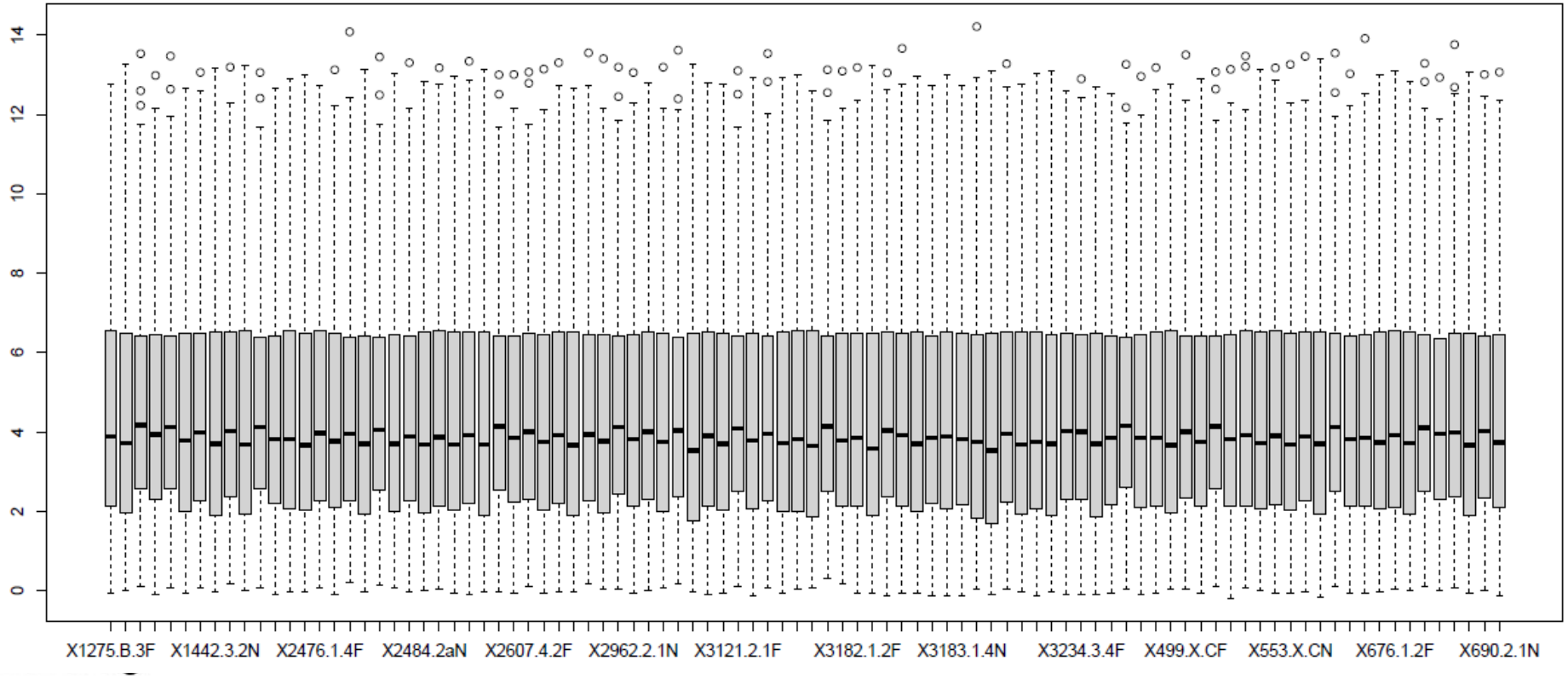
1. Importing raw count data and metadata
2. Creating a “dds object” by merging count and metadata and determining the experimental design (e.g. disease vs. control)
3. Run DESeq on the dds object
4. Export the results – a table with difference and significance of gene expression per gene between the groups
5. (Visualisation and interpretation of the results)

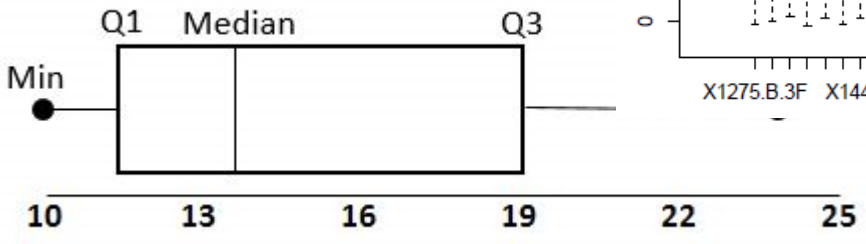
DESeq2 function - in detail

- Starting point: count matrix K with one row for each gene *i* and one column for each sample *j* -> K_{ij}
- **Normalization**: generalized linear model with a negative binomial distribution (sometimes also called a gamma-Poisson distribution) with mean μ_{ij} and dispersion α_i
  - Normalization factors
  - Size factors (can be gene specific: GC content, gene length)
  - Design (experimental design e.g. treated vs. untreated)

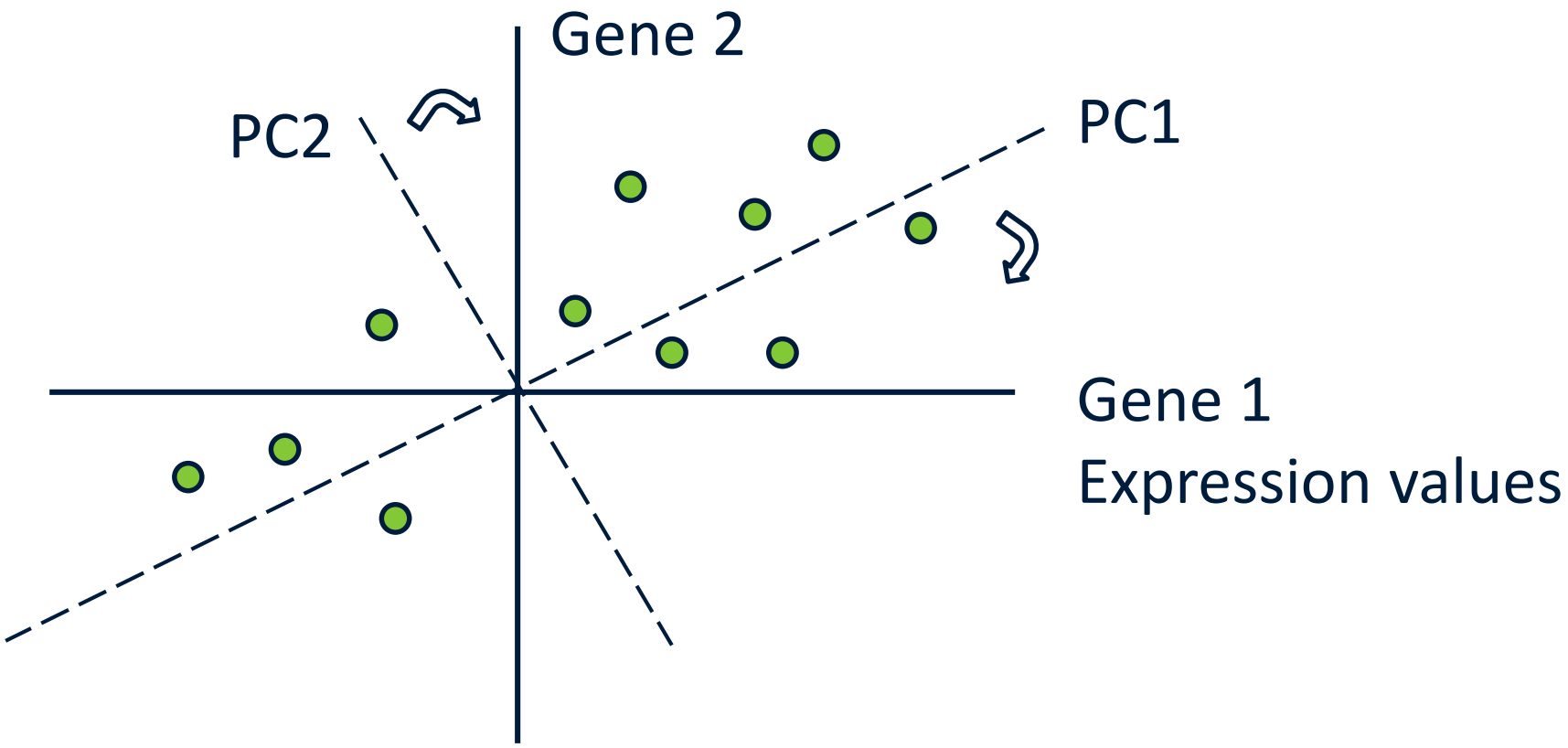
Data boxplot before and after normalization (and filtering)





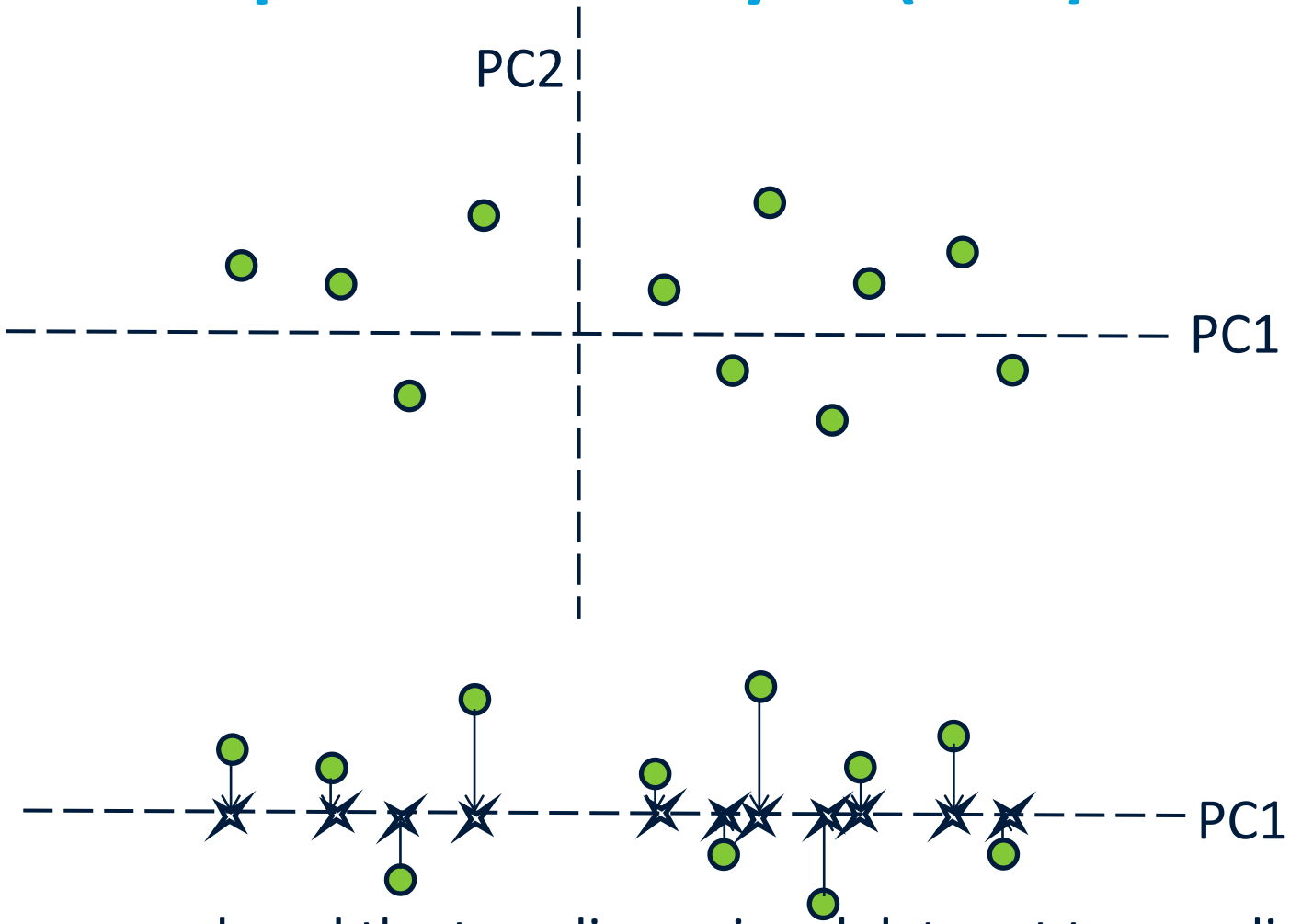


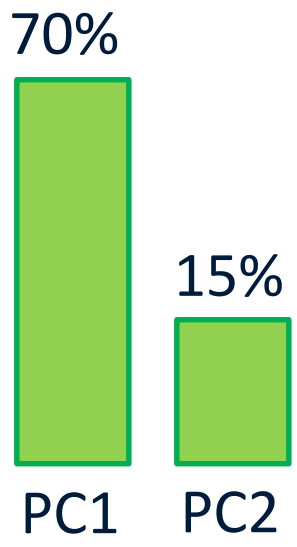
Principle Components Analysis (PCA)



This is a simplified example: in reality we have about 20 000 Genes = 20 000 dimensions!
PCA is used to reduce the dimensions of a multi-dimensional data set to 2 or 3 dimensions

Principle Components Analysis (PCA)

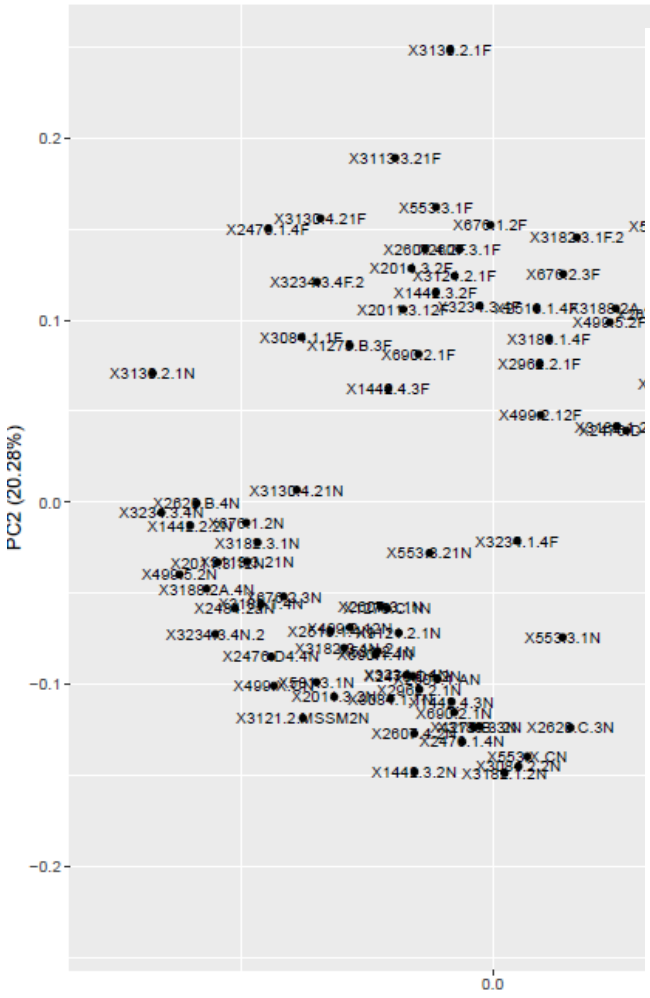


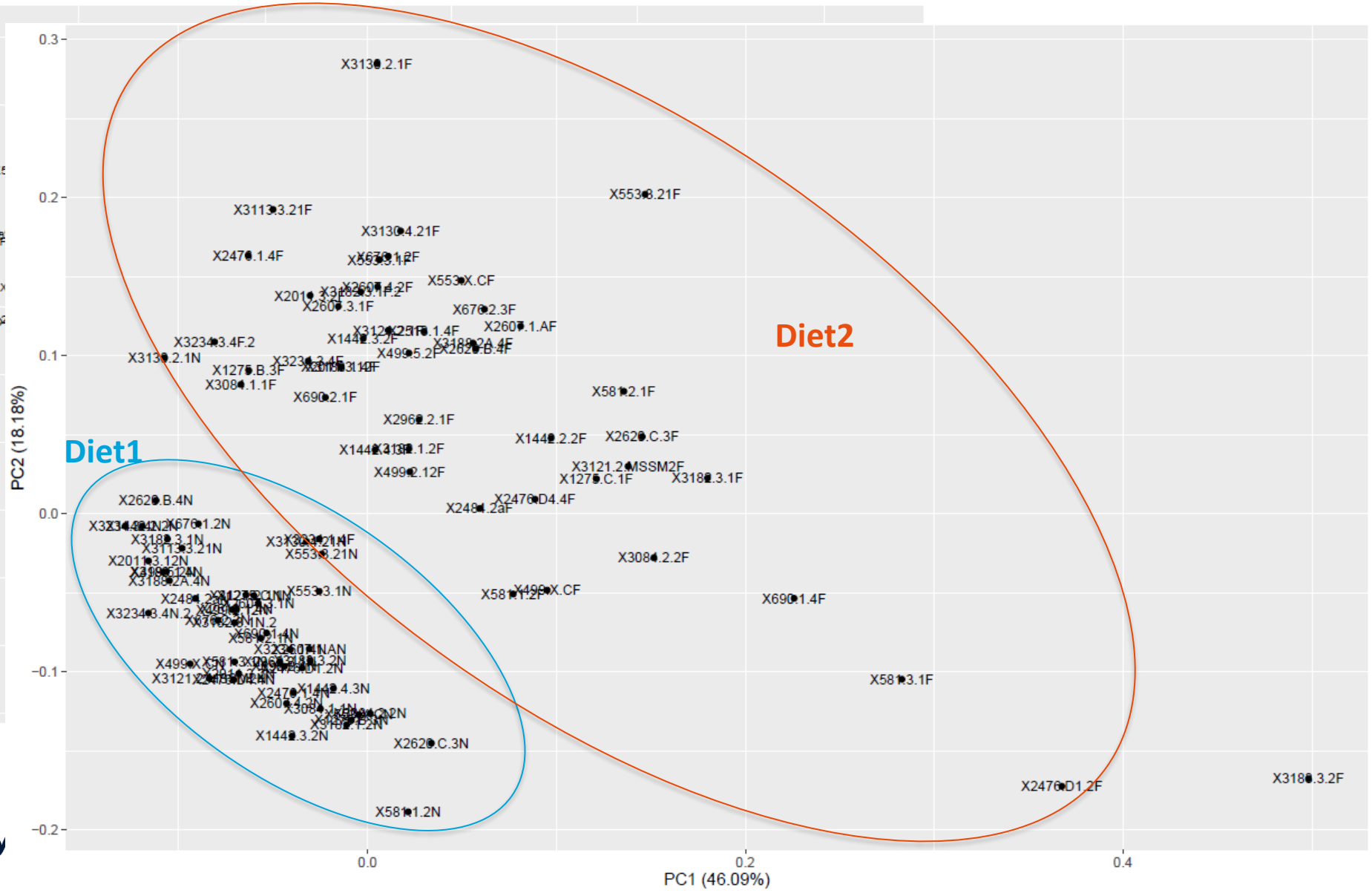


Percentage of variation explained

Now we reduced the two dimensional data set to one dimension, thereby explaining (keeping) 70% of the original variation

PCA (before and after normalization)



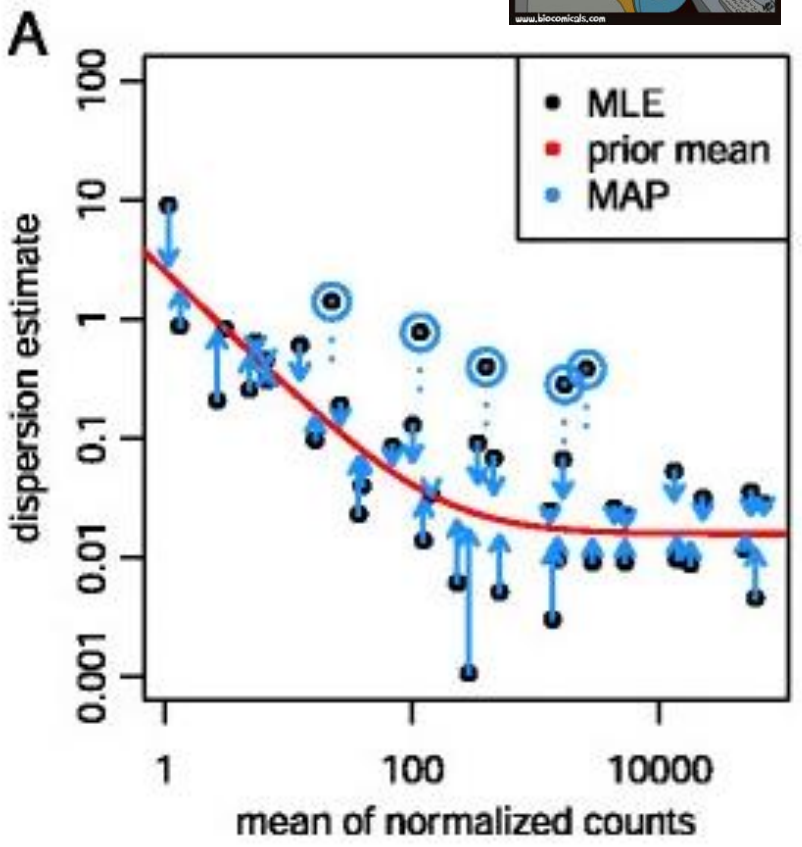


Maastricht University

Variation between genes

- Empirical Bayes shrinkage for dispersion estimation: within group variation
- We assume that genes of similar average expression strength have similar dispersion
- Dispersion parameter is α_i, which describes the variance of counts via

$$\operatorname{Var} K_{ij} = \mu_{ij} + \alpha_i \mu_{ij}^2$$

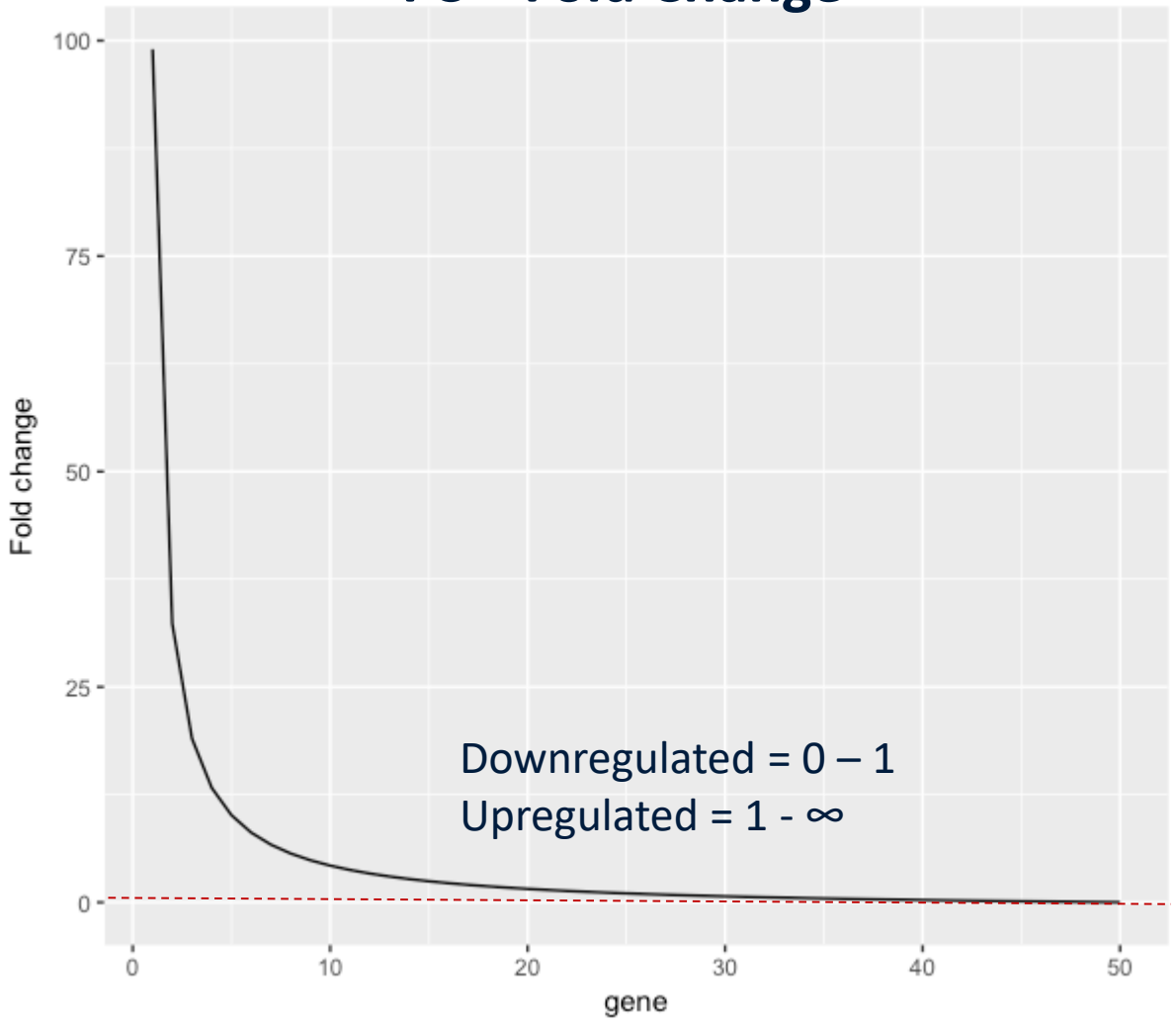


MAP, maximum *a posteriori*; MLE, maximum-likelihood estimate

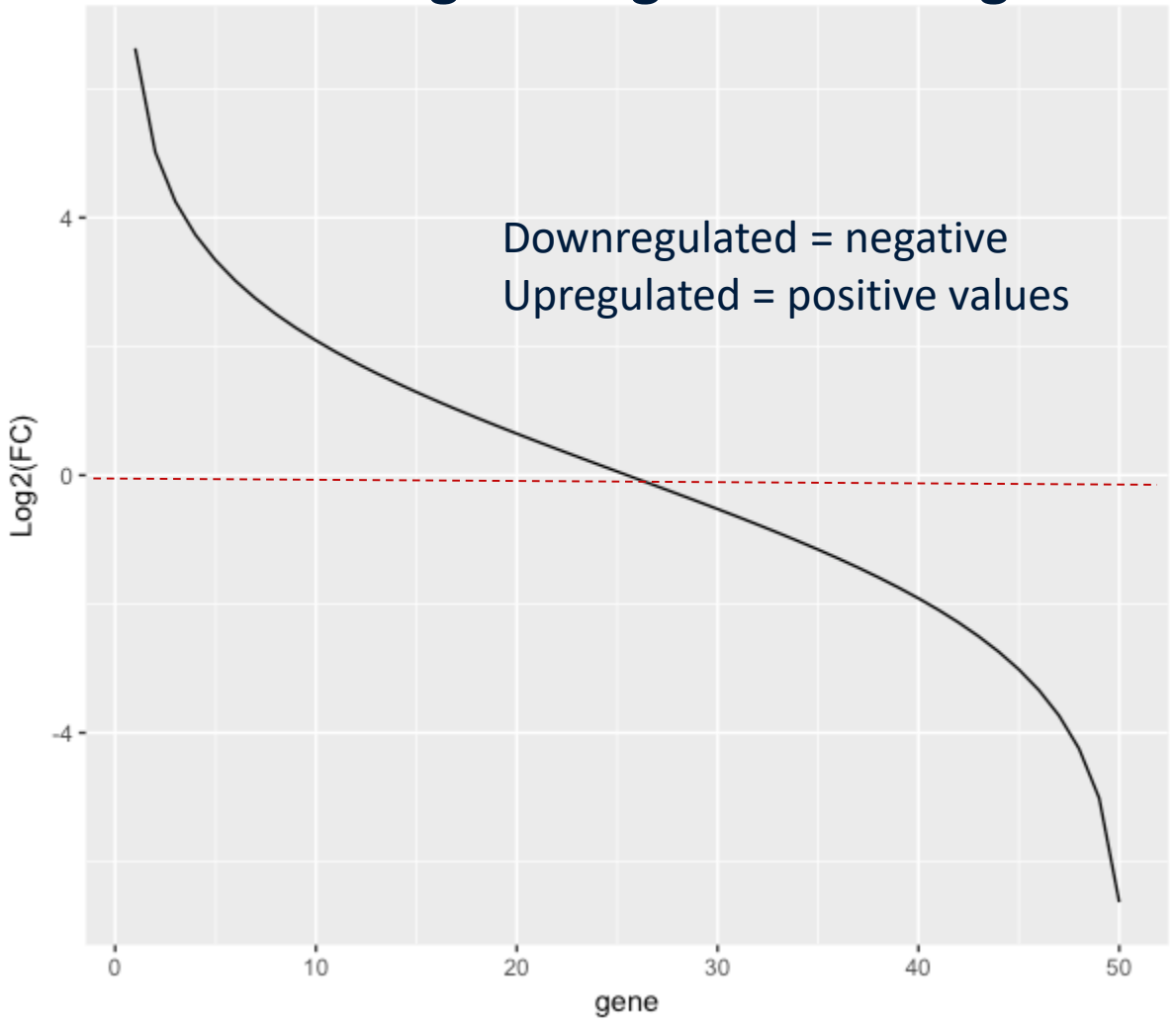
Figure: Love et al. 2014

log2FoldChange – logFC or log2FC

FC – Fold Change



logFC – log 2 Fold Change

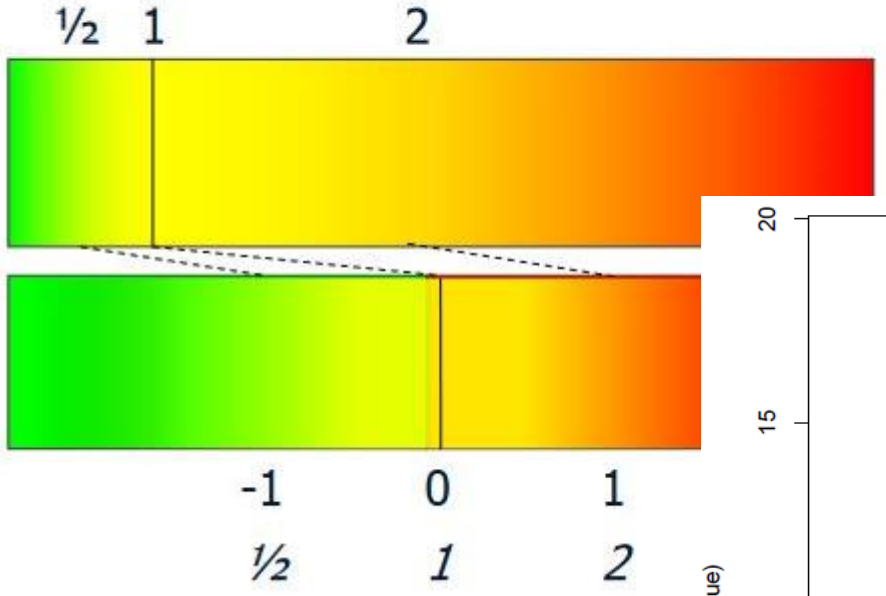


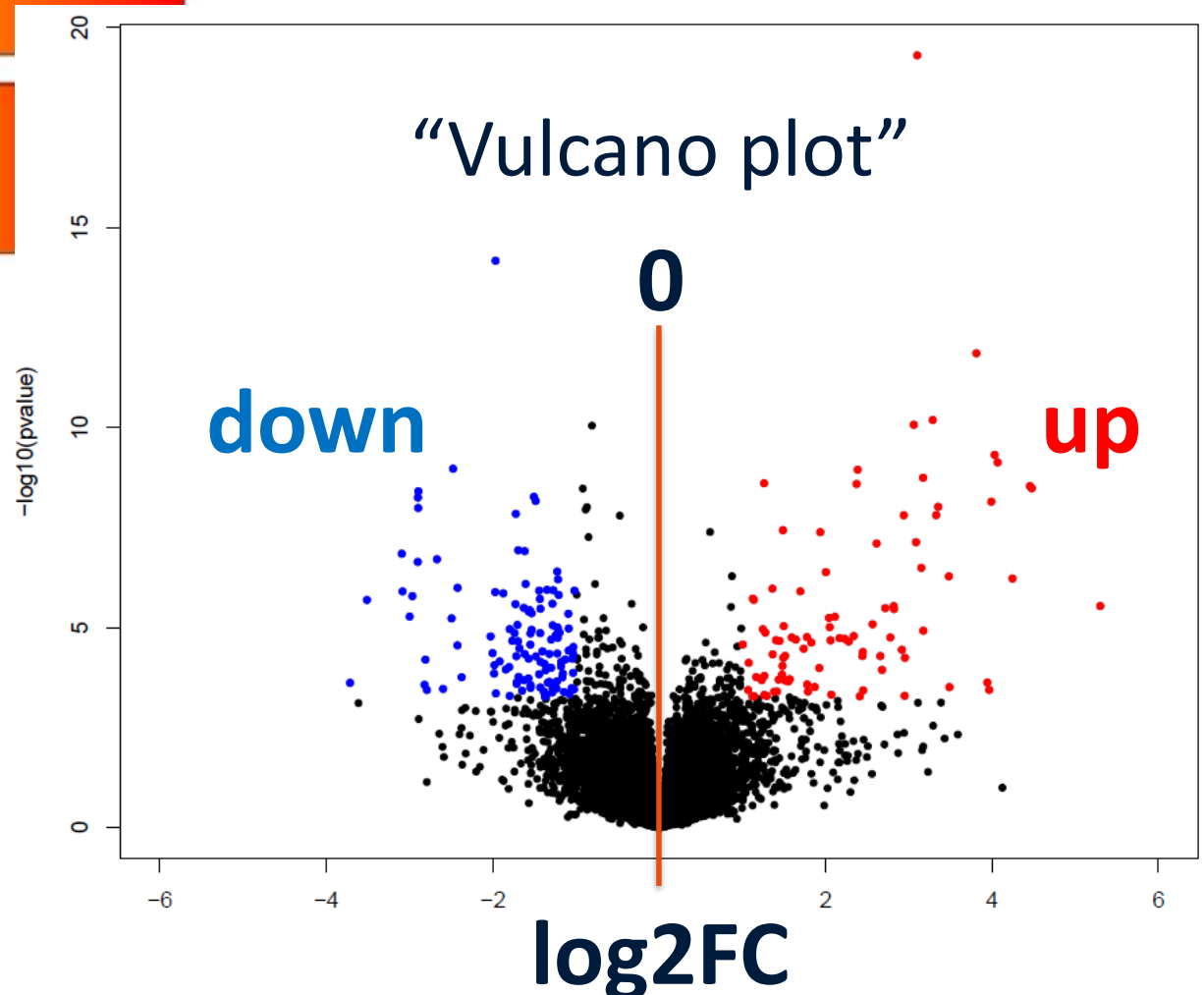
Maastricht University

Source: https://www.biostars.org/p/312980/

- The logFC 'spreads out' the data and offers symmetry

- 'raw' ratio (FC)
- log ratio (logFC)





Hypothesis tests for differential gene expression

P-value (significance) determined by **Wald test:**

The shrunken estimate of log2FC is divided by its standard error, resulting in a *z*-statistic, which is compared to a standard normal distribution.

Adjusted p-value: Correction for multiple testing

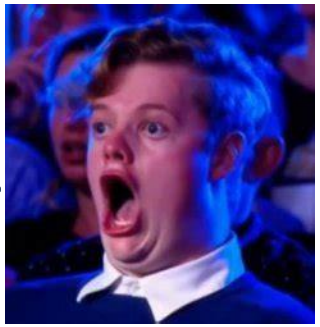
Multiple testing errors

E.g. perform a statistical test with a 0.05 threshold:

- Probability (making a mistake) = 0.05
- Probability (NOT making a mistake) = 0.95

Repeat the test 20 times:

- Probability (not making ANY mistake) = 0.95^{20} = 0.358
- Probability (making at least one mistake) = 1 – 0.358 = 0.642



There is a 64.2% chance of making at least one mistake!

The higher the number of repeats, the higher the chance of making mistakes.

Adjusted p-value – correction for multiple testing

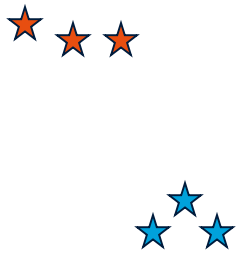
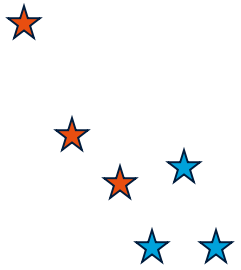
Examples of multiple testing correction methods:

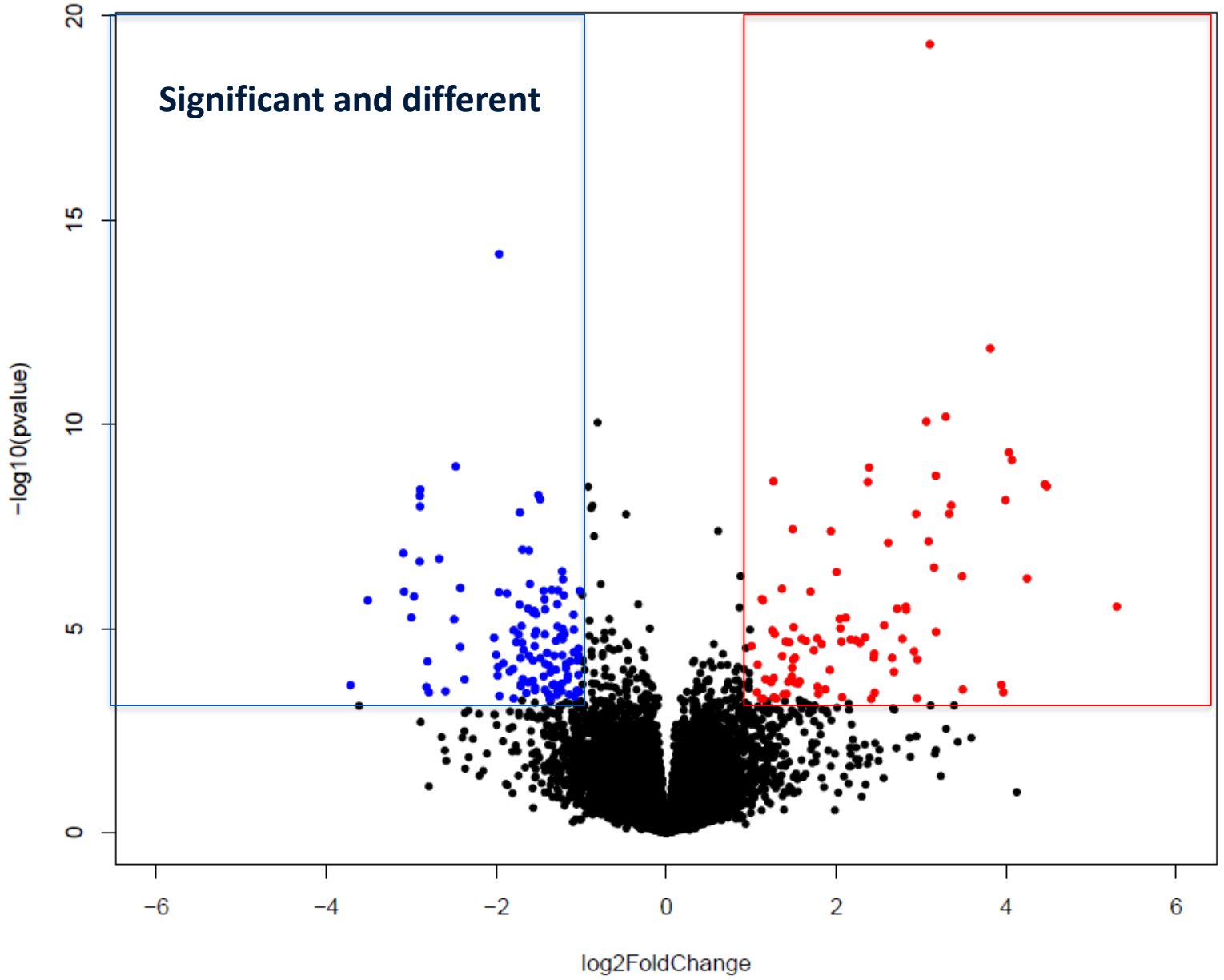
- **Bonferroni** – a very strict correction, very few false positives remain, but we will discount many true positives too.
  - Adjusted p-value = calculated p-value * number of tests done
  - E.g. when we test 100 genes to see if they are different between the two groups. A certain gene gives a p-value of 0.002, the adjusted p-value is: 0.002 * 100 = 0.20 – not significant.

- **Benjamini-Hochberg** – we set the % of results which we can tolerate as false positives (False Discovery Rate or FDR control)

DESeq2 uses FDR/Benjamini-Hochberg: the average expression strength of each gene, across all samples, as its filter criterion

Vulcano plot End diet vs baseline diet

	Different Fold change high	Not different Fold change low
Significant P value < 0.05		
Non-Significant P value > 0.05		



More information

- A simple, straight-forward DESeq2 tutorial: https://lashlock.github.io/compbio/R_presentation.html
- Extensive tutorial covering all eventualities, by the original authors: https://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html
- DESeq2 original paper: Love, M.I., Huber, W., Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**:550. 10.1186/s13059-014-0550-8
- R Tutorial on clustering: https://www.statmethods.net/advstats/cluster.html

How to make sense of omics data?

List of differentially expressed RNA

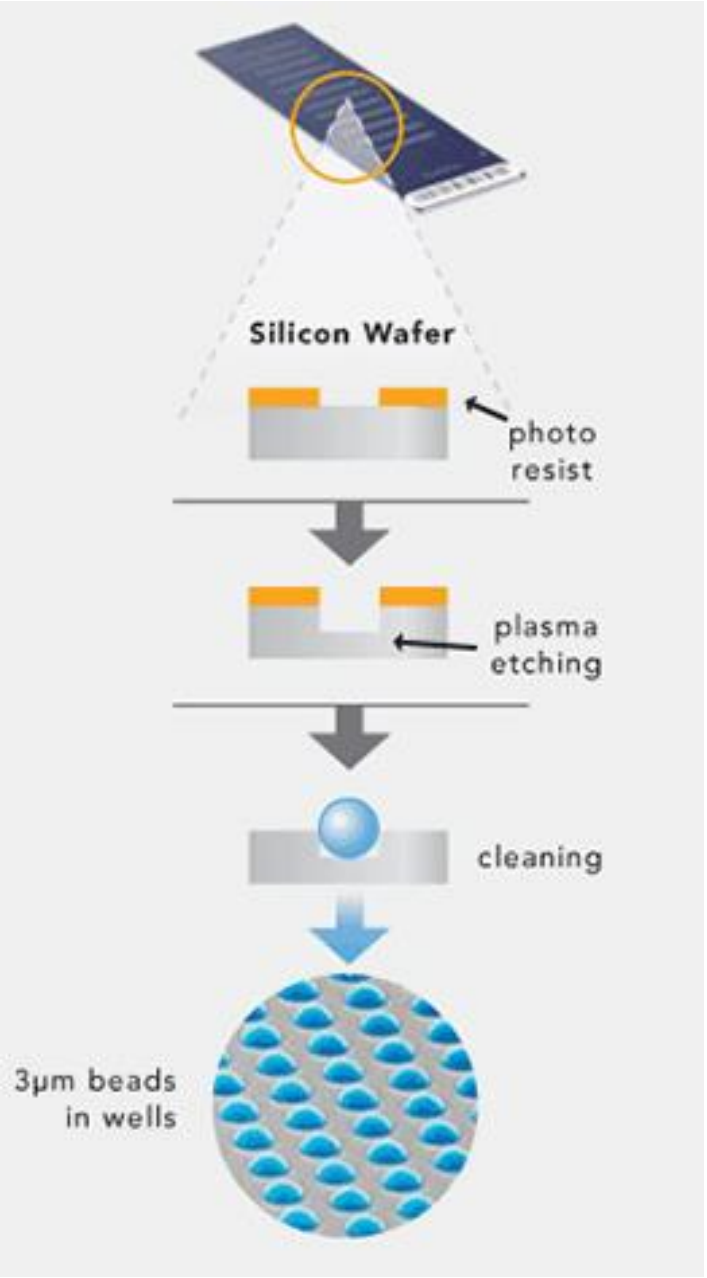
- Clustering
- Correlation
- Classification
- Pathway analysis (Overrepresentation Analysis, ORA)
- Gene Set Enrichment Analysis (GSEA)
- Gene Ontology (GO) analysis
- Network analysis

Biological interpretation  Hypothesis

Microarray technology Illumina beadchip

Illumina bead chip

Illumina microarray technology (also known as BeadArray technology) uses silica microbeads. On the surface of each array, or BeadChip, hundreds of thousands to millions of genes for a single individual can be assayed at once. These tiny silica beads are housed in carefully etched microwells and coated with multiple copies of an oligonucleotide probe targeting a specific locus in the genome.



Expression Beadchip content

Table 1: Expression BeadChip Content

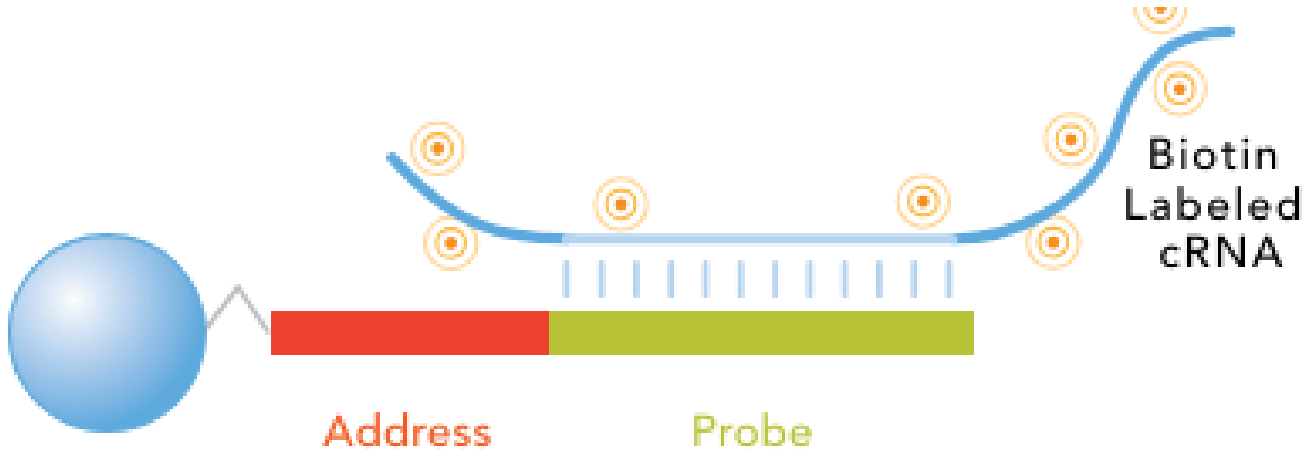
Probes	Description	Human HT-12 v4.0* 12-sample	Mouse WG-6 v2.0 6-sample	Mouse Ref-8 v2.0 8-sample	Rat Ref-12 12-sample	Human WG DASL HT* 12-sample
RefSeq Content						
NM	Coding transcript, well-established annotation	28,688	26,766	24,854	6,277	27,253
XM	Coding transcript, provisional annotation	11,121	6,856	796	15,983	426
NR	Non-coding transcript, well-established annotation	1,752	56	47	1	1,580
XR	Non-coding transcript, provisional annotation	2,209			12	26
Source	RefSeq source release	Human RefSeq Rel 38	Mouse RefSeq Rel 22		Rat RefSeq Rel 16	Human RefSeq Rel 38
Supplementary Content						
UniGene	Experimentally confirmed mRNA sequences that align to EST clusters	3,461			250	
RIKEN FANTOM2	Exemplar protein-coding sequences from the RIKEN FANTOM2 database		5,659			
RefSeq Release 5	Transcripts with NM and XM annotation in RefSeq Release 5 (Build 33.1)		3,573			
MEEBO	Probes to transcripts that do not align with 100% accuracy to RefSeq, but are confirmed as valid mRNA mapping to clusters in Expressed Sequence Tag databases[6]		2,371			
Total		47,231	45,281	25,697	22,523	29,285

* > 99.98% of the bead types are present on any HumanHT-12 array

Maa

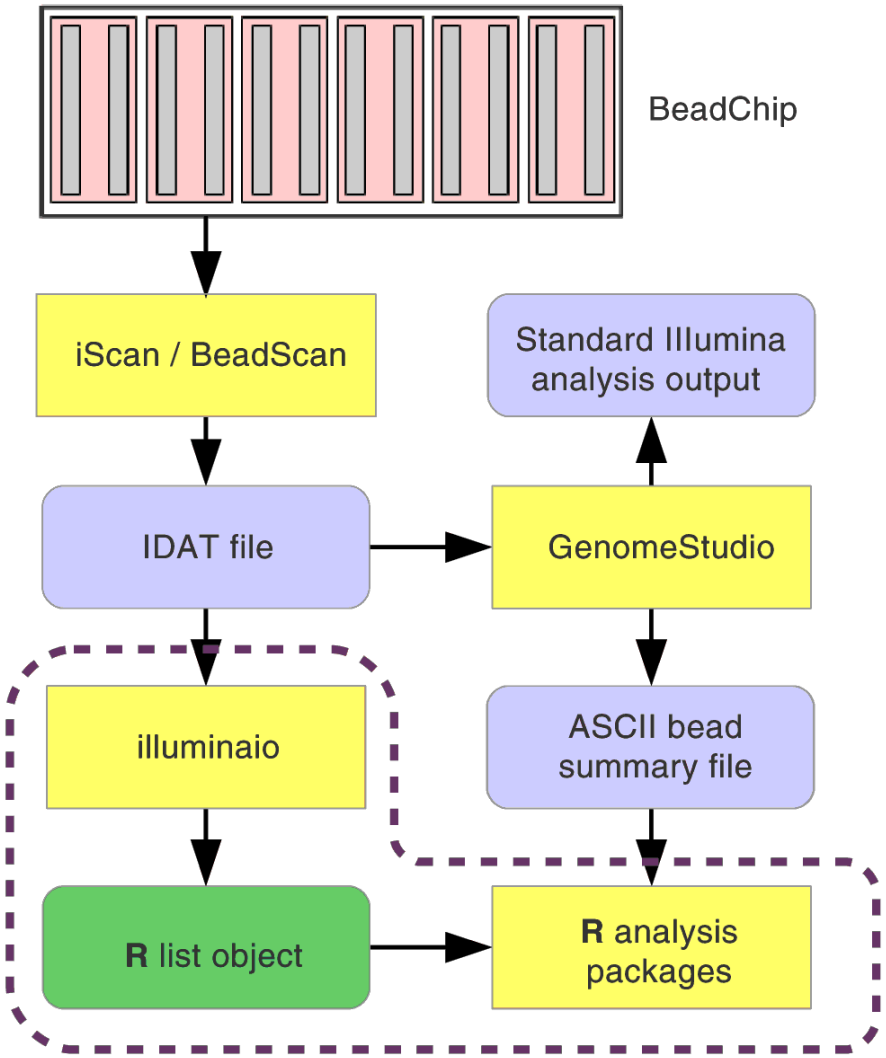
Oligonucleotides immobilized to beads

Figure 2: Direct Hyb Gene Expression Profiling Bead Design



Illumina array-based technologies support a broad array of RNA expression profiling products for a variety of applications.

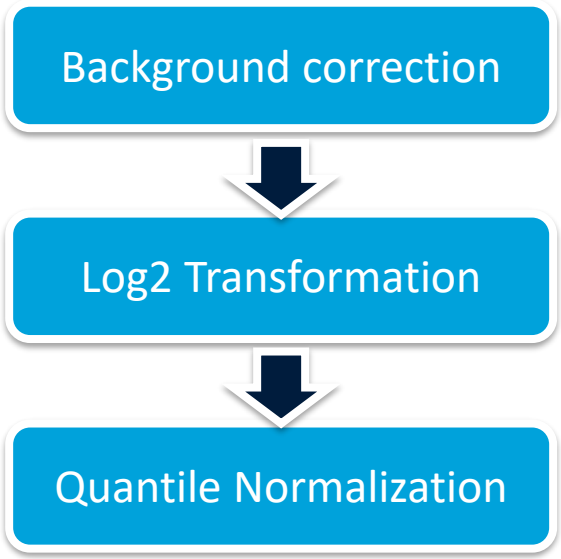
Beadchip Analysis Workflow



: An open source IDAT parsing tool for Illumina microarrays [version 1]. F1000Research 2013, 2:264 (doi: 10.12688/f1000research.2-264.v1)

NEQC function in R

NEQC (normexp + quantile normalization) is a method used to preprocess and normalize data from Illumina BeadChip arrays, which are commonly used for gene expression profiling.
The NEQC method aims to improve the quality of the data and make it more comparable across different samples.



Now…

- The data!

The dataset : original publication

The paper:

ORIGINAL ARTICLE | Open Access

High-protein diet more effectively reduces hepatic fat than low-protein diet despite lower autophagy and FGF21 levels

Chenchen Xu, Mariya Markova, Nicole Seebeck, Anne Loft, Silke Hornemann, Thomas Gantert, Stefan Kabisch, Kathleen Herz, Jennifer Loske, Mario Ost, Verena Coleman ... See all authors

First published: 11 July 2020 | https://doi.org/10.1111/liv.14596 | Citations: 40

Maastricht University find full text

Chenchen Xu, Mariya Markova, Olga Pivovarova-Ramich and Andreas F.H. Pfeiffer contributed equally to this manuscript (CX and MM share the first authorship; OPR and AFHP share the last authorship).
Handling Editor: Stefano Romeo

Experimental design

	LP group $n_{LP} = 10$			HP group $n_{HP} = 9$	
	Week 0	Week 3	P_{LP}	Week 0	Week 3
Anthropometry					
Weight [kg]	133.19 ± 7.48	127.88 ± 6.98	**4.43×10^{-4}**	139.61 ± 8.13	134.92 ± 8.36
BMI [kg/m^2]	45.18 ± 1.24	43.45 ± 1.30	**2.36×10^{-4}**	44.48 ± 1.29	42.94 ± 1.38
Fat mass [%]	53.02 ± 2.60	51.59 ± 2.09	.858	54.84 ± 2.68	55.60 ± 2.80
Fat mass [kg]	69.8 ± 5.9	64.9 ± 4.9	.052	77.4 ± 9.6	77.8 ± 8.1
Fat-free mass [kg]	61.0 ± 4.5	60.9 ± 4.2	.978	62.2 ± 5.2	60.6 ± 4.9
Serum parameters					
AST [U/l]	32.32 ± 6.58	34.68 ± 4.99	.240	24.36 ± 3.44	36.58 ± 11.82
ALT [U/l]	44.31 ± 10.76	43.83 ± 5.85	.944	36.68 ± 8.28	50.28 ± 9.82
GGT [U/l]	66.18 ± 24.08	41.82 ± 12.71	**.001**	34.59 ± 10.91	56.13 ± 31.35

19 participants with morbid obesity

Two hypocaloric (1500-1600 kcal/day) diet groups:

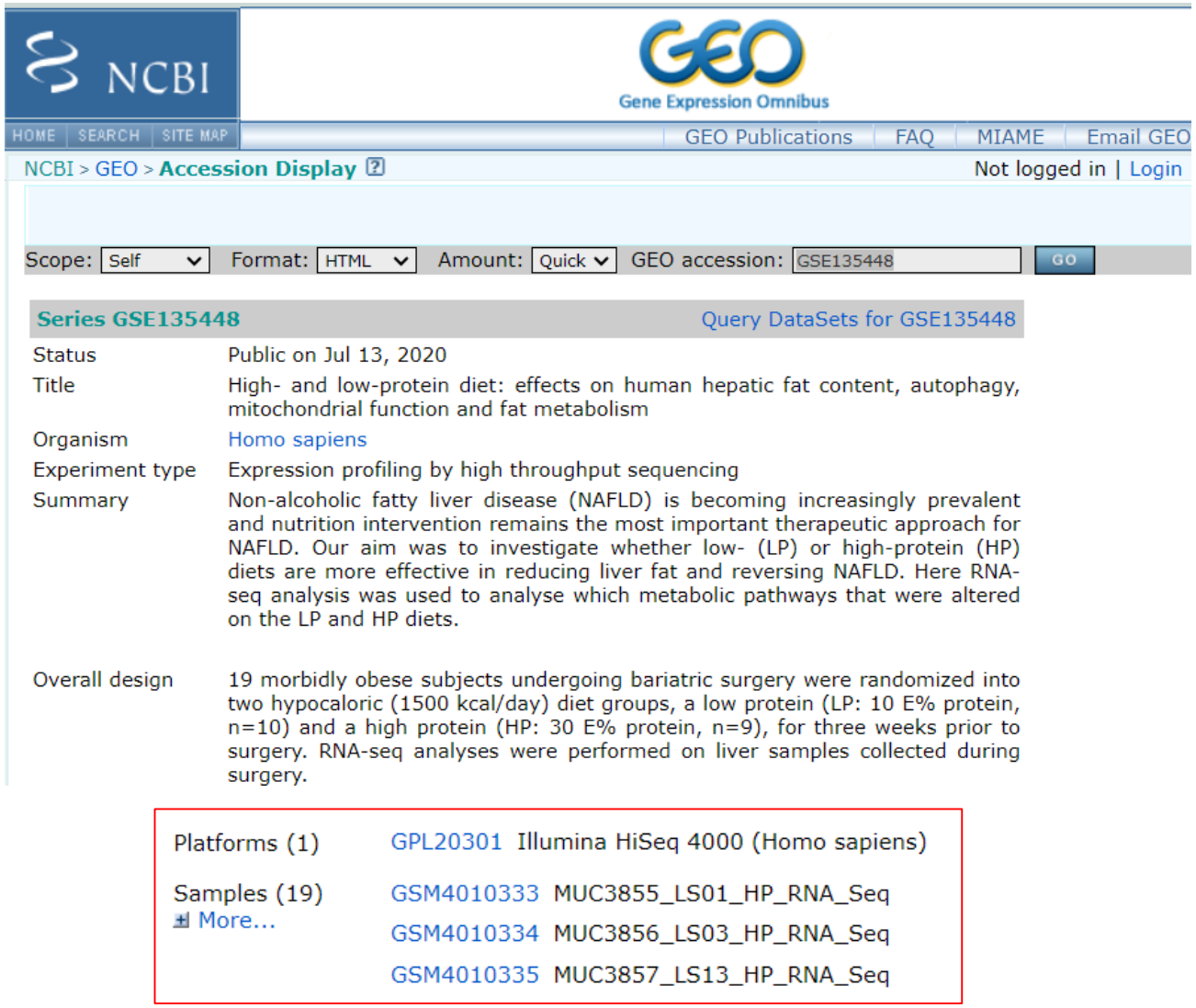
1. High protein (n=9)
   30% protein, 25-30% fat, 35-45% carbohydrates
2. Low protein (n=10)
   10% protein, 25-35% fat, 55-65% carbohydrates

3 weeks of dietary intervention

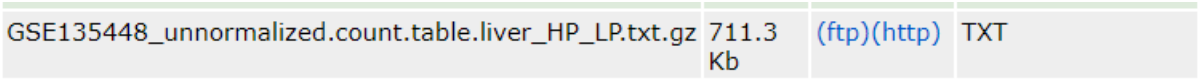
Liver samples -> RNA -> RNA sequencing

The data

GEO database https://www.ncbi.nlm.nih.gov/geo/, accession number **GSE135448:**



raw counts dataset: "GSE135448_geneCounts.csv"



Maastricht University

GSE106589_geneCounts

Gene Symbol

Sample numbers

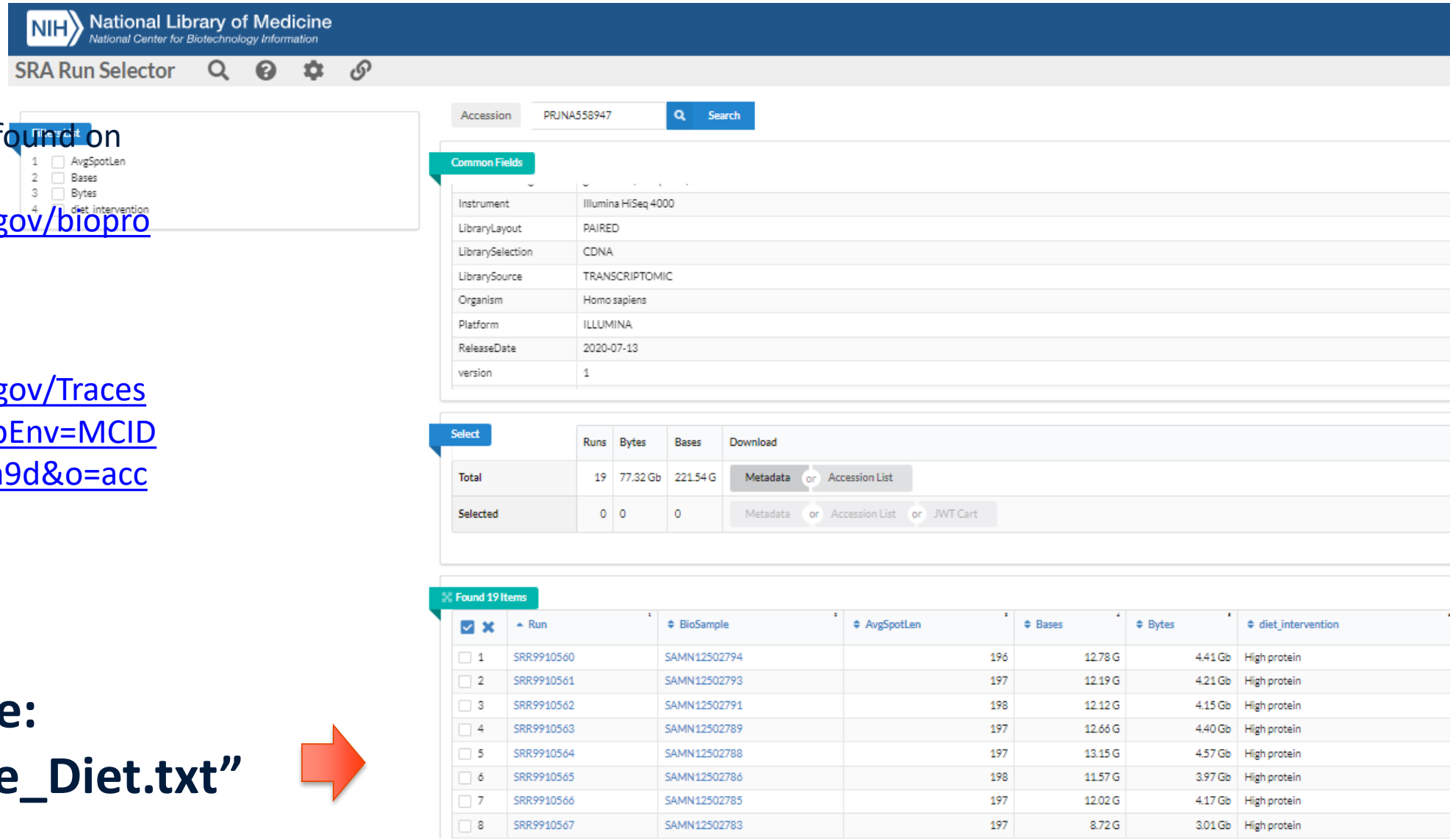
genesymbol	GSM4010333	GSM4010334	GSM4010335	GSM4010336	GSM4010337	GSM4010338	GSM4010339	GSM4010340	GSM4010341	GSM4010342
A1BG	33776	26592	31992	42170	28840	27873	29420	15935	26489	33066
A1CF	14147	17752	17763	21281	18993	24453	15911	9013	18474	8628
A2M	151266	162715	62767	152061	73040	162003	56125	50082	41422	105496
A2ML1	0	0	0	0	0	0	1	0	0	0
A2MP1	4	5	10	4	4	1	1	3	7	0
A3GALT2	0	0	2	2	1	5	1	0	2	0
A4GALT	23	34	25	46	42	48	31	26	15	51
A4GNT	3	5	2	4	3	11	8	2	5	10
AA06	0	0	0	0	0	0	0	0	0	0
AAAS	1173	1259	1157	1149	1264	870	1166	703	1054	1246
AACS	209	156	151	162	212	176	204	111	186	145
AACSP1	1	0	0	0	0	1	0	0	0	0
AADAC	17924	13592	12674	15591	11436	16690	14494	9991	14171	10067
AADACL2	0	0	0	0	0	0	0	0	0	0
AADACL3	0	0	0	0	0	0	0	0	0	0
AADACL4	1	0	0	0	0	0	0	0	2	0
AADAT	2233	2543	2953	2648	2554	3345	2592	1362	3213	1759
AAED1	95	89	61	120	73	105	84	66	83	63
AAGAB	681	600	848	610	679	614	598	387	804	579
AAK1	540	564	657	769	658	674	644	373	765	338
AAMDC	549	489	435	456	609	348	510	364	421	480

Transcript counts per sample and gene

The metadata

Data – sample description found on BioProject:
https://www.ncbi.nlm.nih.gov/bioproject

SRA run selector:
https://www.ncbi.nlm.nih.gov/Traces/study/?query_key=4&WebEnv=MCID_6565af25392c306f7737da9d&o=acc_s%3Aa



Metadata file: "SraRunTable_Diet.txt"

Maastricht University

Sample description in the metadata file

- Please open the "**SraRunTable_Diet_for_inspection.xls**"

Run	Assay Type	AvgSpotLen	Bases	BioProject	BioSample	Bytes	Center Name	Consent	DATASTORE filetype	DATASTORE provider	DATASTORE region	diet_intervention	Experiment	GEO_Accession (
SRR9910560	RNA-Seq	196	12777265395	PRJNA558947	SAMN12502794	4735080640	GEO	public	sra,run.zq,fastq	ncbi,s3,gs	gs.us-east1,ncbi.public,s3.us-east-1	Highprotein	SRX6660008	GSM4010333
SRR9910561	RNA-Seq	197	12194540464	PRJNA558947	SAMN12502793	4522968786	GEO	public	sra,run.zq,fastq	gs,s3,ncbi	gs.us-east1,s3.us-east-1,ncbi.public	Highprotein	SRX6660009	GSM4010334
SRR9910562	RNA-Seq	198	12120405060	PRJNA558947	SAMN12502791	4460448876	GEO	public	fastq,run.zq,sra	s3,ncbi,gs	gs.us-east1,ncbi.public,s3.us-east-1	Highprotein	SRX6660010	GSM4010335
SRR9910563	RNA-Seq	197	12662534057	PRJNA558947	SAMN12502789	4723556165	GEO	public	run.zq,fastq,sra	gs,s3,ncbi	gs.us-east1,ncbi.public,s3.us-east-1	Highprotein	SRX6660011	GSM4010336
SRR9910564	RNA-Seq	197	13148843813	PRJNA558947	SAMN12502788	4909944932	GEO	public	sra,fastq,run.zq	s3,ncbi,gs	s3.us-east-1,gs.us-east1,ncbi.public	Highprotein	SRX6660012	GSM4010337
SRR9910565	RNA-Seq	198	11567180047	PRJNA558947	SAMN12502786	4261698861	GEO	public	sra,fastq,run.zq	s3,gs,ncbi	s3.us-east-1,gs.us-east1,ncbi.public	Highprotein	SRX6660013	GSM4010338
SRR9910566	RNA-Seq	197	12017648869	PRJNA558947	SAMN12502785	4473362646	GEO	public	sra,run.zq,fastq	ncbi,gs,s3	s3.us-east-1,ncbi.public,gs.us-east1	Highprotein	SRX6660014	GSM4010339
SRR9910567	RNA-Seq	197	8718060815	PRJNA558947	SAMN12502783	3229336866	GEO	public	sra,fastq,run.zq	ncbi,s3,gs	gs.us-east1,ncbi.public,s3.us-east-1	Highprotein	SRX6660015	GSM4010340
SRR9910568	RNA-Seq	197	12171340695	PRJNA558947	SAMN12502782	4488223955	GEO	public	sra,fastq,run.zq	s3,gs,ncbi	gs.us-east1,s3.us-east-1,ncbi.public	Highprotein	SRX6660016	GSM4010341
SRR9910569	RNA-Seq	197	11327245139	PRJNA558947	SAMN12502781	4158694996	GEO	public	run.zq,fastq,sra	ncbi,gs,s3	s3.us-east-1,ncbi.public,gs.us-east1	Lowprotein	SRX6660017	GSM4010342
SRR9910570	RNA-Seq	197	11370679910	PRJNA558947	SAMN12502780	4228712889	GEO	public	run.zq,sra,fastq	s3,gs,ncbi	ncbi.public,gs.us-east1,s3.us-east-1	Lowprotein	SRX6660018	GSM4010343
SRR9910571	RNA-Seq	196	13003645401	PRJNA558947	SAMN12502779	4818134376	GEO	public	sra,fastq,run.zq	gs,s3,ncbi	s3.us-east-1,ncbi.public,gs.us-east1	Lowprotein	SRX6660019	GSM4010344
SRR9910572	RNA-Seq	197	10764132292	PRJNA558947	SAMN12502778	4111664095	GEO	public	run.zq,sra,fastq	gs,ncbi,s3	s3.us-east-1,ncbi.public,gs.us-east1	Lowprotein	SRX6660020	GSM4010345
SRR9910573	RNA-Seq	198	12286372973	PRJNA558947	SAMN12502777	4694933544	GEO	public	run.zq,sra,fastq	gs,s3,ncbi	s3.us-east-1,gs.us-east1,ncbi.public	Lowprotein	SRX6660021	GSM4010346
SRR9910574	RNA-Seq	198	9399764553	PRJNA558947	SAMN12502776	3601300223	GEO	public	run.zq,sra,fastq	gs,ncbi,s3	ncbi.public,s3.us-east-1,gs.us-east1	Lowprotein	SRX6660022	GSM4010347
SRR9910575	RNA-Seq	197	12226829567	PRJNA558947	SAMN12502775	4672738778	GEO	public	run.zq,sra,fastq	gs,s3,ncbi	ncbi.public,s3.us-east-1,gs.us-east1	Lowprotein	SRX6660023	GSM4010348
SRR9910576	RNA-Seq	198	11384937737	PRJNA558947	SAMN12502774	4362983194	GEO	public	fastq,sra,run.zq	ncbi,s3,gs	gs.us-east1,ncbi.public,s3.us-east-1	Lowprotein	SRX6660024	GSM4010349
SRR9910577	RNA-Seq	197	11742408932	PRJNA558947	SAMN12502773	4480353786	GEO	public	run.zq,sra,fastq	gs,s3,ncbi	gs.us-east1,s3.us-east-1,ncbi.public	Lowprotein	SRX6660025	GSM4010350
SRR9910578	RNA-Seq	198	10655722072	PRJNA558947	SAMN12502772	4089666054	GEO	public	fastq,sra,run.zq	gs,s3,ncbi	ncbi.public,gs.us-east1,s3.us-east-1	Lowprotein	SRX6660026	GSM4010351

Illuminia Beadchip – Transcriptomic data

Molecular Nutrition & Food Research

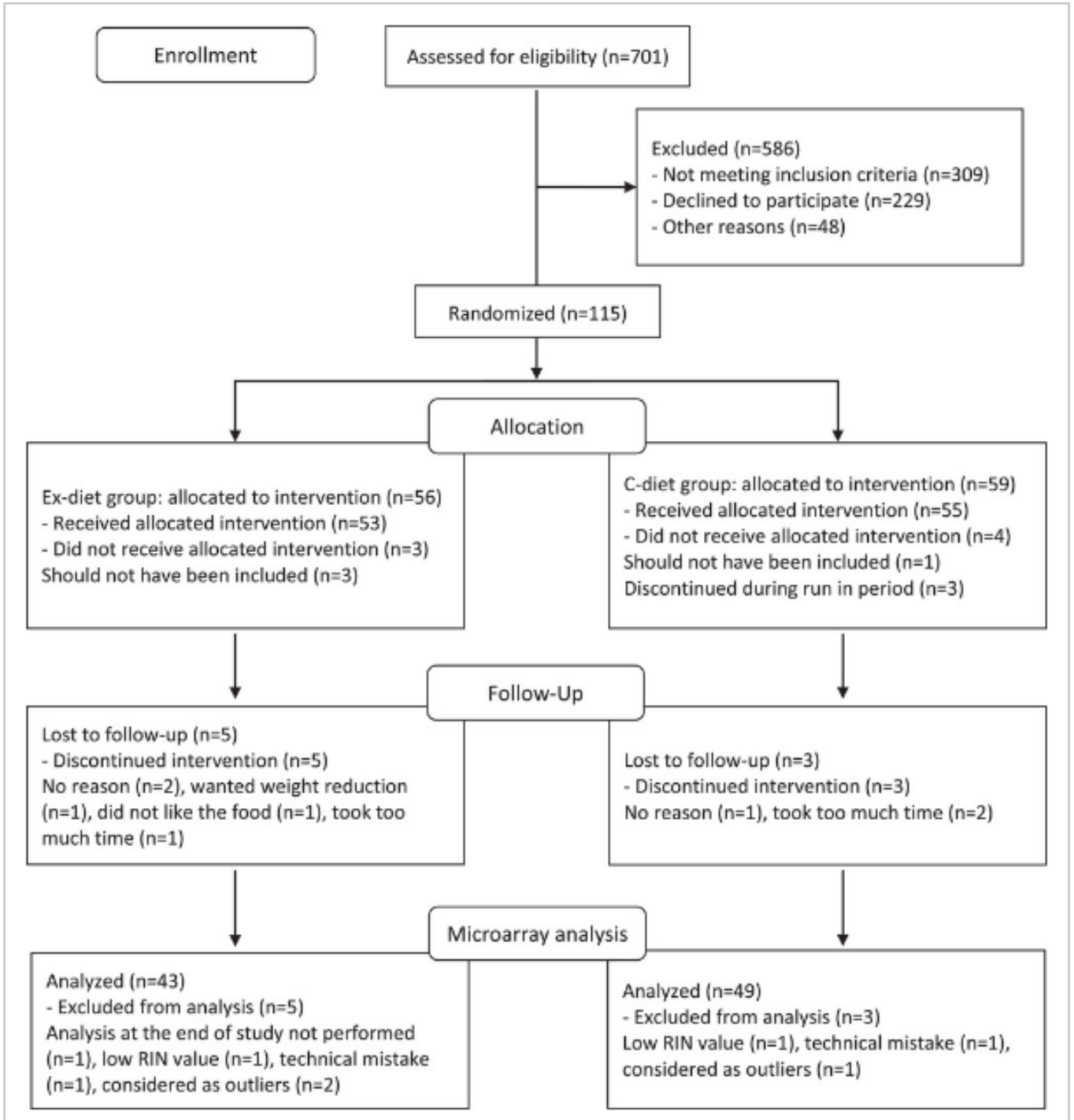
Research Article | Open Access

Replacing Saturated Fat with Polyunsaturated Fat Modulates Peripheral Blood Mononuclear Cell Gene Expression and Pathways Related to Cardiovascular Disease Risk Using a Whole Transcriptome Approach

Sunniva V. Larsen, Kirsten B. Holven, Jacob J. Christensen, Arnar Flatberg, Amanda Rundblad, Lena Leder, Rune Blomhoff, Vibeke Telle-Hansen, Marjukka Kolehmainen, Carsten Carlberg ... See all authors

First published: 27 October 2021 | https://doi.org/10.1002/mnfr.202100633 | Citations: 3

Maastricht University find full text


Enrollment
Assessed for eligibility (n=701)
Excluded (n=586)
- Not meeting inclusion criteria (n=309)
- Declined to participate (n=229)
- Other reasons (n=48)
Randomized (n=115)
Allocation
Ex-diet group: allocated to intervention (n=56)
- Received allocated intervention (n=53)
- Did not receive allocated intervention (n=3)
Should not have been included (n=3)
C-diet group: allocated to intervention (n=59)
- Received allocated intervention (n=55)
- Did not receive allocated intervention (n=4)
Should not have been included (n=1)
Discontinued during run in period (n=3)
Follow-Up
Lost to follow-up (n=5)
- Discontinued intervention (n=5)
No reason (n=2), wanted weight reduction (n=1), did not like the food (n=1), took too much time (n=1)
Lost to follow-up (n=3)
- Discontinued intervention (n=3)
No reason (n=1), took too much time (n=2)
Microarray analysis
Analyzed (n=43)
- Excluded from analysis (n=5)
Analysis at the end of study not performed (n=1), low RIN value (n=1), technical mistake (n=1), considered as outliers (n=2)
Analyzed (n=49)
- Excluded from analysis (n=3)
Low RIN value (n=1), technical mistake (n=1), considered as outliers (n=1)

NCBI

Gene Expression Omnibus

HOME | SEARCH | SITE MAP | GEO Publications | FAQ | MIAME | Email GEO

NCBI > GEO > **Accession Display**

Not logged in | Login

GEO help: Mouse over screen elements for information.

Scope: Self Format: HTML Amount: Quick GEO accession: GSE176043 GO

Series GSE176043

Query DataSets for GSE176043

Status	Public on Jun 03, 2021
Title	Effects of replacing saturated fat with polyunsaturated fat on whole genome gene expression profiles in peripheral blood mononuclear cells - a randomized controlled dietary intervention study
Organism	Homo sapiens
Experiment type	Expression profiling by array
Summary	To further understand the molecular mechanisms underlying the health effects of replacing saturated fat (SFA) with polyunsaturated fat (PUFA), we aimed to explore the whole transcriptome response of peripheral blood mononuclear cells (PBMCs) in a dietary intervention where SFAs were replaced with PUFAs.
Overall design	Moderately hypercholesterolaemic subjects (n = 115) were randomly assigned to a control diet group (c-diet group) or an experimental diet group (ex-diet group) receiving commercially available and regularly consumed food items with different fatty acid composition (replacing SFAs with mostly n-6 PUFAs) for eight weeks. Microarray analyses of RNA isolated from PBMCs at baseline and after eight weeks of intervention were performed using illuminaHT-12 v4 Expression BeadChips. A total of 92 subjects (ex-diet group; n=43, c-diet group; n=49) were included in the statistical analyses of this study. Group A = Ex-diet group Group B = C-diet group
Contributor(s)	Larsen SV, Holven KB, Christensen JJ, Flatberg A, Rundblad A, Leder L, Blomhoff R, Hansen VT, Kolehmainen M, Carlberg C, Myhrstad MC, Thoresen M, Ulven SM

Platforms (1)	GPL10558 Illumina HumanHT-12 V4.0 expression beadchip
Samples (184)	GSM5353517 visit 1 (baseline), group A 1_A1
More...	GSM5353518 visit 2 (end of study), group A 1_A10
	GSM5353519 visit 1 (baseline), group B 1_A11

Relations

BioProject PRJNA734627

Analyze with GEO2R

Download family	Format
SOFT formatted family file(s)	SOFT
MINiML formatted family file(s)	MINiML
Series Matrix File(s)	TXT

Metadata

Supplementary file	Size	Download	File type/resource
GSE176043_RAW.tar	26.2 Mb	(http)(custom)	TAR
GSE176043_non-normalized.txt.gz	49.6 Mb	(ftp)(http)	TXT

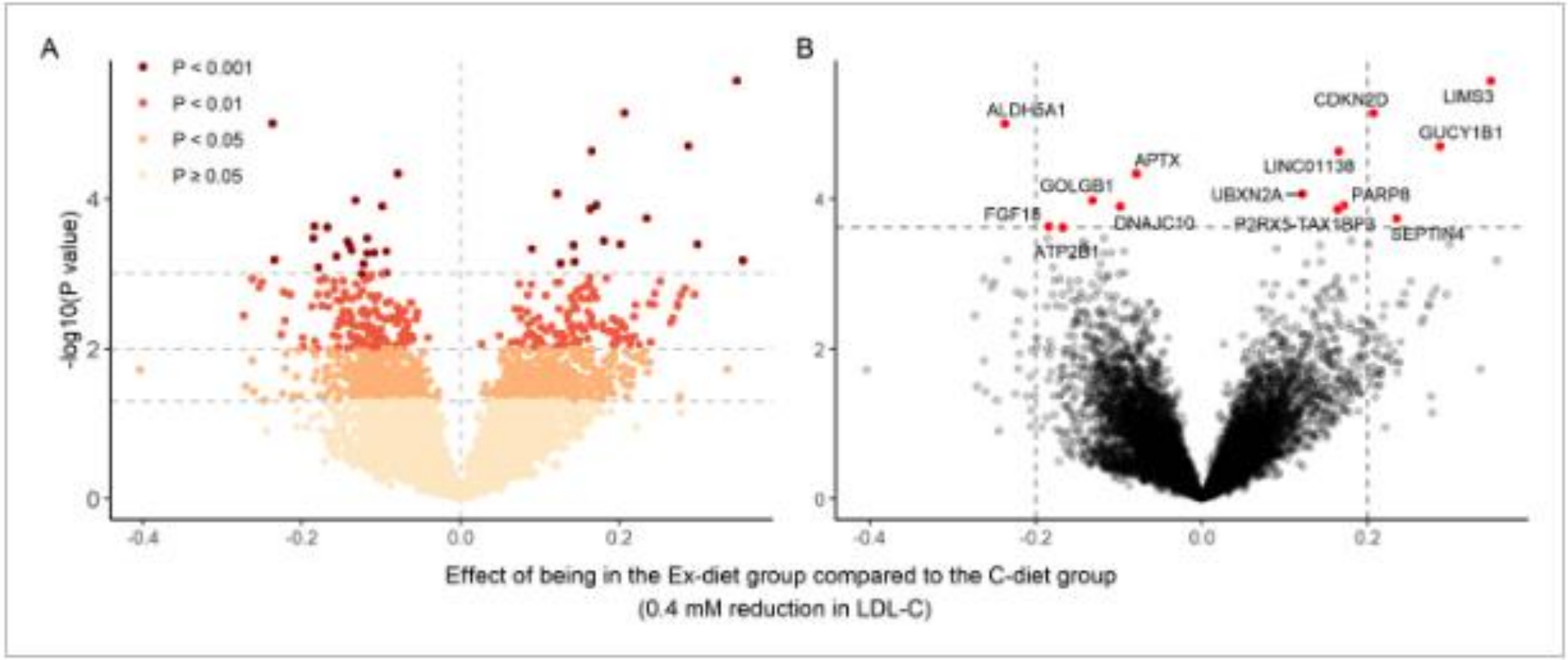
Expression data

Processed data included within Sample table

GSE176043_non-normalized

ID Ref	1_A1.AVG_Signal	1_A1.Detection Pval	1_A10.AVG_Signal	1_A10.Detection Pval	1_A11.AVG_Signal	1_A11.Detection Pval
ILMN_1762337	89.15668	0.7363636	90.90633	0.4025974	97.96025	0.1896104
ILMN_2055271	102.6775	0.06103896	111.7579	0.006493506	100.7383	0.0948052
ILMN_1736007	90.20063	0.651948	81.54804	0.9662338	105.5627	0.02857143
ILMN_2383229	92.38062	0.4727273	87.94315	0.6363636	84.79773	0.9454545
ILMN_1806310	89.96393	0.6753247	98.10756	0.1402597	94.4104	0.3376623
ILMN_1779670	98.10552	0.1662338	91.50809	0.3649351	95.45519	0.2948052
ILMN_1653355	89.09814	0.7402598	97.89408	0.1467533	95.12685	0.3077922
ILMN_1717783	88.41372	0.7909091	82.10333	0.9506493	93.16858	0.4038961
ILMN_1705025	90.98981	0.5805195	87.98139	0.6311688	94.13724	0.3493507
ILMN_1814316	90.83102	0.5922078	84.68837	0.8688312	93.88177	0.3662338
ILMN_2359168	83.19699	0.9922078	88.73565	0.5558441	87.23669	0.8441558
ILMN_1731507	86.49041	0.8948052	86.35243	0.7571428	86.3761	0.8922078

Volcano plot



Pathway analysis