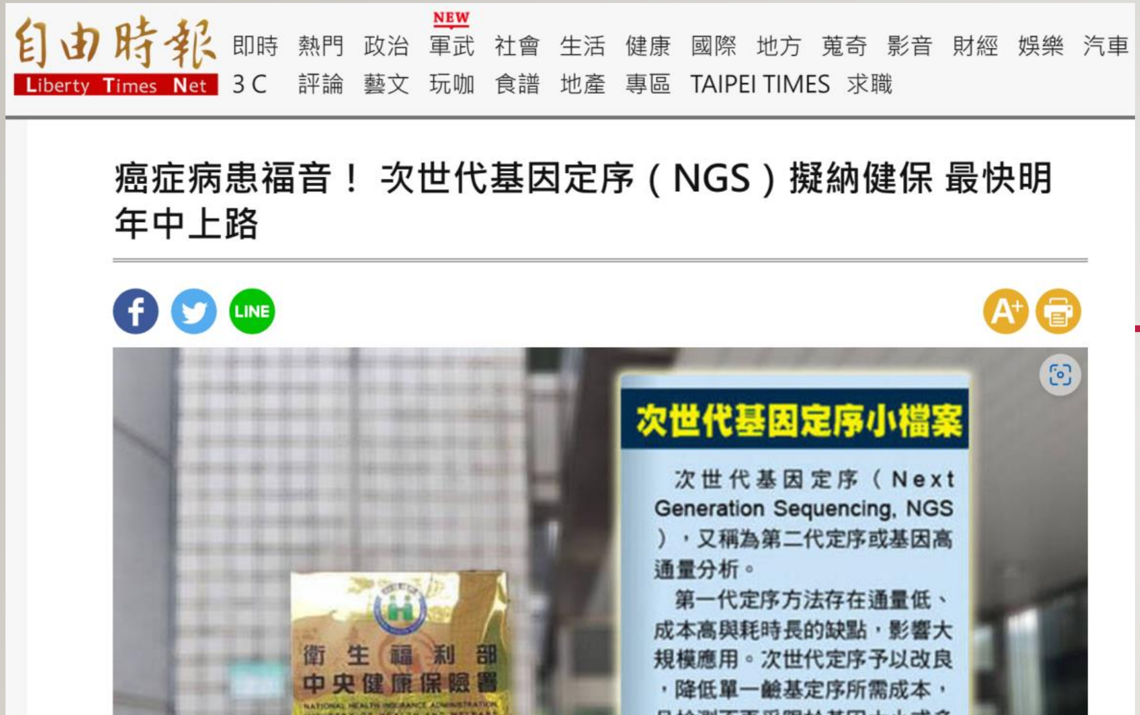
# 大數據實戰分享：大師成長之路』之五：

## 實戰基因大數據(一):次世代高通量基因定序分析平台於精準藥物之應用

大數據中心生統小組
生醫資訊實驗室
蔡秉興 高級助理研究員

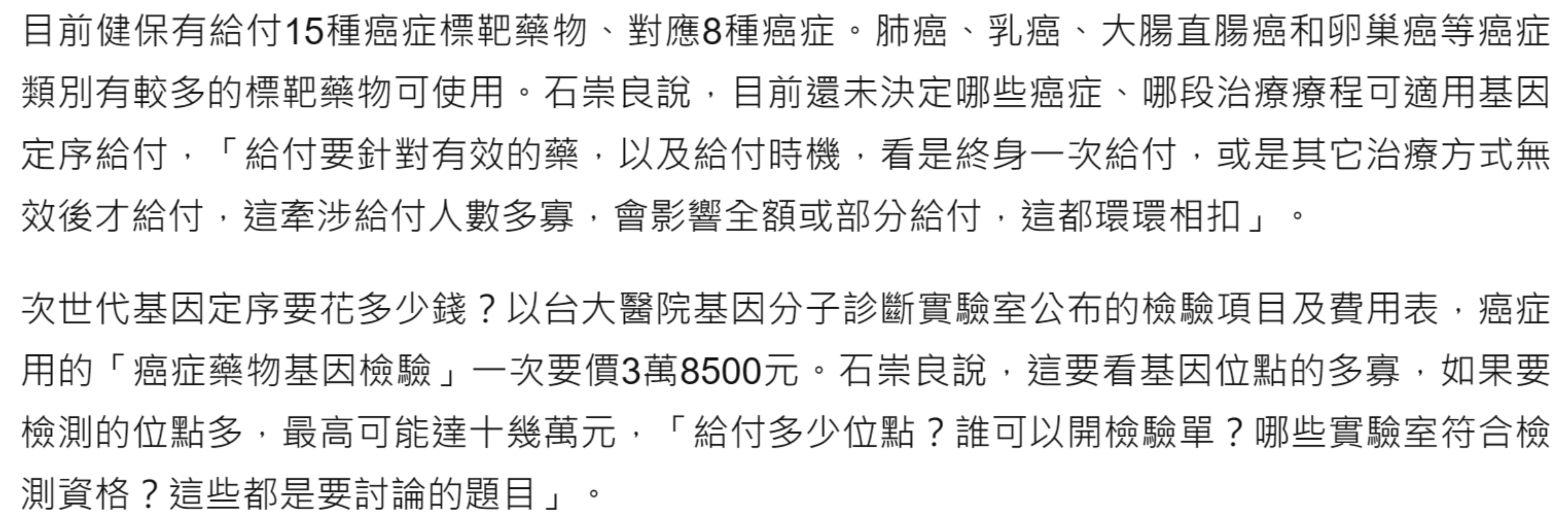

自由時報
Liberty Times Net
即時 熱門 政治 NEW 軍武 社會 生活 健康 國際 地方 蒐奇 影音 財經 娛樂 汽車
3C 評論 藝文 玩咖 食譜 地產 專區 TAIPEI TIMES 求職
LINE
衛生福利部
中央健康保險署

# 癌症病患福音！ 次世代基因定序（NGS）擬納健保 最快明年中上路

**次世代基因定序小檔案**

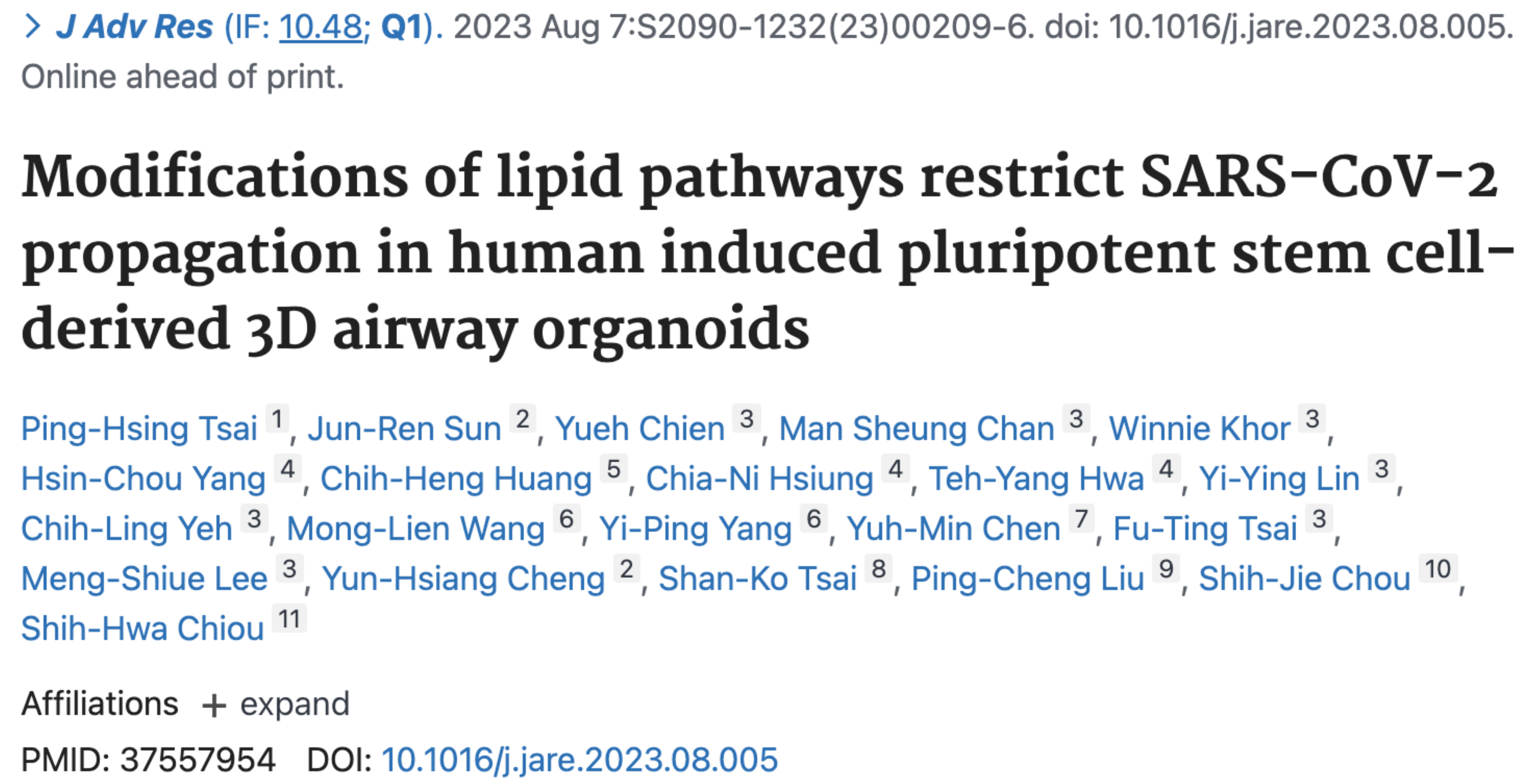
次世代基因定序（Next Generation Sequencing, NGS），又稱為第二代定序或基因高通量分析。

第一代定序方法存在通量低、成本高與耗時長的缺點，影響大規模應用。次世代定序予以改良，降低單一鹼基定序所需成本，



目前健保有給付15種癌症標靶藥物、對應8種癌症。肺癌、乳癌、大腸直腸癌和卵巢癌等癌症類別有較多的標靶藥物可使用。石崇良說，目前還未決定哪些癌症、哪段治療療程可適用基因定序給付，「給付要針對有效的藥，以及給付時機，看是終身一次給付，或是其它治療方式無效後才給付，這牽涉給付人數多寡，會影響全額或部分給付，這都環環相扣」。

次世代基因定序要花多少錢？以台大醫院基因分子診斷實驗室公布的檢驗項目及費用表，癌症用的「癌症藥物基因檢驗」一次要價3萬8500元。石崇良說，這要看基因位點的多寡，如果要檢測的位點多，最高可能達十幾萬元，「給付多少位點？誰可以開檢驗單？哪些實驗室符合檢測資格？這些都是要討論的題目」。



> ***J Adv Res*** (IF: 10.48; **Q1**). 2023 Aug 7:S2090-1232(23)00209-6. doi: 10.1016/j.jare.2023.08.005.
Online ahead of print.

# Modifications of lipid pathways restrict SARS-CoV-2 propagation in human induced pluripotent stem cell-derived 3D airway organoids

Ping-Hsing Tsai [1], Jun-Ren Sun [2], Yueh Chien [3], Man Sheung Chan [3], Winnie Khor [3], Hsin-Chou Yang [4], Chih-Heng Huang [5], Chia-Ni Hsiung [4], Teh-Yang Hwa [4], Yi-Ying Lin [3], Chih-Ling Yeh [3], Mong-Lien Wang [6], Yi-Ping Yang [6], Yuh-Min Chen [7], Fu-Ting Tsai [3], Meng-Shiue Lee [3], Yun-Hsiang Cheng [2], Shan-Ko Tsai [8], Ping-Cheng Liu [9], Shih-Jie Chou [10], Shih-Hwa Chiou [11]

Affiliations + expand

PMID: 37557954 DOI: 10.1016/j.jare.2023.08.005

# STATINS: THE CHOLESTEROL BUSTING PILL

## FOR

- Helps prevent up to 2% of heart attacks and strokes for most people
- But this rises to up to 7% among the most vulnerable patients
- Helps bring down Britain's billion-pound cardiovascular health bill
- Commonly complained of 'side effects' could actually be down to old age
- Some studies show pills could also be used to improve eye-health and treat cancer

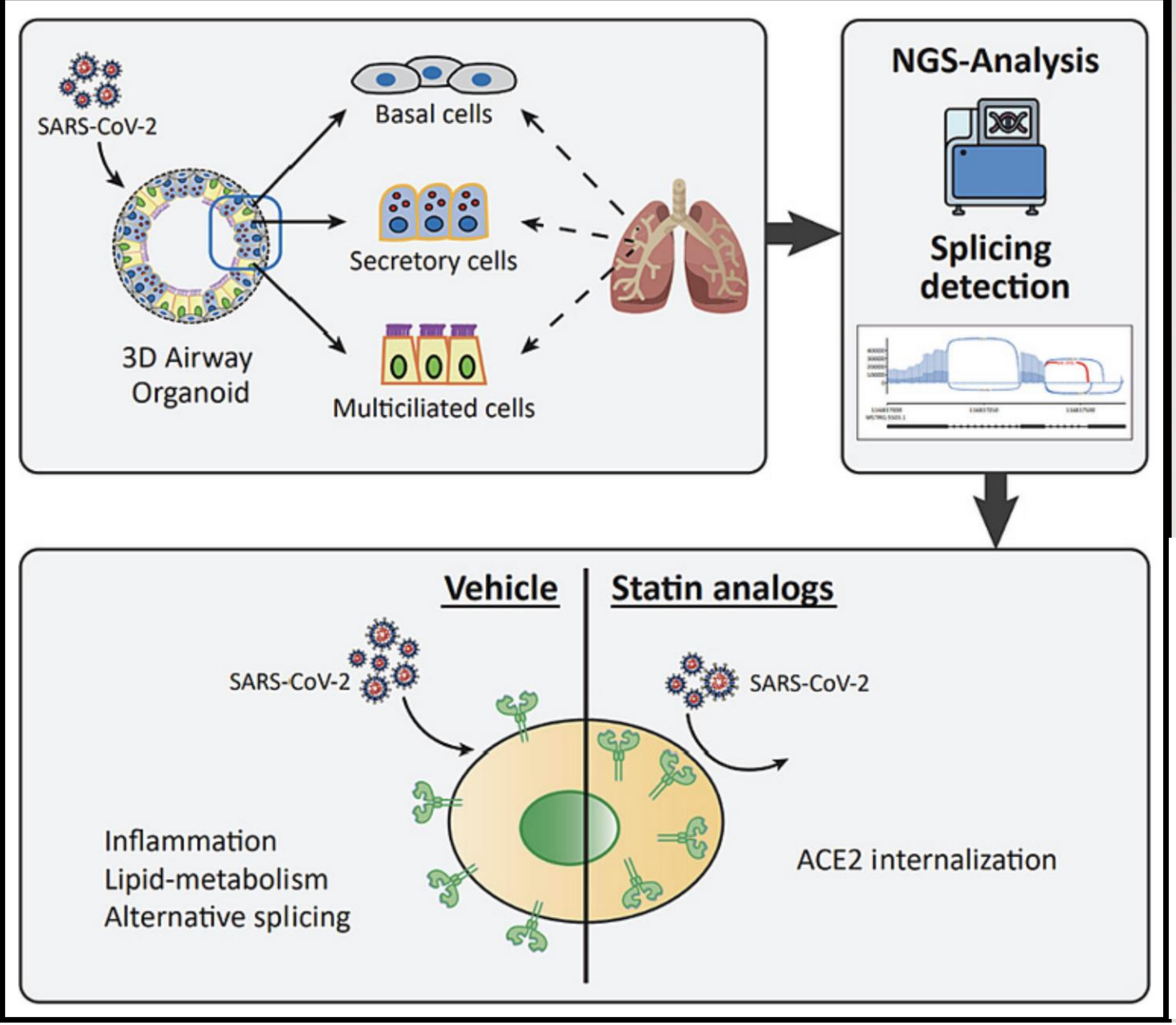
史他汀 (Statins) 類藥物包含Rosuvastatin（冠脂妥）、Atorvastatin（立普妥）、Fluvastatin（益脂可）、Simvastatin（維妥力）等，是醫師最常開立的處方藥，主要抑制肝臟再製造膽固醇，進而減少血中的膽固醇與三酸甘油脂的含量。

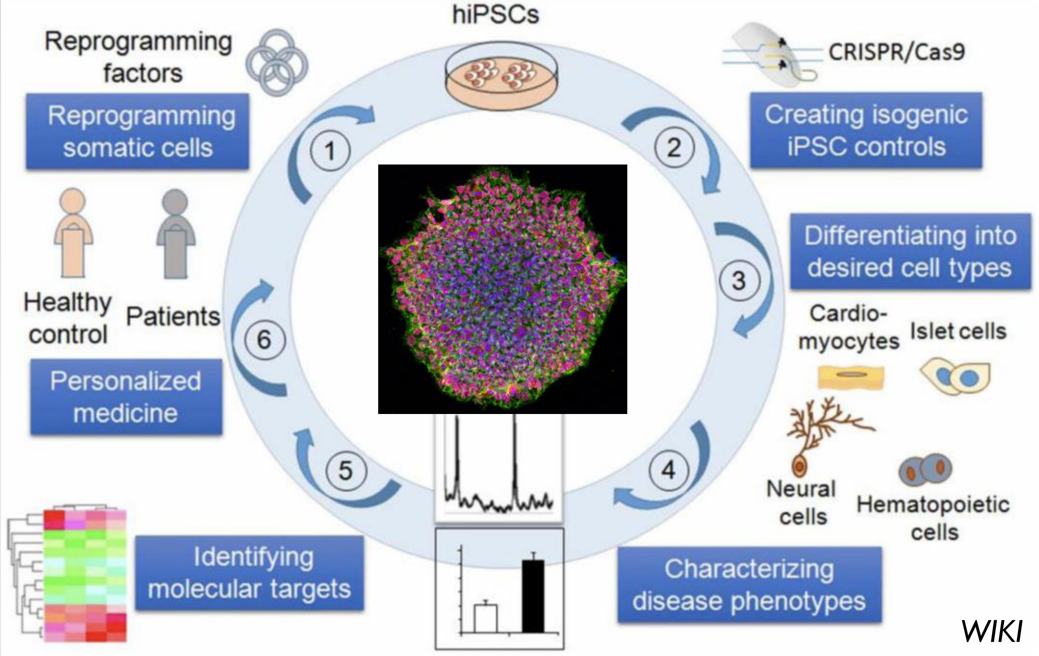
STATIN
Tablets
STATIN 40 mg

- Taken daily by 8million Brits and 32million Americans
- Can help reduce bad cholesterol by up to 50%
- Costs as little as 2p per pill to manufacture

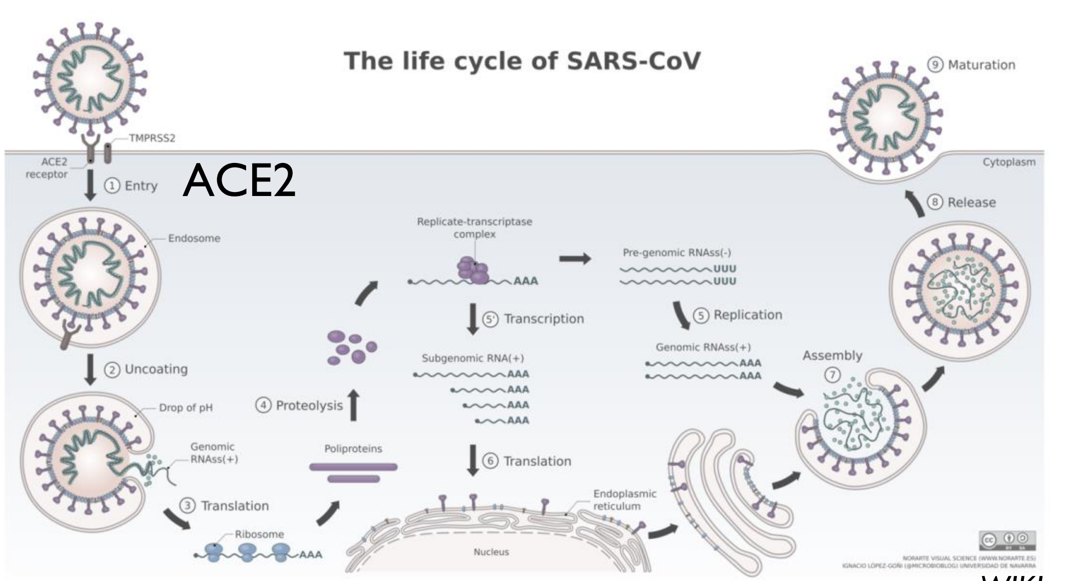
## AGAINST

- Pills have been blamed for increasing aches and pains by some Britons
- Currently costs £100m to dish out the pills in England alone
- Some experts claim patients are skipping exercise and relying on statins for heart health
- Some studies have linked them to increased risk of cataracts and diabetes


SARS-CoV-2
3D Airway Organoid
Basal cells
Secretory cells
Multiciliated cells
NGS-Analysis
Splicing detection
Vehicle
Statin analogs
SARS-CoV-2
SARS-CoV-2
Inflammation
Lipid-metabolism
Alternative splicing
ACE2 internalization

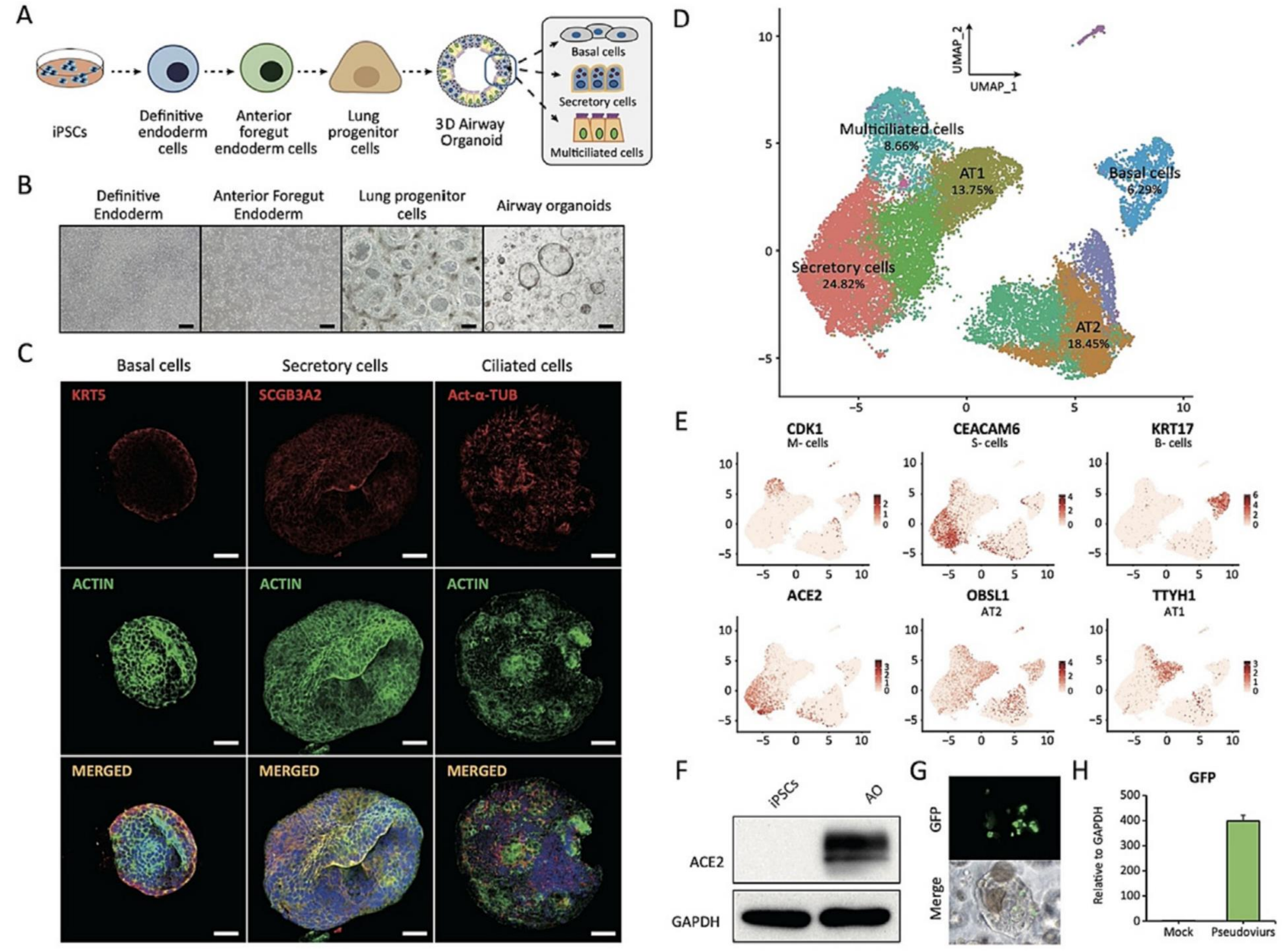

hiPSCs
Reprogramming factors
Reprogramming somatic cells
CRISPR/Cas9
Creating isogenic iPSC controls
Healthy control
Patients
Differentiating into desired cell types
Cardio-myocytes
Islet cells
Neural cells
Hematopoietic cells
Personalized medicine
Identifying molecular targets
Characterizing disease phenotypes
1
2
3
4
5
6

WIKI

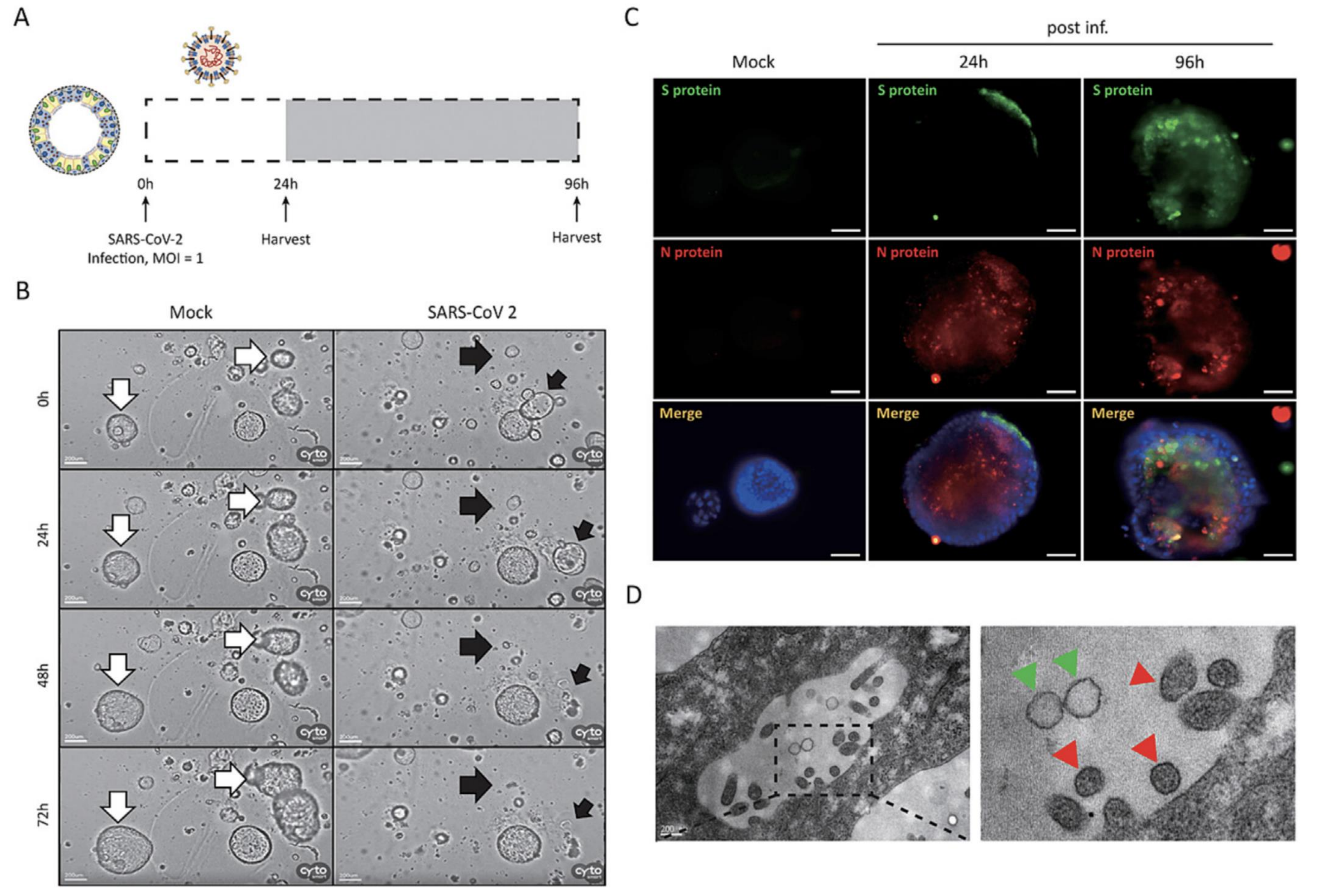

The life cycle of SARS-CoV
ACE2
TMPRSS2
ACE2 receptor
① Entry
Endosome
② Uncoating
Drop of pH
Genomic RNAss(+)
③ Translation
Ribosome
④ Proteolysis
Poliproteins
Replicate-transcriptase complex
⑤ Transcription
Subgenomic RNA(+)
⑥ Translation
Pre-genomic RNAss(-)
⑤ Replication
Genomic RNAss(+)
Endoplasmic reticulum
Nucleus
Assembly ⑦
⑧ Release
⑨ Maturation
Cytoplasm

WIKI

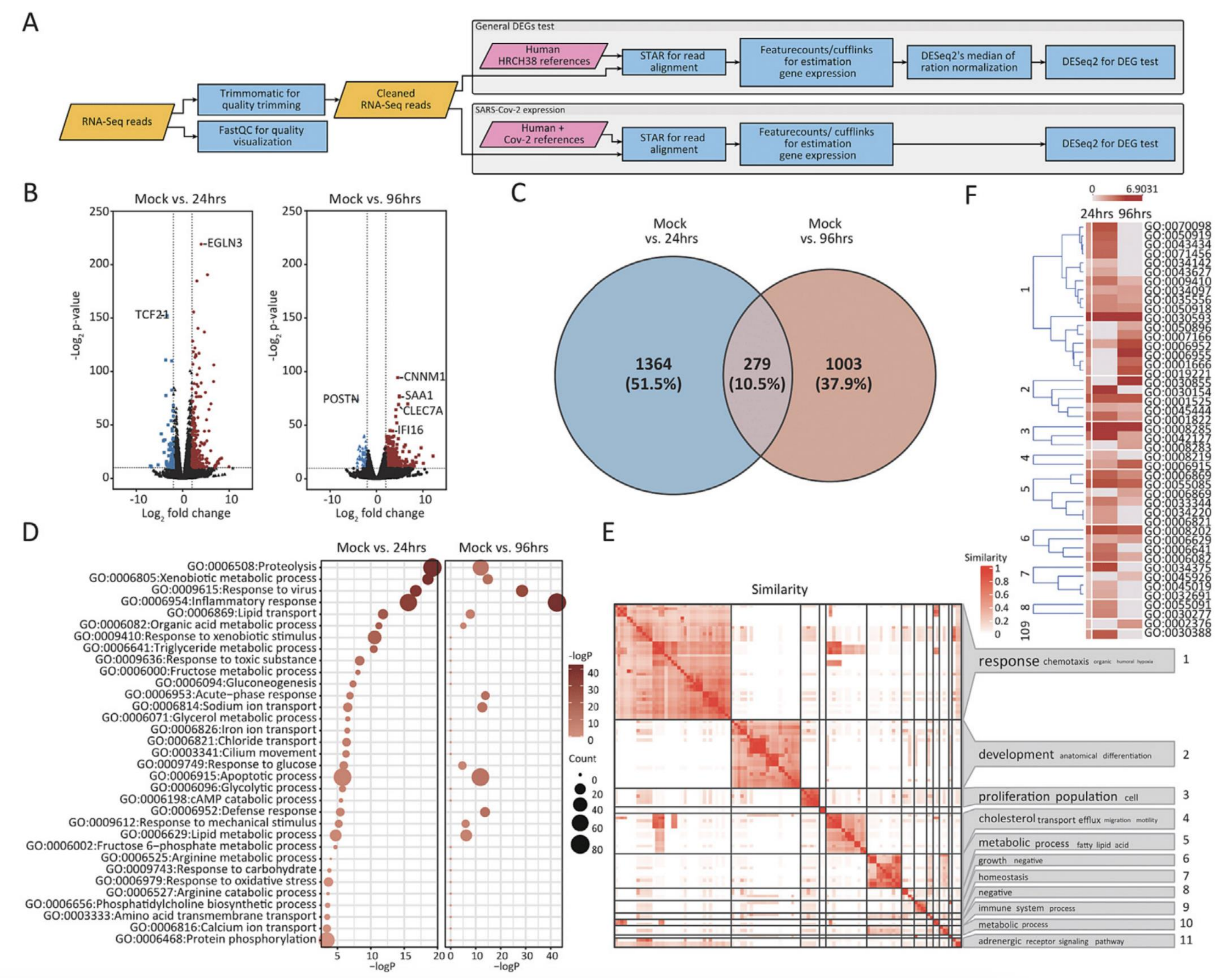
# Modifications of lipid pathways restrict SARS-CoV-2 propagation in hIPSC-derived 3D airway organoids


A
iPSCs
Definitive endoderm cells
Anterior foregut endoderm cells
Lung progenitor cells
3D Airway Organoid
Basal cells
Secretory cells
Multiciliated cells
B
Definitive Endoderm
Anterior Foregut Endoderm
Lung progenitor cells
Airway organoids
C
Basal cells
Secretory cells
Ciliated cells
KRT5
SCGB3A2
Act-α-TUB
ACTIN
ACTIN
ACTIN
MERGED
MERGED
MERGED
D
UMAP_2
UMAP_1
Multiciliated cells 8.66%
AT1 13.75%
Basal cells 6.29%
Secretory cells 24.82%
AT2 18.45%
10
5
0
-5
-5
0
5
10
E
CDK1
M- cells
CEACAM6
S- cells
KRT17
B- cells
ACE2
OBSL1
AT2
TTYH1
AT1
F
iPSCs
AO
ACE2
GAPDH
G
GFP
Merge
H
GFP
Relative to GAPDH
500
400
300
200
100
0
Mock
Pseudoviurs

# IPSC-derived airway organoids exhibit ACE2 expression

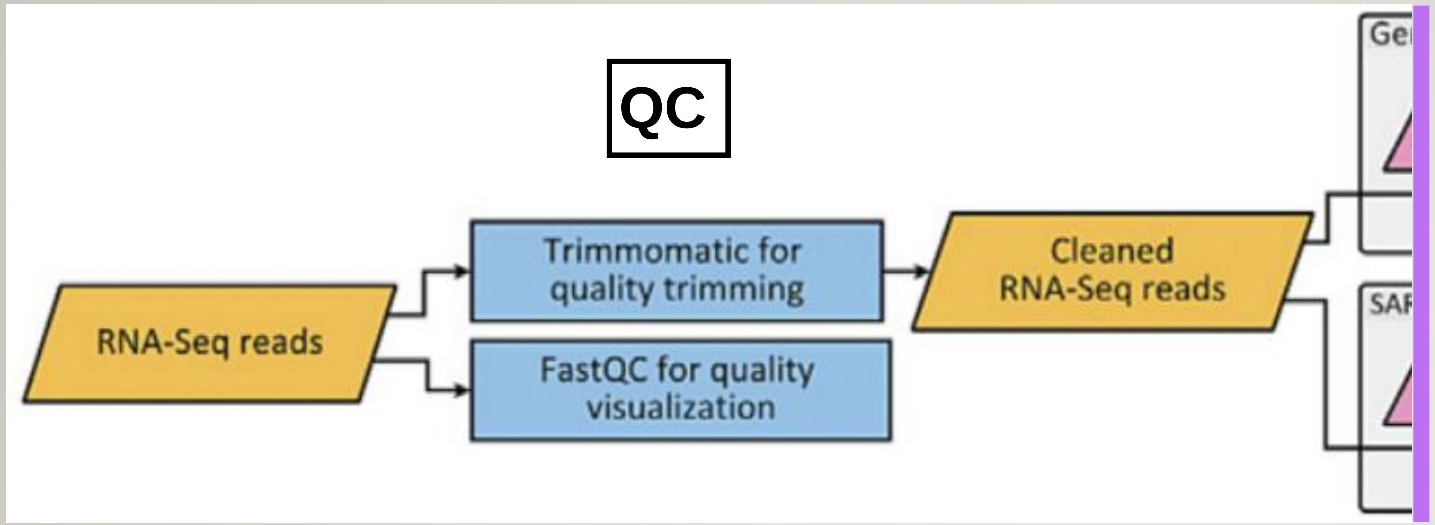

A
0h
24h
96h
SARS-CoV-2
Infection, MOI = 1
Harvest
Harvest
B
Mock
SARS-CoV 2
0h
24h
48h
72h
C
post inf.
Mock
24h
96h
S protein
N protein
Merge
D

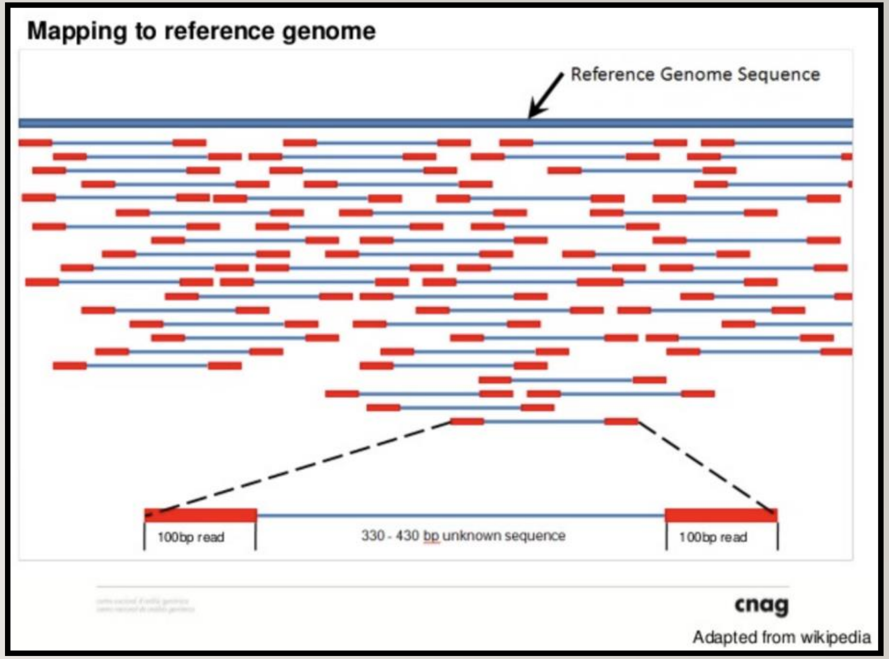
# Vulnerability of Airway Organoids to SARS-CoV-2 Infection



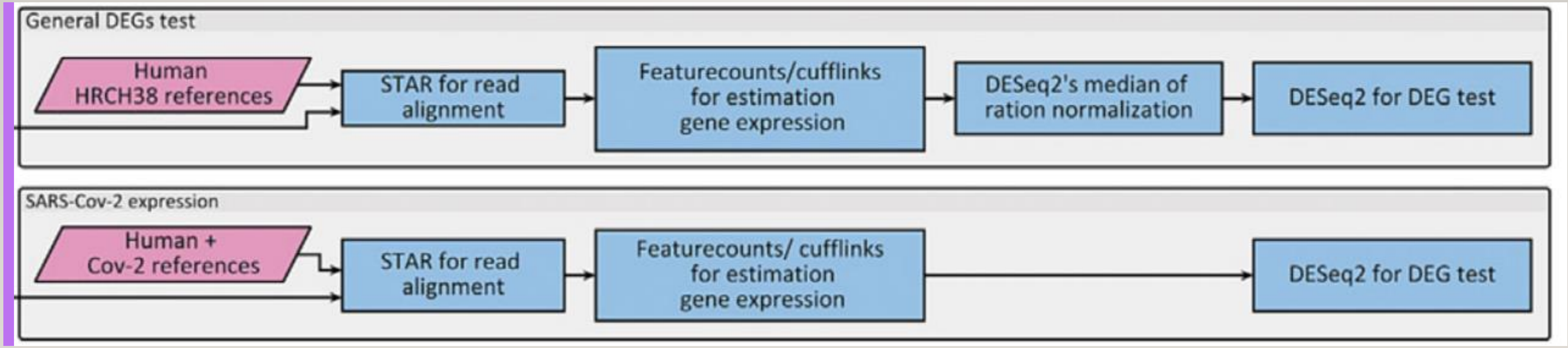
# Next-generation sequencing analysis of SARS-CoV-2 infection.

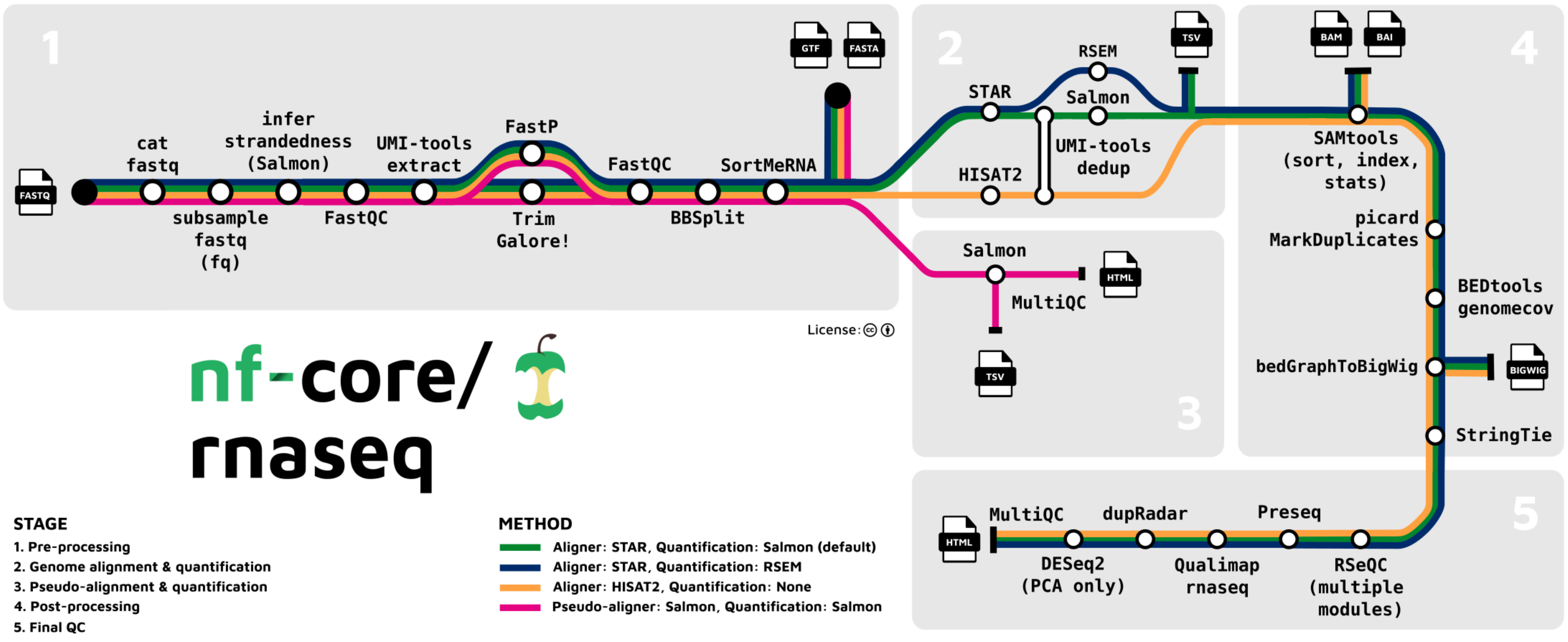
# Next-generation sequencing analysis of SARS-CoV-2 infection.


QC
RNA-Seq reads
Trimmomatic for quality trimming
FastQC for quality visualization
Cleaned RNA-Seq reads


Mapping to reference genome
Reference Genome Sequence
100bp read
330 - 430 bp unknown sequence
100bp read
cnag

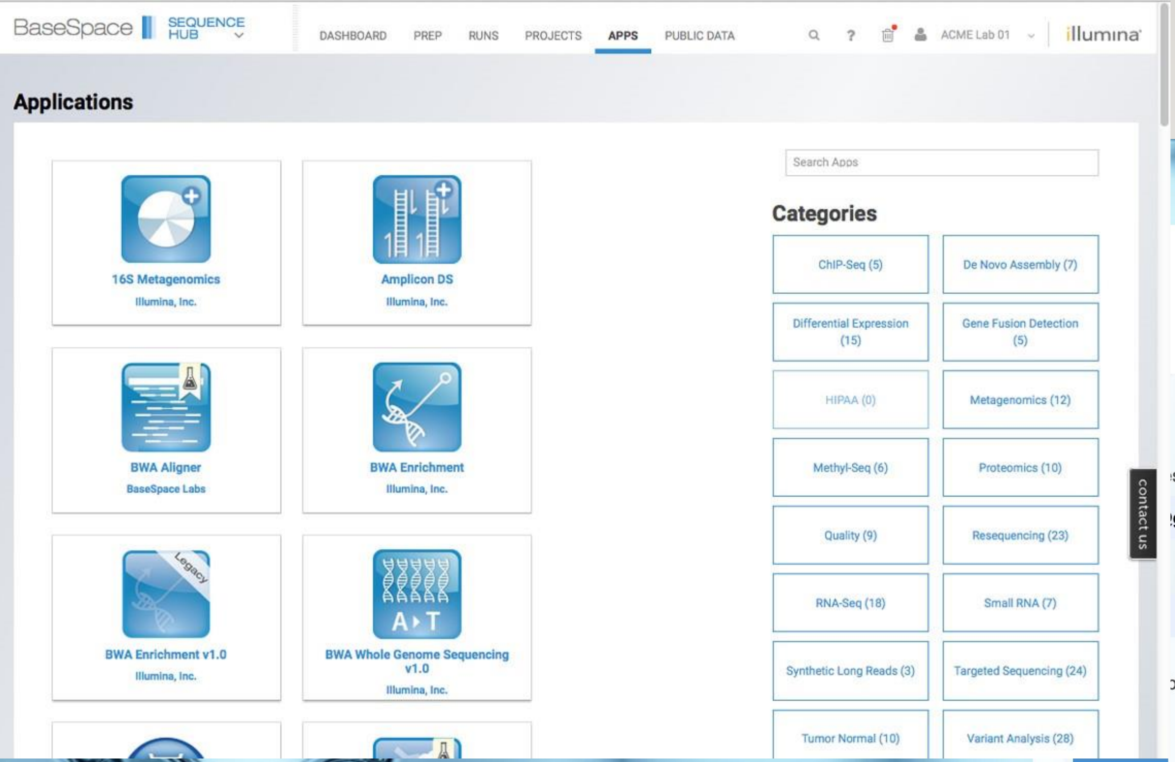
Adapted from wikipedia


General DEGs test
Human HRCH38 references
STAR for read alignment
Featurecounts/cufflinks for estimation gene expression
DESeq2's median of ration normalization
DESeq2 for DEG test
SARS-Cov-2 expression
Human + Cov-2 references
STAR for read alignment
Featurecounts/ cufflinks for estimation gene expression
DESeq2 for DEG test


1
FASTQ
cat fastq
subsample fastq (fq)
infer strandedness (Salmon)
FastQC
UMI-tools extract
FastP
Trim Galore!
FastQC
BBSplit
SortMeRNA
GTF
FASTA
2
STAR
RSEM
Salmon
TSV
HISAT2
UMI-tools dedup
Salmon
HTML
MultiQC
TSV
3
BAM
BAI
4
SAMtools (sort, index, stats)
picard MarkDuplicates
BEDtools genomecov
bedGraphToBigWig
BIGWIG
StringTie
5
HTML
MultiQC
dupRadar
Preseq
DESeq2 (PCA only)
Qualimap rnaseq
RSeQC (multiple modules)
License:
nf-core/ rnaseq
STAGE
1. Pre-processing
2. Genome alignment & quantification
3. Pseudo-alignment & quantification
4. Post-processing
5. Final QC
METHOD
🟩 Aligner: STAR, Quantification: Salmon (default)
🟦 Aligner: STAR, Quantification: RSEM
🟧 Aligner: HISAT2, Quantification: None
🟪 Pseudo-aligner: Salmon, Quantification: Salmon

**nf-core/rnaseq** is a bioinformatics pipeline

STAR + Salmon


BaseSpace SEQUENCE HUB
DASHBOARD
PREP
RUNS
PROJECTS
APPS
PUBLIC DATA
ACME Lab 01
illumina
contact us

## Applications

**16S Metagenomics**
Illumina, Inc.

**Amplicon DS**
Illumina, Inc.

**BWA Aligner**
BaseSpace Labs

**BWA Enrichment**
Illumina, Inc.

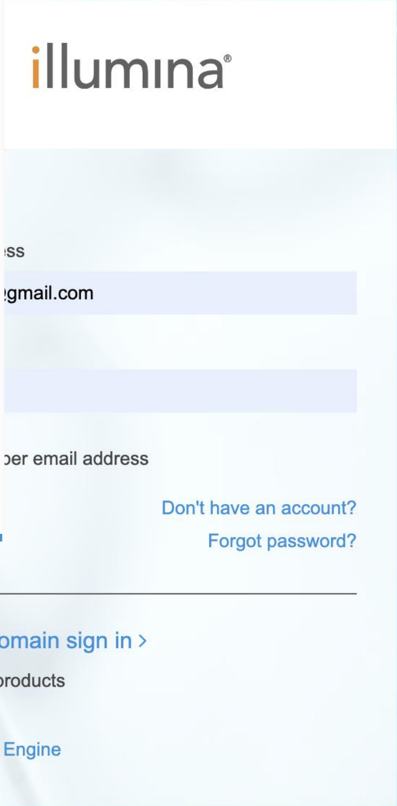
**BWA Enrichment v1.0**
Illumina, Inc.

**BWA Whole Genome Sequencing v1.0**
Illumina, Inc.

Search Apps

**Categories**

| | |
|---|---|
| ChIP-Seq (5) | De Novo Assembly (7) |
| Differential Expression (15) | Gene Fusion Detection (5) |
| HIPAA (0) | Metagenomics (12) |
| Methyl-Seq (6) | Proteomics (10) |
| Quality (9) | Resequencing (23) |
| RNA-Seq (18) | Small RNA (7) |
| Synthetic Long Reads (3) | Targeted Sequencing (24) |
| Tumor Normal (10) | Variant Analysis (28) |



ss

gmail.com

per email address

Don't have an account?

Forgot password?

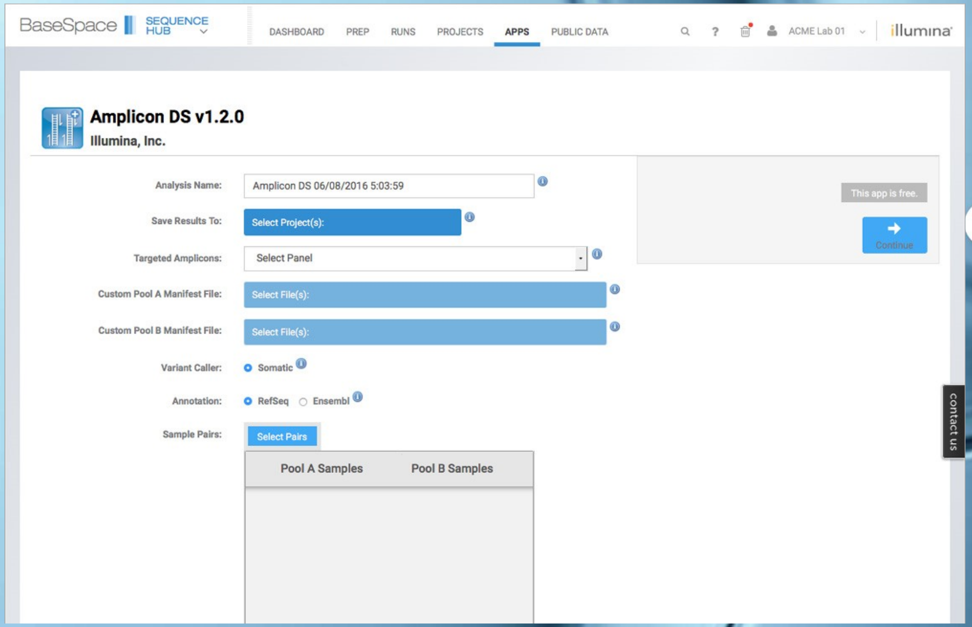
Private domain sign in ›

Additional products

Verinata

Correlation Engine

# 免費試用一個月


BaseSpace SEQUENCE HUB
DASHBOARD
PREP
RUNS
PROJECTS
APPS
PUBLIC DATA
ACME Lab 01
illumina
This app is free.
Continue
contact us

**Amplicon DS v1.2.0**
Illumina, Inc.

Analysis Name: Amplicon DS 06/08/2016 5:03:59

Save Results To: Select Project(s):

Targeted Amplicons: Select Panel

Custom Pool A Manifest File: Select File(s):

Custom Pool B Manifest File: Select File(s):

Variant Caller: ● Somatic

Annotation: ● RefSeq ○ Ensembl

Sample Pairs: Select Pairs

| Pool A Samples | Pool B Samples |
|---|---|

**BaseSpace** Sequence Hub幫助使用者彈性管理NGS數據，提供遠端查看定序流程, 儲存定序分析資料, 分享數據, 分析軟體及公共的演示數據等多項功能，多項分析軟體可針對不同研究領域作應用，簡單易懂的圖形介面設計以及自動化分析流程，可有效縮短時間與分析所需人力成本。

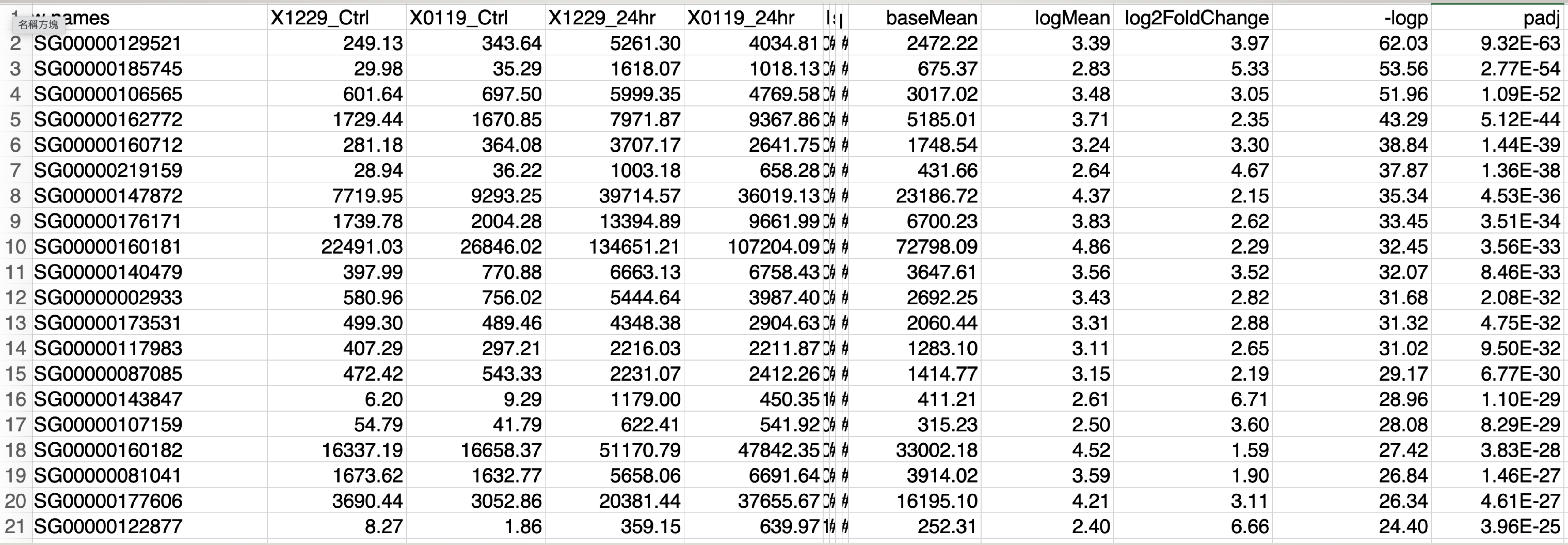
# 次世代高通量基因定序(NGS)分析操作

- 如何挑出有差異的基因？
- 如何解釋差異？
- 可以不要很明顯的現象嗎？
- 就愛這一個功能可以嗎？
- 憑空可以生資料嗎？

# 如何挑出有差異的基因？

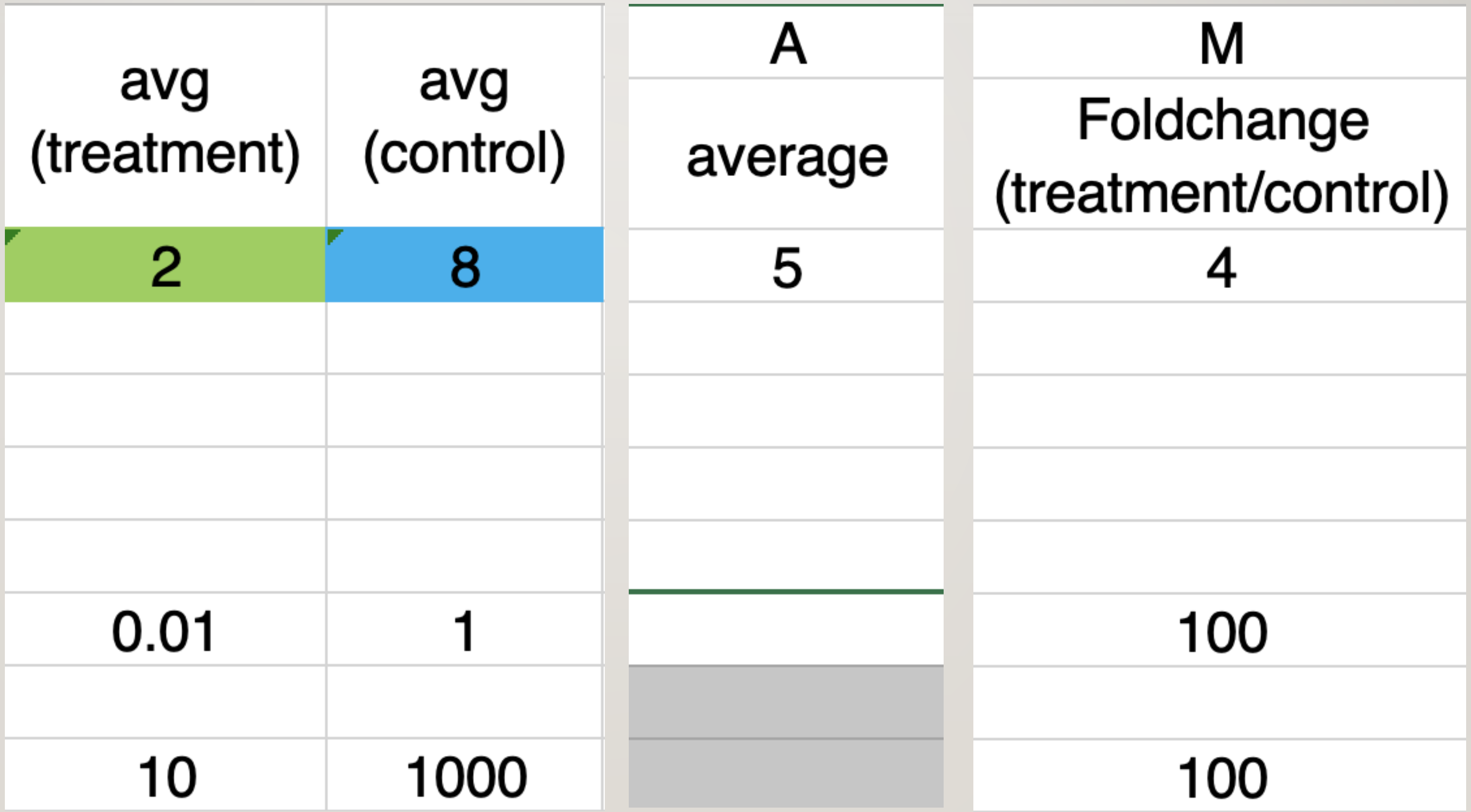
- **MA plot**
- **Volcano plot**

# 次世代高通量基因定序EXCEL操作



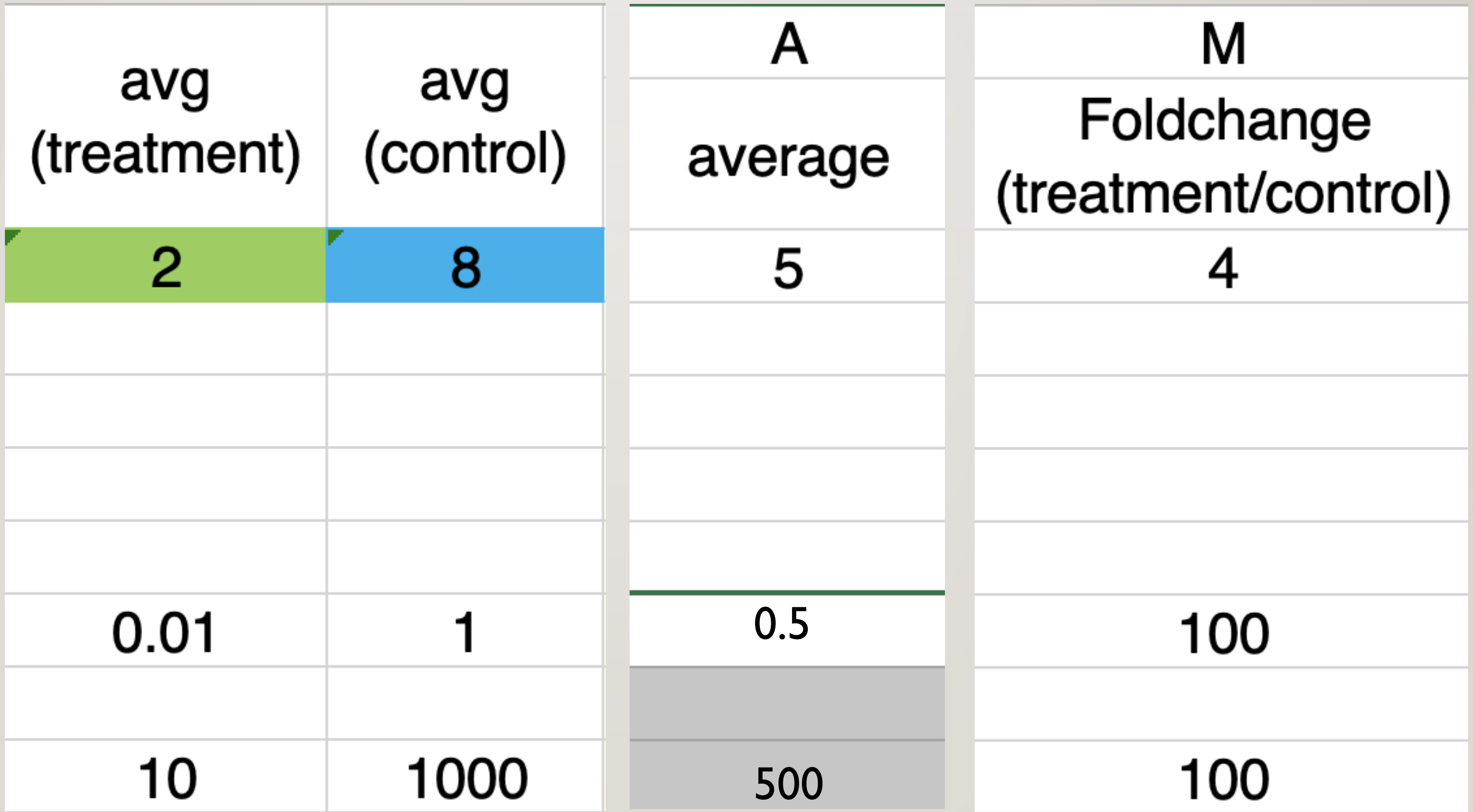
| 1 | w names | X1229_Ctrl | X0119_Ctrl | X1229_24hr | X0119_24hr | l | s | \| | baseMean | logMean | log2FoldChange | -logp | padj |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | SG00000129521 | 249.13 | 343.64 | 5261.30 | 4034.81 | 0 | # | # | 2472.22 | 3.39 | 3.97 | 62.03 | 9.32E-63 |
| 3 | SG00000185745 | 29.98 | 35.29 | 1618.07 | 1018.13 | 0 | # | # | 675.37 | 2.83 | 5.33 | 53.56 | 2.77E-54 |
| 4 | SG00000106565 | 601.64 | 697.50 | 5999.35 | 4769.58 | 0 | # | # | 3017.02 | 3.48 | 3.05 | 51.96 | 1.09E-52 |
| 5 | SG00000162772 | 1729.44 | 1670.85 | 7971.87 | 9367.86 | 0 | # | # | 5185.01 | 3.71 | 2.35 | 43.29 | 5.12E-44 |
| 6 | SG00000160712 | 281.18 | 364.08 | 3707.17 | 2641.75 | 0 | # | # | 1748.54 | 3.24 | 3.30 | 38.84 | 1.44E-39 |
| 7 | SG00000219159 | 28.94 | 36.22 | 1003.18 | 658.28 | 0 | # | # | 431.66 | 2.64 | 4.67 | 37.87 | 1.36E-38 |
| 8 | SG00000147872 | 7719.95 | 9293.25 | 39714.57 | 36019.13 | 0 | # | # | 23186.72 | 4.37 | 2.15 | 35.34 | 4.53E-36 |
| 9 | SG00000176171 | 1739.78 | 2004.28 | 13394.89 | 9661.99 | 0 | # | # | 6700.23 | 3.83 | 2.62 | 33.45 | 3.51E-34 |
| 10 | SG00000160181 | 22491.03 | 26846.02 | 134651.21 | 107204.09 | 0 | # | # | 72798.09 | 4.86 | 2.29 | 32.45 | 3.56E-33 |
| 11 | SG00000140479 | 397.99 | 770.88 | 6663.13 | 6758.43 | 0 | # | # | 3647.61 | 3.56 | 3.52 | 32.07 | 8.46E-33 |
| 12 | SG00000002933 | 580.96 | 756.02 | 5444.64 | 3987.40 | 0 | # | # | 2692.25 | 3.43 | 2.82 | 31.68 | 2.08E-32 |
| 13 | SG00000173531 | 499.30 | 489.46 | 4348.38 | 2904.63 | 0 | # | # | 2060.44 | 3.31 | 2.88 | 31.32 | 4.75E-32 |
| 14 | SG00000117983 | 407.29 | 297.21 | 2216.03 | 2211.87 | 0 | # | # | 1283.10 | 3.11 | 2.65 | 31.02 | 9.50E-32 |
| 15 | SG00000087085 | 472.42 | 543.33 | 2231.07 | 2412.26 | 0 | # | # | 1414.77 | 3.15 | 2.19 | 29.17 | 6.77E-30 |
| 16 | SG00000143847 | 6.20 | 9.29 | 1179.00 | 450.35 | 1 | # | # | 411.21 | 2.61 | 6.71 | 28.96 | 1.10E-29 |
| 17 | SG00000107159 | 54.79 | 41.79 | 622.41 | 541.92 | 0 | # | # | 315.23 | 2.50 | 3.60 | 28.08 | 8.29E-29 |
| 18 | SG00000160182 | 16337.19 | 16658.37 | 51170.79 | 47842.35 | 0 | # | # | 33002.18 | 4.52 | 1.59 | 27.42 | 3.83E-28 |
| 19 | SG00000081041 | 1673.62 | 1632.77 | 5658.06 | 6691.64 | 0 | # | # | 3914.02 | 3.59 | 1.90 | 26.84 | 1.46E-27 |
| 20 | SG00000177606 | 3690.44 | 3052.86 | 20381.44 | 37655.67 | 0 | # | # | 16195.10 | 4.21 | 3.11 | 26.34 | 4.61E-27 |
| 21 | SG00000122877 | 8.27 | 1.86 | 359.15 | 639.97 | 1 | # | # | 252.31 | 2.40 | 6.66 | 24.40 | 3.96E-25 |

# How to chose important genes?



| | avg (treatment) | avg (control) | A | M |
|---|---|---|---|---|
| | | | average | Foldchange (treatment/control) |
| | 2 | 8 | 5 | 4 |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| **Gene A** | 0.01 | 1 | | 100 |
| | | | | |
| **Gene B** | 10 | 1000 | | 100 |

# How to chose important genes?



|  |  |  | A | M |
|---|---|---|---|---|
|  | avg (treatment) | avg (control) | average | Foldchange (treatment/control) |
|  | 2 | 8 | 5 | 4 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| **Gene A** | 0.01 | 1 | 0.5 | 100 |
|  |  |  |  |  |
| **Gene B** | 10 | 1000 | 500 | 100 |

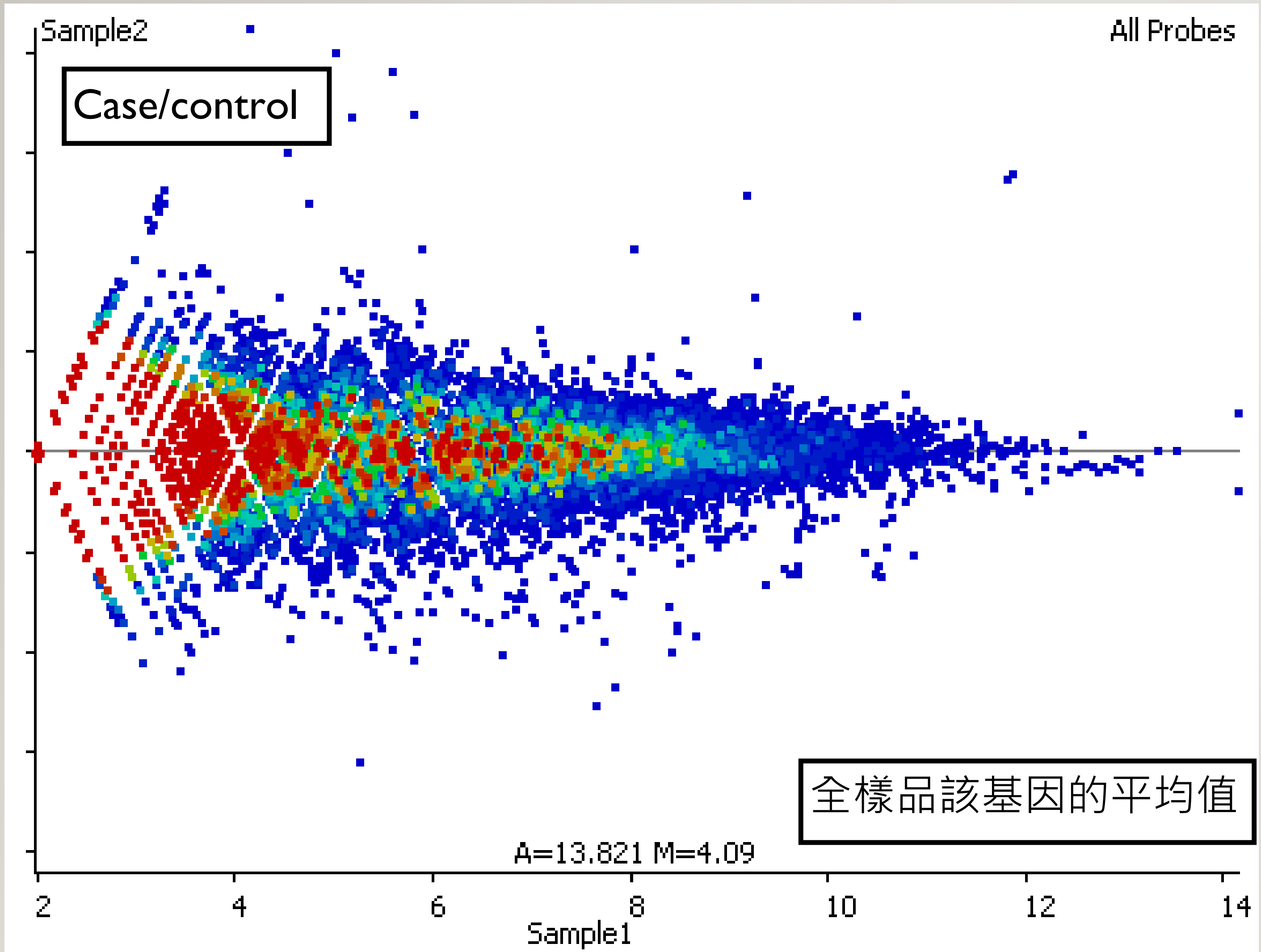
# MA plot

ratio average (RA) plot
Bland–Altman plot

MA PLOT 展示兩組間 "基因表現量" 和 "FOLD CHANGE" 的分佈，

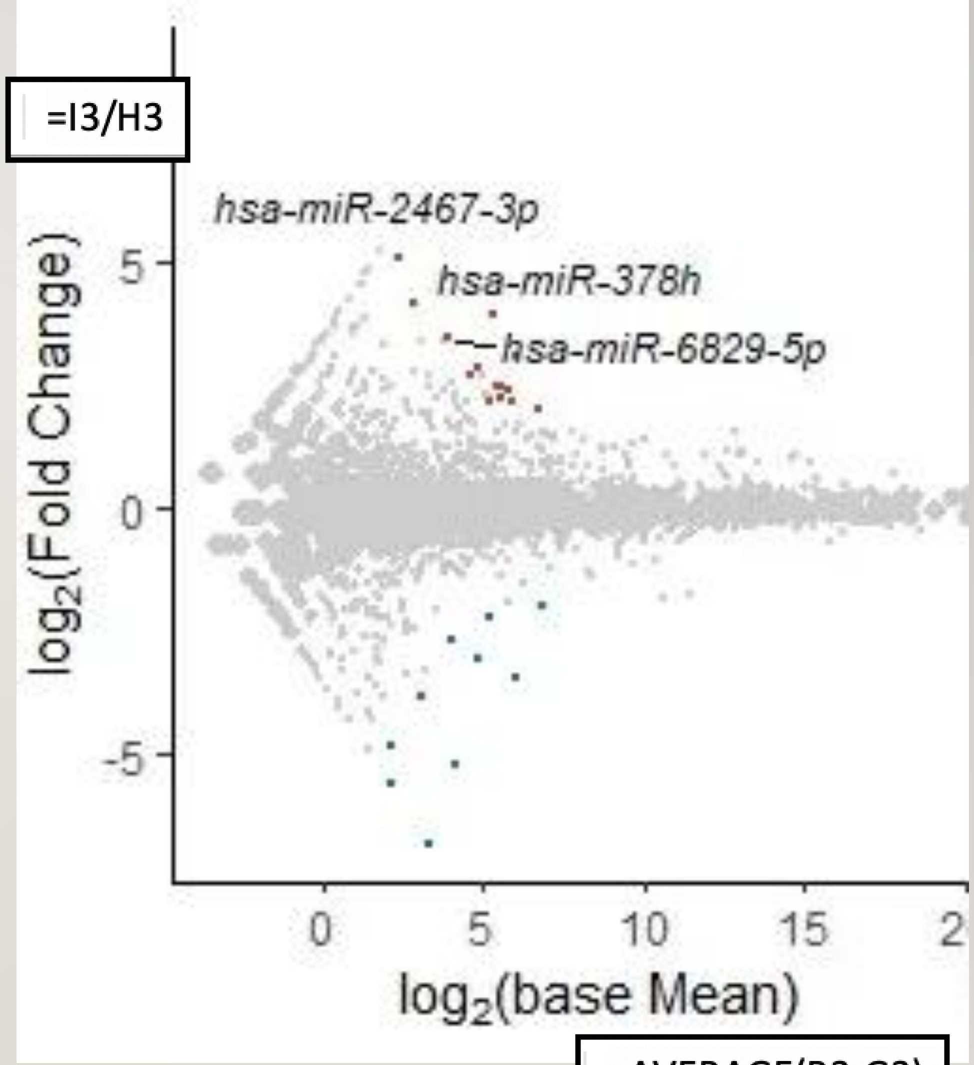
$$M = \log_2(R/G) = \log_2(R) - \log_2(G)$$

$$A = \frac{1}{2}\log_2(RG) = \frac{1}{2}(\log_2(R) + \log_2(G))$$



Case/control

全樣品該基因的平均值



=I3/H3

=AVERAGE(B3:G3)

# Volcano plot

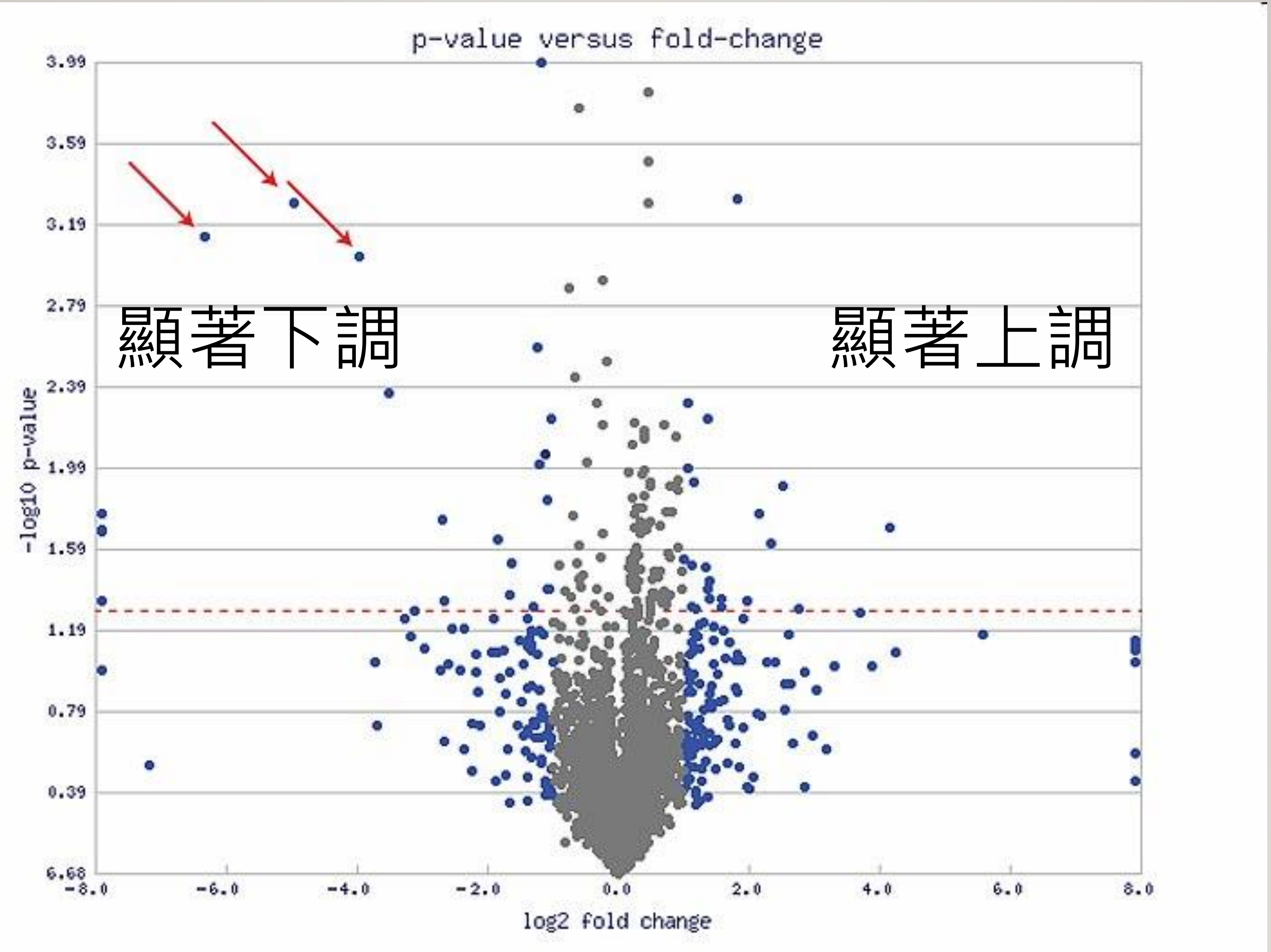
火山圖將統計檢定的統計顯著性測量（P 值）與變化的幅度(FC)結合起來，

$$FC = \log_2(R/G) = \log_2(R) - \log_2(G)$$

$$P = -log\,(p\text{-value})$$

$p1=0.01 \Rightarrow 2$

$p2=0.0001 \Rightarrow 4$


p-value versus fold-change
-log10 p-value
3.99
3.59
3.19
2.79
2.39
1.99
1.59
1.19
0.79
0.39
6.68
-8.0
-6.0
-4.0
-2.0
0.0
2.0
4.0
6.0
8.0
log2 fold change

顯著下調

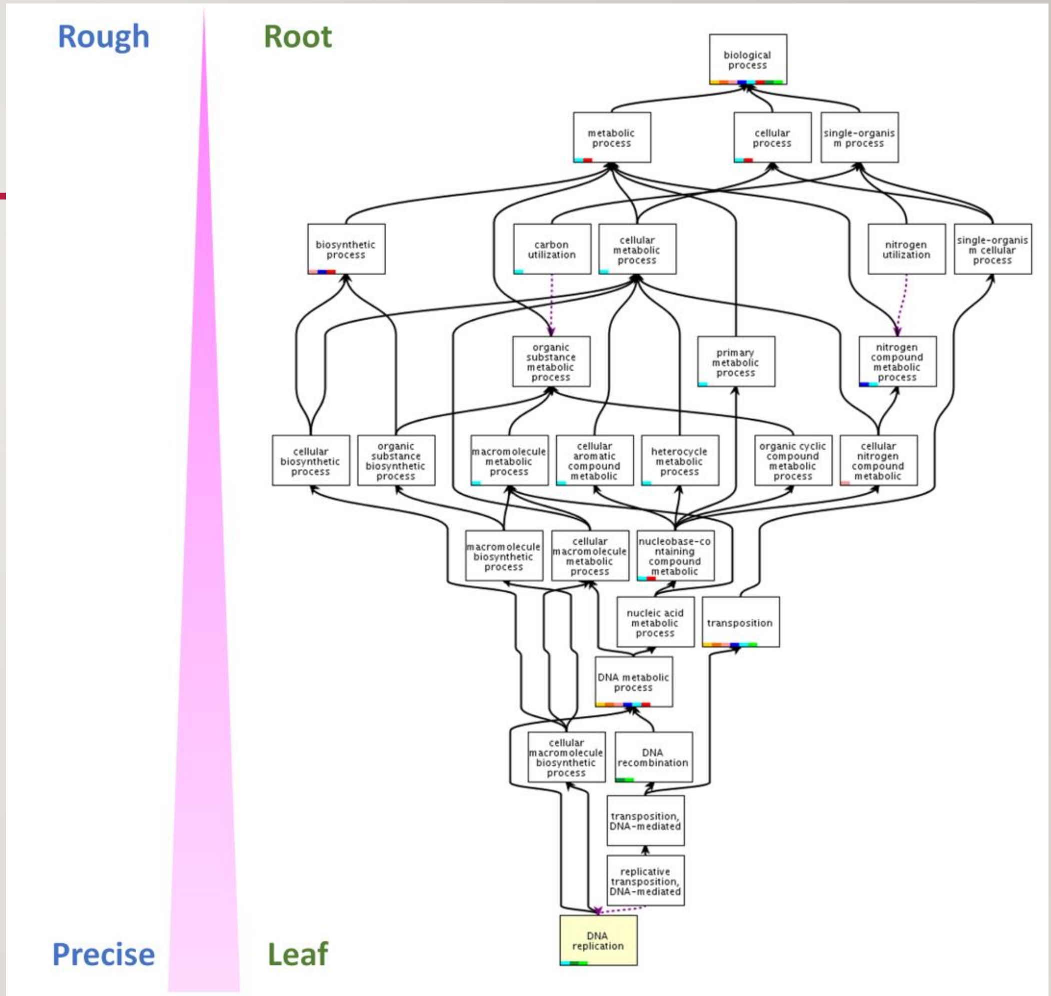
顯著上調

# 如何解釋差異？

資料拓墣學

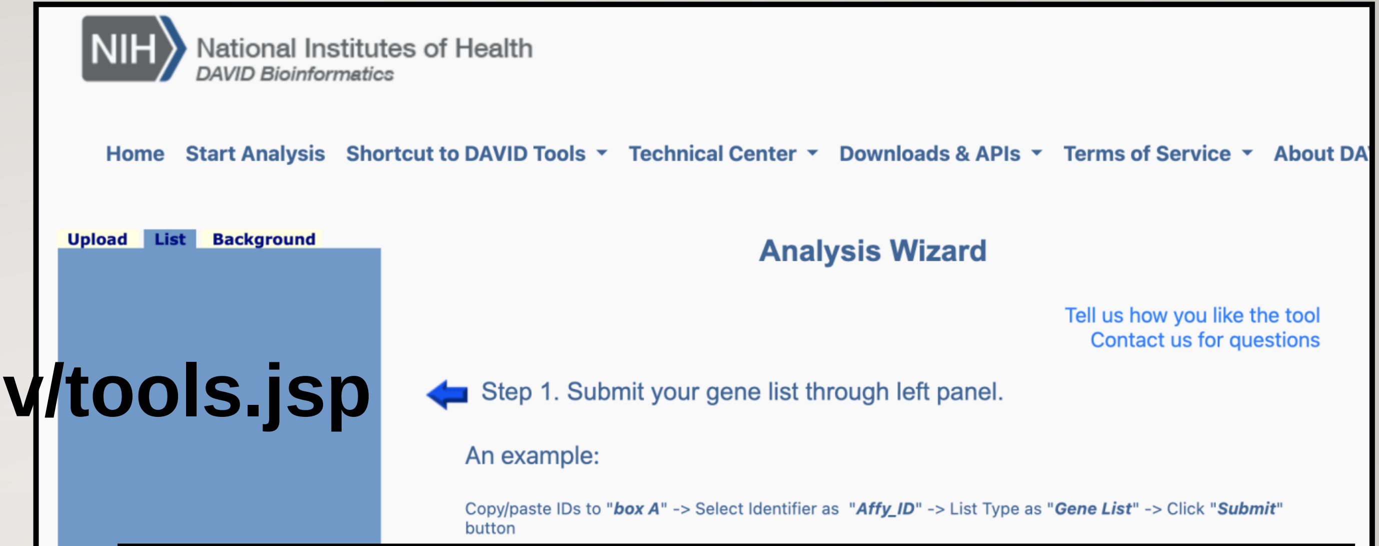
Gene ontology (GO)

Geneset enrichment


Rough
Root
biological process
metabolic process
cellular process
single-organism process
biosynthetic process
carbon utilization
cellular metabolic process
nitrogen utilization
single-organism cellular process
organic substance metabolic process
primary metabolic process
nitrogen compound metabolic process
cellular biosynthetic process
organic substance biosynthetic process
macromolecule metabolic process
cellular aromatic compound metabolic
heterocycle metabolic process
organic cyclic compound metabolic process
cellular nitrogen compound metabolic
macromolecule biosynthetic process
cellular macromolecule metabolic process
nucleobase-containing compound metabolic
nucleic acid metabolic process
transposition
DNA metabolic process
cellular macromolecule biosynthetic process
DNA recombination
transposition, DNA-mediated
replicative transposition, DNA-mediated
DNA replication
Precise
Leaf

# DAVID
## 基因知識拓樸學

https://david.ncifcrf.gov/tools.jsp


NIH National Institutes of Health
DAVID Bioinformatics
Home
Start Analysis
Shortcut to DAVID Tools
Technical Center
Downloads & APIs
Terms of Service
About DA
Upload
List
Background

Analysis Wizard

Tell us how you like the tool
Contact us for questions

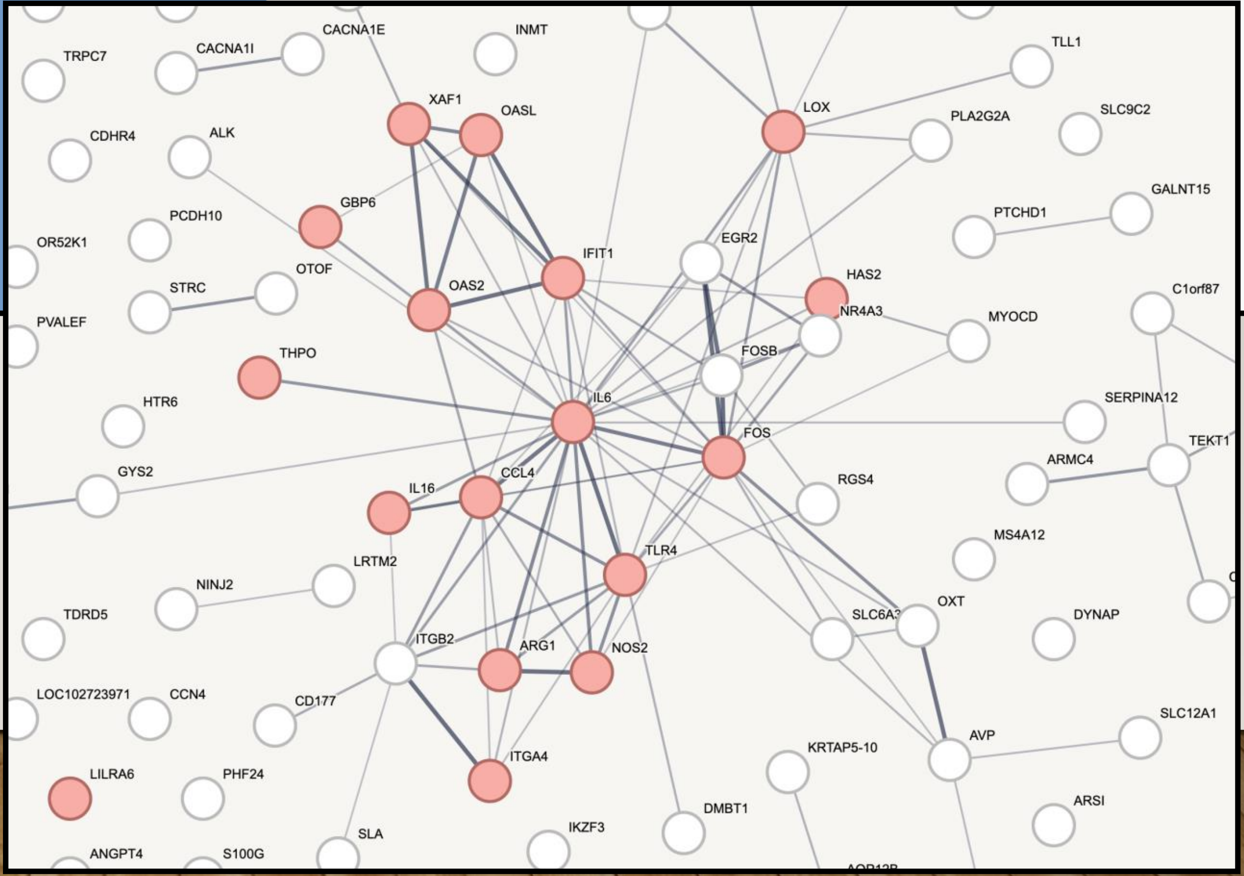
Step 1. Submit your gene list through left panel.

An example:

Copy/paste IDs to "***box A***" -> Select Identifier as "***Affy_ID***" -> List Type as "***Gene List***" -> Click "***Submit***" button

# String
## 基因交互作用

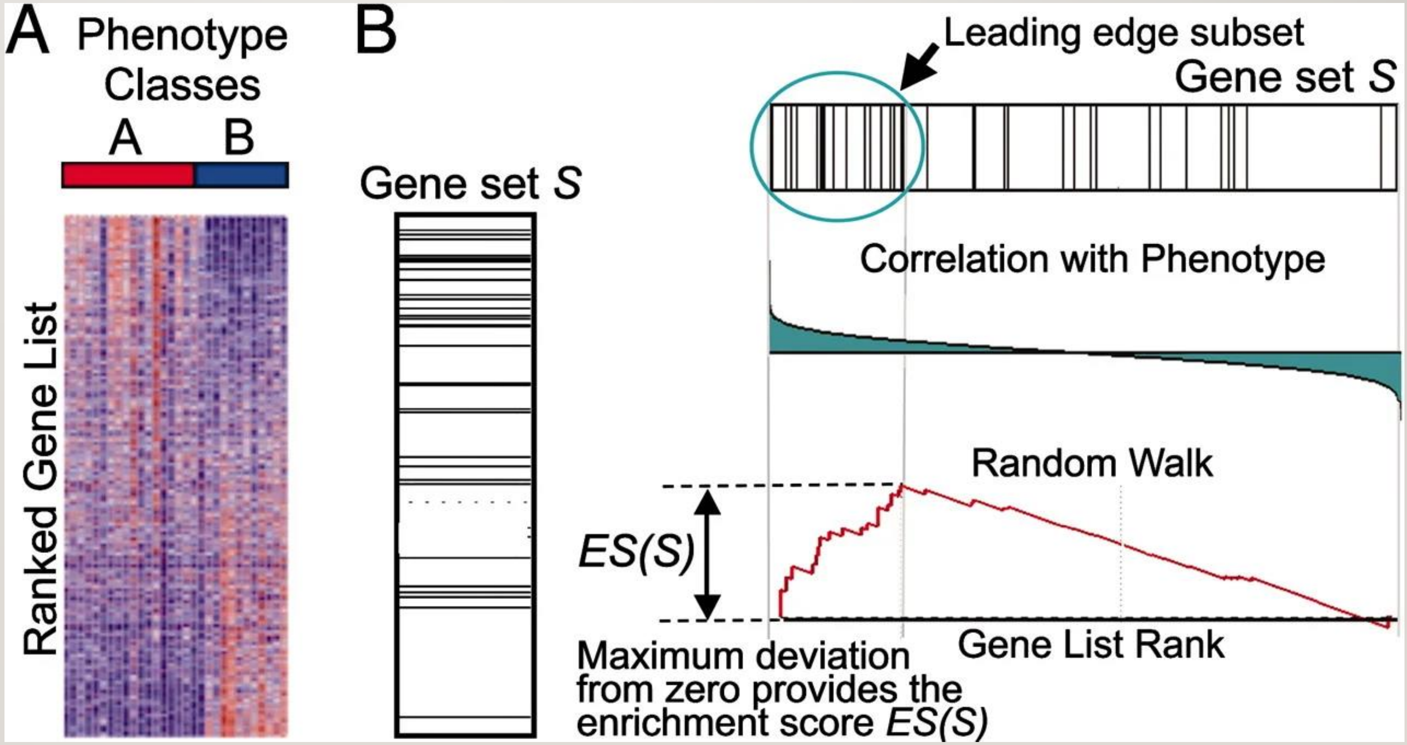
https://string-db.org

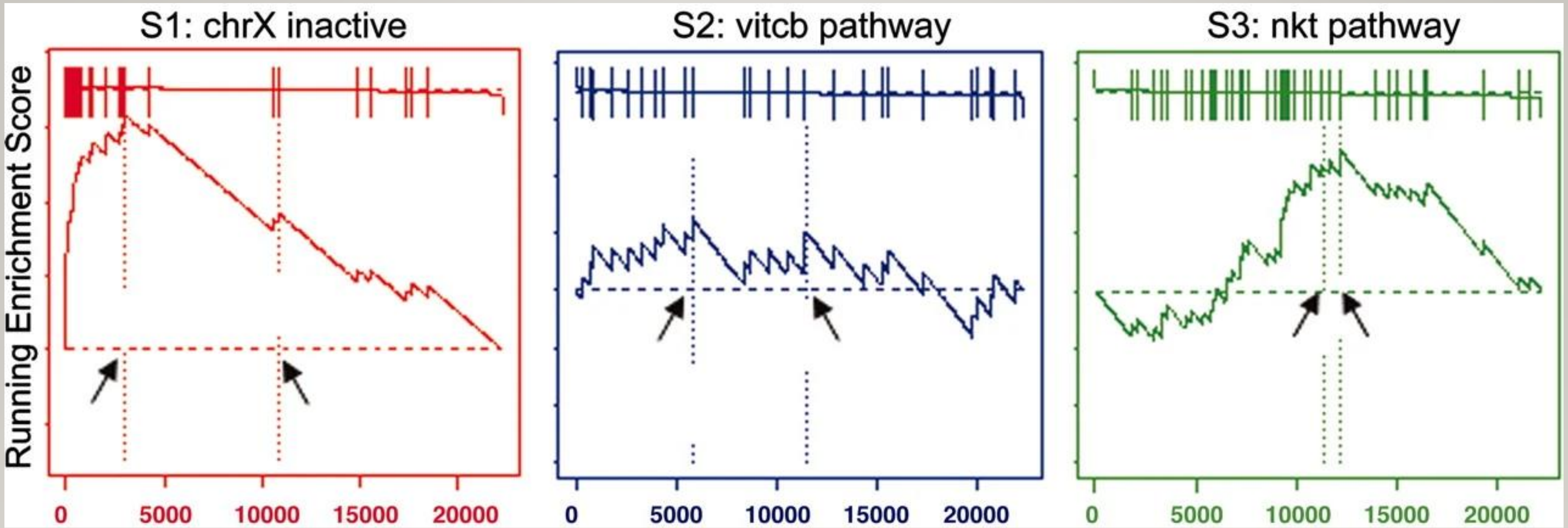

TRPC7
CACNA1I
CACNA1E
INMT
TLL1
CDHR4
ALK
XAF1
OASL
LOX
PLA2G2A
SLC9C2
GBP6
PCDH10
GALNT15
PTCHD1
OR52K1
EGR2
IFIT1
HAS2
OTOF
STRC
OAS2
NR4A3
C1orf87
PVALEF
MYOCD
FOSB
THPO
IL6
SERPINA12
HTR6
FOS
TEKT1
ARMC4
GYS2
CCL4
IL16
RGS4
MS4A12
LRTM2
TLR4
NINJ2
OXT
TDRD5
SLC6A
DYNAP
ITGB2
ARG1
NOS2
LOC102723971
CCN4
CD177
SLC12A1
AVP
KRTAP5-10
ITGA4
LILRA6
PHF24
ARSI
DMBT1
IKZF3
SLA
ANGPT4
S100G

# 可以不要只看很明顯的現象嗎？

**GSEA**

**https://www.gsea-msigdb.org/gsea/index.jsp**


A
Phenotype Classes
A
B
Ranked Gene List
B
Gene set S
Leading edge subset
Gene set S
Correlation with Phenotype
Random Walk
ES(S)
Gene List Rank
Maximum deviation from zero provides the enrichment score ES(S)

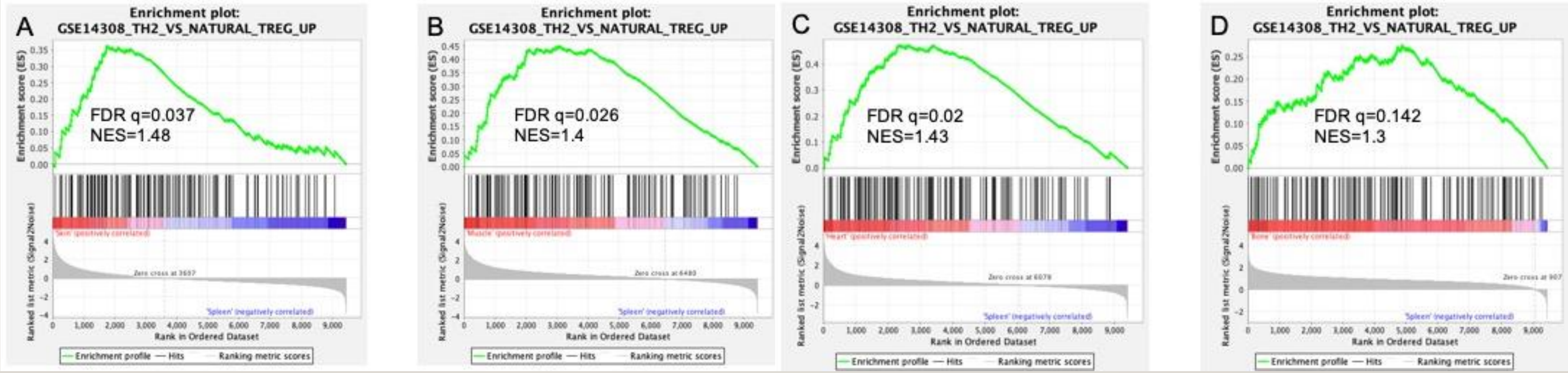


X軸通常表示基因強度序位，通常是按照它們在基因表現量中的重要性·

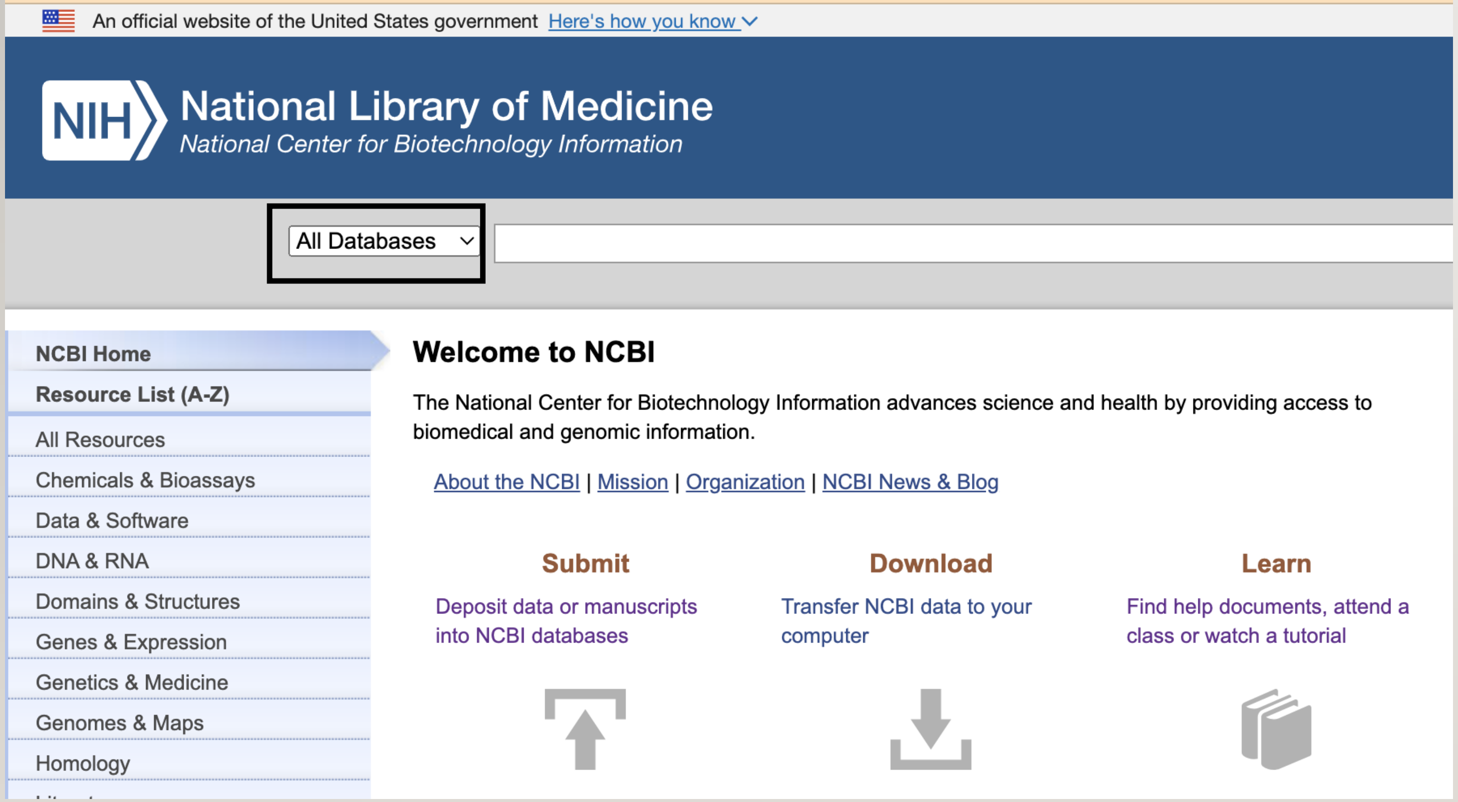
Y軸通常表示基因集合中的基因與該生物功能條件之間的關係有多强。

- 就愛這一個功能可以嗎？

## GSEA可以



# 憑空可以生資料嗎？


An official website of the United States government Here's how you know
NIH
National Library of Medicine
National Center for Biotechnology Information
All Databases
NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology

**Welcome to NCBI**

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | NCBI News & Blog

**Submit**

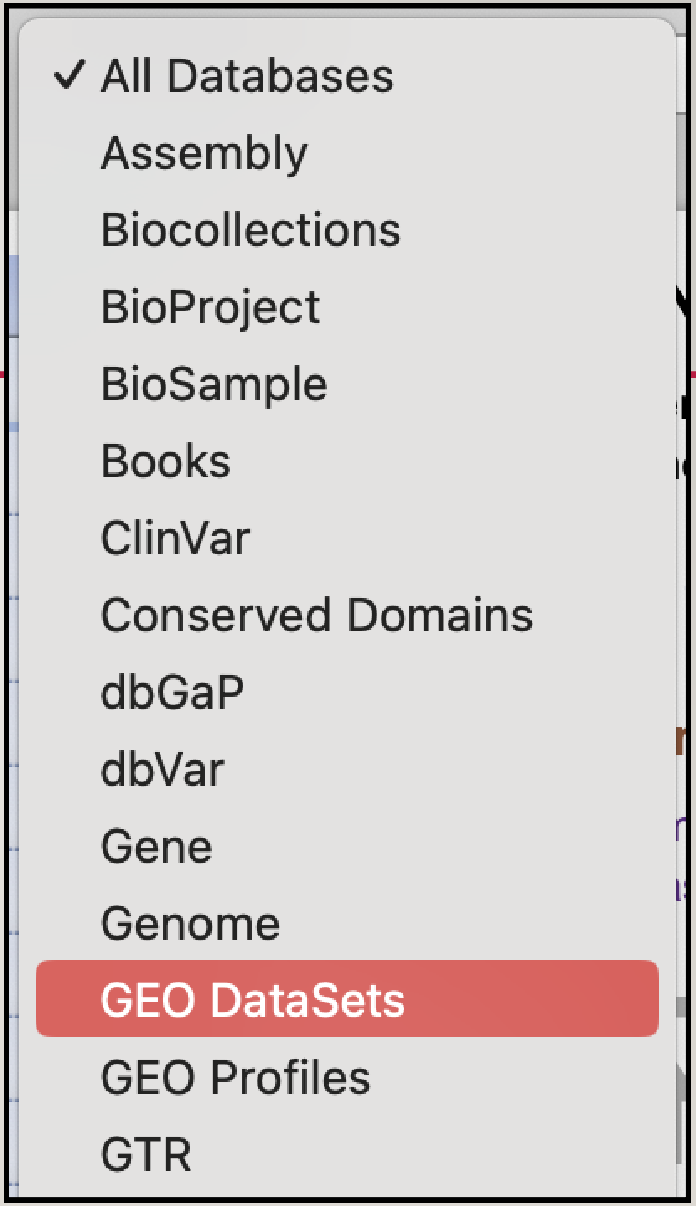
Deposit data or manuscripts into NCBI databases

**Download**

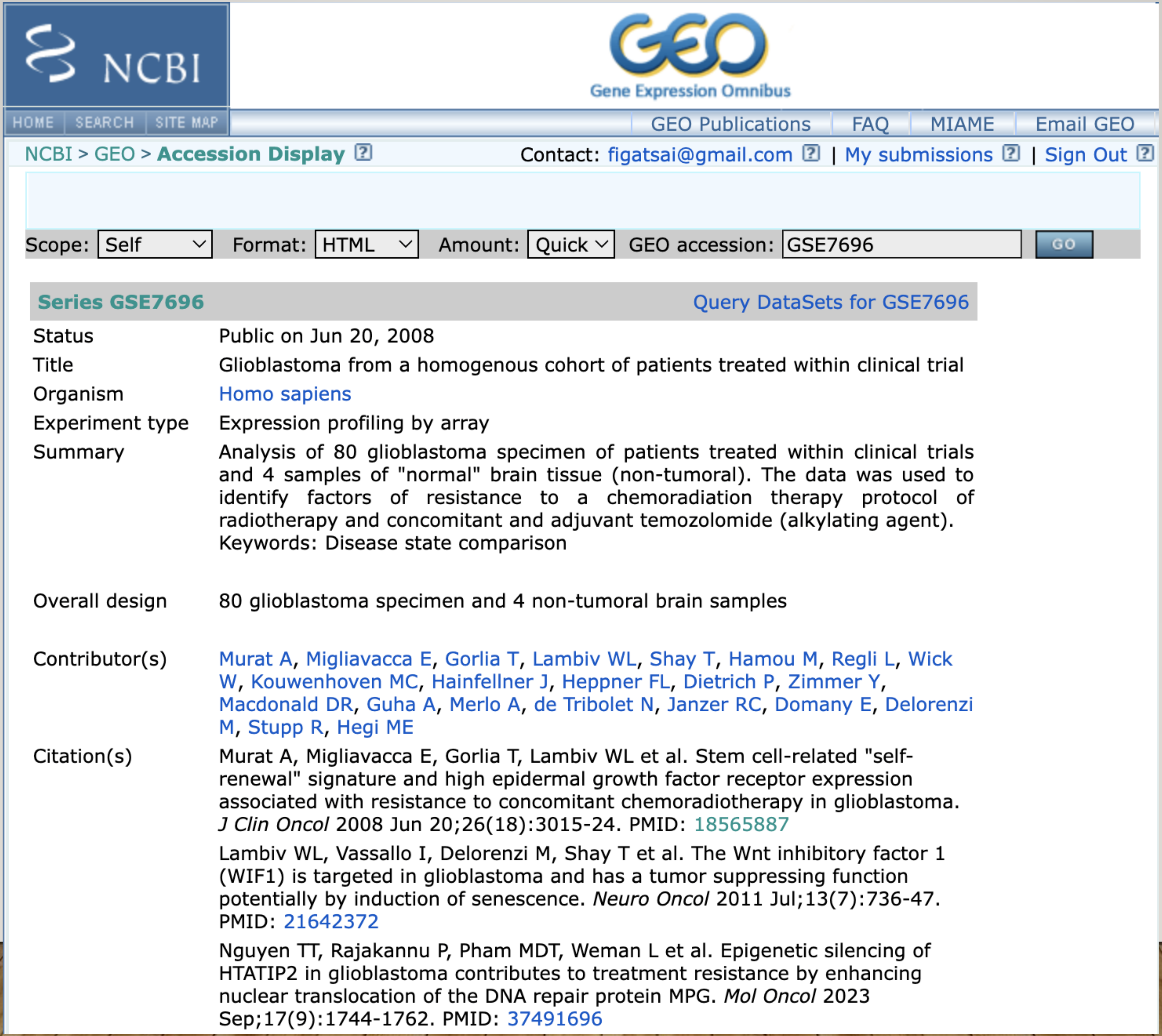
Transfer NCBI data to your computer

**Learn**

Find help documents, attend a class or watch a tutorial


✓ All Databases
Assembly
Biocollections
BioProject
BioSample
Books
ClinVar
Conserved Domains
dbGaP
dbVar
Gene
Genome
GEO DataSets
GEO Profiles
GTR

# 文章發表細節

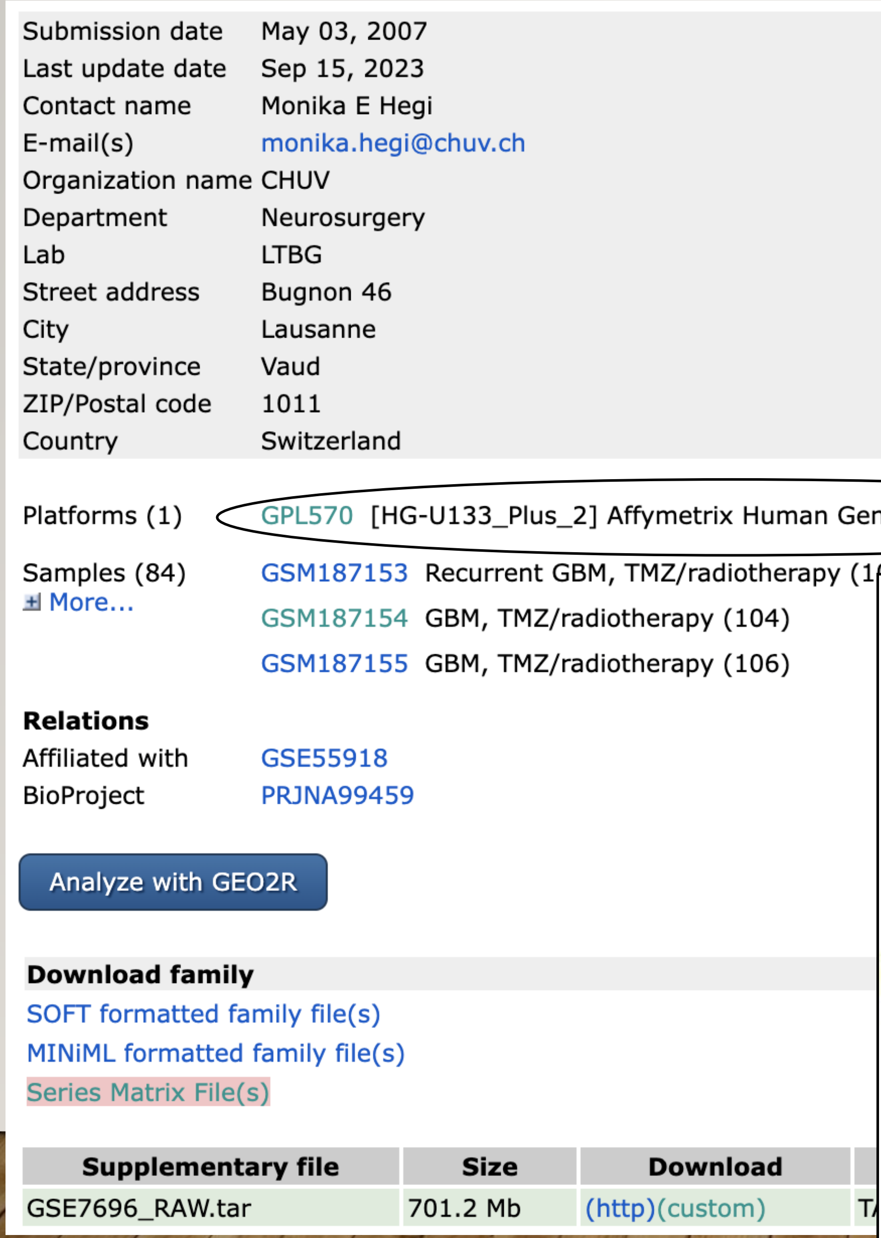


NCBI > GEO > **Accession Display** Contact: figatsai@gmail.com | My submissions | Sign Out

Scope: Self Format: HTML Amount: Quick GEO accession: GSE7696

**Series GSE7696** Query DataSets for GSE7696

| | |
|---|---|
| Status | Public on Jun 20, 2008 |
| Title | Glioblastoma from a homogenous cohort of patients treated within clinical trial |
| Organism | Homo sapiens |
| Experiment type | Expression profiling by array |
| Summary | Analysis of 80 glioblastoma specimen of patients treated within clinical trials and 4 samples of "normal" brain tissue (non-tumoral). The data was used to identify factors of resistance to a chemoradiation therapy protocol of radiotherapy and concomitant and adjuvant temozolomide (alkylating agent).<br>Keywords: Disease state comparison |
| Overall design | 80 glioblastoma specimen and 4 non-tumoral brain samples |
| Contributor(s) | Murat A, Migliavacca E, Gorlia T, Lambiv WL, Shay T, Hamou M, Regli L, Wick W, Kouwenhoven MC, Hainfellner J, Heppner FL, Dietrich P, Zimmer Y, Macdonald DR, Guha A, Merlo A, de Tribolet N, Janzer RC, Domany E, Delorenzi M, Stupp R, Hegi ME |
| Citation(s) | Murat A, Migliavacca E, Gorlia T, Lambiv WL et al. Stem cell-related "self-renewal" signature and high epidermal growth factor receptor expression associated with resistance to concomitant chemoradiotherapy in glioblastoma. *J Clin Oncol* 2008 Jun 20;26(18):3015-24. PMID: 18565887<br>Lambiv WL, Vassallo I, Delorenzi M, Shay T et al. The Wnt inhibitory factor 1 (WIF1) is targeted in glioblastoma and has a tumor suppressing function potentially by induction of senescence. *Neuro Oncol* 2011 Jul;13(7):736-47. PMID: 21642372<br>Nguyen TT, Rajakannu P, Pham MDT, Weman L et al. Epigenetic silencing of HTATIP2 in glioblastoma contributes to treatment resistance by enhancing nuclear translocation of the DNA repair protein MPG. *Mol Oncol* 2023 Sep;17(9):1744-1762. PMID: 37491696 |



| | |
|---|---|
| Submission date | May 03, 2007 |
| Last update date | Sep 15, 2023 |
| Contact name | Monika E Hegi |
| E-mail(s) | monika.hegi@chuv.ch |
| Organization name | CHUV |
| Department | Neurosurgery |
| Lab | LTBG |
| Street address | Bugnon 46 |
| City | Lausanne |
| State/province | Vaud |
| ZIP/Postal code | 1011 |
| Country | Switzerland |

Platforms (1) GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array

Samples (84) GSM187153 Recurrent GBM, TMZ/radiotherapy (100R)
More...
GSM187154 GBM, TMZ/radiotherapy (104)
GSM187155 GBM, TMZ/radiotherapy (106)

**Relations**

Affiliated with GSE55918
BioProject PRJNA99459

Analyze with GEO2R

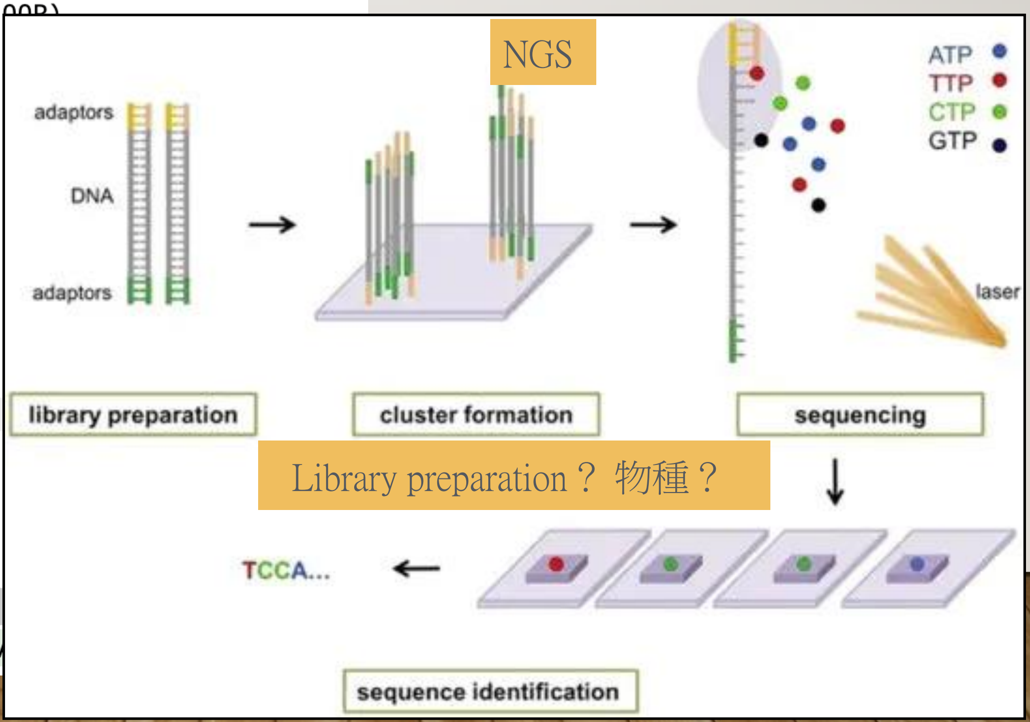
**Download family**

SOFT formatted family file(s)
MINiML formatted family file(s)
Series Matrix File(s)

| Supplementary file | Size | Download |
|---|---|---|
| GSE7696_RAW.tar | 701.2 Mb | (http)(custom) |

Microarray

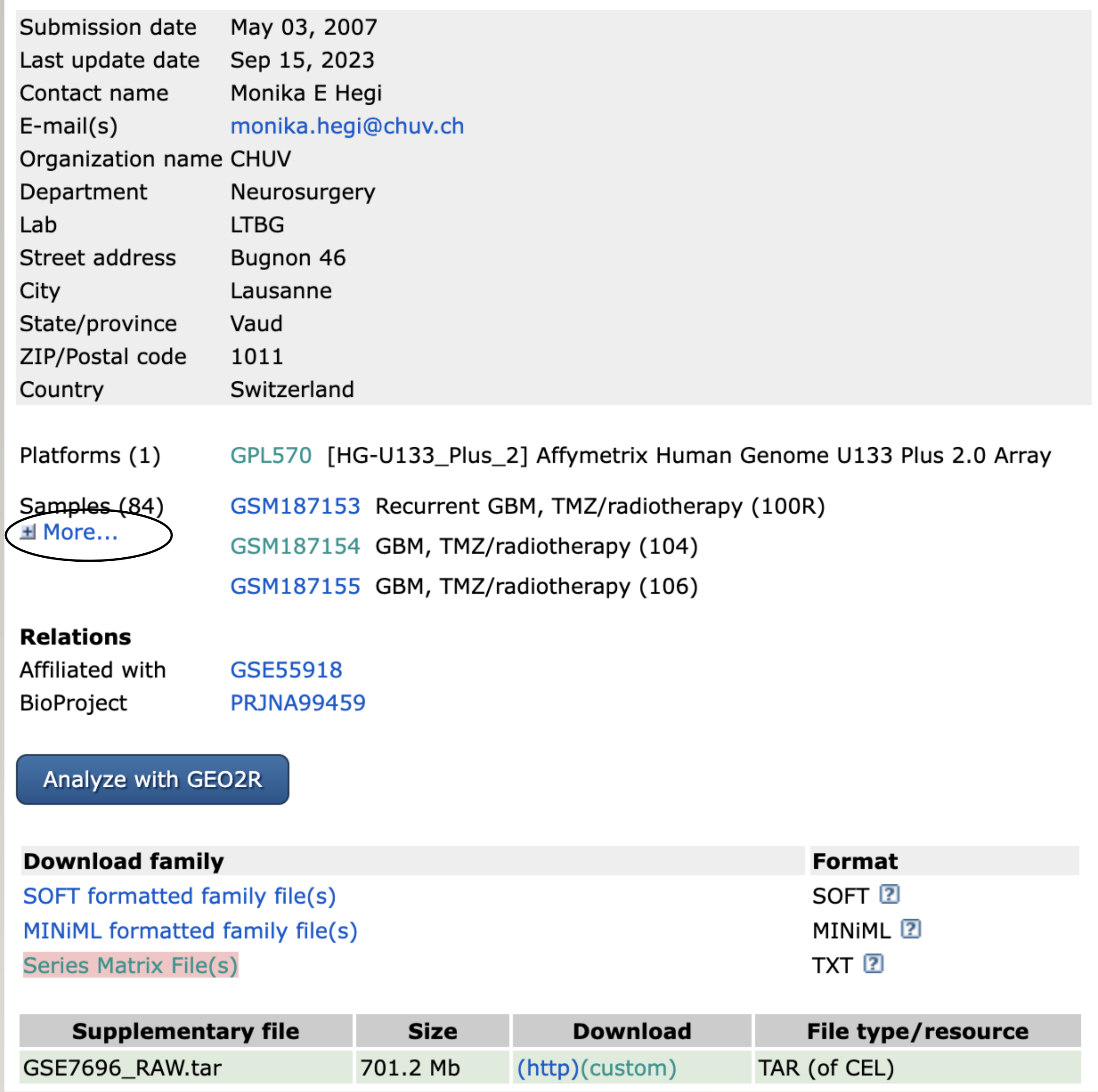
廠牌？ 型號？ 物種？



NGS

adaptors
DNA
adaptors
ATP
TTP
CTP
GTP
laser
library preparation
cluster formation
sequencing
TCCA...
sequence identification

Library preparation？ 物種？



| | |
|---|---|
| Submission date | May 03, 2007 |
| Last update date | Sep 15, 2023 |
| Contact name | Monika E Hegi |
| E-mail(s) | monika.hegi@chuv.ch |
| Organization name | CHUV |
| Department | Neurosurgery |
| Lab | LTBG |
| Street address | Bugnon 46 |
| City | Lausanne |
| State/province | Vaud |
| ZIP/Postal code | 1011 |
| Country | Switzerland |

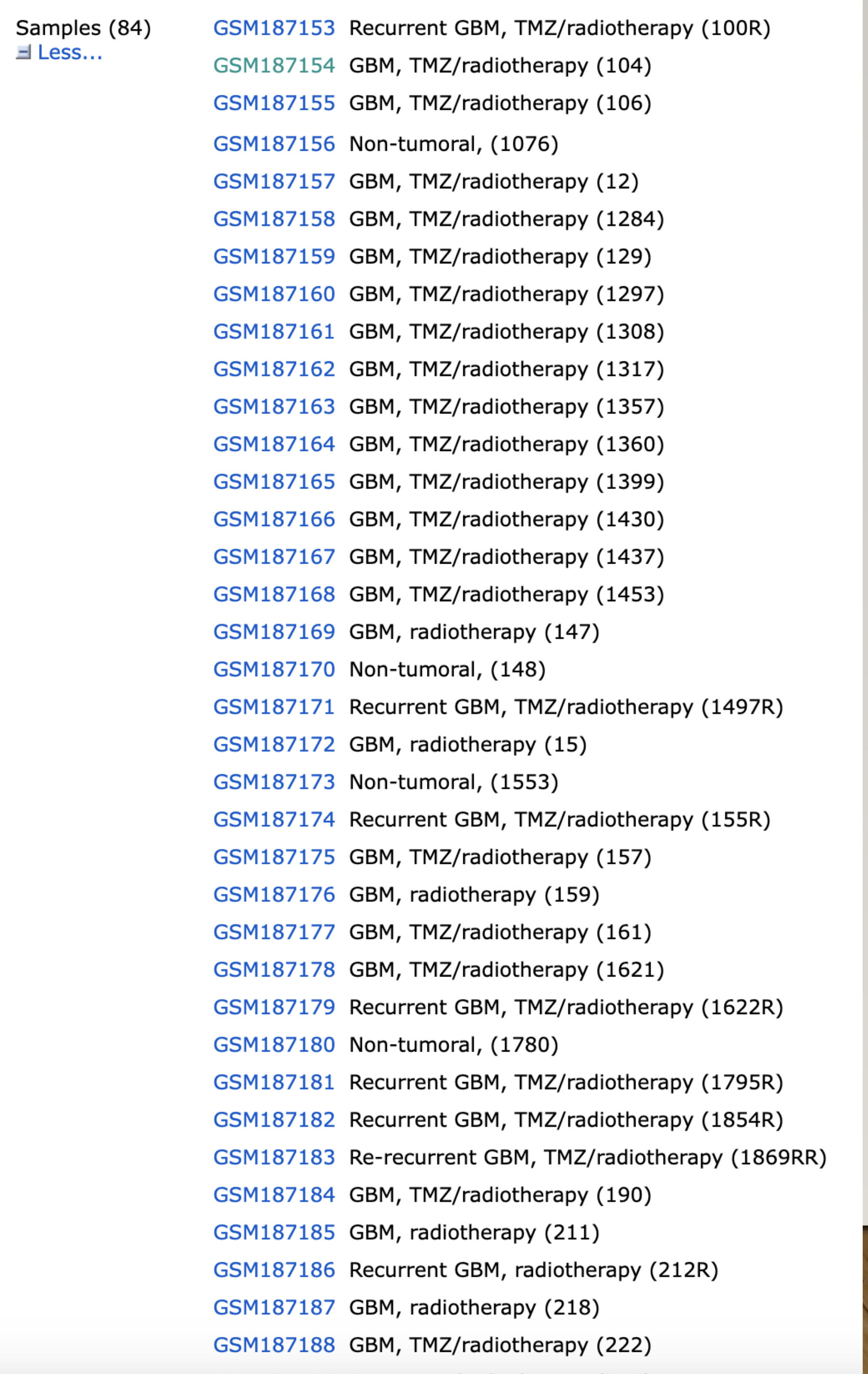
| | | |
|---|---|---|
| Platforms (1) | GPL570 | [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array |
| Samples (84)<br>More... | GSM187153 | Recurrent GBM, TMZ/radiotherapy (100R) |
| | GSM187154 | GBM, TMZ/radiotherapy (104) |
| | GSM187155 | GBM, TMZ/radiotherapy (106) |

**Relations**

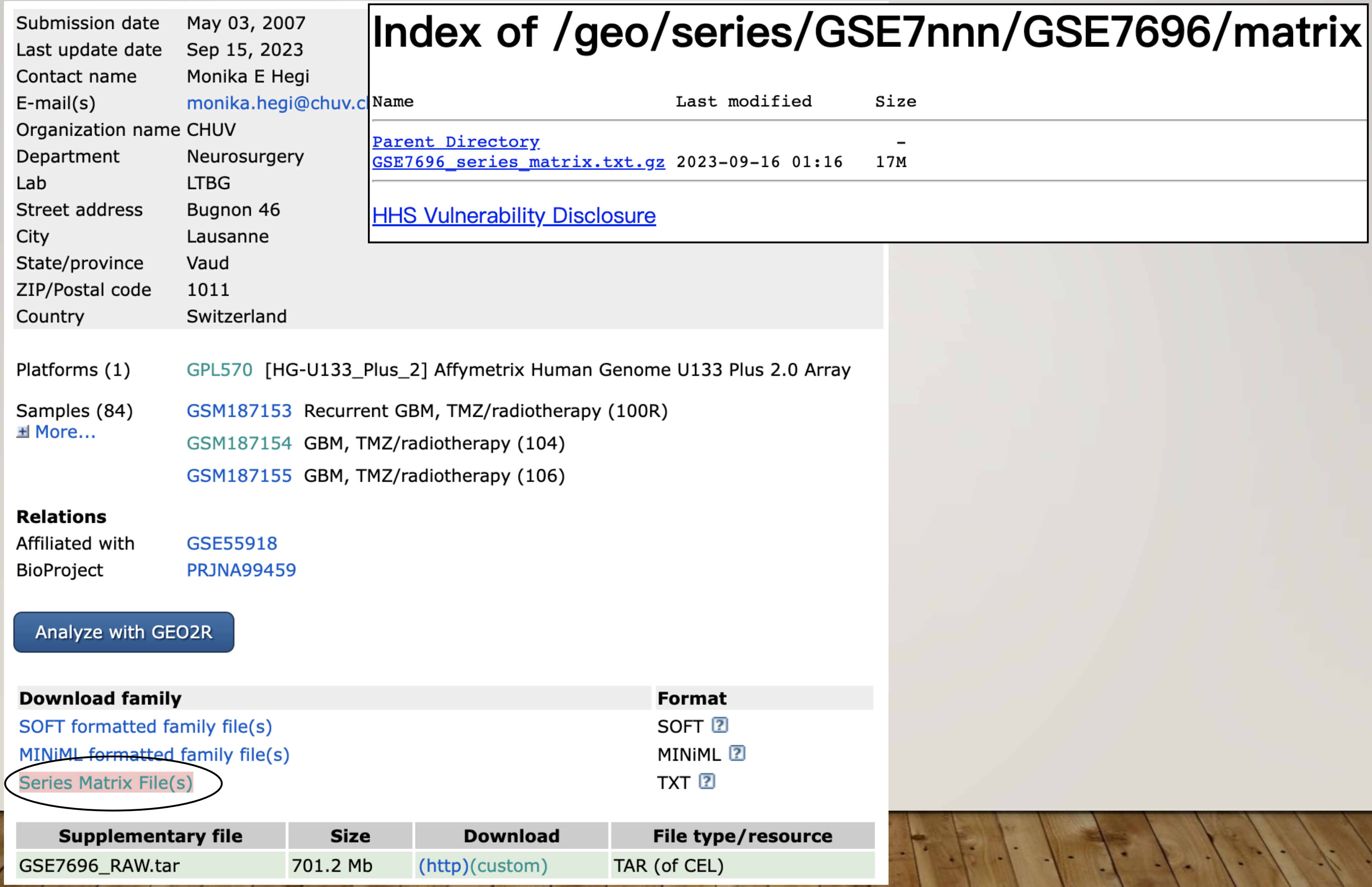
| | |
|---|---|
| Affiliated with | GSE55918 |
| BioProject | PRJNA99459 |

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

| Supplementary file | Size | Download | File type/resource |
|---|---|---|---|
| GSE7696_RAW.tar | 701.2 Mb | (http)(custom) | TAR (of CEL) |



| | | |
|---|---|---|
| Samples (84)<br>Less... | GSM187153 | Recurrent GBM, TMZ/radiotherapy (100R) |
| | GSM187154 | GBM, TMZ/radiotherapy (104) |
| | GSM187155 | GBM, TMZ/radiotherapy (106) |
| | GSM187156 | Non-tumoral, (1076) |
| | GSM187157 | GBM, TMZ/radiotherapy (12) |
| | GSM187158 | GBM, TMZ/radiotherapy (1284) |
| | GSM187159 | GBM, TMZ/radiotherapy (129) |
| | GSM187160 | GBM, TMZ/radiotherapy (1297) |
| | GSM187161 | GBM, TMZ/radiotherapy (1308) |
| | GSM187162 | GBM, TMZ/radiotherapy (1317) |
| | GSM187163 | GBM, TMZ/radiotherapy (1357) |
| | GSM187164 | GBM, TMZ/radiotherapy (1360) |
| | GSM187165 | GBM, TMZ/radiotherapy (1399) |
| | GSM187166 | GBM, TMZ/radiotherapy (1430) |
| | GSM187167 | GBM, TMZ/radiotherapy (1437) |
| | GSM187168 | GBM, TMZ/radiotherapy (1453) |
| | GSM187169 | GBM, radiotherapy (147) |
| | GSM187170 | Non-tumoral, (148) |
| | GSM187171 | Recurrent GBM, TMZ/radiotherapy (1497R) |
| | GSM187172 | GBM, radiotherapy (15) |
| | GSM187173 | Non-tumoral, (1553) |
| | GSM187174 | Recurrent GBM, TMZ/radiotherapy (155R) |
| | GSM187175 | GBM, TMZ/radiotherapy (157) |
| | GSM187176 | GBM, radiotherapy (159) |
| | GSM187177 | GBM, TMZ/radiotherapy (161) |
| | GSM187178 | GBM, TMZ/radiotherapy (1621) |
| | GSM187179 | Recurrent GBM, TMZ/radiotherapy (1622R) |
| | GSM187180 | Non-tumoral, (1780) |
| | GSM187181 | Recurrent GBM, TMZ/radiotherapy (1795R) |
| | GSM187182 | Recurrent GBM, TMZ/radiotherapy (1854R) |
| | GSM187183 | Re-recurrent GBM, TMZ/radiotherapy (1869RR) |
| | GSM187184 | GBM, TMZ/radiotherapy (190) |
| | GSM187185 | GBM, radiotherapy (211) |
| | GSM187186 | Recurrent GBM, radiotherapy (212R) |
| | GSM187187 | GBM, radiotherapy (218) |
| | GSM187188 | GBM, TMZ/radiotherapy (222) |



| | |
|---|---|
| Submission date | May 03, 2007 |
| Last update date | Sep 15, 2023 |
| Contact name | Monika E Hegi |
| E-mail(s) | monika.hegi@chuv.c |
| Organization name | CHUV |
| Department | Neurosurgery |
| Lab | LTBG |
| Street address | Bugnon 46 |
| City | Lausanne |
| State/province | Vaud |
| ZIP/Postal code | 1011 |
| Country | Switzerland |

| | |
|---|---|
| Platforms (1) | GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array |
| Samples (84)<br>⊞ More... | GSM187153 Recurrent GBM, TMZ/radiotherapy (100R)<br>GSM187154 GBM, TMZ/radiotherapy (104)<br>GSM187155 GBM, TMZ/radiotherapy (106) |

**Relations**

| | |
|---|---|
| Affiliated with | GSE55918 |
| BioProject | PRJNA99459 |

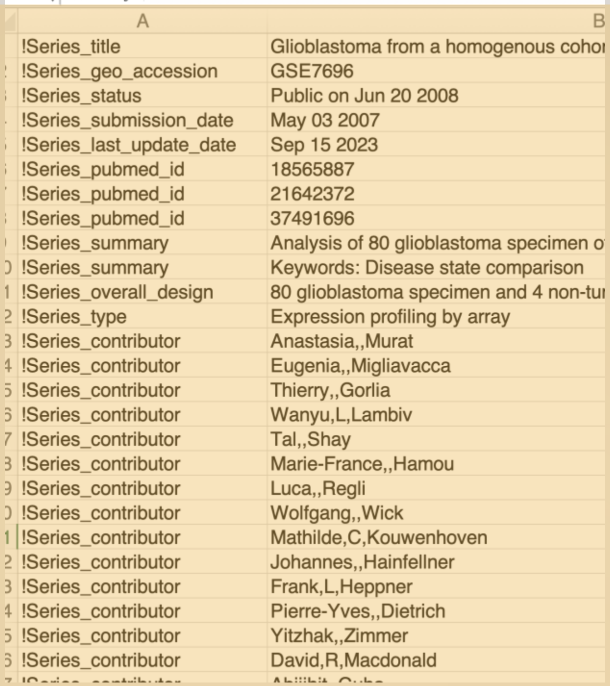
| **Download family** | **Format** |
|---|---|
| SOFT formatted family file(s) | SOFT ? |
| MINiML formatted family file(s) | MINiML ? |
| Series Matrix File(s) | TXT ? |

| Supplementary file | Size | Download | File type/resource |
|---|---|---|---|
| GSE7696_RAW.tar | 701.2 Mb | (http)(custom) | TAR (of CEL) |

# Index of /geo/series/GSE7nnn/GSE7696/matrix

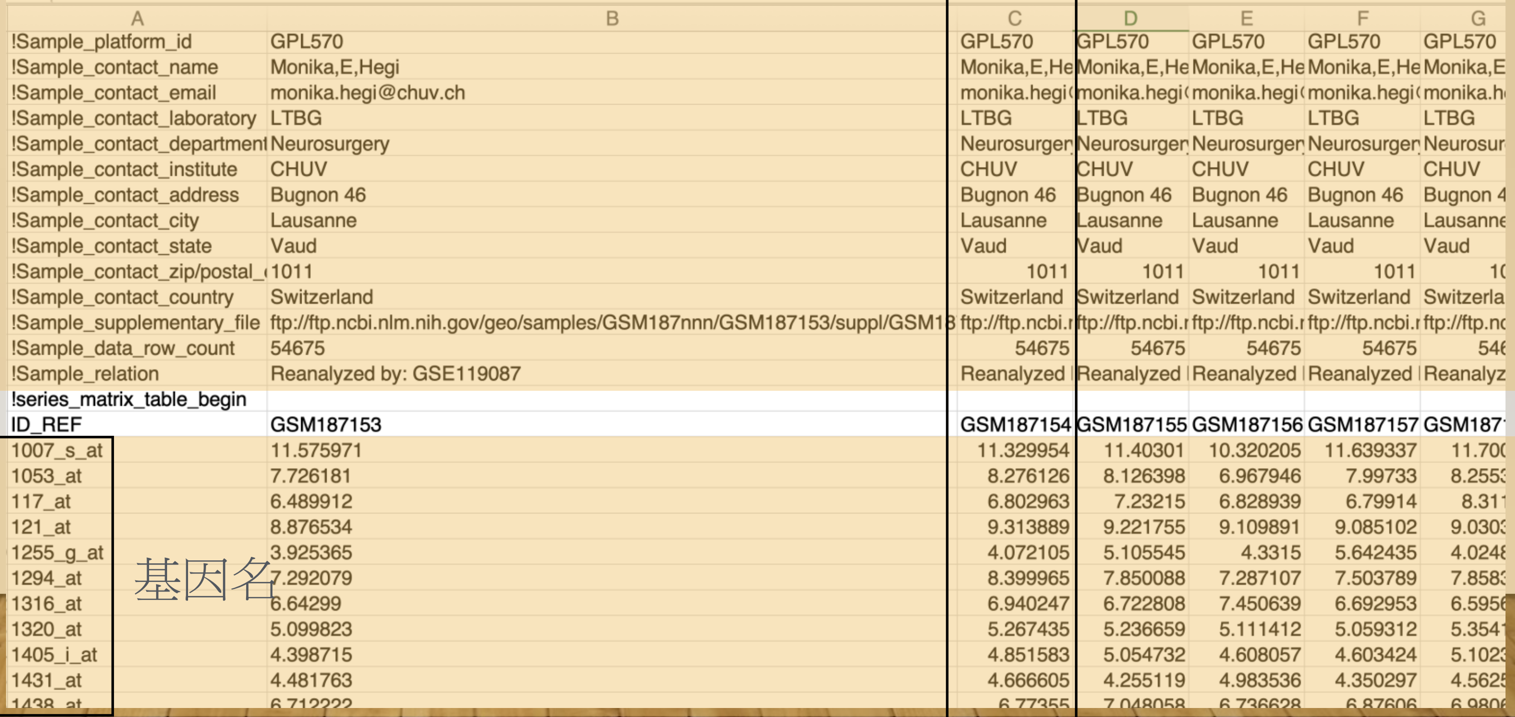
```
Name                          Last modified      Size

Parent Directory                                   -
GSE7696_series_matrix.txt.gz  2023-09-16 01:16   17M
```

HHS Vulnerability Disclosure



| A | B |
|---|---|
| !Series_title | Glioblastoma from a homogenous cohoı |
| !Series_geo_accession | GSE7696 |
| !Series_status | Public on Jun 20 2008 |
| !Series_submission_date | May 03 2007 |
| !Series_last_update_date | Sep 15 2023 |
| !Series_pubmed_id | 18565887 |
| !Series_pubmed_id | 21642372 |
| !Series_pubmed_id | 37491696 |
| !Series_summary | Analysis of 80 glioblastoma specimen o |
| !Series_summary | Keywords: Disease state comparison |
| !Series_overall_design | 80 glioblastoma specimen and 4 non-tu |
| !Series_type | Expression profiling by array |
| !Series_contributor | Anastasia,,Murat |
| !Series_contributor | Eugenia,,Migliavacca |
| !Series_contributor | Thierry,,Gorlia |
| !Series_contributor | Wanyu,L,Lambiv |
| !Series_contributor | Tal,,Shay |
| !Series_contributor | Marie-France,,Hamou |
| !Series_contributor | Luca,,Regli |
| !Series_contributor | Wolfgang,,Wick |
| !Series_contributor | Mathilde,C,Kouwenhoven |
| !Series_contributor | Johannes,,Hainfellner |
| !Series_contributor | Frank,L,Heppner |
| !Series_contributor | Pierre-Yves,,Dietrich |
| !Series_contributor | Yitzhak,,Zimmer |
| !Series_contributor | David,R,Macdonald |
| !Series_contributor | Abhijit,,Guha |

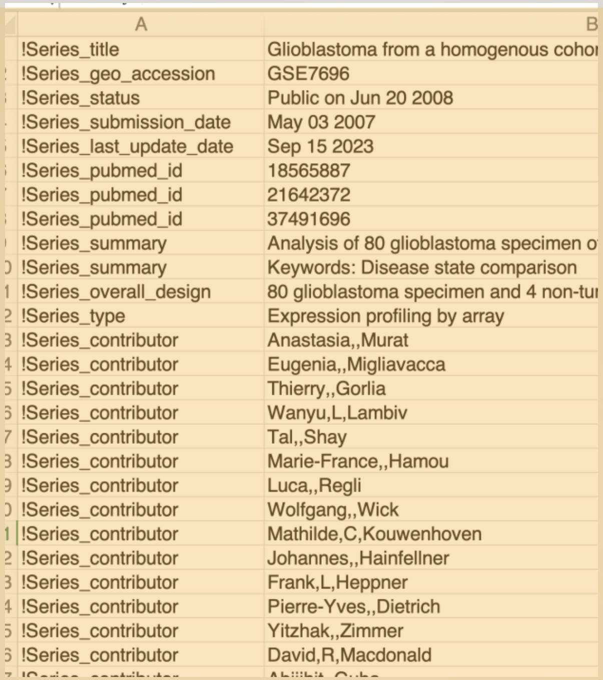
病人特徵描述



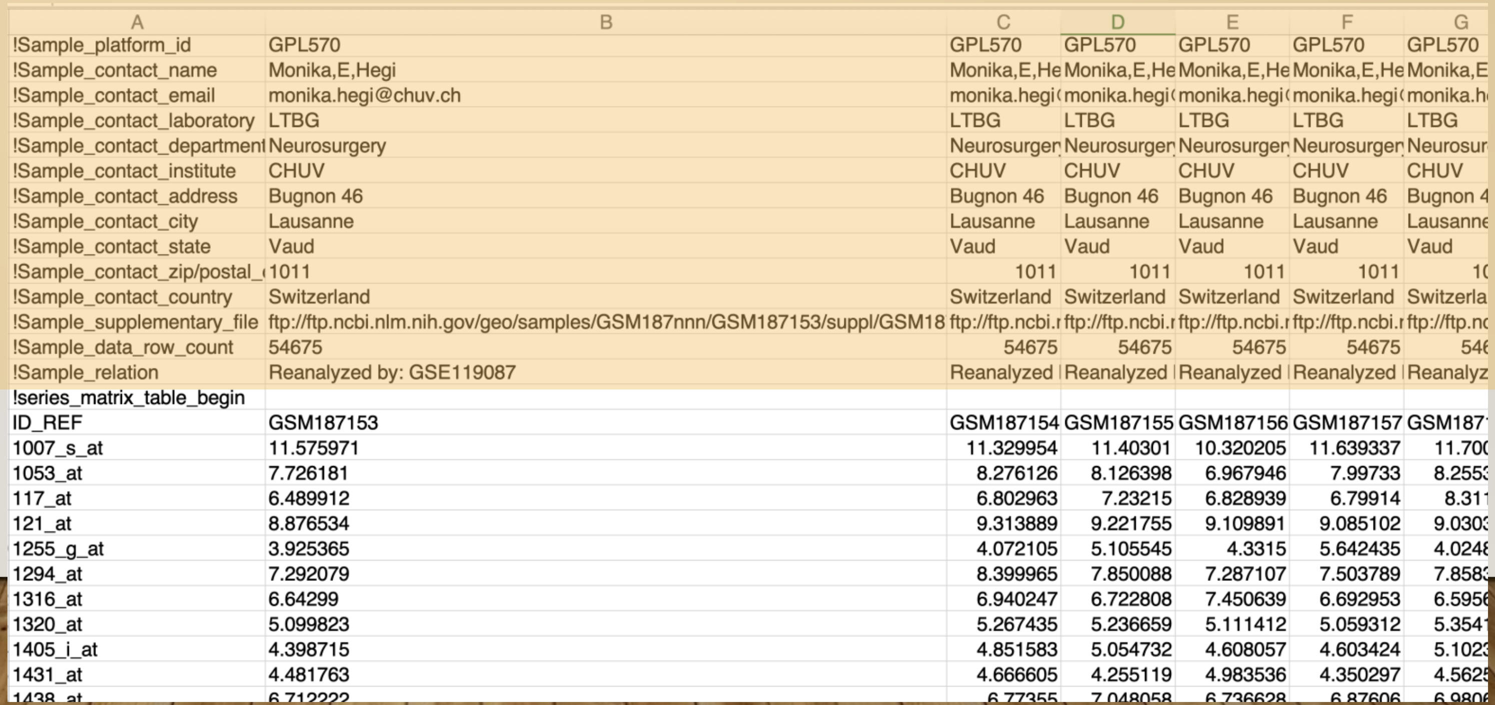
| A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|
| !Sample_platform_id | GPL570 | GPL570 | GPL570 | GPL570 | GPL570 | GPL570 |
| !Sample_contact_name | Monika,E,Hegi | Monika,E,He | Monika,E,He | Monika,E,He | Monika,E,He | Monika,E |
| !Sample_contact_email | monika.hegi@chuv.ch | monika.hegi( | monika.hegi( | monika.hegi( | monika.hegi( | monika.h |
| !Sample_contact_laboratory | LTBG | LTBG | LTBG | LTBG | LTBG | LTBG |
| !Sample_contact_department | Neurosurgery | Neurosurger | Neurosurger | Neurosurger | Neurosurger | Neurosur |
| !Sample_contact_institute | CHUV | CHUV | CHUV | CHUV | CHUV | CHUV |
| !Sample_contact_address | Bugnon 46 | Bugnon 46 | Bugnon 46 | Bugnon 46 | Bugnon 46 | Bugnon 4 |
| !Sample_contact_city | Lausanne | Lausanne | Lausanne | Lausanne | Lausanne | Lausanne |
| !Sample_contact_state | Vaud | Vaud | Vaud | Vaud | Vaud | Vaud |
| !Sample_contact_zip/postal_( | 1011 | 1011 | 1011 | 1011 | 1011 | 10 |
| !Sample_contact_country | Switzerland | Switzerland | Switzerland | Switzerland | Switzerland | Switzerla |
| !Sample_supplementary_file | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM187153/suppl/GSM18 | ftp://ftp.ncbi.r | ftp://ftp.ncbi.r | ftp://ftp.ncbi.r | ftp://ftp.ncbi.r | ftp://ftp.nc |
| !Sample_data_row_count | 54675 | 54675 | 54675 | 54675 | 54675 | 546 |
| !Sample_relation | Reanalyzed by: GSE119087 | Reanalyzed | Reanalyzed | Reanalyzed | Reanalyzed | Reanalyz |
| !series_matrix_table_begin | | | | | | |
| ID_REF | GSM187153 | GSM187154 | GSM187155 | GSM187156 | GSM187157 | GSM187 |
| 1007_s_at | 11.575971 | 11.329954 | 11.40301 | 10.320205 | 11.639337 | 11.70 |
| 1053_at | 7.726181 | 8.276126 | 8.126398 | 6.967946 | 7.99733 | 8.255 |
| 117_at | 6.489912 | 6.802963 | 7.23215 | 6.828939 | 6.79914 | 8.31 |
| 121_at | 8.876534 | 9.313889 | 9.221755 | 9.109891 | 9.085102 | 9.030 |
| 1255_g_at | 3.925365 | 4.072105 | 5.105545 | 4.3315 | 5.642435 | 4.024 |
| 1294_at | 7.292079 | 8.399965 | 7.850088 | 7.287107 | 7.503789 | 7.858 |
| 1316_at | 6.64299 | 6.940247 | 6.722808 | 7.450639 | 6.692953 | 6.595 |
| 1320_at | 5.099823 | 5.267435 | 5.236659 | 5.111412 | 5.059312 | 5.354 |
| 1405_i_at | 4.398715 | 4.851583 | 5.054732 | 4.608057 | 4.603424 | 5.102 |
| 1431_at | 4.481763 | 4.666605 | 4.255119 | 4.983536 | 4.350297 | 4.562 |
| 1438_at | 6.712222 | 6.77355 | 7.048058 | 6.736628 | 6.87606 | 6.980 |

基因名

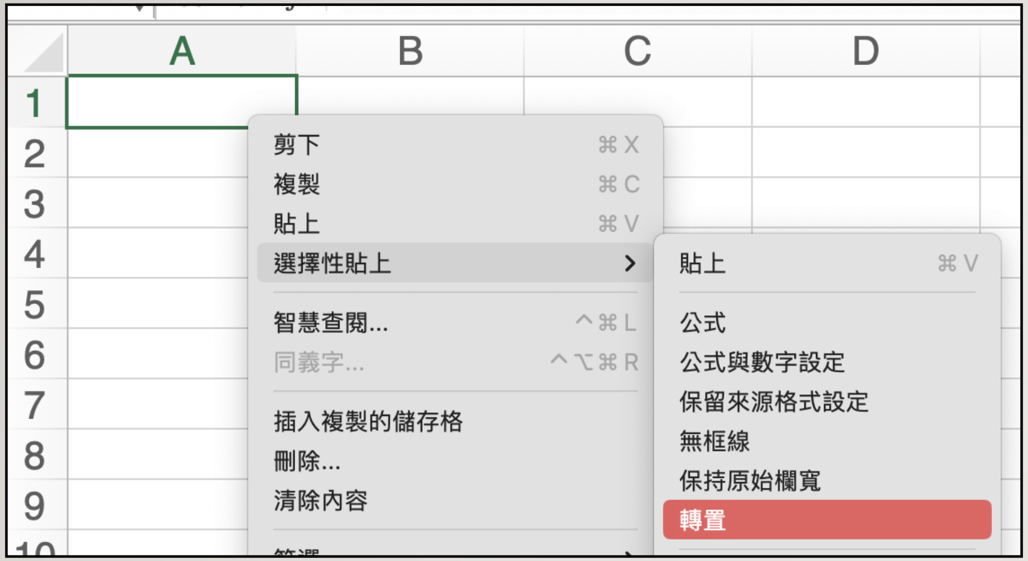
# 挑選病人

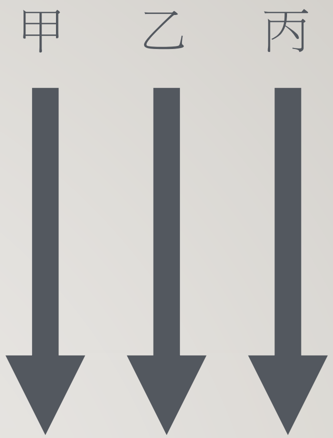


| A | B |
|---|---|
| !Series_title | Glioblastoma from a homogenous cohor |
| !Series_geo_accession | GSE7696 |
| !Series_status | Public on Jun 20 2008 |
| !Series_submission_date | May 03 2007 |
| !Series_last_update_date | Sep 15 2023 |
| !Series_pubmed_id | 18565887 |
| !Series_pubmed_id | 21642372 |
| !Series_pubmed_id | 37491696 |
| !Series_summary | Analysis of 80 glioblastoma specimen o |
| !Series_summary | Keywords: Disease state comparison |
| !Series_overall_design | 80 glioblastoma specimen and 4 non-tu |
| !Series_type | Expression profiling by array |
| !Series_contributor | Anastasia,,Murat |
| !Series_contributor | Eugenia,,Migliavacca |
| !Series_contributor | Thierry,,Gorlia |
| !Series_contributor | Wanyu,L,Lambiv |
| !Series_contributor | Tal,,Shay |
| !Series_contributor | Marie-France,,Hamou |
| !Series_contributor | Luca,,Regli |
| !Series_contributor | Wolfgang,,Wick |
| !Series_contributor | Mathilde,C,Kouwenhoven |
| !Series_contributor | Johannes,,Hainfellner |
| !Series_contributor | Frank,L,Heppner |
| !Series_contributor | Pierre-Yves,,Dietrich |
| !Series_contributor | Yitzhak,,Zimmer |
| !Series_contributor | David,R,Macdonald |

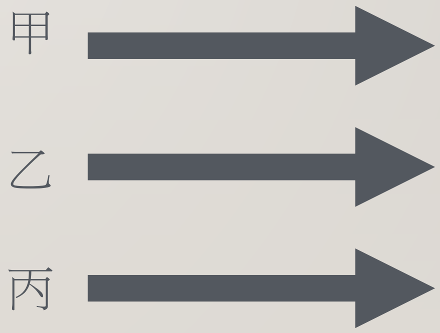


| A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|
| !Sample_platform_id | GPL570 | GPL570 | GPL570 | GPL570 | GPL570 | GPL570 |
| !Sample_contact_name | Monika,E,Hegi | Monika,E,He | Monika,E,He | Monika,E,He | Monika,E,He | Monika,E |
| !Sample_contact_email | monika.hegi@chuv.ch | monika.hegi( | monika.hegi( | monika.hegi( | monika.hegi( | monika.h |
| !Sample_contact_laboratory | LTBG | LTBG | LTBG | LTBG | LTBG | LTBG |
| !Sample_contact_department | Neurosurgery | Neurosurger | Neurosurger | Neurosurger | Neurosurger | Neurosur |
| !Sample_contact_institute | CHUV | CHUV | CHUV | CHUV | CHUV | CHUV |
| !Sample_contact_address | Bugnon 46 | Bugnon 46 | Bugnon 46 | Bugnon 46 | Bugnon 46 | Bugnon 4 |
| !Sample_contact_city | Lausanne | Lausanne | Lausanne | Lausanne | Lausanne | Lausanne |
| !Sample_contact_state | Vaud | Vaud | Vaud | Vaud | Vaud | Vaud |
| !Sample_contact_zip/postal_( | 1011 | 1011 | 1011 | 1011 | 1011 | 1( |
| !Sample_contact_country | Switzerland | Switzerland | Switzerland | Switzerland | Switzerland | Switzerla |
| !Sample_supplementary_file | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM187153/suppl/GSM18 | ftp://ftp.ncbi.r | ftp://ftp.ncbi.r | ftp://ftp.ncbi.r | ftp://ftp.ncbi.r | ftp://ftp.nc |
| !Sample_data_row_count | 54675 | 54675 | 54675 | 54675 | 54675 | 546 |
| !Sample_relation | Reanalyzed by: GSE119087 | Reanalyzed | Reanalyzed | Reanalyzed | Reanalyzed | Reanalyz |
| !series_matrix_table_begin | | | | | | |
| ID_REF | GSM187153 | GSM187154 | GSM187155 | GSM187156 | GSM187157 | GSM187 |
| 1007_s_at | 11.575971 | 11.329954 | 11.40301 | 10.320205 | 11.639337 | 11.700 |
| 1053_at | 7.726181 | 8.276126 | 8.126398 | 6.967946 | 7.99733 | 8.2553 |
| 117_at | 6.489912 | 6.802963 | 7.23215 | 6.828939 | 6.79914 | 8.311 |
| 121_at | 8.876534 | 9.313889 | 9.221755 | 9.109891 | 9.085102 | 9.0303 |
| 1255_g_at | 3.925365 | 4.072105 | 5.105545 | 4.3315 | 5.642435 | 4.0248 |
| 1294_at | 7.292079 | 8.399965 | 7.850088 | 7.287107 | 7.503789 | 7.8583 |
| 1316_at | 6.64299 | 6.940247 | 6.722808 | 7.450639 | 6.692953 | 6.5956 |
| 1320_at | 5.099823 | 5.267435 | 5.236659 | 5.111412 | 5.059312 | 5.3541 |
| 1405_i_at | 4.398715 | 4.851583 | 5.054732 | 4.608057 | 4.603424 | 5.1023 |
| 1431_at | 4.481763 | 4.666605 | 4.255119 | 4.983536 | 4.350297 | 4.5625 |
| 1438_at | 6.712222 | 6.77355 | 7.048058 | 6.736628 | 6.87606 | 6.9806 |

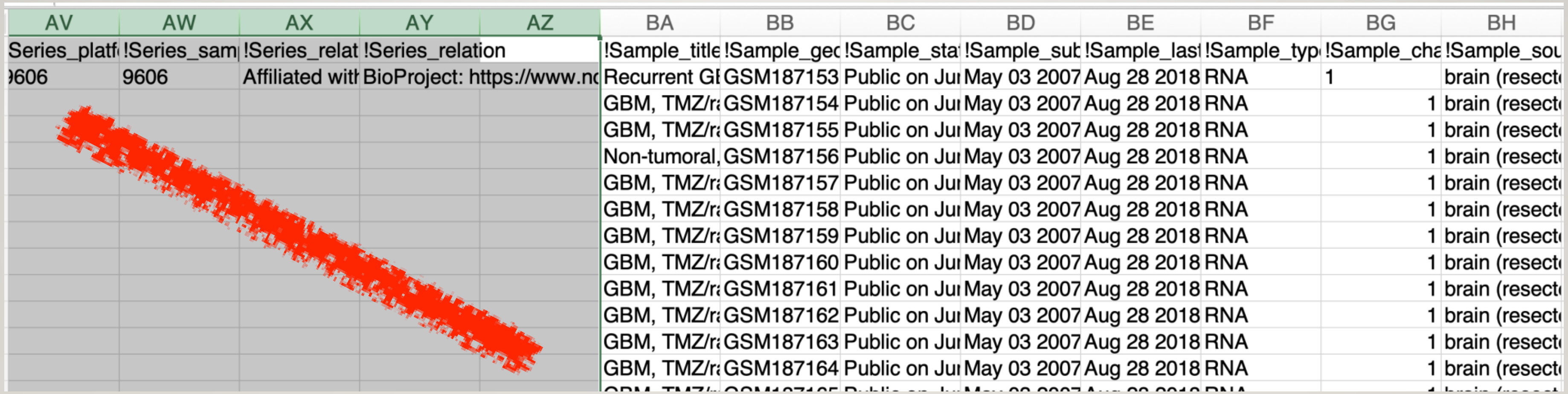

A
B
C
D
1
2
3
4
5
6
7
8
9
剪下 ⌘X
複製 ⌘C
貼上 ⌘V
選擇性貼上 >
智慧查閱... ^⌘L
同義字... ^⌥⌘R
插入複製的儲存格
刪除...
清除內容
貼上 ⌘V
公式
公式與數字設定
保留來源格式設定
無框線
保持原始欄寬
轉置


甲
乙
丙

## 轉置

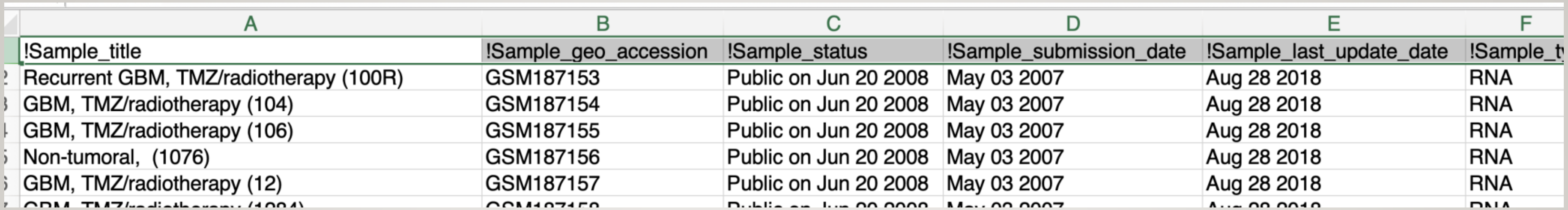

甲
乙
丙

# 刪掉NCBI描述



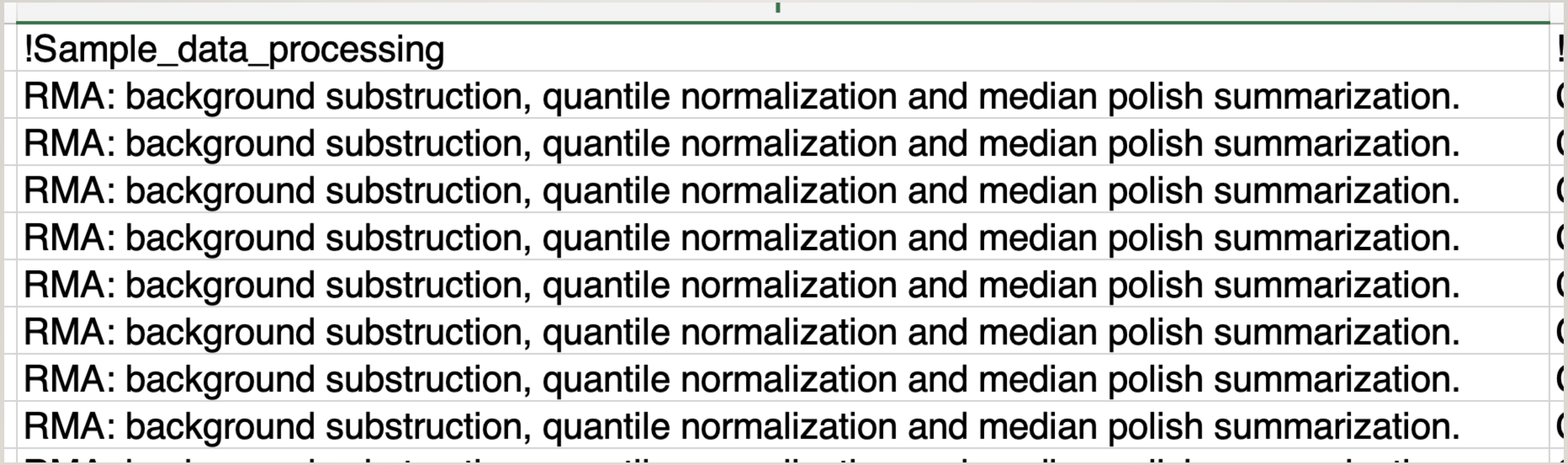
| AV | AW | AX | AY | AZ | BA | BB | BC | BD | BE | BF | BG | BH |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Series_platf | !Series_sam | !Series_relat | !Series_relation | | !Sample_title | !Sample_gec | !Sample_stat | !Sample_sub | !Sample_last | !Sample_typ | !Sample_cha | !Sample_sou |
| 9606 | 9606 | Affiliated with | BioProject: https://www.nc | | Recurrent G | GSM187153 | Public on Ju | May 03 2007 | Aug 28 2018 | RNA | 1 | brain (resect |
| | | | | | GBM, TMZ/r | GSM187154 | Public on Ju | May 03 2007 | Aug 28 2018 | RNA | 1 | brain (resect |
| | | | | | GBM, TMZ/r | GSM187155 | Public on Ju | May 03 2007 | Aug 28 2018 | RNA | 1 | brain (resect |
| | | | | | Non-tumoral, | GSM187156 | Public on Ju | May 03 2007 | Aug 28 2018 | RNA | 1 | brain (resect |
| | | | | | GBM, TMZ/r | GSM187157 | Public on Ju | May 03 2007 | Aug 28 2018 | RNA | 1 | brain (resect |
| | | | | | GBM, TMZ/r | GSM187158 | Public on Ju | May 03 2007 | Aug 28 2018 | RNA | 1 | brain (resect |
| | | | | | GBM, TMZ/r | GSM187159 | Public on Ju | May 03 2007 | Aug 28 2018 | RNA | 1 | brain (resect |
| | | | | | GBM, TMZ/r | GSM187160 | Public on Ju | May 03 2007 | Aug 28 2018 | RNA | 1 | brain (resect |
| | | | | | GBM, TMZ/r | GSM187161 | Public on Ju | May 03 2007 | Aug 28 2018 | RNA | 1 | brain (resect |
| | | | | | GBM, TMZ/r | GSM187162 | Public on Ju | May 03 2007 | Aug 28 2018 | RNA | 1 | brain (resect |
| | | | | | GBM, TMZ/r | GSM187163 | Public on Ju | May 03 2007 | Aug 28 2018 | RNA | 1 | brain (resect |
| | | | | | GBM, TMZ/r | GSM187164 | Public on Ju | May 03 2007 | Aug 28 2018 | RNA | 1 | brain (resect |

# 留下你關心的資訊



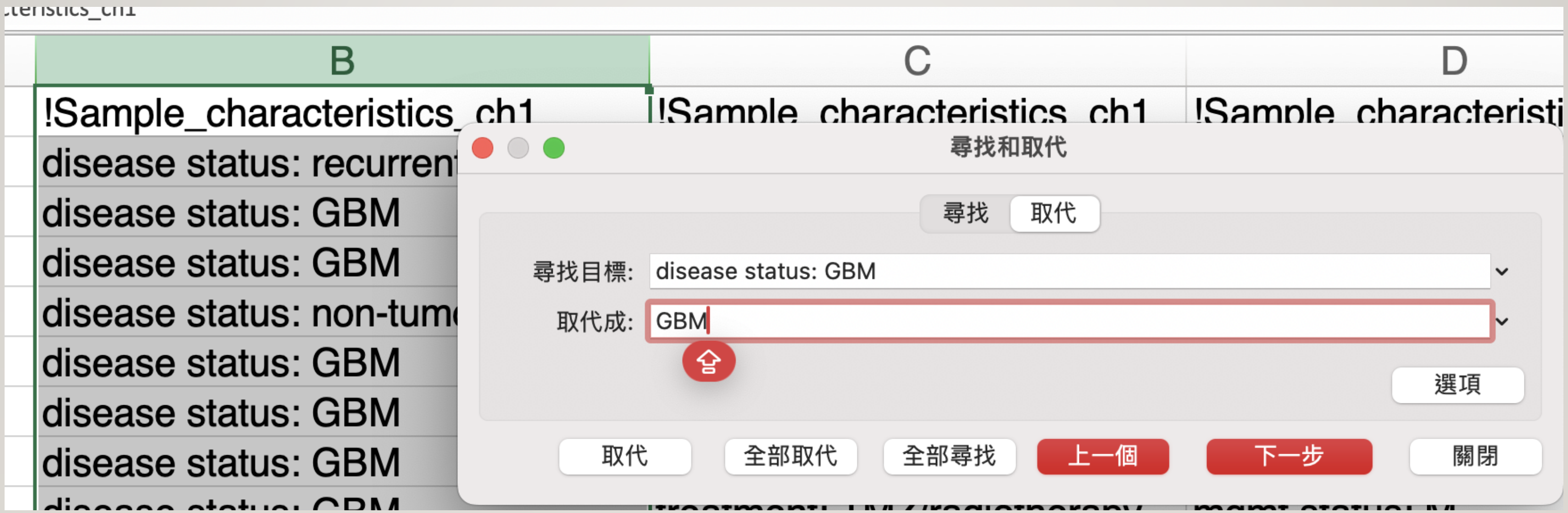
| A | B | C | D | E | F |
|---|---|---|---|---|---|
| !Sample_title | !Sample_geo_accession | !Sample_status | !Sample_submission_date | !Sample_last_update_date | !Sample_t |
| Recurrent GBM, TMZ/radiotherapy (100R) | GSM187153 | Public on Jun 20 2008 | May 03 2007 | Aug 28 2018 | RNA |
| GBM, TMZ/radiotherapy (104) | GSM187154 | Public on Jun 20 2008 | May 03 2007 | Aug 28 2018 | RNA |
| GBM, TMZ/radiotherapy (106) | GSM187155 | Public on Jun 20 2008 | May 03 2007 | Aug 28 2018 | RNA |
| Non-tumoral,  (1076) | GSM187156 | Public on Jun 20 2008 | May 03 2007 | Aug 28 2018 | RNA |
| GBM, TMZ/radiotherapy (12) | GSM187157 | Public on Jun 20 2008 | May 03 2007 | Aug 28 2018 | RNA |

# 已處理Batch effect。可用



| !Sample_data_processing |
|---|
| RMA: background substruction, quantile normalization and median polish summarization. |
| RMA: background substruction, quantile normalization and median polish summarization. |
| RMA: background substruction, quantile normalization and median polish summarization. |
| RMA: background substruction, quantile normalization and median polish summarization. |
| RMA: background substruction, quantile normalization and median polish summarization. |
| RMA: background substruction, quantile normalization and median polish summarization. |
| RMA: background substruction, quantile normalization and median polish summarization. |
| RMA: background substruction, quantile normalization and median polish summarization. |

Ctrl + F


B
C
D
尋找和取代
尋找
取代
選項
取代
全部取代
全部尋找
上一個
下一步
關閉

!Sample_characteristics_ch1

disease status: recurrent

disease status: GBM

disease status: GBM

disease status: non-tum

disease status: GBM

disease status: GBM

disease status: GBM

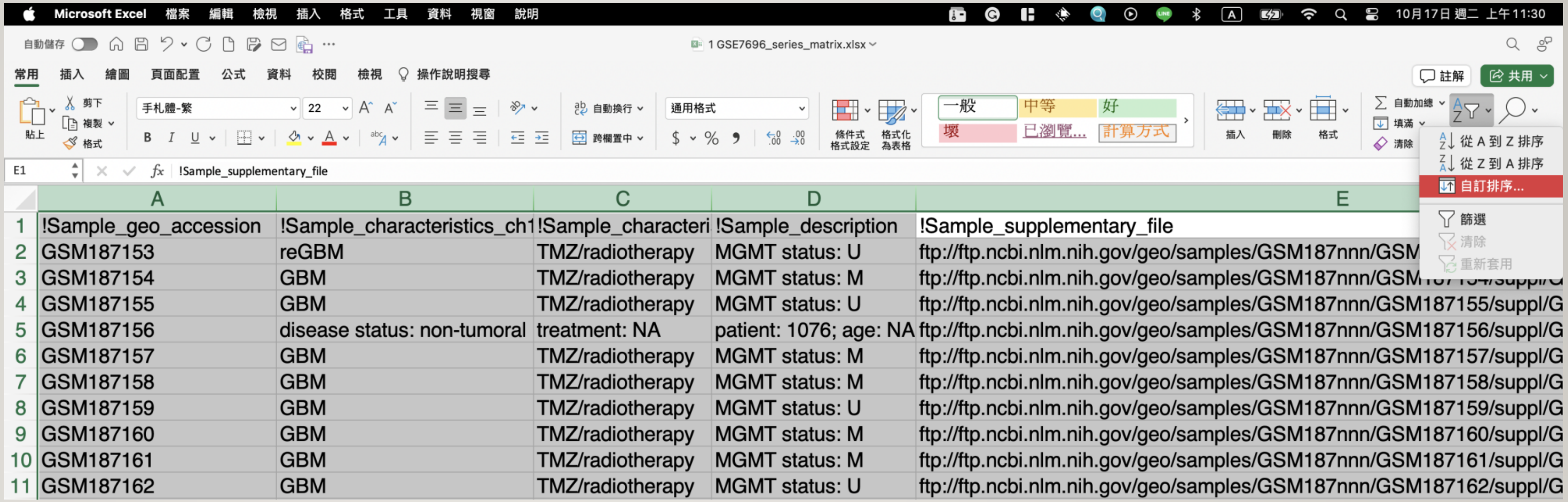
!Sample_characteristics_ch1

!Sample_characteristi

尋找目標: disease status: GBM

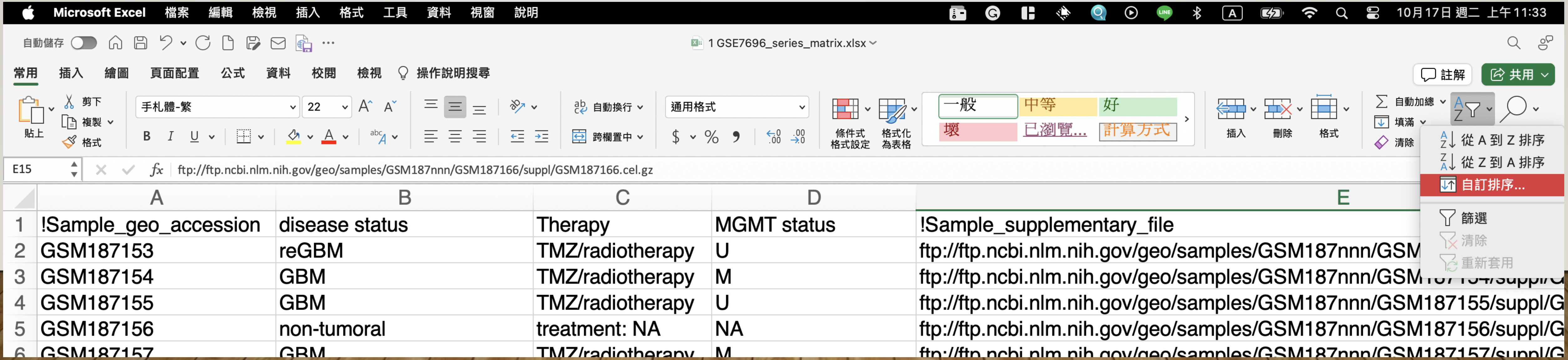
取代成: GBM

# 自定排序


Microsoft Excel
檔案
編輯
檢視
插入
格式
工具
資料
視窗
說明
10月17日 週二 上午11:30
自動儲存
1 GSE7696_series_matrix.xlsx
常用
插入
繪圖
頁面配置
公式
資料
校閱
檢視
操作說明搜尋
註解
共用
手札體-繁
22
自動換行
跨欄置中
通用格式
條件式格式設定
格式化為表格
一般
中等
好
壞
已瀏覽...
計算方式
插入
刪除
格式
自動加總
填滿
清除
從 A 到 Z 排序
從 Z 到 A 排序
自訂排序...
篩選
清除
重新套用

E1 !Sample_supplementary_file

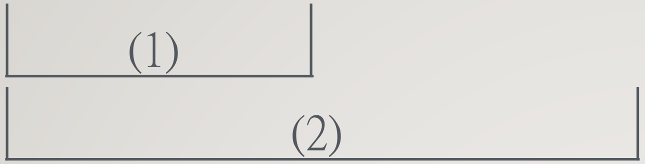
| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | !Sample_geo_accession | !Sample_characteristics_ch1 | !Sample_characteri | !Sample_description | !Sample_supplementary_file |
| 2 | GSM187153 | reGBM | TMZ/radiotherapy | MGMT status: U | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM |
| 3 | GSM187154 | GBM | TMZ/radiotherapy | MGMT status: M | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM |
| 4 | GSM187155 | GBM | TMZ/radiotherapy | MGMT status: U | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM187155/suppl/G |
| 5 | GSM187156 | disease status: non-tumoral | treatment: NA | patient: 1076; age: NA | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM187156/suppl/G |
| 6 | GSM187157 | GBM | TMZ/radiotherapy | MGMT status: M | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM187157/suppl/G |
| 7 | GSM187158 | GBM | TMZ/radiotherapy | MGMT status: M | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM187158/suppl/G |
| 8 | GSM187159 | GBM | TMZ/radiotherapy | MGMT status: U | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM187159/suppl/G |
| 9 | GSM187160 | GBM | TMZ/radiotherapy | MGMT status: M | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM187160/suppl/G |
| 10 | GSM187161 | GBM | TMZ/radiotherapy | MGMT status: M | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM187161/suppl/G |
| 11 | GSM187162 | GBM | TMZ/radiotherapy | MGMT status: U | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM187162/suppl/G |

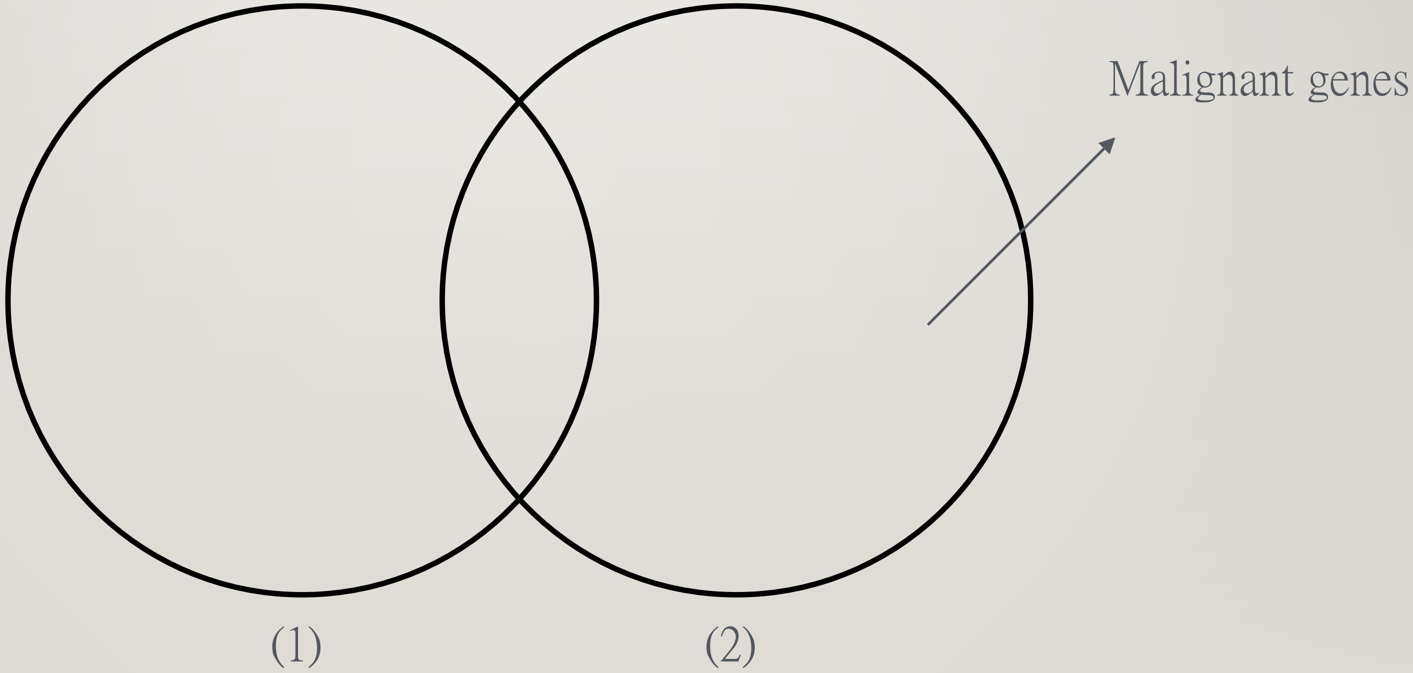

Microsoft Excel
檔案
編輯
檢視
插入
格式
工具
資料
視窗
說明
10月17日 週二 上午11:33
自動儲存
1 GSE7696_series_matrix.xlsx
常用
插入
繪圖
頁面配置
公式
資料
校閱
檢視
操作說明搜尋
註解
共用
剪下
複製
格式
貼上
手札體-繁
22
自動換行
跨欄置中
通用格式
條件式格式設定
格式化為表格
一般
中等
好
壞
已瀏覽...
計算方式
插入
刪除
格式
自動加總
填滿
清除
從 A 到 Z 排序
從 Z 到 A 排序
自訂排序...
篩選
清除
重新套用

E15 ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM187166/suppl/GSM187166.cel.gz

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | !Sample_geo_accession | disease status | Therapy | MGMT status | !Sample_supplementary_file |
| 2 | GSM187153 | reGBM | TMZ/radiotherapy | U | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM |
| 3 | GSM187154 | GBM | TMZ/radiotherapy | M | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM |
| 4 | GSM187155 | GBM | TMZ/radiotherapy | U | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM187155/suppl/G |
| 5 | GSM187156 | non-tumoral | treatment: NA | NA | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM187156/suppl/G |
| 6 | GSM187157 | GBM | TMZ/radiotherapy | M | ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM187nnn/GSM187157/suppl/G |

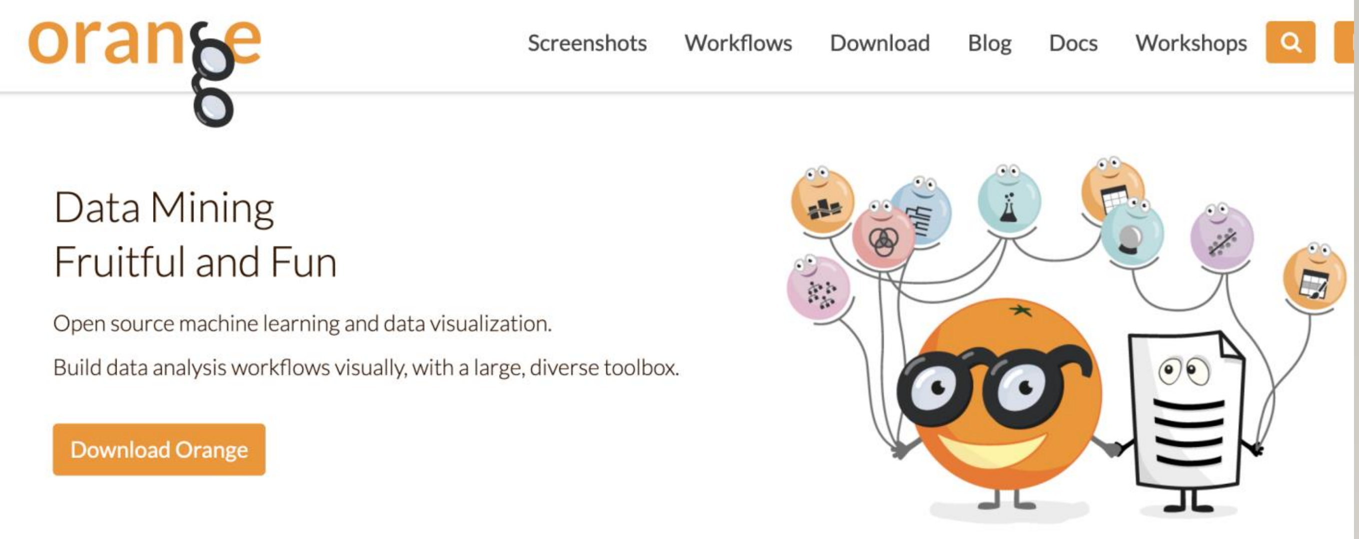
| Normal | GBM | recurrent_GBM |
|---|---|---|




Malignant genes
(1)
(2)

# 推薦工具

**Orange**


orange
Screenshots
Workflows
Download
Blog
Docs
Workshops

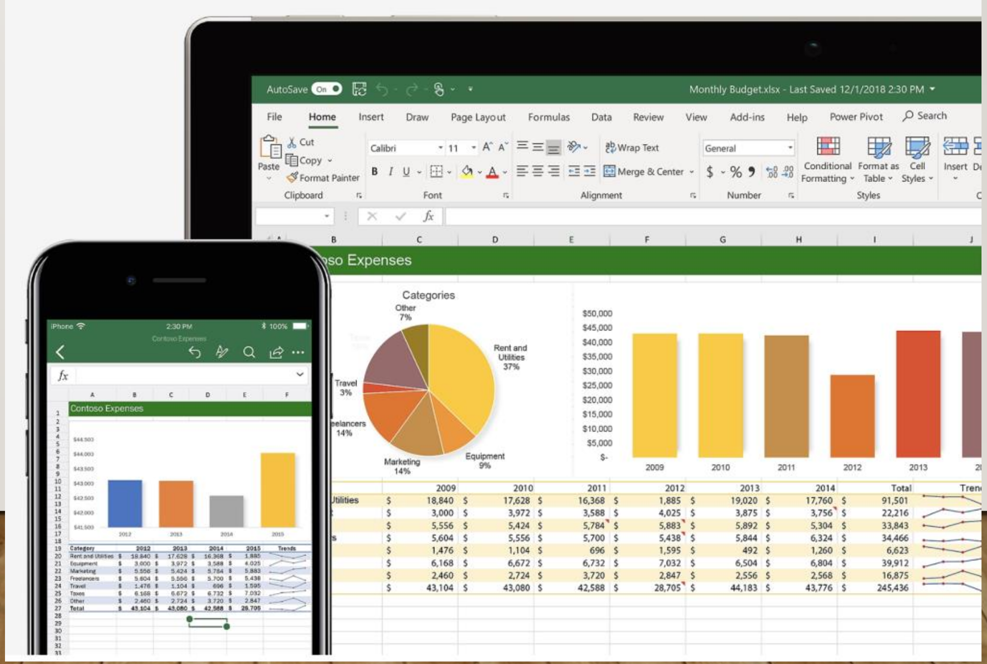
Data Mining
Fruitful and Fun

Open source machine learning and data visualization.

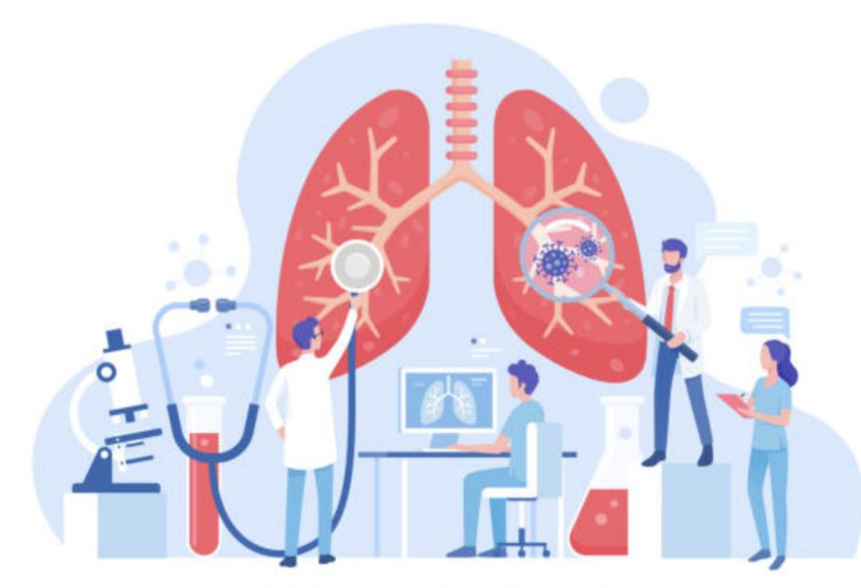
Build data analysis workflows visually, with a large, diverse toolbox.

Download Orange

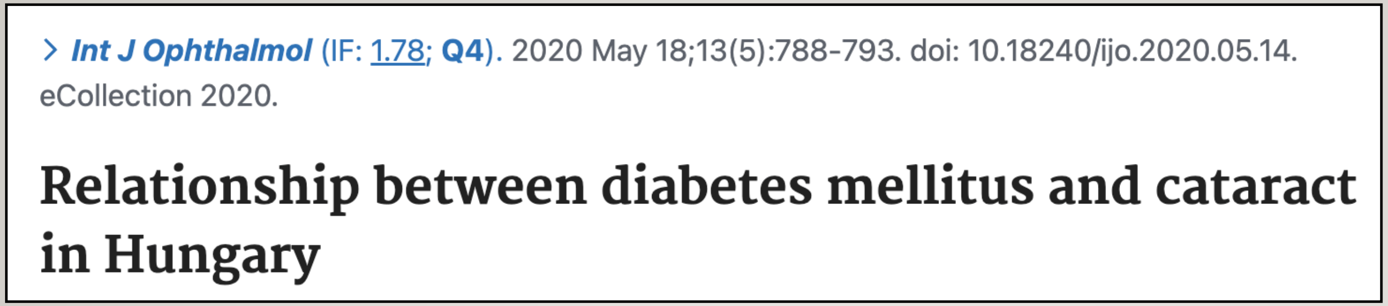
**Excel**


AutoSave On
Monthly Budget.xlsx - Last Saved 12/1/2018 2:30 PM
File
Home
Insert
Draw
Page Layout
Formulas
Data
Review
View
Add-ins
Help
Power Pivot
Search
Paste
Cut
Copy
Format Painter
Clipboard
Calibri
11
Font
Wrap Text
Merge & Center
Alignment
General
Number
Conditional Formatting
Format as Table
Cell Styles
Styles
Insert
oso Expenses
Categories
Other 7%
Rent and Utilities 37%
Travel 3%
eelancers 14%
Marketing 14%
Equipment 9%
$50,000
$45,000
$40,000
$35,000
$30,000
$25,000
$20,000
$15,000
$10,000
$5,000
$-
2009
2010
2011
2012
2013
Contoso Expenses

| | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | Total | Tren |
|---|---|---|---|---|---|---|---|---|
| Utilities | $ 18,840 | $ 17,628 | $ 16,368 | $ 1,885 | $ 19,020 | $ 17,760 | $ 91,501 | |
| | $ 3,000 | $ 3,972 | $ 3,588 | $ 4,025 | $ 3,875 | $ 3,756 | $ 22,216 | |
| | $ 5,556 | $ 5,424 | $ 5,784 | $ 5,883 | $ 5,892 | $ 5,304 | $ 33,843 | |
| s | $ 5,604 | $ 5,556 | $ 5,700 | $ 5,438 | $ 5,844 | $ 6,324 | $ 34,466 | |
| | $ 1,476 | $ 1,104 | $ 696 | $ 1,595 | $ 492 | $ 1,260 | $ 6,623 | |
| | $ 6,168 | $ 6,672 | $ 6,732 | $ 7,032 | $ 6,504 | $ 6,804 | $ 39,912 | |
| | $ 2,460 | $ 2,724 | $ 3,720 | $ 2,847 | $ 2,556 | $ 2,568 | $ 16,875 | |
| | $ 43,104 | $ 43,080 | $ 42,588 | $ 28,705 | $ 44,183 | $ 43,776 | $ 245,436 | |



# 如何提高科學的影響力?



> ***Int J Ophthalmol*** (IF: 1.78; **Q4**). 2020 May 18;13(5):788-793. doi: 10.18240/ijo.2020.05.14. eCollection 2020.

**Relationship between diabetes mellitus and cataract in Hungary**

# 1. Characterization

# SCI:1.78

**But, 能做的题目早做光了**

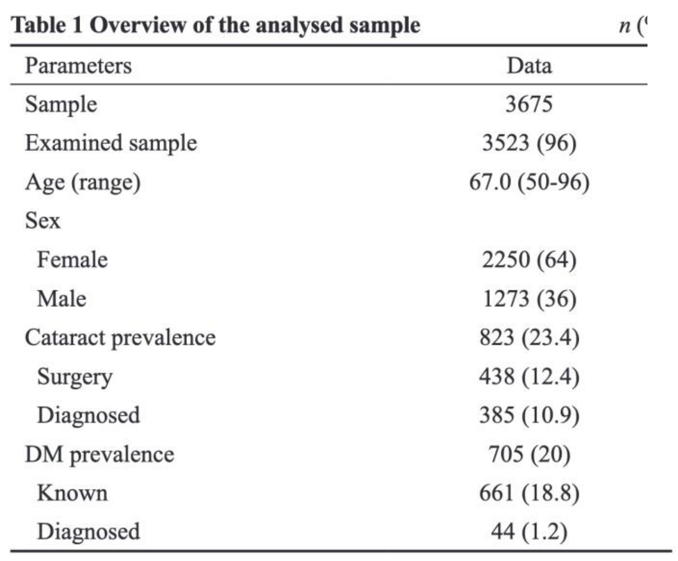
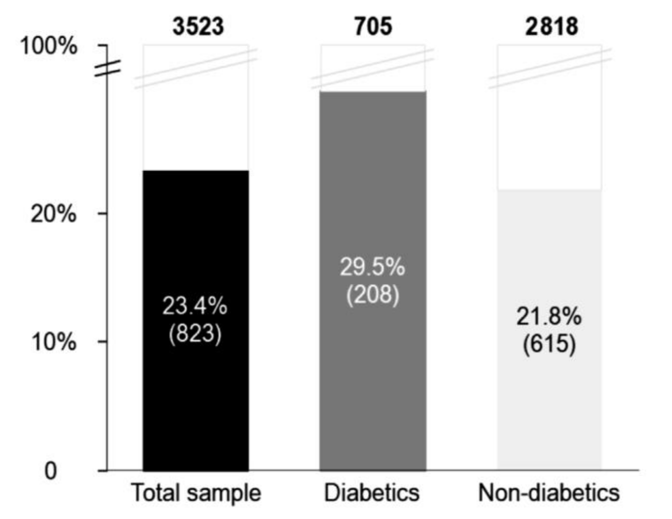
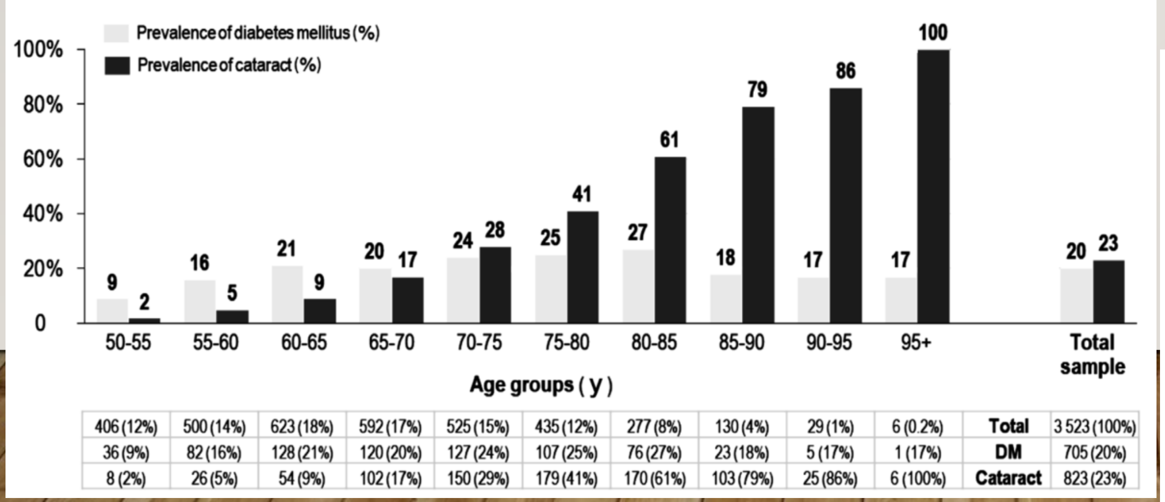


**Table 1 Overview of the analysed sample** *n* (

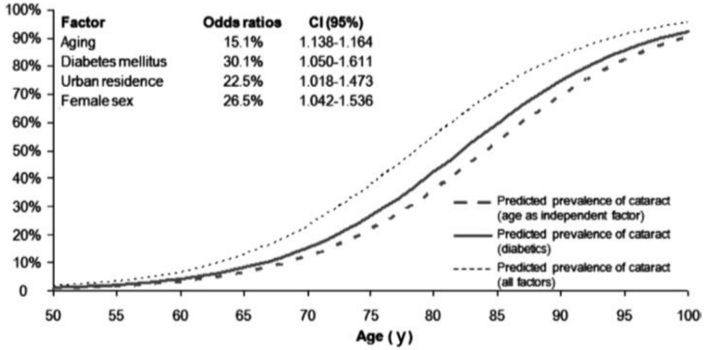
| Parameters | Data |
|---|---|
| Sample | 3675 |
| Examined sample | 3523 (96) |
| Age (range) | 67.0 (50-96) |
| Sex | |
| Female | 2250 (64) |
| Male | 1273 (36) |
| Cataract prevalence | 823 (23.4) |
| Surgery | 438 (12.4) |
| Diagnosed | 385 (10.9) |
| DM prevalence | 705 (20) |
| Known | 661 (18.8) |
| Diagnosed | 44 (1.2) |



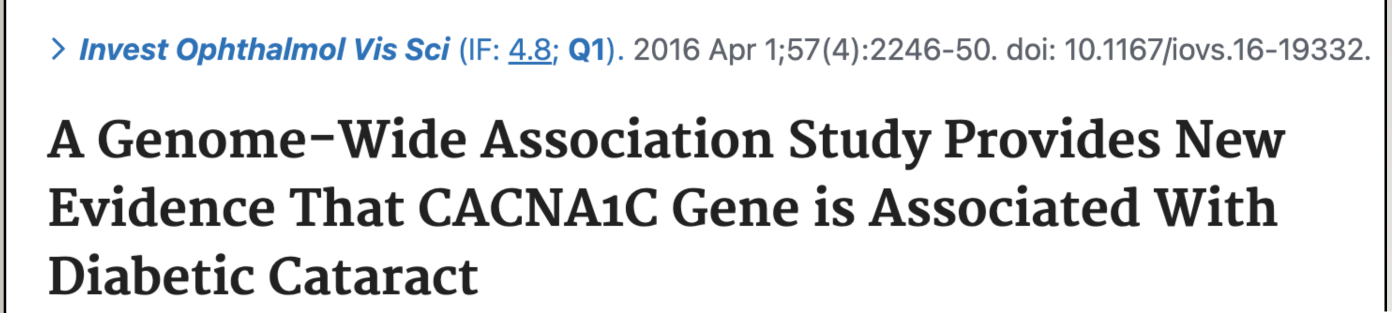


| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 406 (12%) | 500 (14%) | 623 (18%) | 592 (17%) | 525 (15%) | 435 (12%) | 277 (8%) | 130 (4%) | 29 (1%) | 6 (0.2%) | **Total** | 3 523 (100%) |
| 36 (9%) | 82 (16%) | 128 (21%) | 120 (20%) | 127 (24%) | 107 (25%) | 76 (27%) | 23 (18%) | 5 (17%) | 1 (17%) | **DM** | 705 (20%) |
| 8 (2%) | 26 (5%) | 54 (9%) | 102 (17%) | 150 (29%) | 179 (41%) | 170 (61%) | 103 (79%) | 25 (86%) | 6 (100%) | **Cataract** | 823 (23%) |

**Figure 2 Co-existence of cataract and DM and prevalence of cataract in sub-groups.**



| Factor | Odds ratios | CI (95%) |
|---|---|---|
| Aging | 15.1% | 1.138-1.164 |
| Diabetes mellitus | 30.1% | 1.050-1.611 |
| Urban residence | 22.5% | 1.018-1.473 |
| Female sex | 26.5% | 1.042-1.536 |


> *Invest Ophthalmol Vis Sci* (IF: 4.8; **Q1**). 2016 Apr 1;57(4):2246-50. doi: 10.1167/iovs.16-19332.

# A Genome-Wide Association Study Provides New Evidence That CACNA1C Gene is Associated With Diabetic Cataract

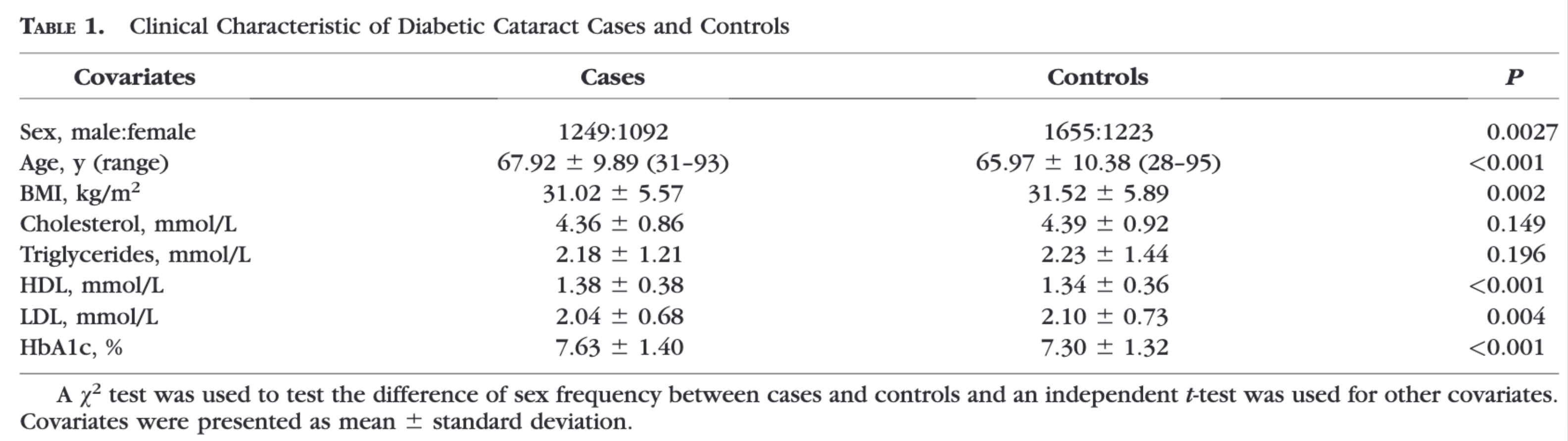


**Table 1.** Clinical Characteristic of Diabetic Cataract Cases and Controls

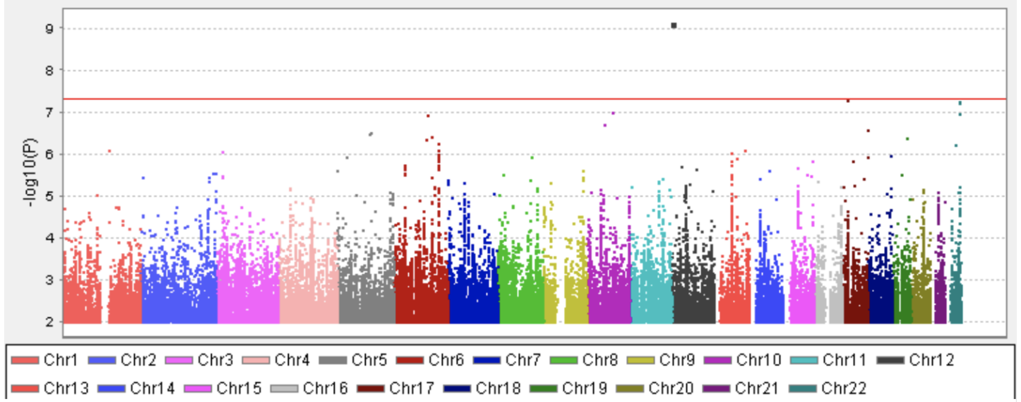
| Covariates | Cases | Controls | *P* |
|---|---|---|---|
| Sex, male:female | 1249:1092 | 1655:1223 | 0.0027 |
| Age, y (range) | 67.92 ± 9.89 (31-93) | 65.97 ± 10.38 (28-95) | <0.001 |
| BMI, kg/m$^2$ | 31.02 ± 5.57 | 31.52 ± 5.89 | 0.002 |
| Cholesterol, mmol/L | 4.36 ± 0.86 | 4.39 ± 0.92 | 0.149 |
| Triglycerides, mmol/L | 2.18 ± 1.21 | 2.23 ± 1.44 | 0.196 |
| HDL, mmol/L | 1.38 ± 0.38 | 1.34 ± 0.36 | <0.001 |
| LDL, mmol/L | 2.04 ± 0.68 | 2.10 ± 0.73 | 0.004 |
| HbA1c, % | 7.63 ± 1.40 | 7.30 ± 1.32 | <0.001 |

A $\chi^2$ test was used to test the difference of sex frequency between cases and controls and an independent *t*-test was used for other covariates. Covariates were presented as mean ± standard deviation.

1. **Characterization**
2. **Gene information**

**SCI:4.8**

**But, 題目須受基因影響的疾病**



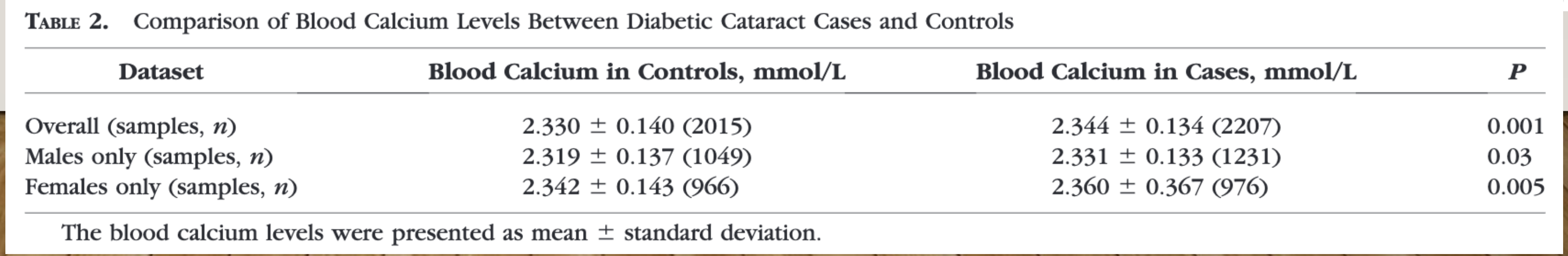
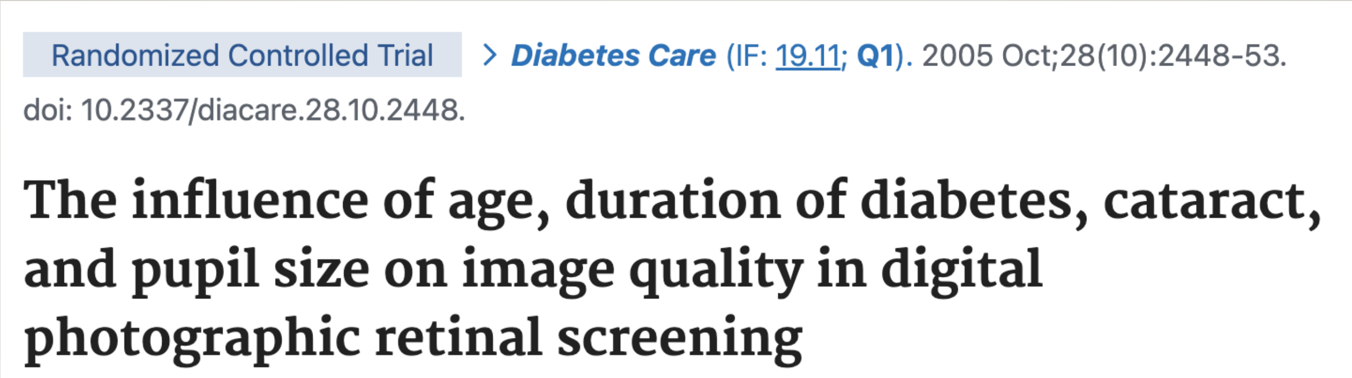


**Table 2.** Comparison of Blood Calcium Levels Between Diabetic Cataract Cases and Controls

| Dataset | Blood Calcium in Controls, mmol/L | Blood Calcium in Cases, mmol/L | *P* |
|---|---|---|---|
| Overall (samples, *n*) | 2.330 ± 0.140 (2015) | 2.344 ± 0.134 (2207) | 0.001 |
| Males only (samples, *n*) | 2.319 ± 0.137 (1049) | 2.331 ± 0.133 (1231) | 0.03 |
| Females only (samples, *n*) | 2.342 ± 0.143 (966) | 2.360 ± 0.367 (976) | 0.005 |

The blood calcium levels were presented as mean ± standard deviation.



Randomized Controlled Trial > *Diabetes Care* (IF: 19.11; Q1). 2005 Oct;28(10):2448-53.
doi: 10.2337/diacare.28.10.2448.

**The influence of age, duration of diabetes, cataract, and pupil size on image quality in digital photographic retinal screening**

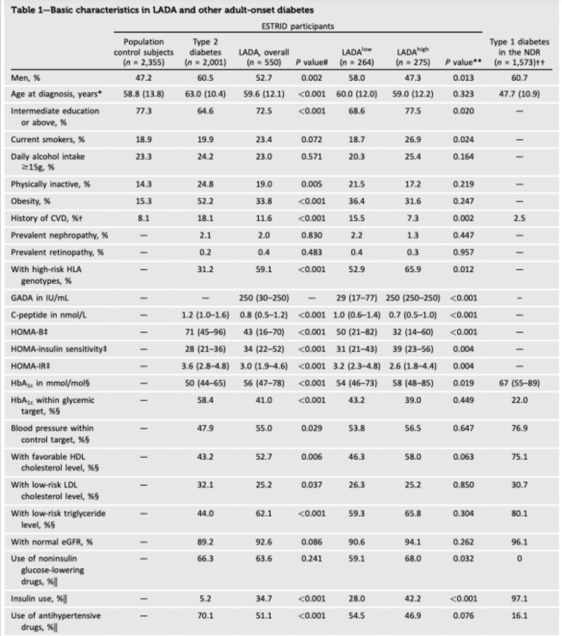
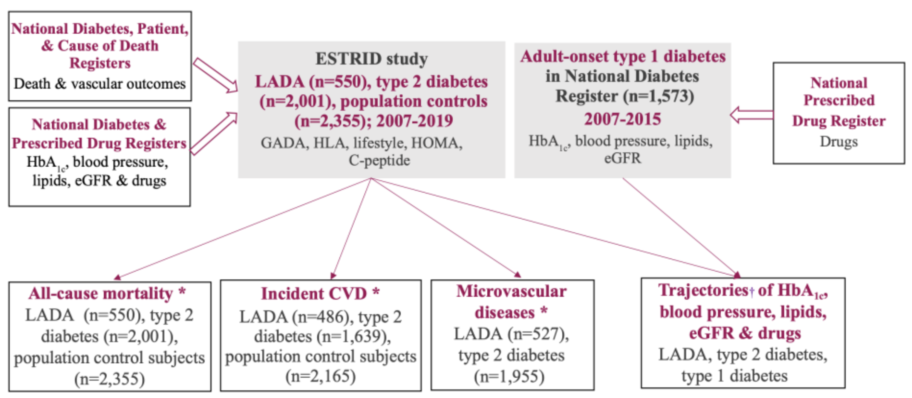


**Table 1—Basic characteristics in LADA and other adult-onset diabetes**

| | ESTRID participants | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Population control subjects (*n* = 2,355) | Type 2 diabetes (*n* = 2,001) | LADA, overall (*n* = 550) | *P* value# | $LADA^{low}$ (*n* = 264) | $LADA^{high}$ (*n* = 275) | *P* value** | Type 1 diabetes in the NDR (*n* = 1,573)†† |
| Men, % | 47.2 | 60.5 | 52.7 | 0.002 | 58.0 | 47.3 | 0.013 | 60.7 |
| Age at diagnosis, years* | 58.8 (13.8) | 63.0 (10.4) | 59.6 (12.1) | <0.001 | 60.0 (12.0) | 59.0 (12.2) | 0.323 | 47.7 (10.9) |
| Intermediate education or above, % | 77.3 | 64.6 | 72.5 | <0.001 | 68.6 | 77.5 | 0.020 | — |
| Current smokers, % | 18.9 | 19.9 | 23.4 | 0.072 | 18.7 | 26.9 | 0.024 | — |
| Daily alcohol intake ≥15g, % | 23.3 | 24.2 | 23.0 | 0.571 | 20.3 | 25.4 | 0.164 | — |
| Physically inactive, % | 14.3 | 24.8 | 19.0 | 0.005 | 21.5 | 17.2 | 0.219 | — |
| Obesity, % | 15.3 | 52.2 | 33.8 | <0.001 | 36.4 | 31.6 | 0.247 | — |
| History of CVD, %† | 8.1 | 18.1 | 11.6 | <0.001 | 15.5 | 7.3 | 0.002 | 2.5 |
| Prevalent nephropathy, % | — | 2.1 | 2.0 | 0.830 | 2.2 | 1.3 | 0.447 | — |
| Prevalent retinopathy, % | — | 0.2 | 0.4 | 0.483 | 0.4 | 0.3 | 0.957 | — |
| With high-risk HLA genotypes, % | — | 31.2 | 59.1 | <0.001 | 52.9 | 65.9 | 0.012 | — |
| GADA in IU/mL | — | — | 250 (30–250) | — | 29 (17–77) | 250 (250–250) | <0.001 | - |
| C-peptide in nmol/L | — | 1.2 (1.0–1.6) | 0.8 (0.5–1.2) | <0.001 | 1.0 (0.6–1.4) | 0.7 (0.5–1.0) | <0.001 | — |
| HOMA-B‡ | — | 71 (45–96) | 43 (16–70) | <0.001 | 50 (21–82) | 32 (14–60) | <0.001 | — |
| HOMA-insulin sensitivity‡ | — | 28 (21–36) | 34 (22–52) | <0.001 | 31 (21–43) | 39 (23–56) | 0.004 | — |
| HOMA-IR‡ | — | 3.6 (2.8–4.8) | 3.0 (1.9–4.6) | <0.001 | 3.2 (2.3–4.8) | 2.6 (1.8–4.4) | 0.004 | — |
| $HbA_{1c}$ in mmol/mol§ | — | 50 (44–65) | 56 (47–78) | <0.001 | 54 (46–73) | 58 (48–85) | 0.019 | 67 (55–89) |
| $HbA_{1c}$ within glycemic target, %§ | — | 58.4 | 41.0 | <0.001 | 43.2 | 39.0 | 0.449 | 22.0 |
| Blood pressure within control target, %§ | — | 47.9 | 55.0 | 0.029 | 53.8 | 56.5 | 0.647 | 76.9 |
| With favorable HDL cholesterol level, %§ | — | 43.2 | 52.7 | 0.006 | 46.3 | 58.0 | 0.063 | 75.1 |
| With low-risk LDL cholesterol level, %§ | — | 32.1 | 25.2 | 0.037 | 26.3 | 25.2 | 0.850 | 30.7 |
| With low-risk triglyceride level, %§ | — | 44.0 | 62.1 | <0.001 | 59.3 | 65.8 | 0.304 | 80.1 |
| With normal eGFR, % | — | 89.2 | 92.6 | 0.086 | 90.6 | 94.1 | 0.262 | 96.1 |
| Use of noninsulin glucose-lowering drugs, %‖ | — | 66.3 | 63.6 | 0.241 | 59.1 | 68.0 | 0.032 | 0 |
| Insulin use, %‖ | — | 5.2 | 34.7 | <0.001 | 28.0 | 42.2 | <0.001 | 97.1 |
| Use of antihypertensive drugs, %‖ | — | 70.1 | 51.1 | <0.001 | 54.5 | 46.9 | 0.076 | 16.1 |

# 1. Characterization
# 2. Gene information
# 3. Longitudinal Record

# SCI:19.11



**National Diabetes, Patient, & Cause of Death Registers**
Death & vascular outcomes

**National Diabetes & Prescribed Drug Registers**
$HbA_{1c}$, blood pressure, lipids, eGFR & drugs

ESTRID study
**LADA (n=550), type 2 diabetes (n=2,001), population controls (n=2,355); 2007-2019**
GADA, HLA, lifestyle, HOMA, C-peptide

**Adult-onset type 1 diabetes** in National Diabetes Register (n=1,573)
**2007-2015**
$HbA_{1c}$, blood pressure, lipids, eGFR

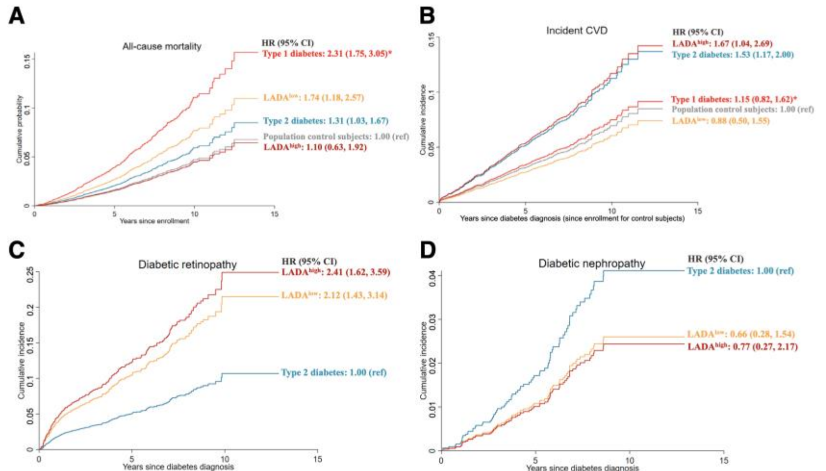
**National Prescribed Drug Register**
Drugs

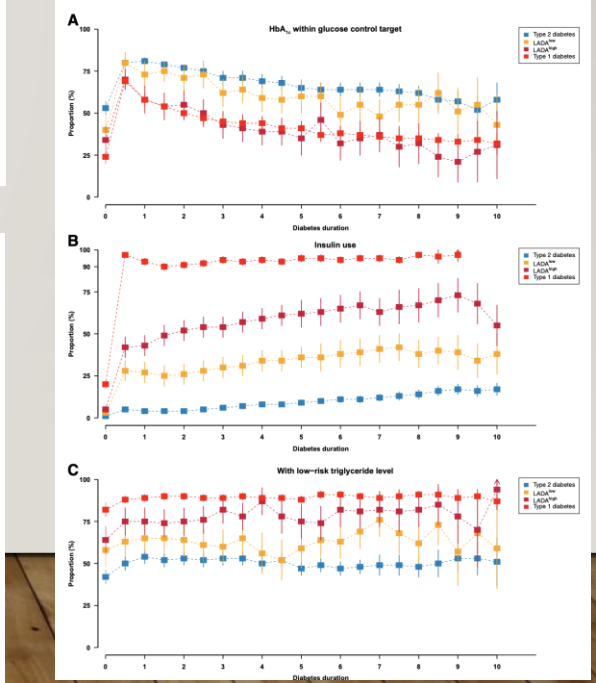
**All-cause mortality ***
LADA (n=550), type 2 diabetes (n=2,001), population control subjects (n=2,355)

**Incident CVD ***
LADA (n=486), type 2 diabetes (n=1,639), population control subjects (n=2,165)

**Microvascular diseases ***
LADA (n=527), type 2 diabetes (n=1,955)

**Trajectories† of $HbA_{1c}$, blood pressure, lipids, eGFR & drugs**
LADA, type 2 diabetes, type 1 diabetes

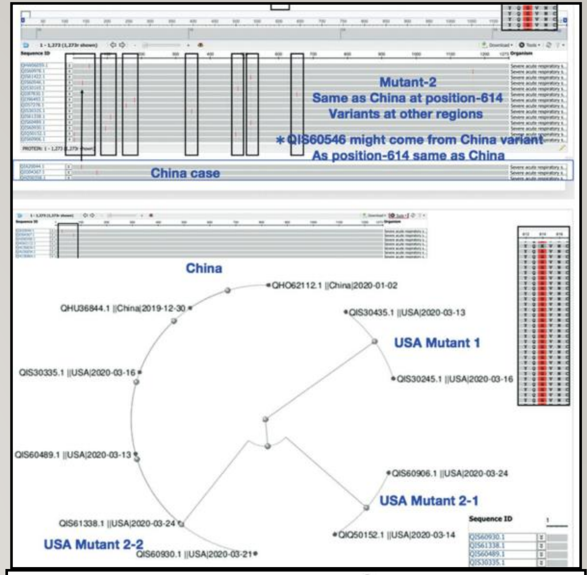


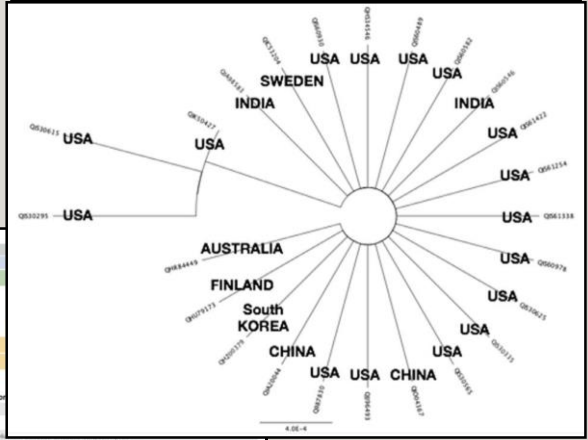


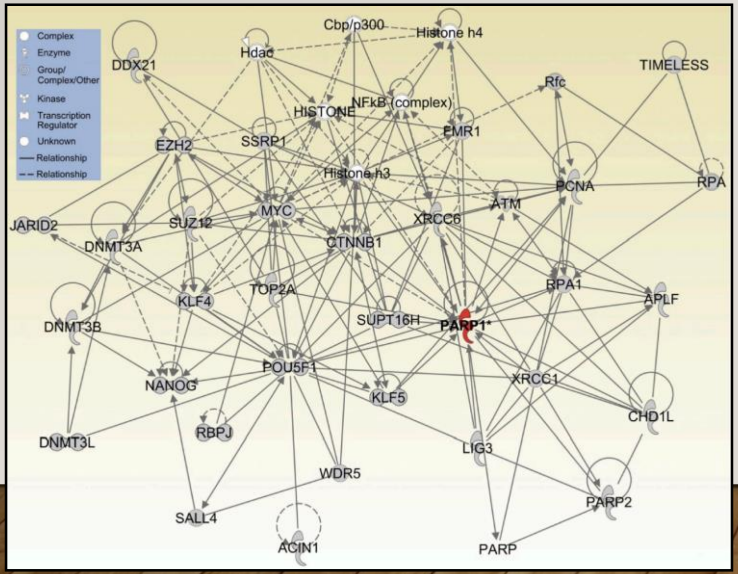
受基因影響的疾病
But，人數要夠多 夠久

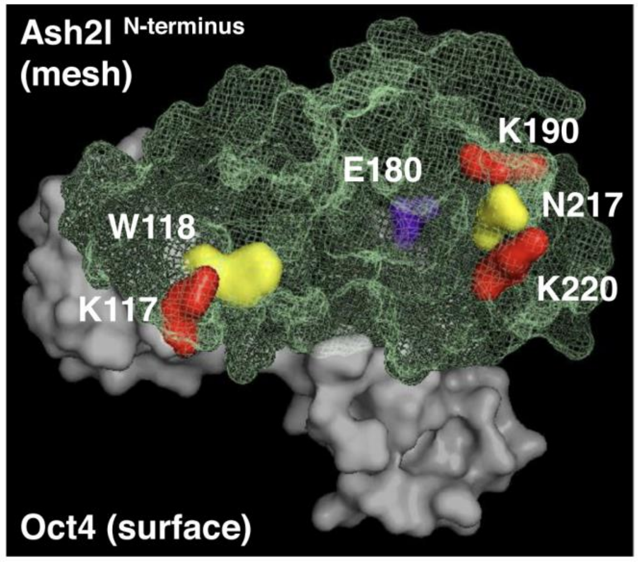
# 11/22

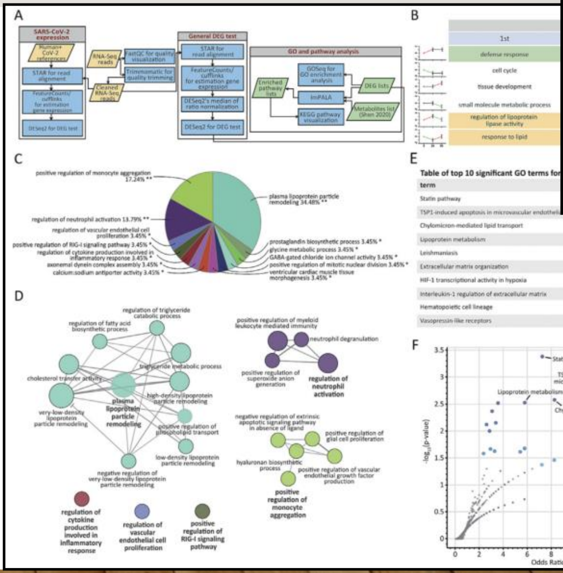
# 大數據實戰分享：大師成長之路』之六：


Mutant-2
Same as China at position-614
Variants at other regions
* QIS60546 might come from China variant
As position-614 same as China
China case
China
QHO62112.1 ||China|2020-01-02
QHU36844.1 ||China|2019-12-30
QIS30435.1 ||USA|2020-03-13
USA Mutant 1
QIS30335.1 ||USA|2020-03-16
QIS30245.1 ||USA|2020-03-16
QIS60489.1 ||USA|2020-03-13
QIS60906.1 ||USA|2020-03-24
USA Mutant 2-1
QIS61338.1 ||USA|2020-03-24
USA Mutant 2-2
QIQ50152.1 ||USA|2020-03-14
QIS60930.1 ||USA|2020-03-21
Sequence ID

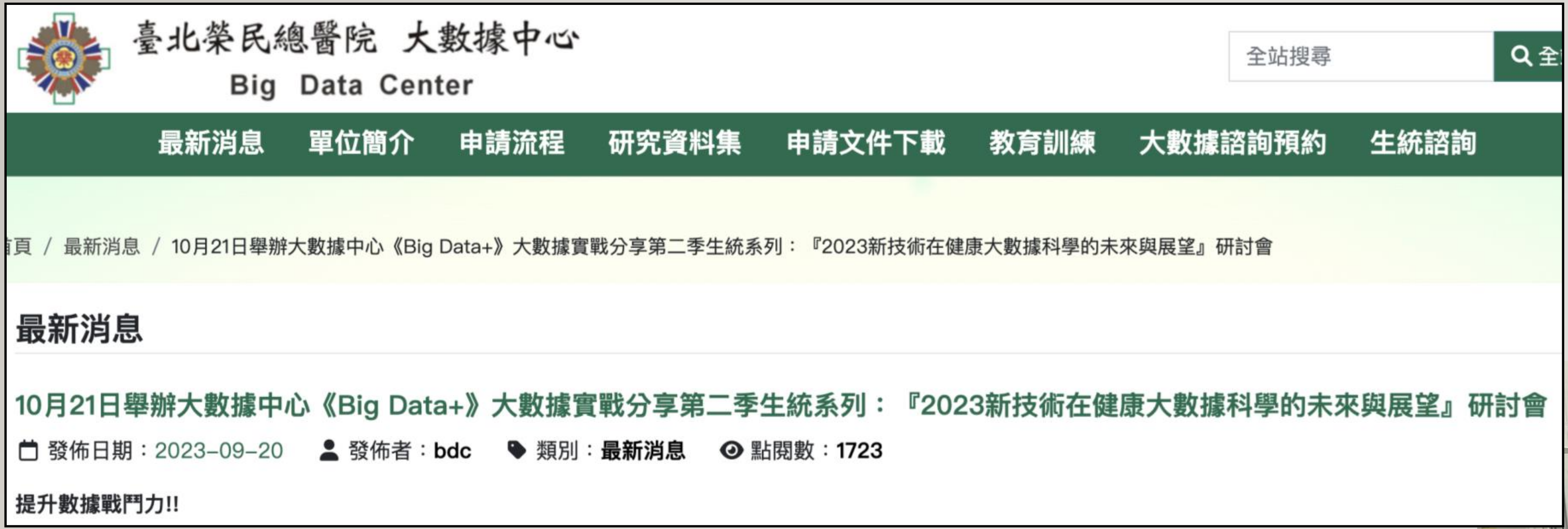

USA
USA
USA
USA
SWEDEN
INDIA
INDIA
USA
USA
USA
USA
USA
AUSTRALIA
FINLAND
South KOREA
CHINA
USA
USA
CHINA
USA
USA
USA
USA
4.0E-4


Complex
Enzyme
Group/Complex/Other
Kinase
Transcription Regulator
Unknown
Relationship
Relationship
Cbp/p300
Histone h4
Hdac
DDX21
TIMELESS
Rfc
NFkB (complex)
HISTONE
EMR1
EZH2
SSRP1
Histone h3
PCNA
RPA
MYC
ATM
XRCC6
JARID2
SUZ12
CTNNB1
DNMT3A
RPA1
TOP2A
APLF
KLF4
SUPT16H
PARP1*
DNMT3B
POU5F1
XRCC1
NANOG
KLF5
CHD1L
RBPJ
LIG3
DNMT3L
WDR5
PARP2
SALL4
ACIN1
PARP


Ash2l N-terminus (mesh)
K190
E180
N217
W118
K220
K117
Oct4 (surface)



# 北榮生統小組可以幫幫你


全站搜尋
全
最新消息
單位簡介
申請流程
研究資料集
申請文件下載
教育訓練
大數據諮詢預約
生統諮詢

臺北榮民總醫院 大數據中心

Big Data Center

頁 / 最新消息 / 10月21日舉辦大數據中心《Big Data+》大數據實戰分享第二季生統系列：『2023新技術在健康大數據科學的未來與展望』研討會

**最新消息**

**10月21日舉辦大數據中心《Big Data+》大數據實戰分享第二季生統系列：『2023新技術在健康大數據科學的未來與展望』研討會**

發佈日期：2023-09-20 發佈者：**bdc** 類別：**最新消息** 點閱數：**1723**

**提升數據戰鬥力!!**

蔡秉興 高級助理研究員 (院內分機：2620)

醫學科技大樓一樓 生統小組辦公室

biostat@vghtpe.gov.tw