Bioinformatics and R: Visualising Genomic Data

Yari Ciani, PhD

Laboratory of Computational and Functional Oncology

Department for Cellular, Computational and Integrative Biology - CIBIO
University of Trento

yari.ciani@unitn.it

UNIVERSITÀ DI TRENTO | Department of Cellular, Computational and Integrative Biology - CIBIO

30/05/2025

DEMICHELIS LABORATORY

Laboratory of Computational and Functional Oncology

The Topic

Precision Oncology and Biomarker discovery

Diagnostic, prognostic, treatment response markers for tumor stratification and precision oncology
Hypothesis driven and agnostic studies

The Team

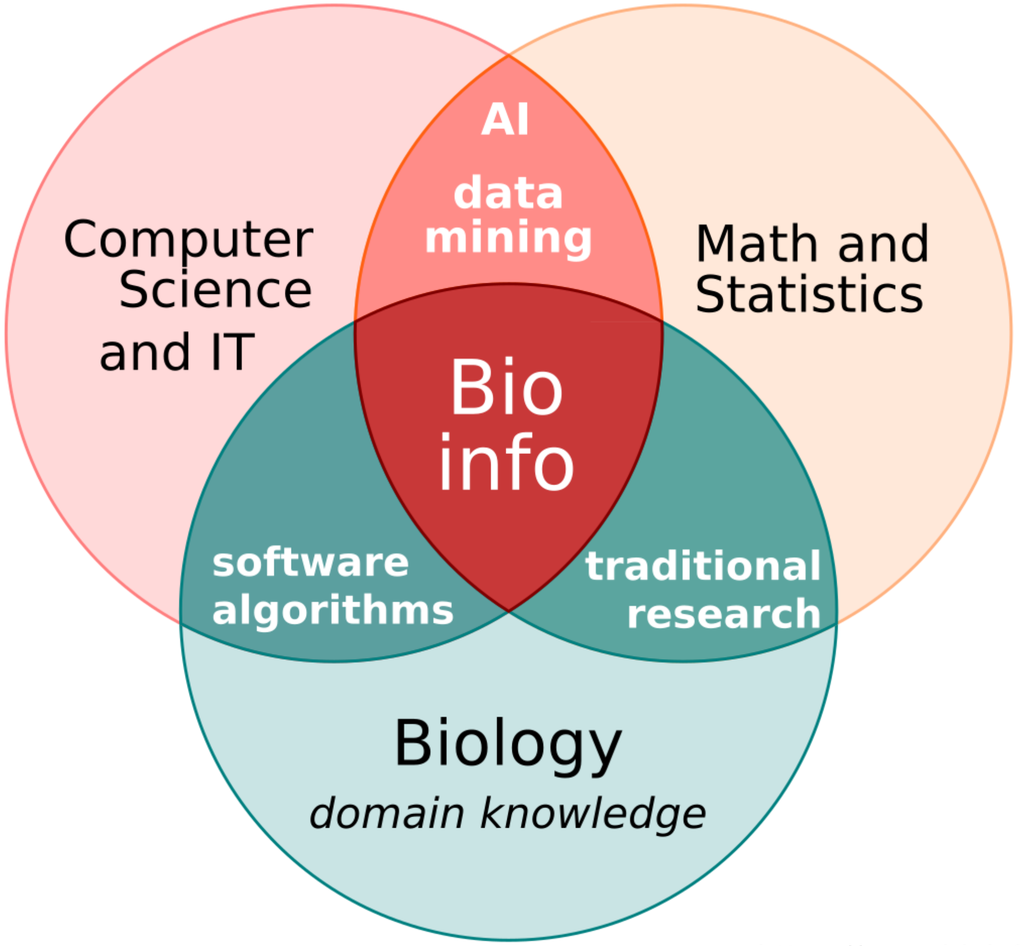
The Funding

MIUR
FONDAZIONE CARITRO
CASSA DI RISPARMIO DI TRENTO E ROVERETO
Weill Cornell Medicine
P50CA211024
erc European Research Council
FONDAZIONE SAMUEL & BARBARA STERNBERG PER LA CURA E LA RICERCA ONCOLOGICA
Prostate Cancer Foundation
Curing Together.
F.T.R.T.
FONDAZIONE VRT VALORIZZAZIONE RICERCA TRENTINA
CANCER RESEARCH UK
AIRC
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
UNIVERSITÀ DI TRENTO

What is Bioinformatics?

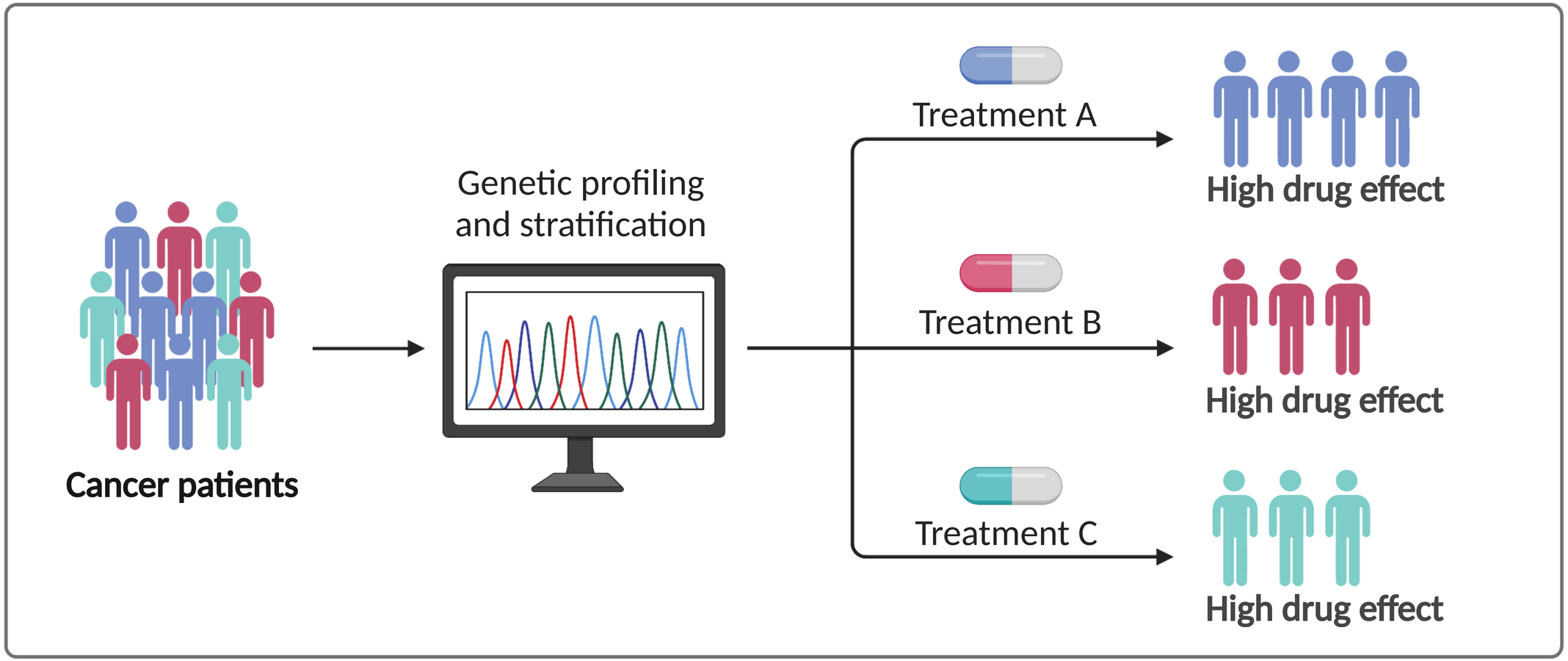
Bioinformatics: *applies principles of information sciences and technologies to make the vast, diverse, and complex life sciences data understandable and useful.*

e.g., Text mining,
Processing raw data,
Artificial Intelligence,
Software development,
but also **experiment design**
and **results interpretation**



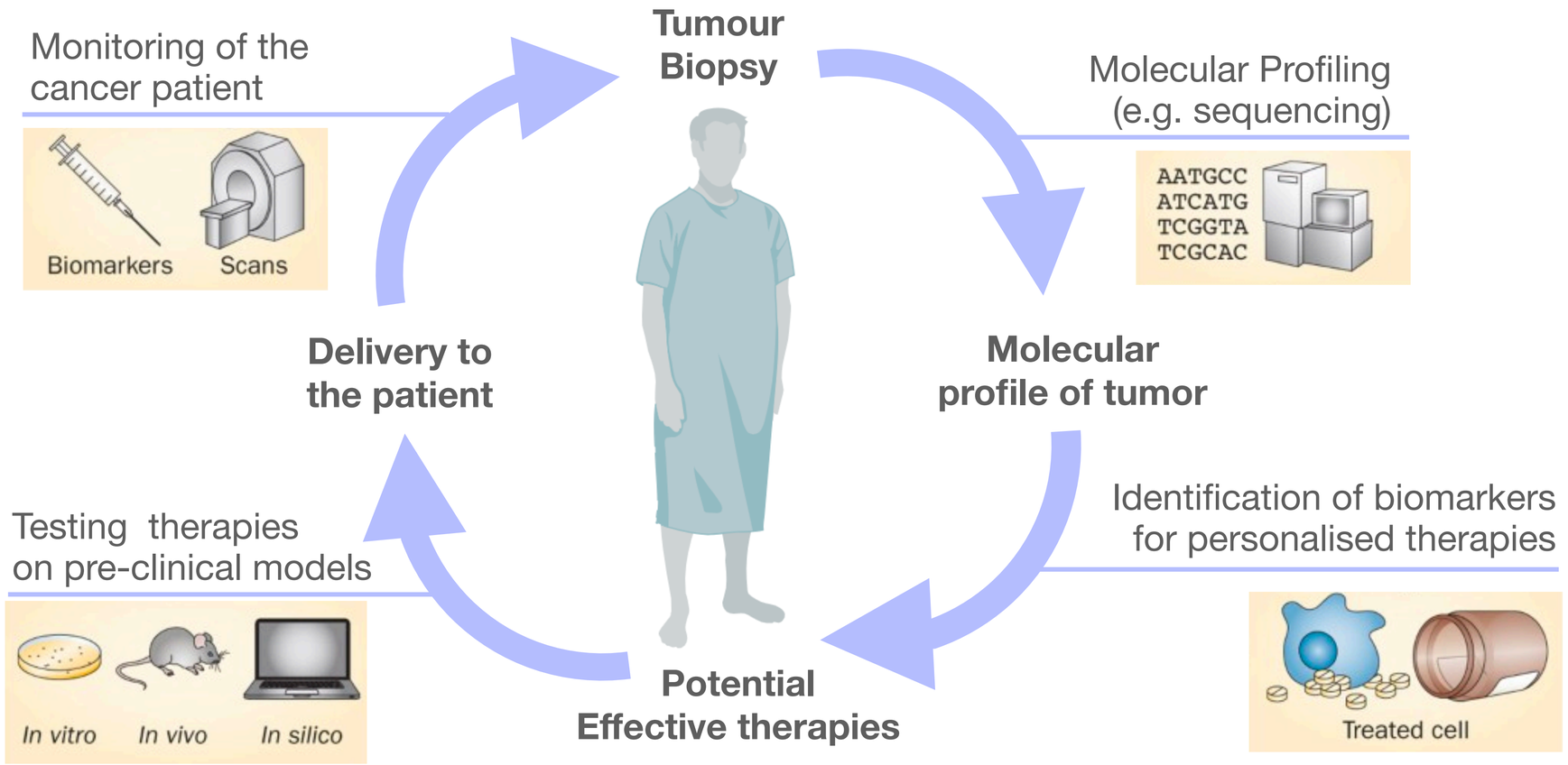
Precision Oncology

Precision cancer therapy



Designed by Yari Ciani

The Precision Drug Discovery Cycle

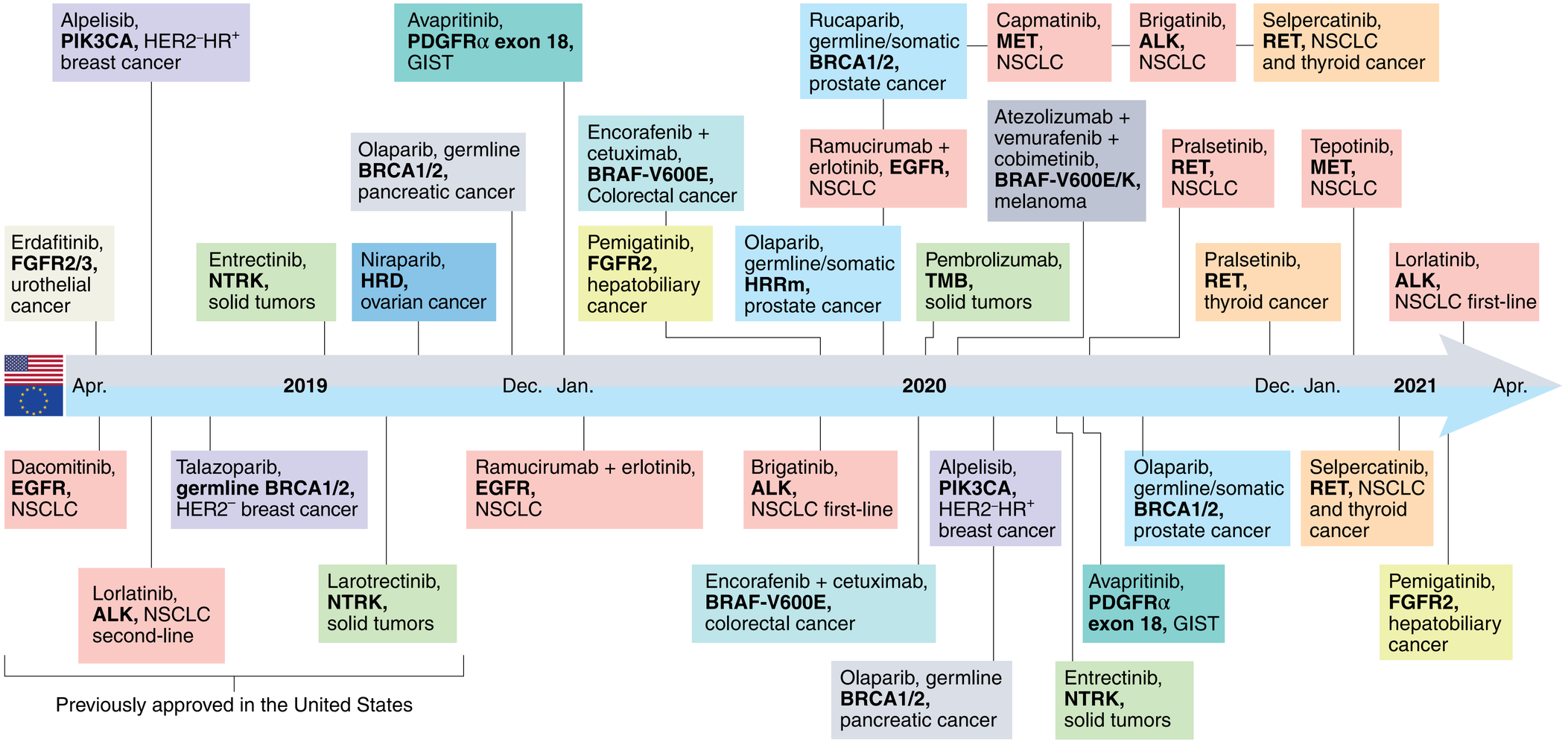


The cost of developing a single FDA/EMA-approved drug:
1B $
and
10 - 15 years

FDA: Food and Drug Administration
EMA: European Medicines Agency

Adapted by T Cantore from Shrager, Jeff, and Jay M. Tenenbaum. "Rapid learning for precision oncology." *Nature reviews Clinical oncology* 11.2 (2014): 109-118.

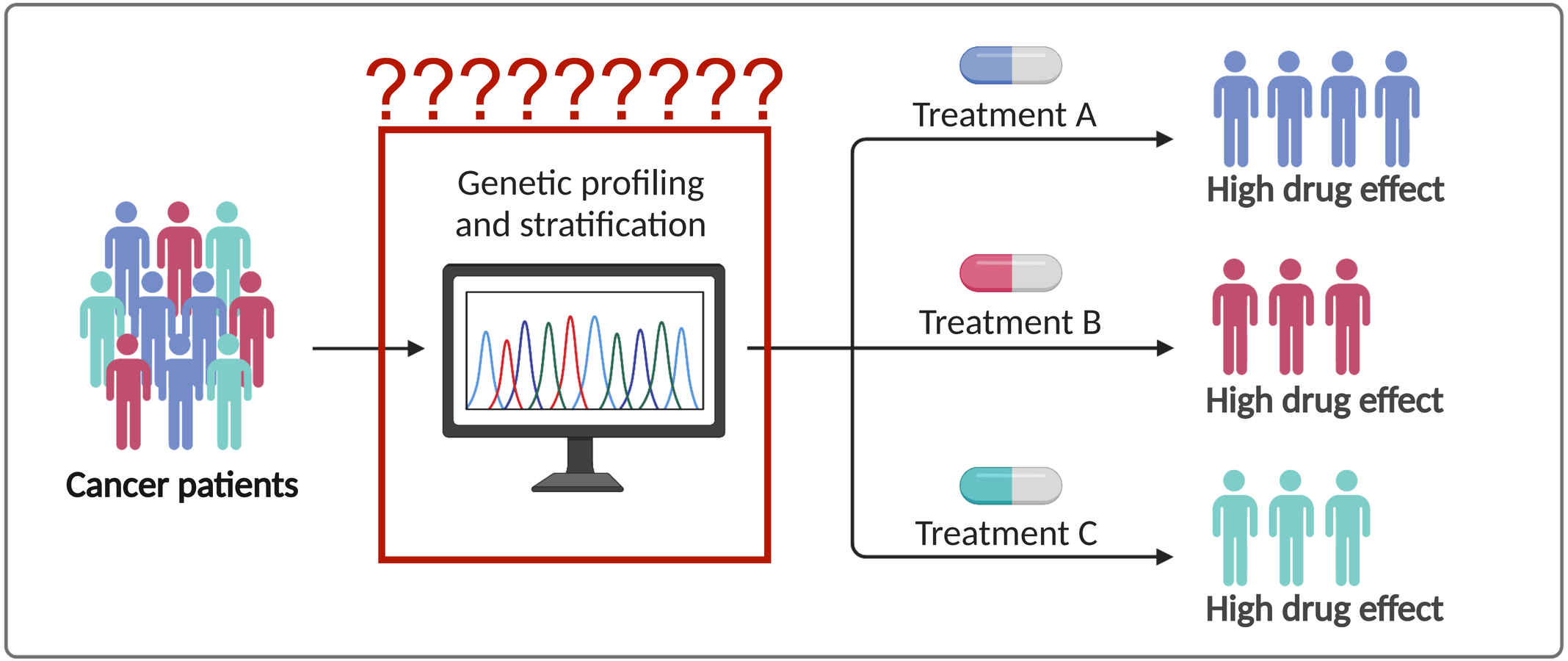
…but it is worth



Delivering precision oncology to patients with cancer, Mateo et al, Nature Medicine 2022

Precision Oncology

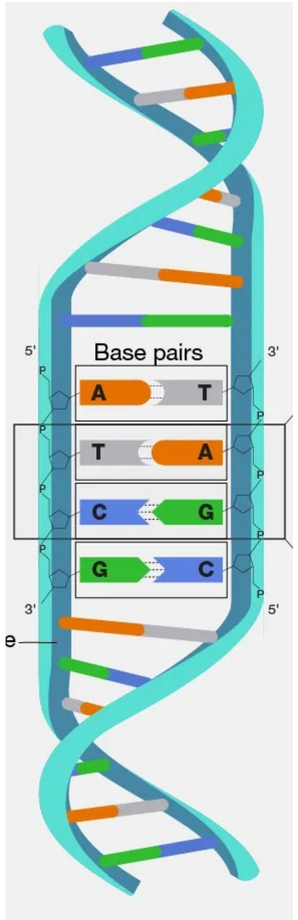
Precision cancer therapy



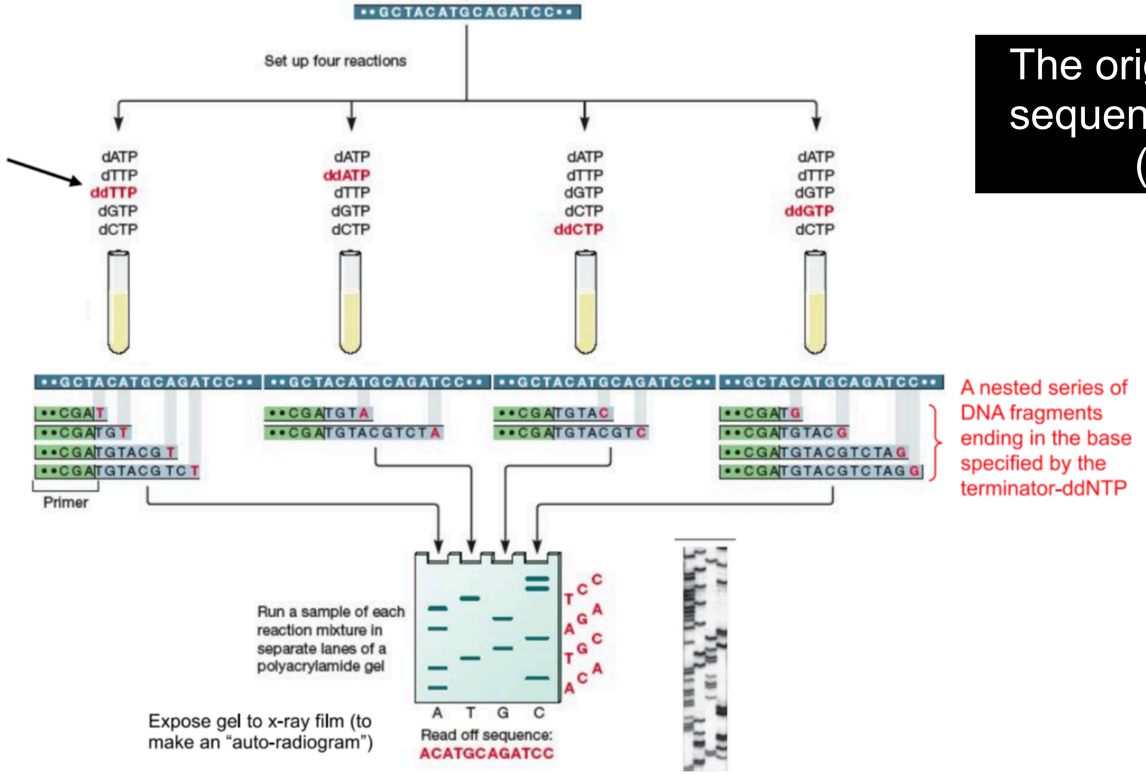
Designed by Yari Ciani

The origin of DNA sequencing

DNA Sequencing is figuring out the order of DNA nucleotides, or bases (A T G C), in a genome that make up an organism's DNA.



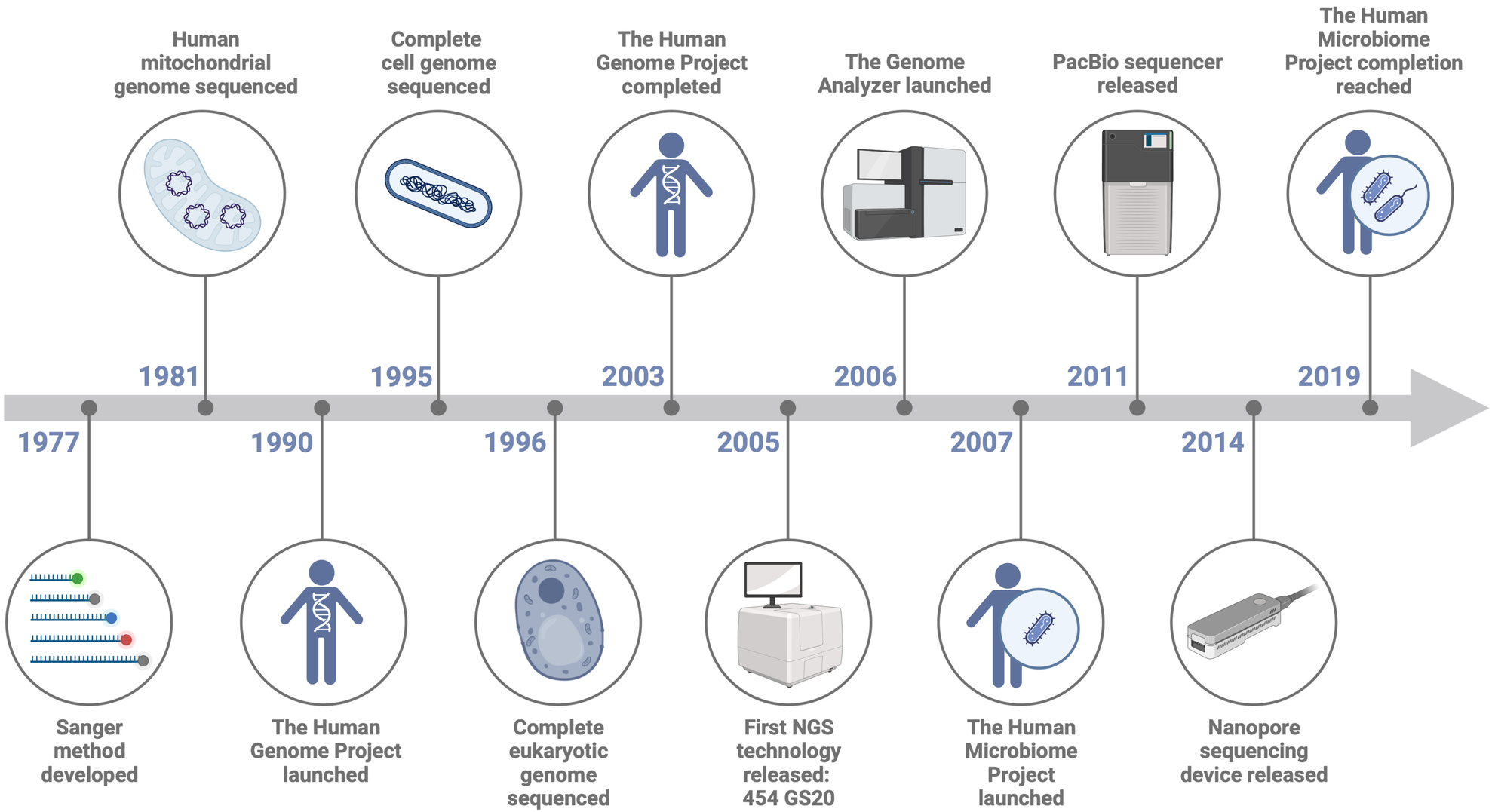
ddNTP are at low concentrations to permit elongation of fragments



The original Sanger sequencing method (1977)

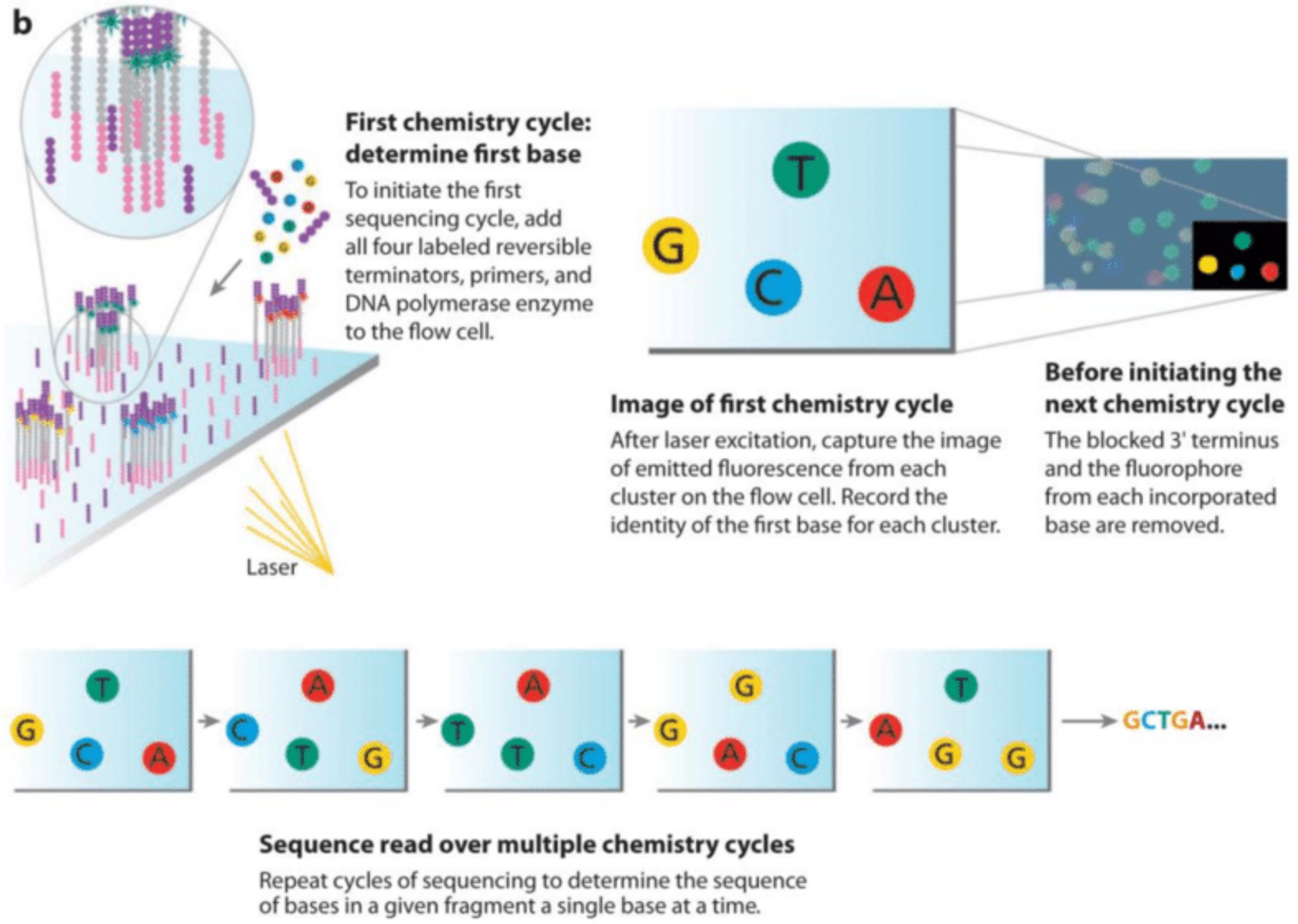
Credits to Curtis Huttenhower, https://slideplayer.com/slide/10893306/

The NGS revolution



NGS approaches in a nutshell

Illumina - Sequencing by synthesis



https://www.researchgate.net/profile/Kamila-Knapik/publication/340535984/figure/fig10/AS:878431982063629@1586445913332/The-principle-of-Illumina-sequencing-process-A-DNA-is-converted-into-an-Illumina.png

How to get from NGS data to biological interpretation?

Up to 8B read pairs (2x150bp)

BASECALLING

ACGTCGATCGATCGATCGATCG

TCGATCGCGCGAGATGGCTGAA

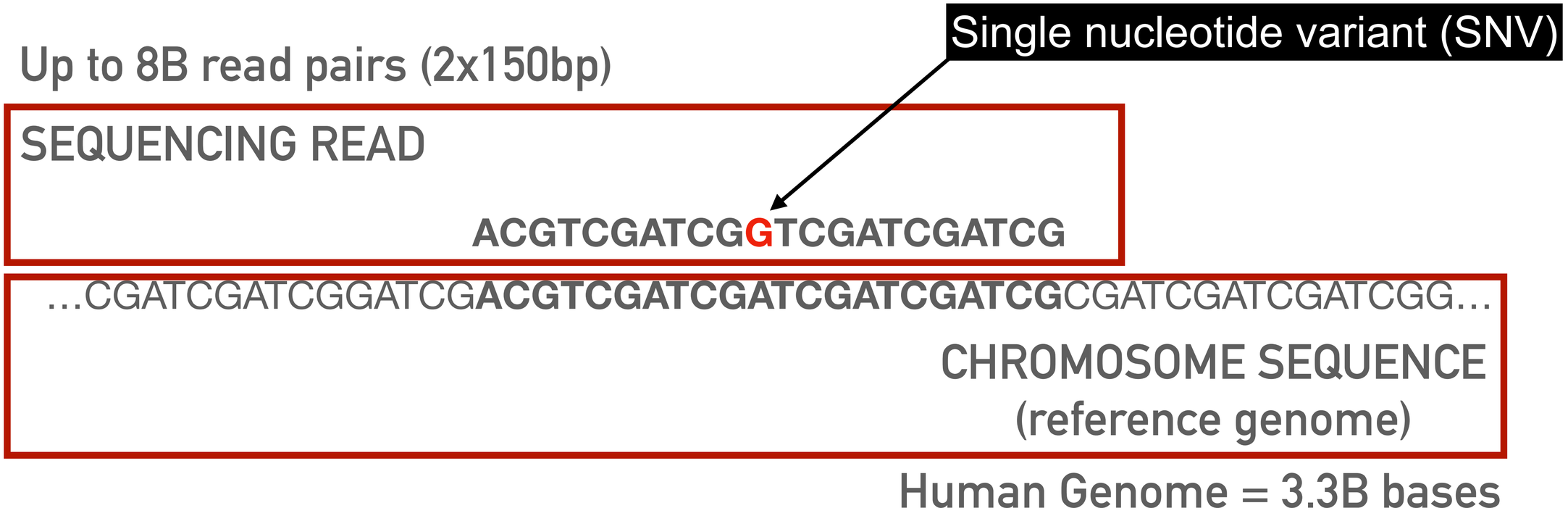
CGAGCTAGCTAGCTGGCTAGAGCT

CAGCGAGCTAGCTAGCATCGAT

CGATGCTAGCTAGCTAGCTAGC

- Sequencing produces high-resolution TIFF images
- 100 tiles per lane, 8 lanes per flow cell, 100 cycles
- 4 images (A,G,C,T) per tile per cycle = 320,000 images
- Each *TIFF* image ~ 7Mb = 2,240,000 Mb of data (**2.24TB**)

How to get from NGS data to biological interpretation?

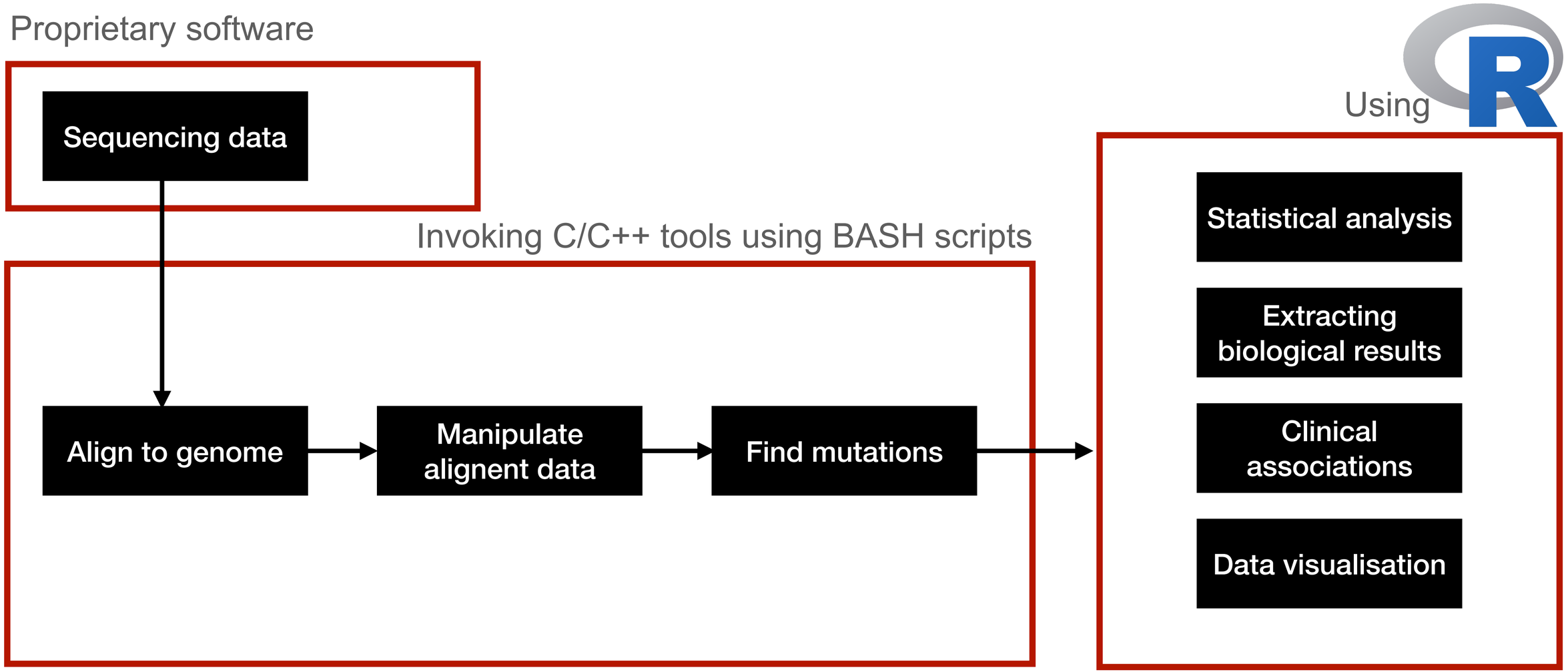


Approximate String Matching with Bounded Edit Distance, where the goal is to find all (locally) similar substrings of a large reference that align to a given read, allowing for a limited number of edits.

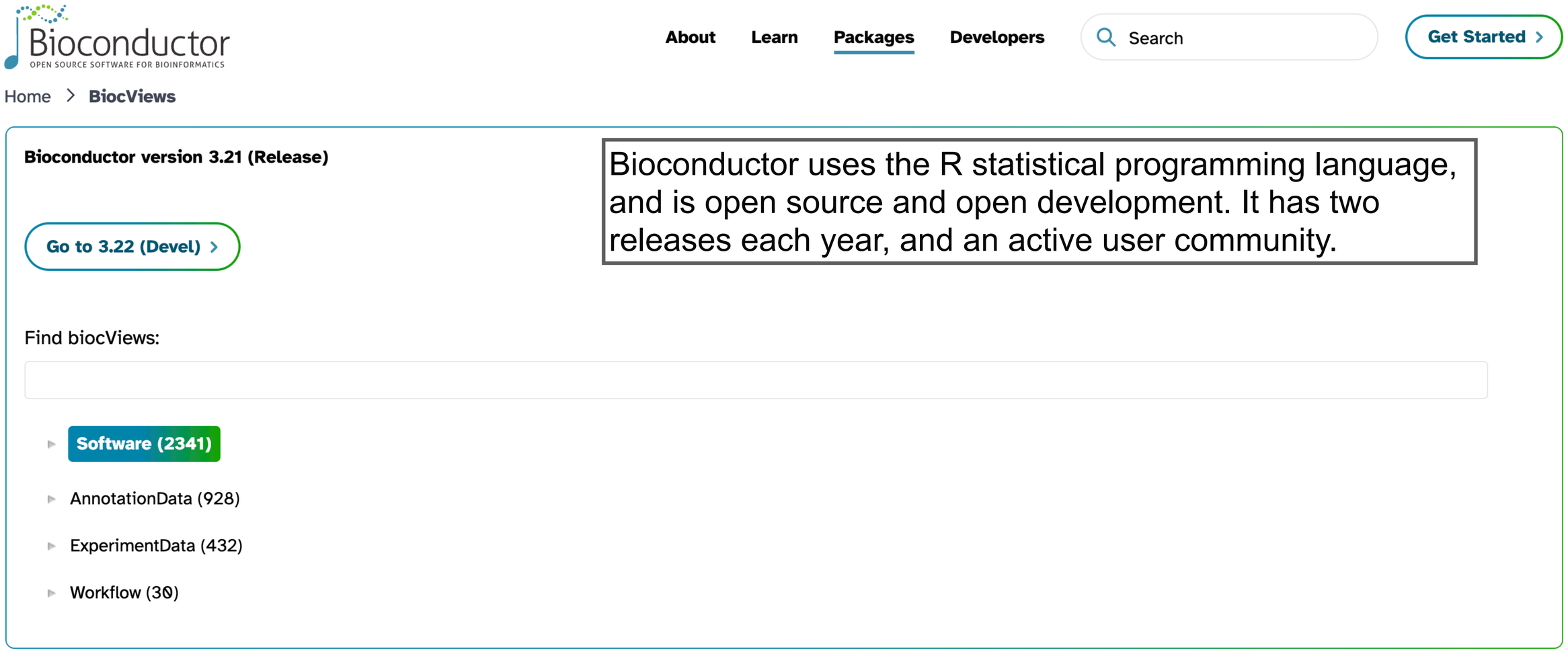
R is not optimized for low-level memory access or speed.

Genomic-scale alignment is typically done in **C/C++** (e.g., BWA, minimap2) due to performance constraints.

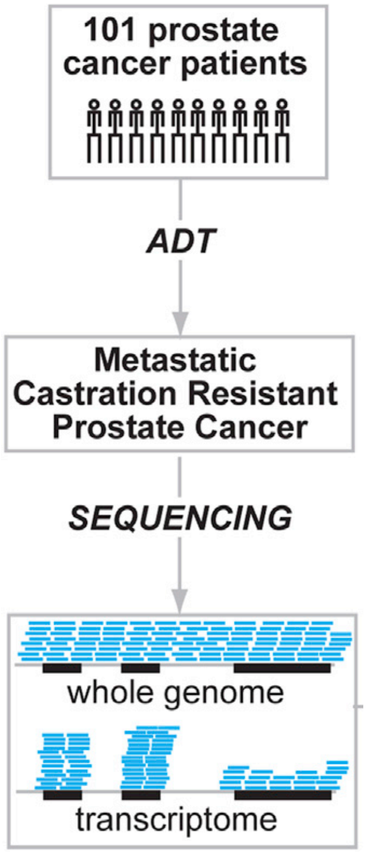
The role of R in Bioinformatics

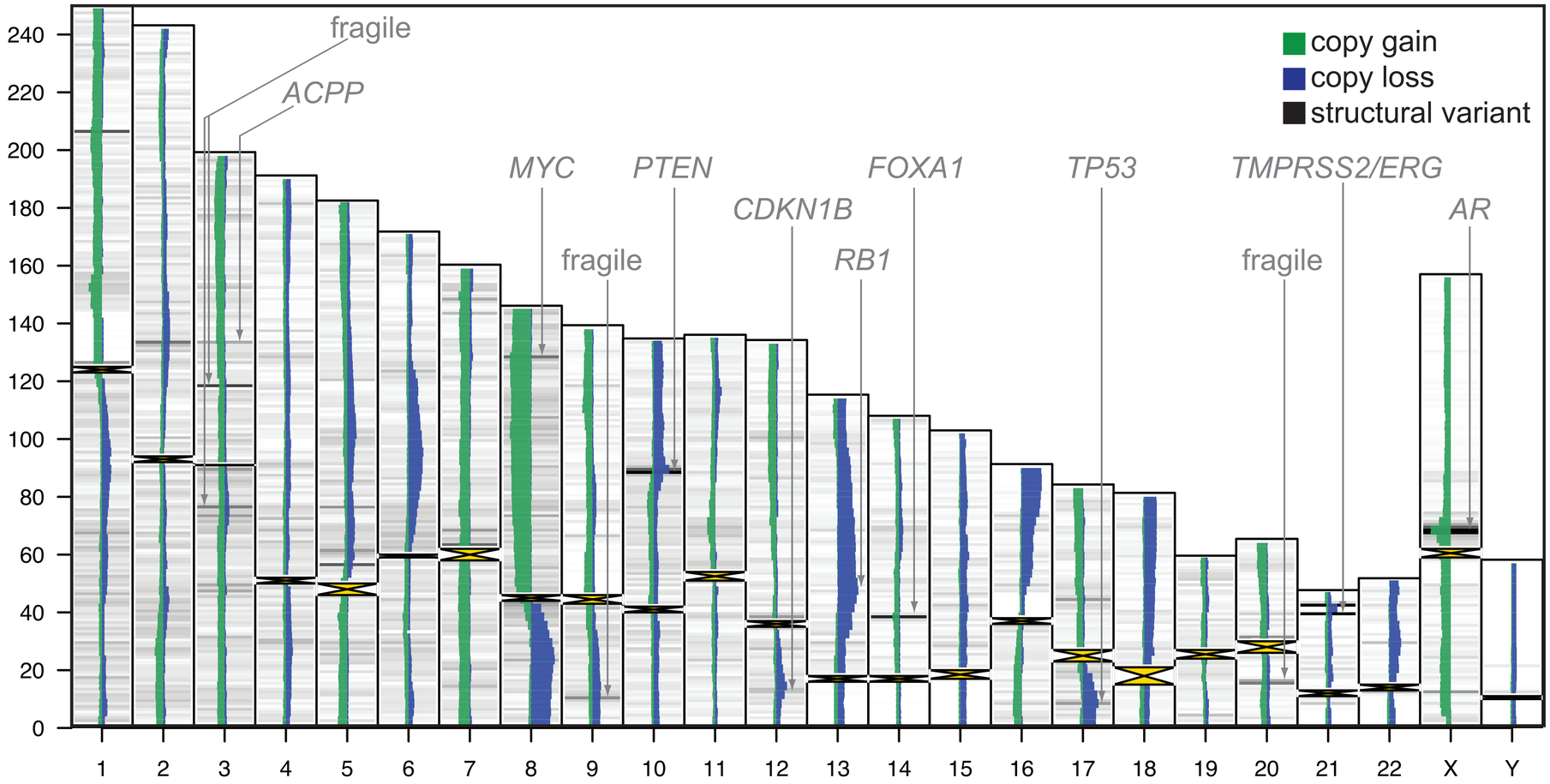


The R Bioinformatics community



Visualising the Genome: an example

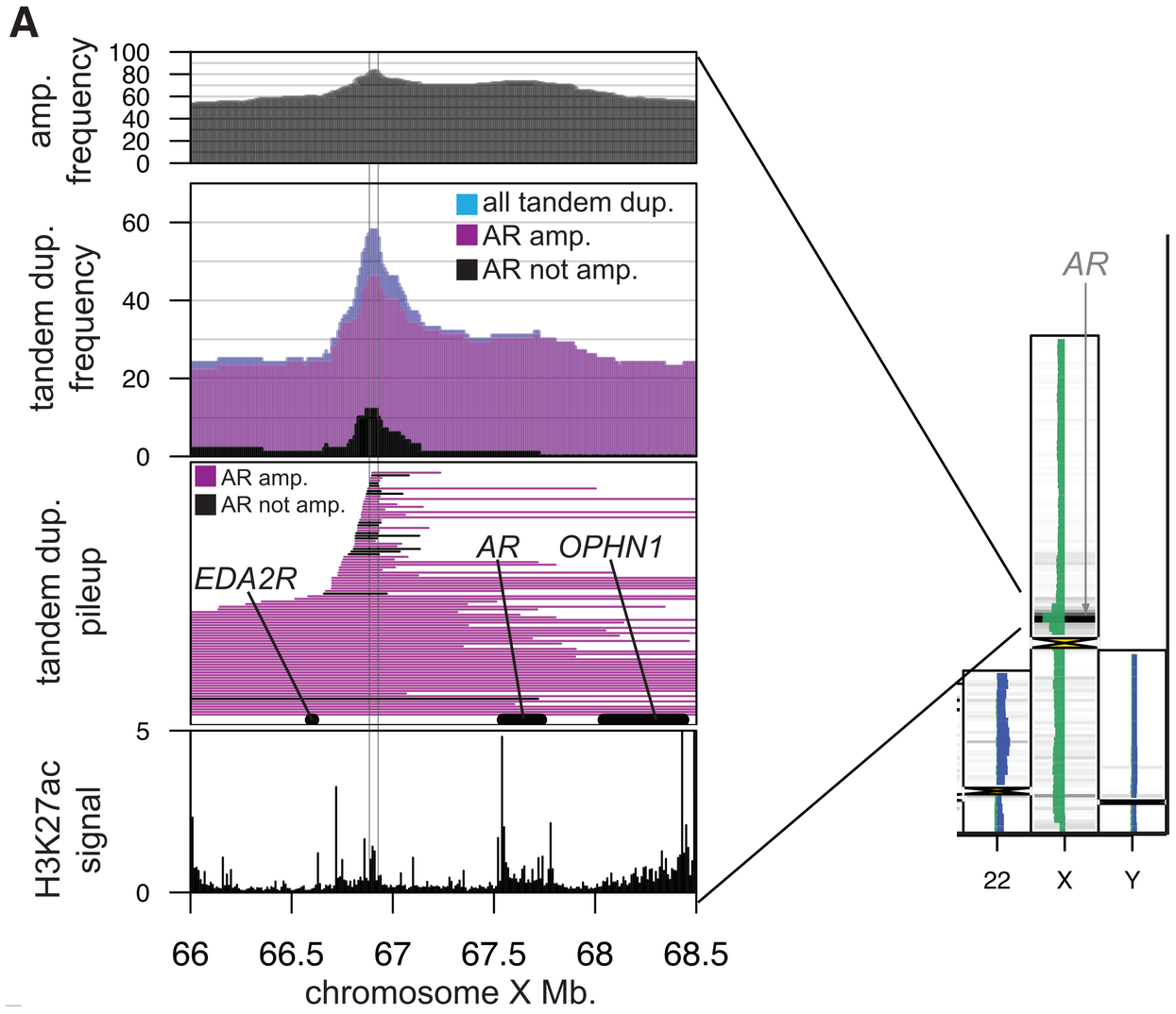




Showing frequencies of mutations of an entire cohort across the entire genome

Quigley et al., 2018, Cell

From Overview to Detail



AR = androgen receptor. Gene involved in prostate cancer

Quigley et al., 2018, Cell

From Overview to Detail

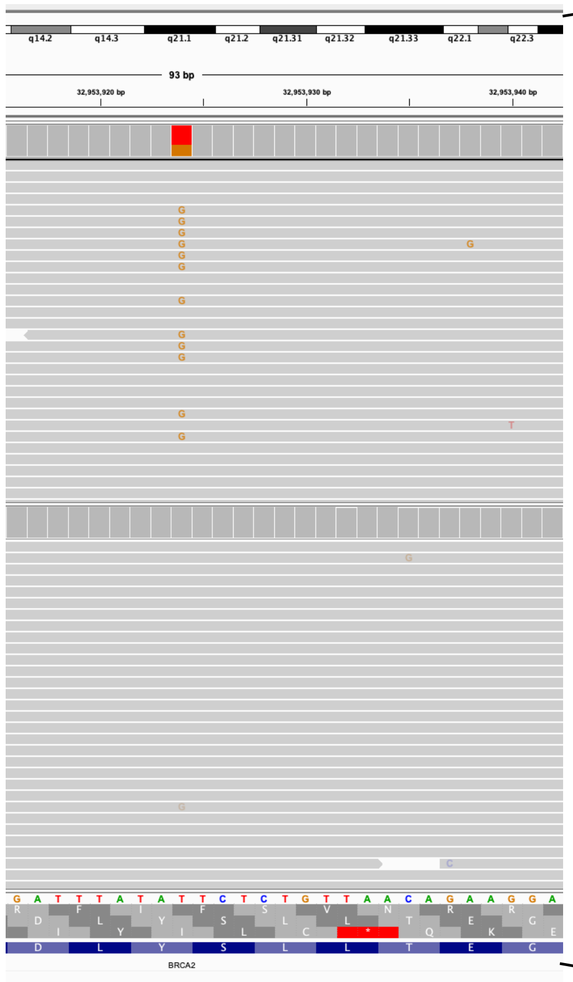
CONS:

-graphical interface

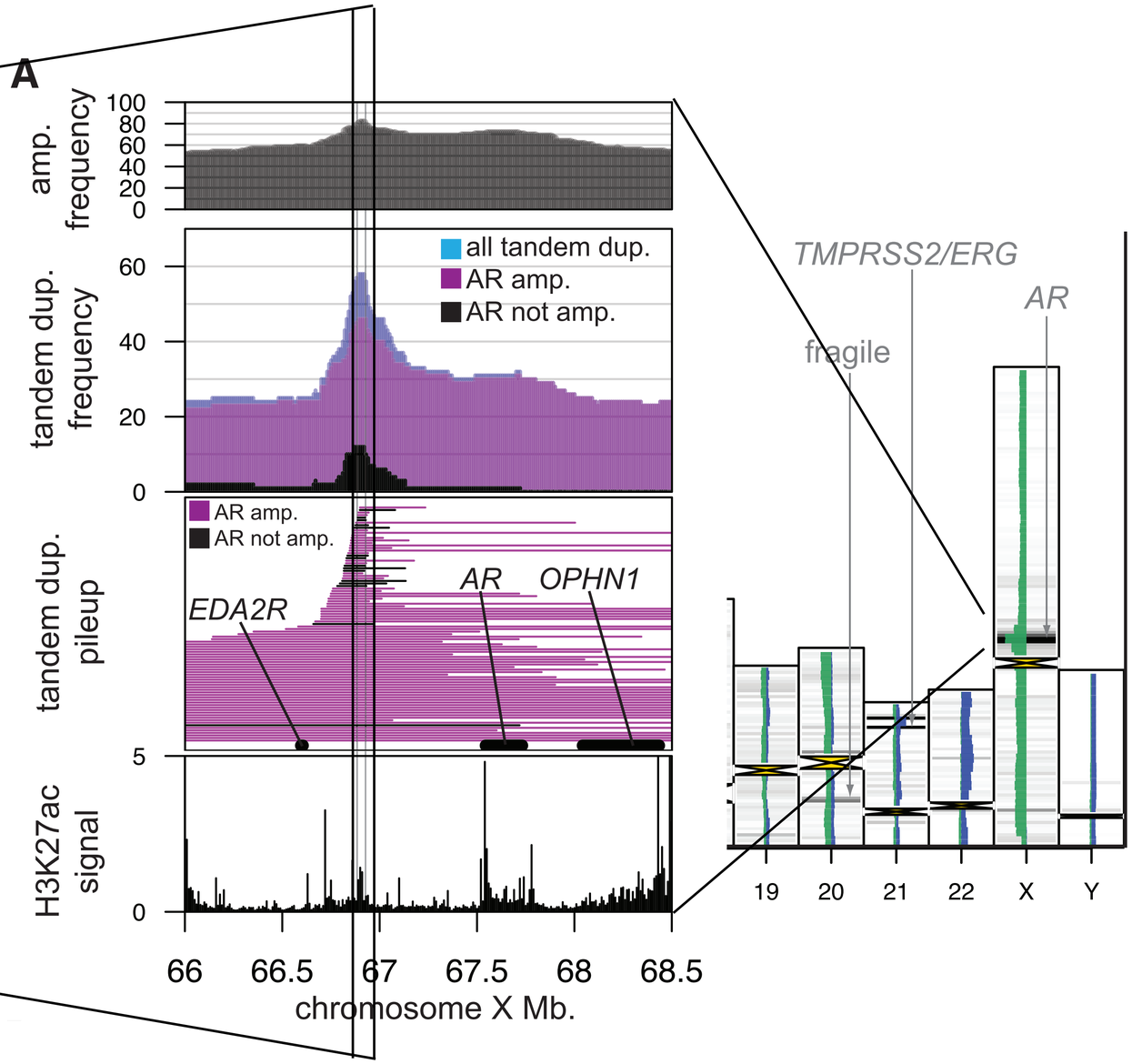
-slow

-graphical customisation

-JAVA!

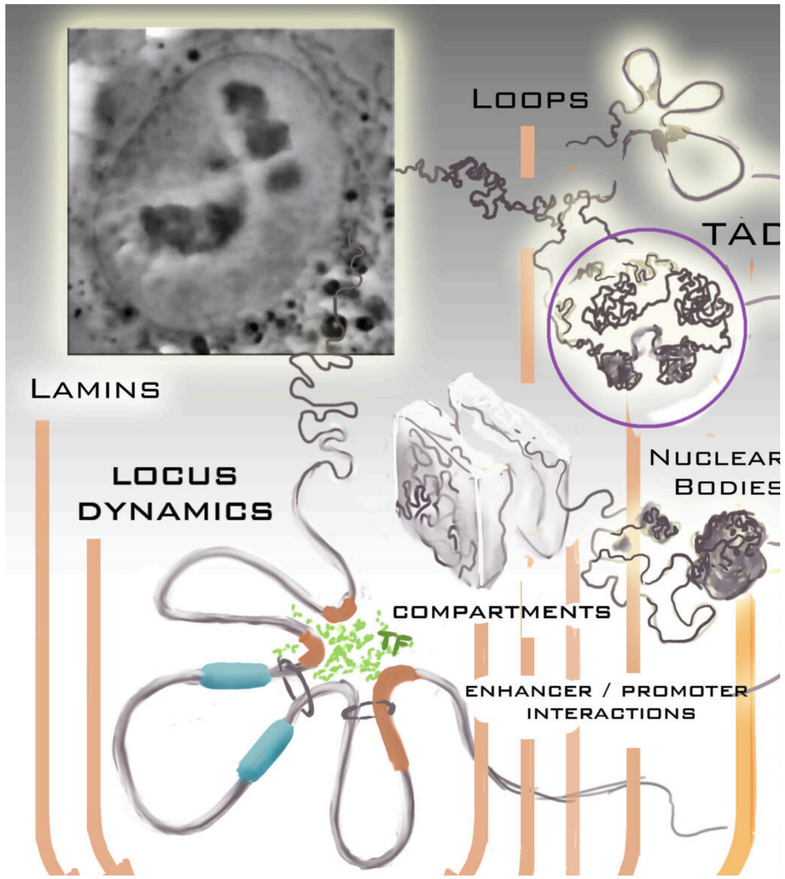


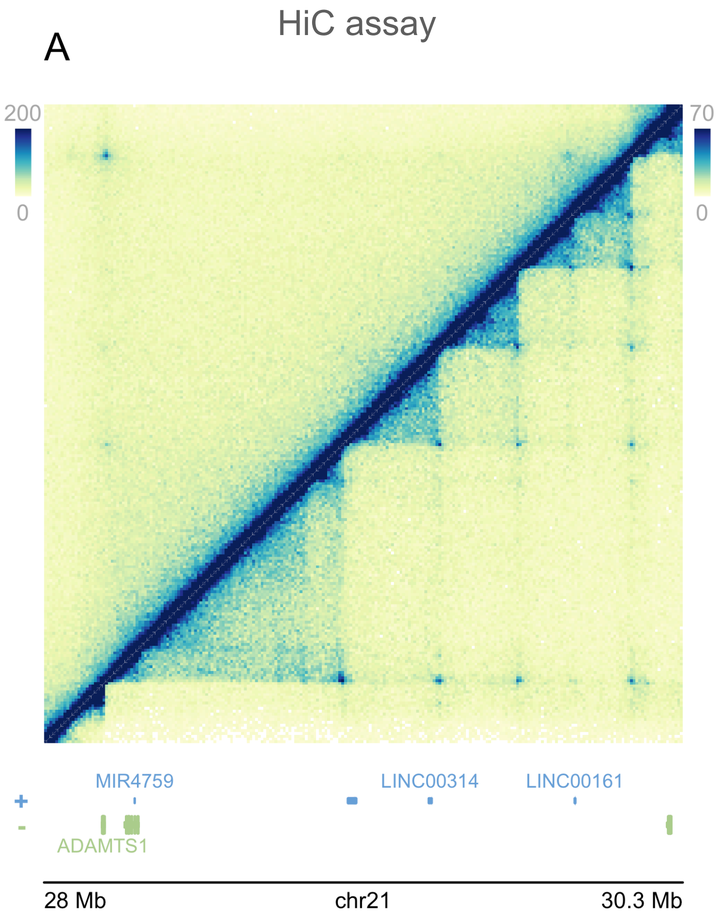
Screenshot from IGV, a Java based alignment visualisation tool

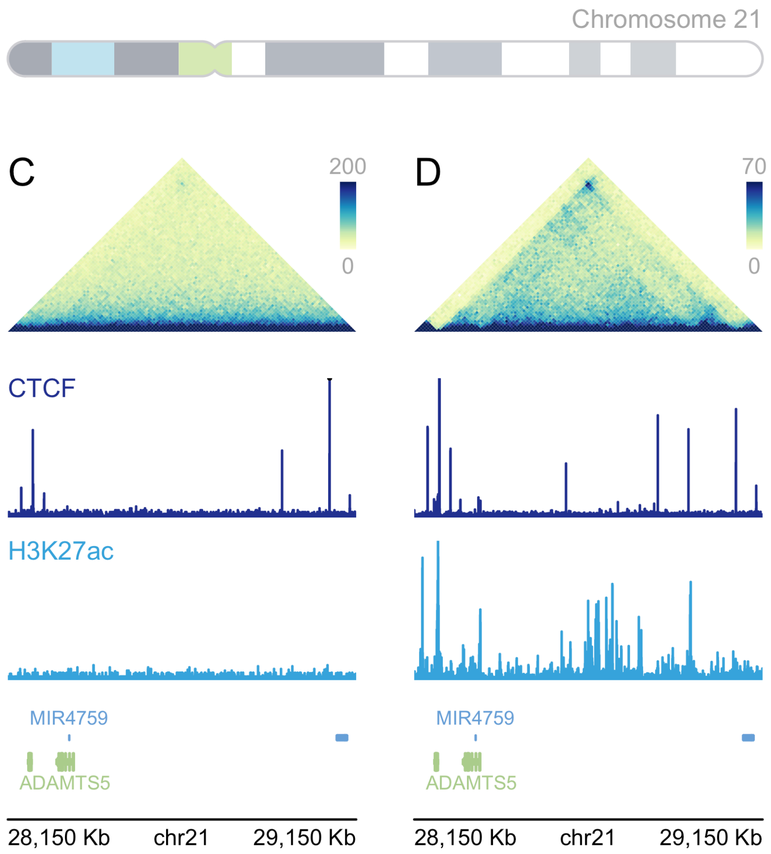


Quigley et al., 2018, Cell

Long Range Interactions (physical)







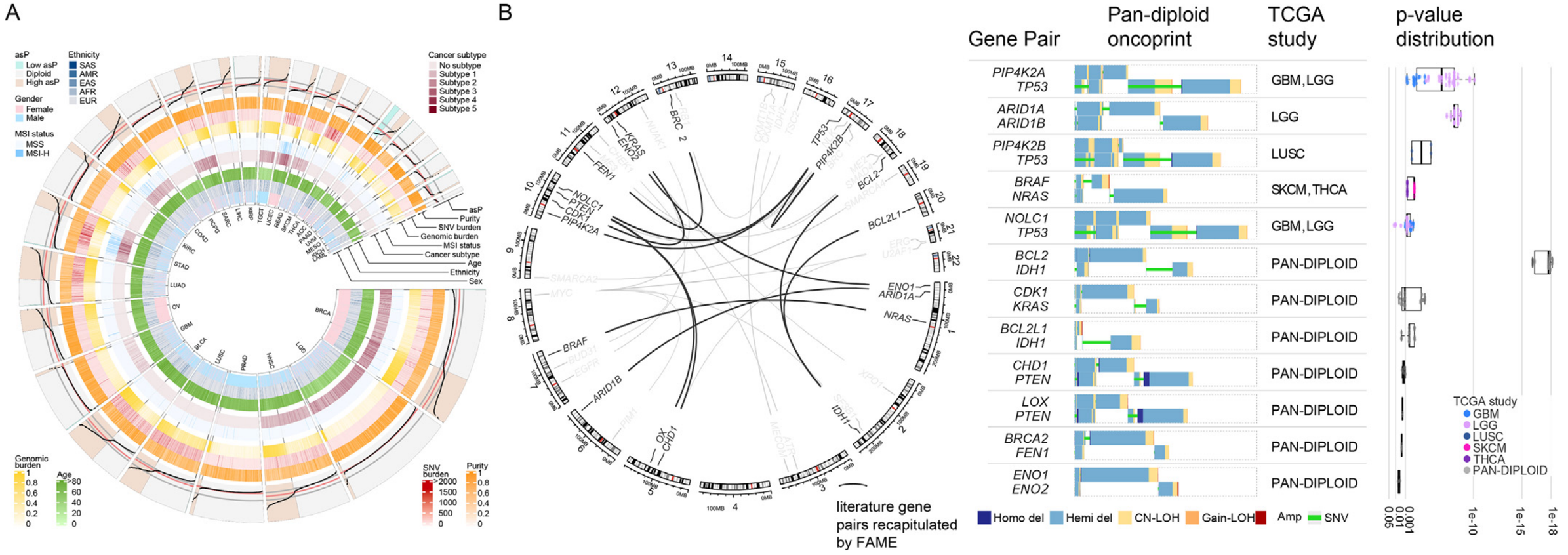
https://4dnucleome.org/

https://phanstiellab.github.io/plotgardener/

Long Range Interactions (functional)

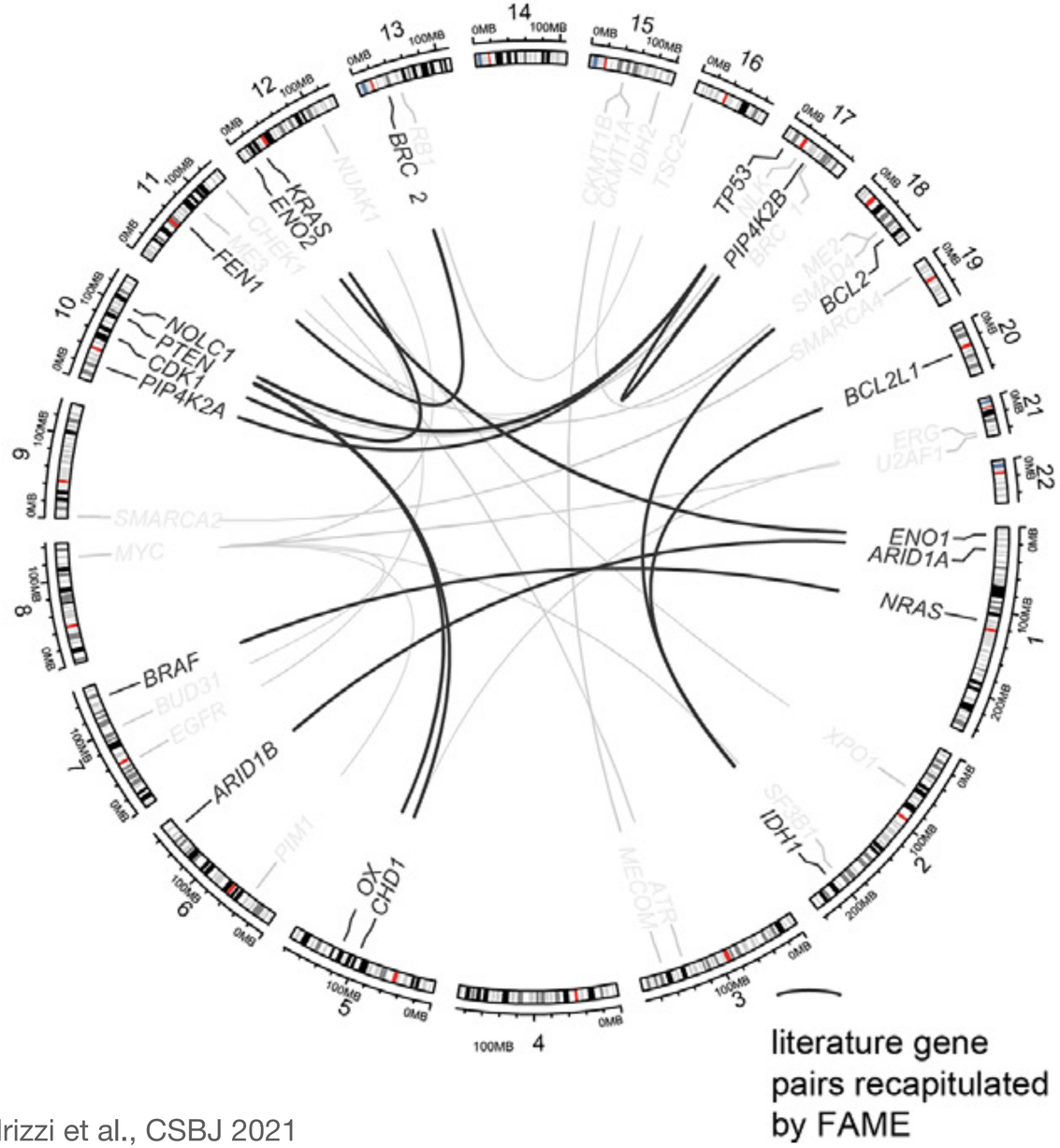
T. Fedrizzi, Y. Ciani, F. Lorenzin et al.

Computational and Structural Biotechnology Journal 19 (2021) 4394–4403



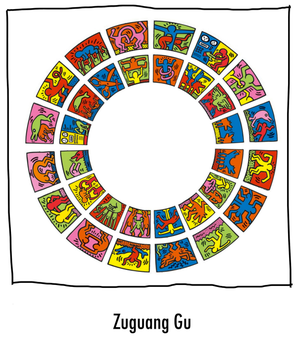
Designed by Tarcisio Fedrizzi, ex-group member

Long Range Interactions (functional)



Fedrizzi et al., CSBJ 2021

***circlize* implements and enhances circular visualization in R**
Bioinformatics, 2014



https://jokergoo.github.io/circlize_book/book/index.html

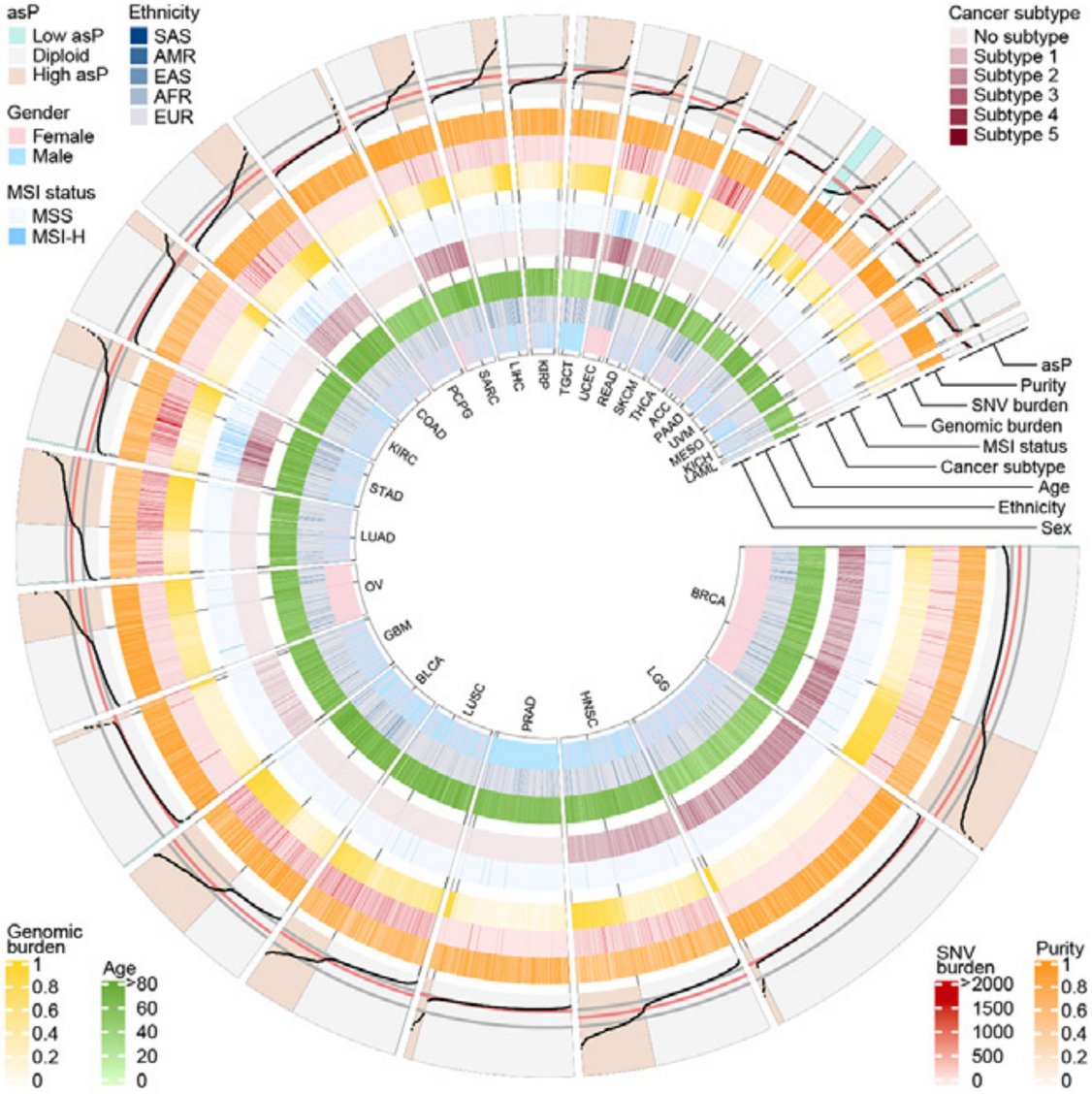
```
circos.initializeWithIdeogram()
circos.genomicLink(bed1, bed2)
```

bed1

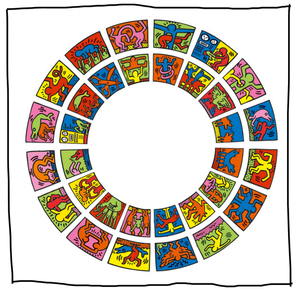
chr	start	end	value1
chr6	102324459	147617643	-0.50418830
chr17	65167455	77619820	-0.10264963
chr11	13366995	32331617	0.42482152
chr8	93343457	96256710	0.65620649
chr16	64403195	65047798	0.04966380

Summarising Entire Cohorts

~4800 patients



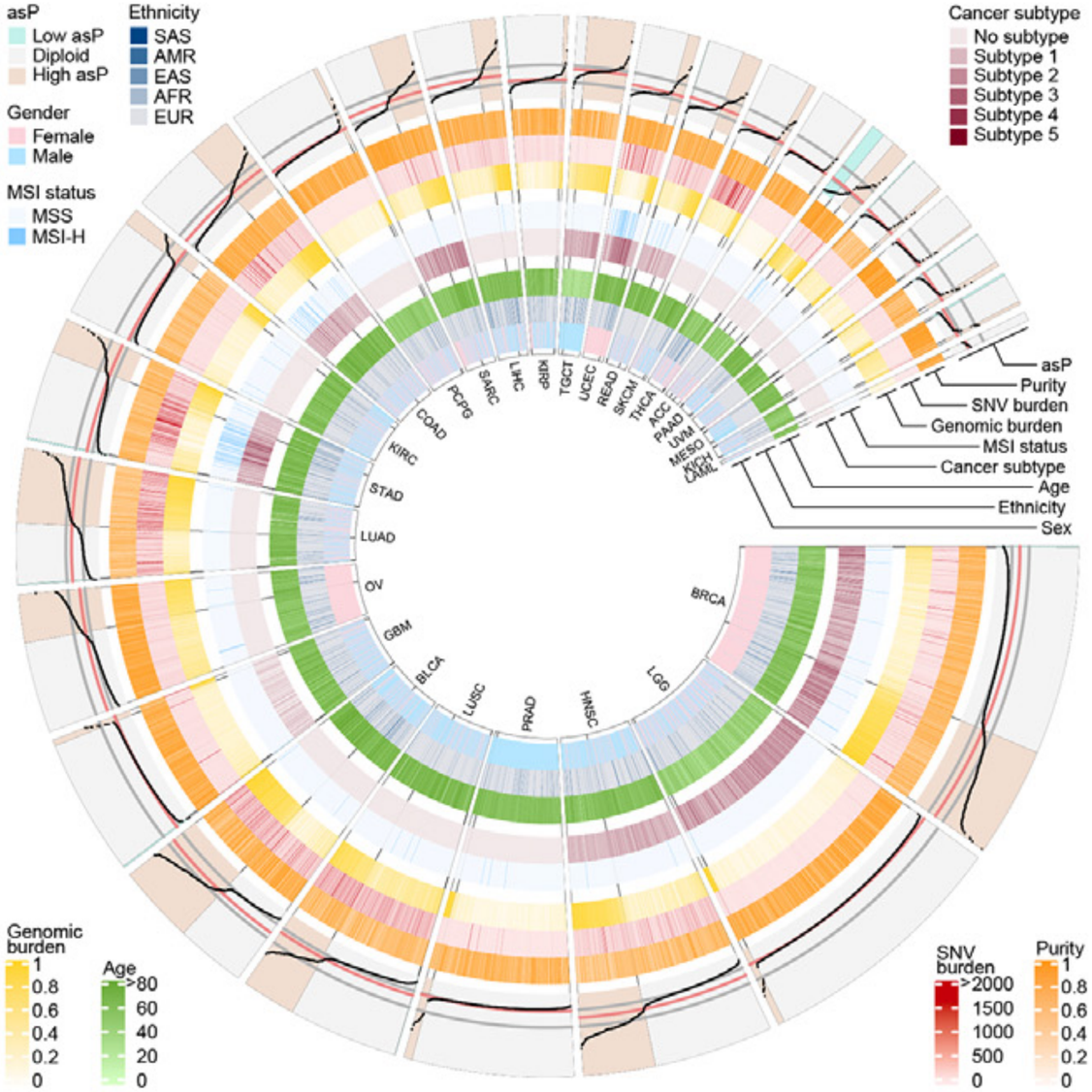
***circlize* implements and enhances circular visualization in R**
Bioinformatics, 2014



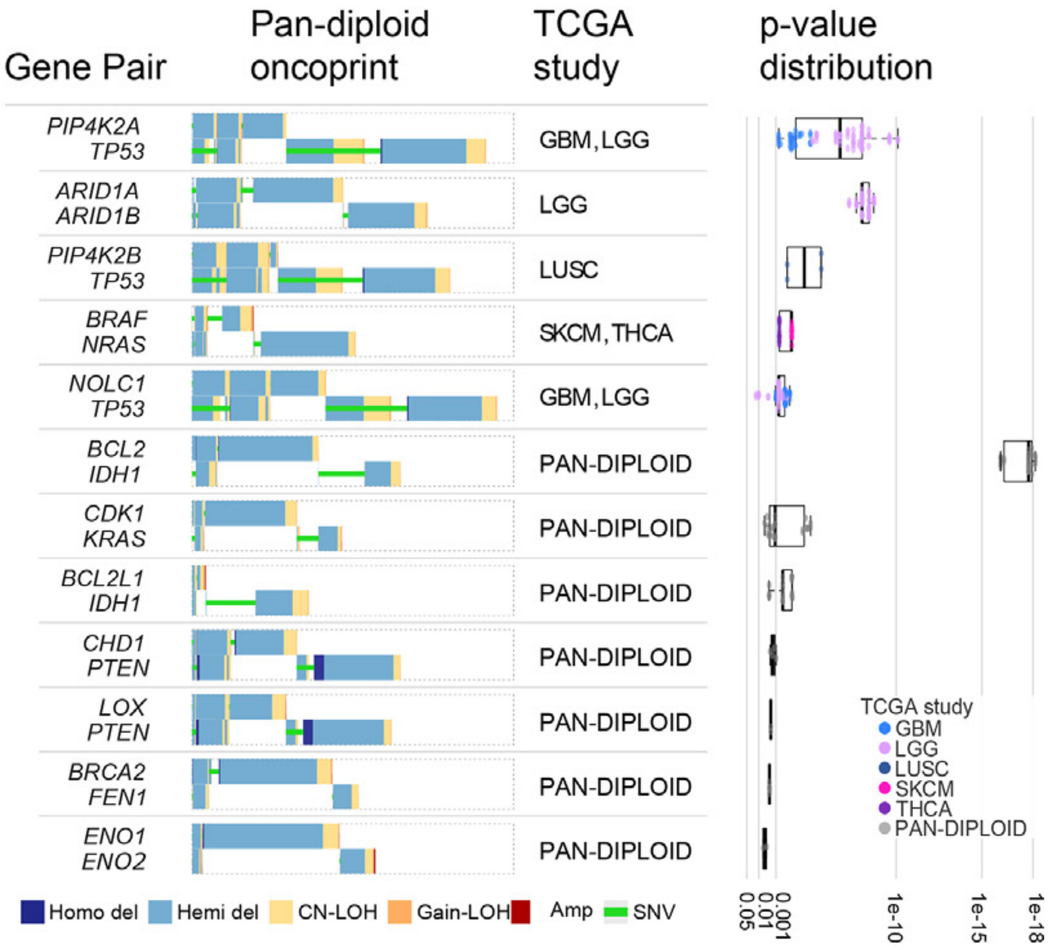
https://jokergoo.github.io/circlize_book/book/index.html

Summarising Entire Cohorts

~4800 patients

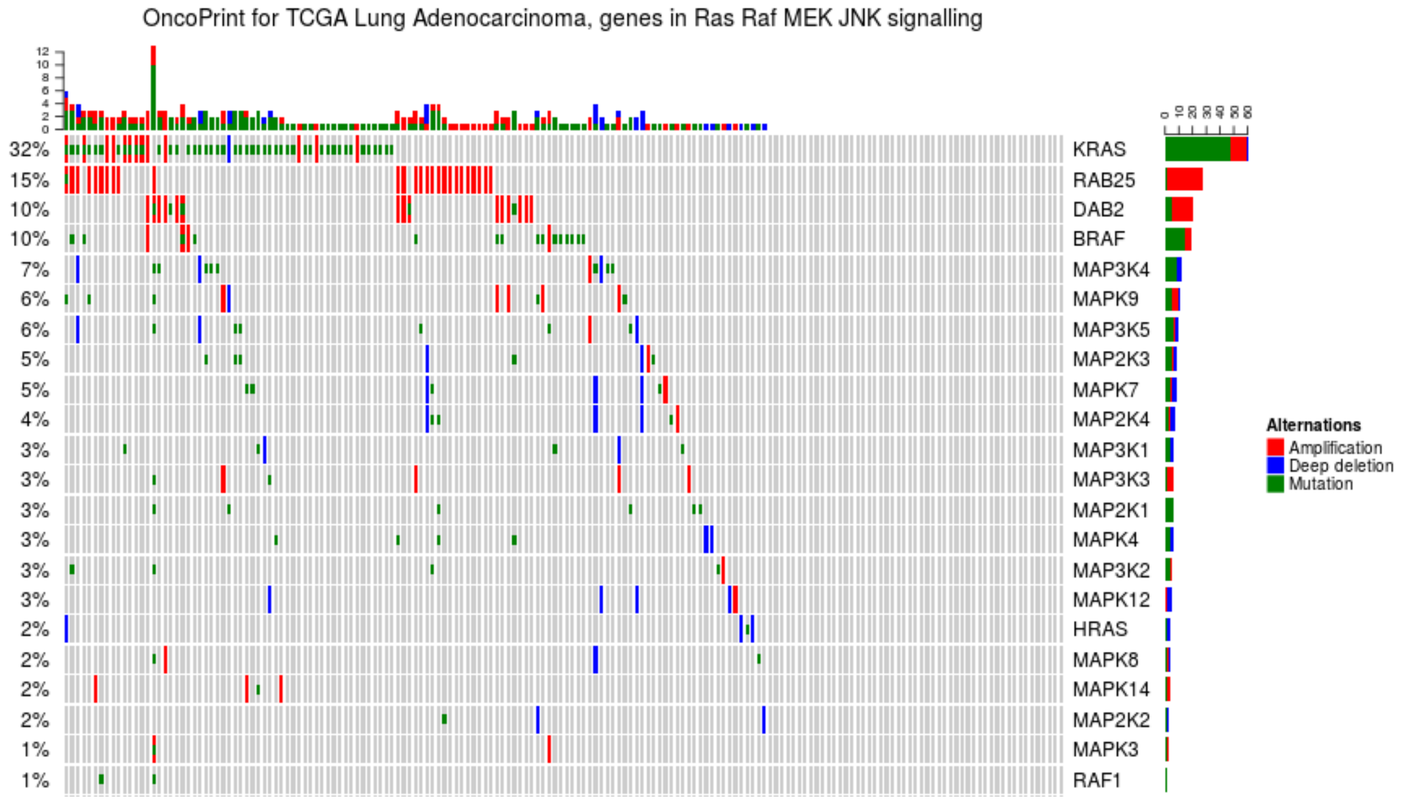


Highlighting mutual exclusivity

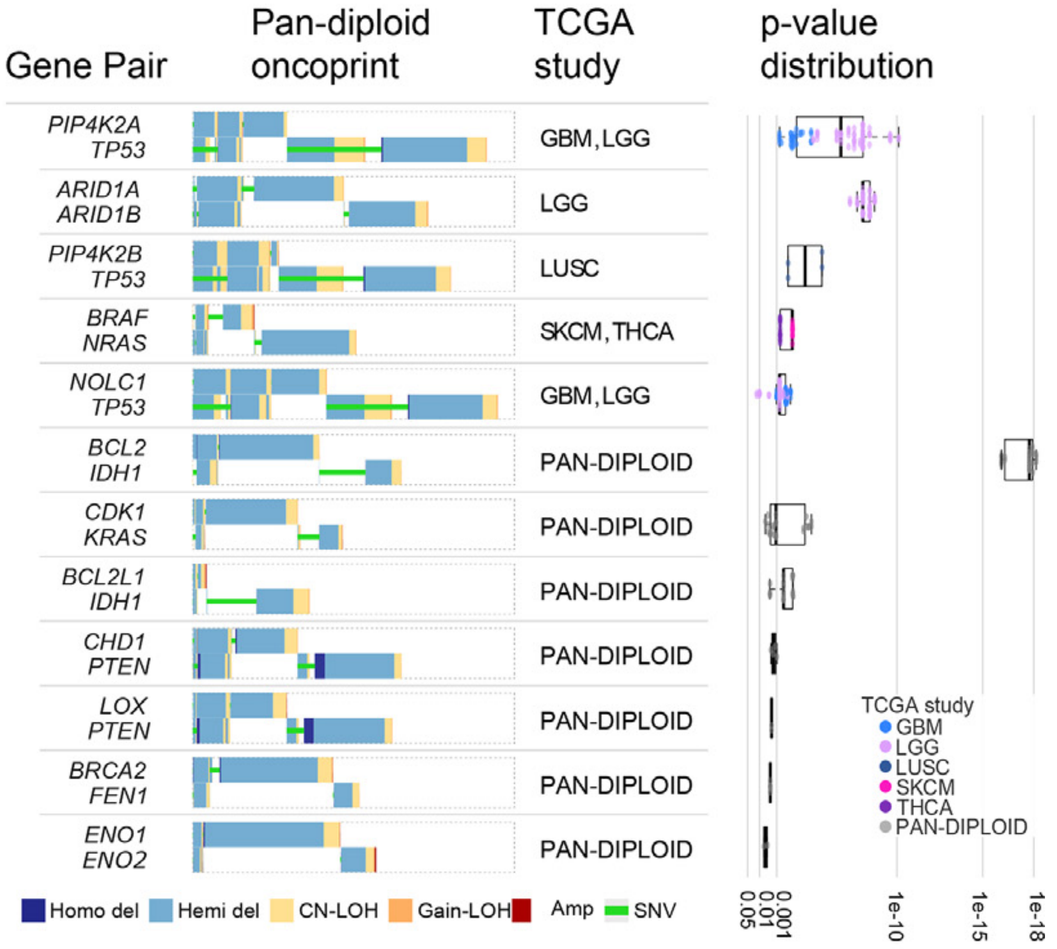


Summarising Entire Cohorts

Based on the ComplexHeatmap package

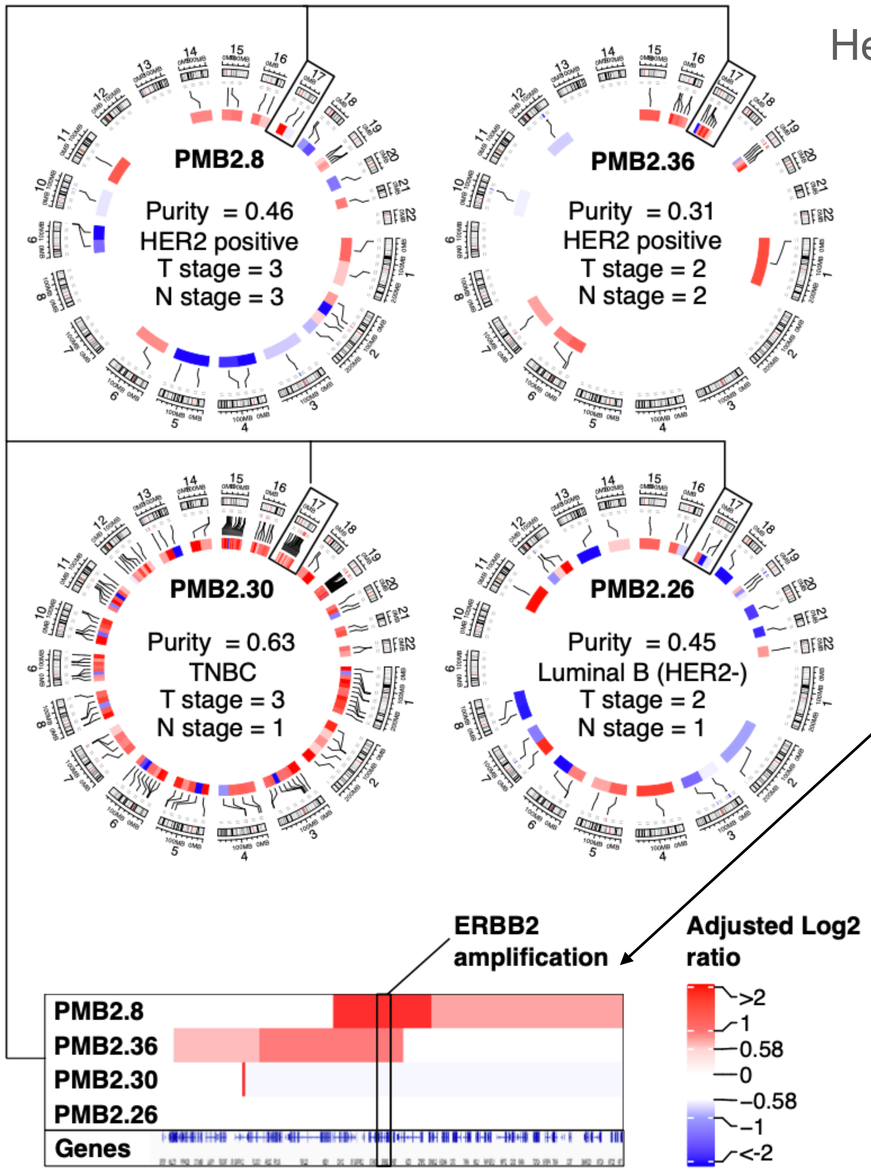


Highlighting mutual exclusivity



"Memo" Sorting, designed by B. Arman Aksoy to highlight mutual exclusivity

From Overview to Detail



Heatmap + Circos visualization

Image from IGV, a Java based alignment visualisation tool

Received: 26 April 2023 | Revised: 19 July 2023 | Accepted: 11 August 2023

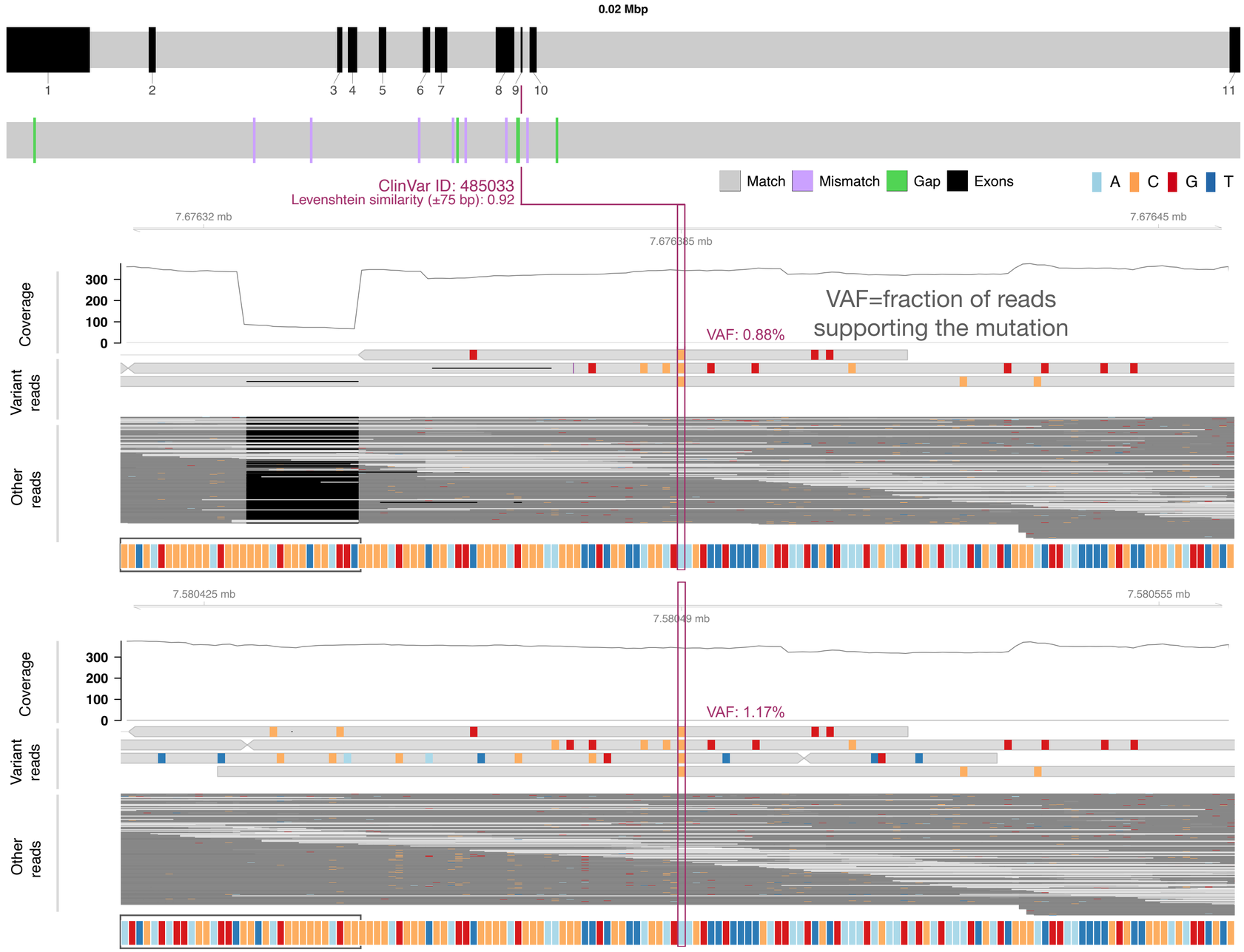
DOI: 10.1002/jex2.108

RESEARCH ARTICLE

Integrating extracellular vesicle and circulating cell-free DNA analysis using a single plasma aliquot improves the detection of HER2 positivity in breast cancer patients

Vera Mugoni[1] | Yari Ciani[1] | Orsetta Quaini[1] | Simone Tomasini[1] | Michela Notarangelo[1] | Federico Vannuccini[1] | Alessia Marinelli[1] | Elena Leonardi[2] | Stefano Pontalti[3] | Angela Martinelli[1] | Daniele Rossetto[1] | Isabella Pesce[1] | Sheref S. Mansy[1] | Mattia Barbareschi[2] | Antonella Ferro[3] | Orazio Caffo[3] | Gerhardt Attard[4] | Dolores Di Vizio[5] | Vito Giuseppe D'Agostino[1] | Caterina Nardella[1] | Francesca Demichelis[1]

Focusing on single bases



Based on packages:
ggplot2
ggrepel
Gviz
seqvisr
GenomicAlignments

Designed by Ilaria Cherchi, PhD student

Unpublished data do not post

Conclusions

- Bioinformatics have high requirements in terms of visualisation. Starting from huge amounts of data, we need to show broad overview of results but also precise details.

- DNA is a 1D entity of 3.3B points. At the same time it’s a dynamic 3D physical object. Each DNA base is interesting on its own but also in relationship with the others.

- R provides access to visualisation packages that are pivotal for our comprehension of biology and for the dissemination of our results.

Visualisation packages used in this presentation:

ggplot2, ComplexHeatmap, Circlize, patchwork, Gviz, ggrepel, seqvisr, GenomicAlignments

UNIVERSITÀ DI TRENTO | Department of Cellular, Computational and Integrative Biology - CIBIO

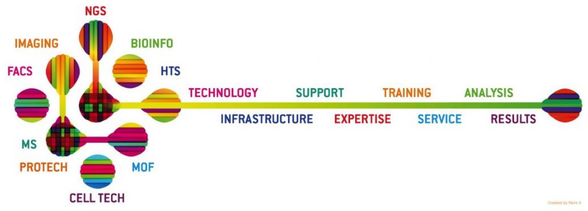
ACKNOWLEDGEMENTS

Francesca Demichelis Lab:
Liquid Biopsies Team
Prof. Francesca Demichelis
Francesco Orlando
Caterina Nardella
Orsetta Quaini
Federico Vannuccini
Alessia Marinelli
Marta Paoli
Ilaria Cherchi

Previous Members:
Vera Mugoni
Gian Marco Franceschini
Thomas Cantore
Davide Prandi
Tarcisio Fedrizzi
Giacomo D'Amato

All the other members of the group!

Core Facilities, CIBIO Department, University of Trento
Next generation Sequencing (NGS)



Funded by CANCER RESEARCH UK

FONDAZIONE CARITRO
CASSA DI RISPARMIO DI TRENTO E ROVERETO

AIRC

FONDAZIONE VRT
VALORIZZAZIONE RICERCA TRENTINA

Many thanks to all patients and families!