From womb to tomb sequencing

Prof. Ioannis Xenarios
Center for Integrative Genomics (UNIL)
Departement of Biochemistry (UNIGE)
SIB Swiss Institute of Bioinformatics / Swiss-Prot&Vital-IT groups

Platform for Advanced Scientific Computing (PASC)

PoSeNoGap : Portable Scalable Concurrency for Genomic Data Processing.

The main aim of the project is to develop for the Swiss Platform for Advanced Scientific Computing a new computation node composed of **heterogeneous** hardware, a new compression format for genomic data and a software infrastructure that enables emerging applications such as Genome Analysis to be able to process extremely large volumes of genome data in an efficient and timely way for scientific and diagnostic purposes (clinical application).

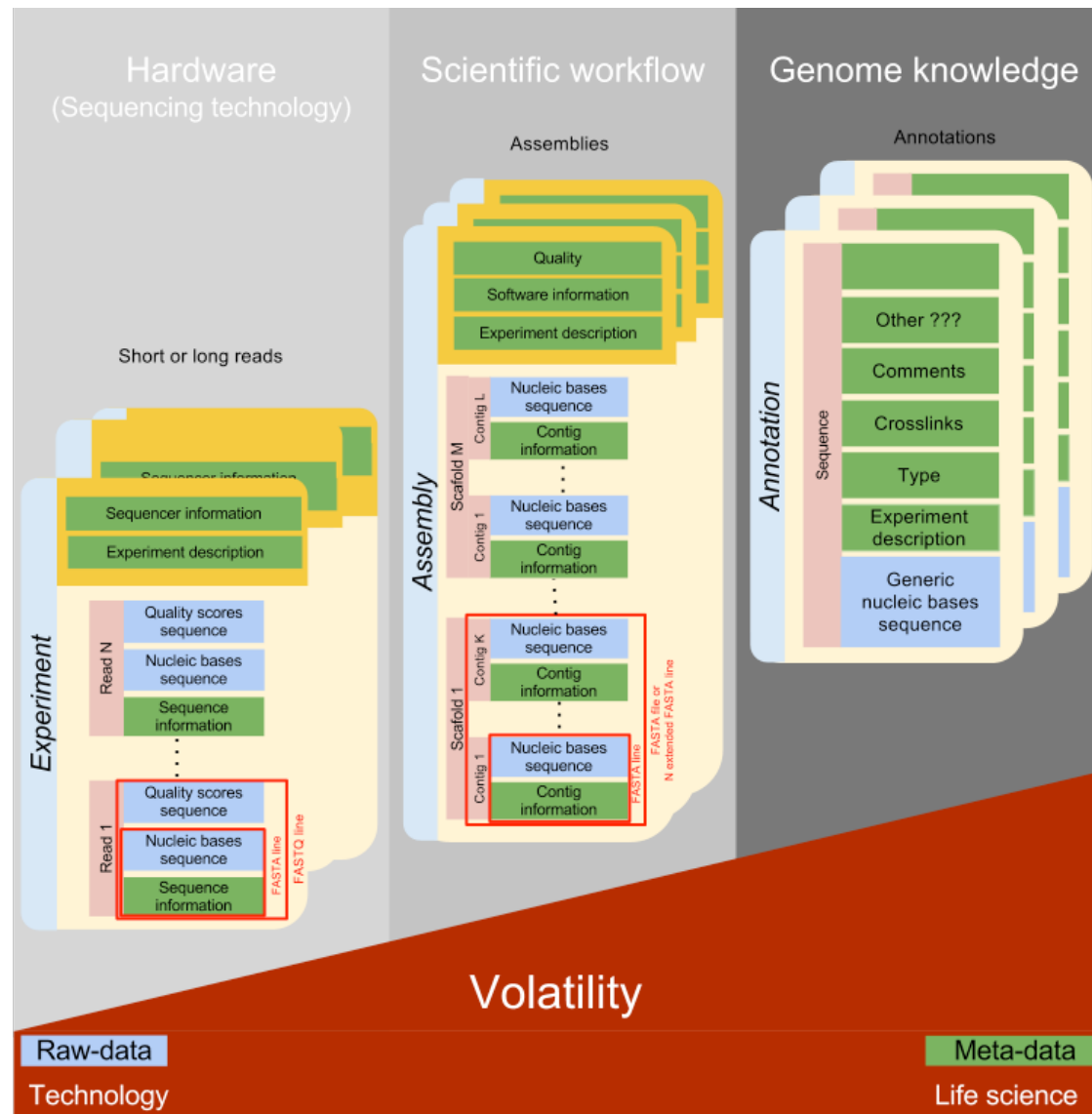Ioannis Xenarios
SIB- (Vital-IT)
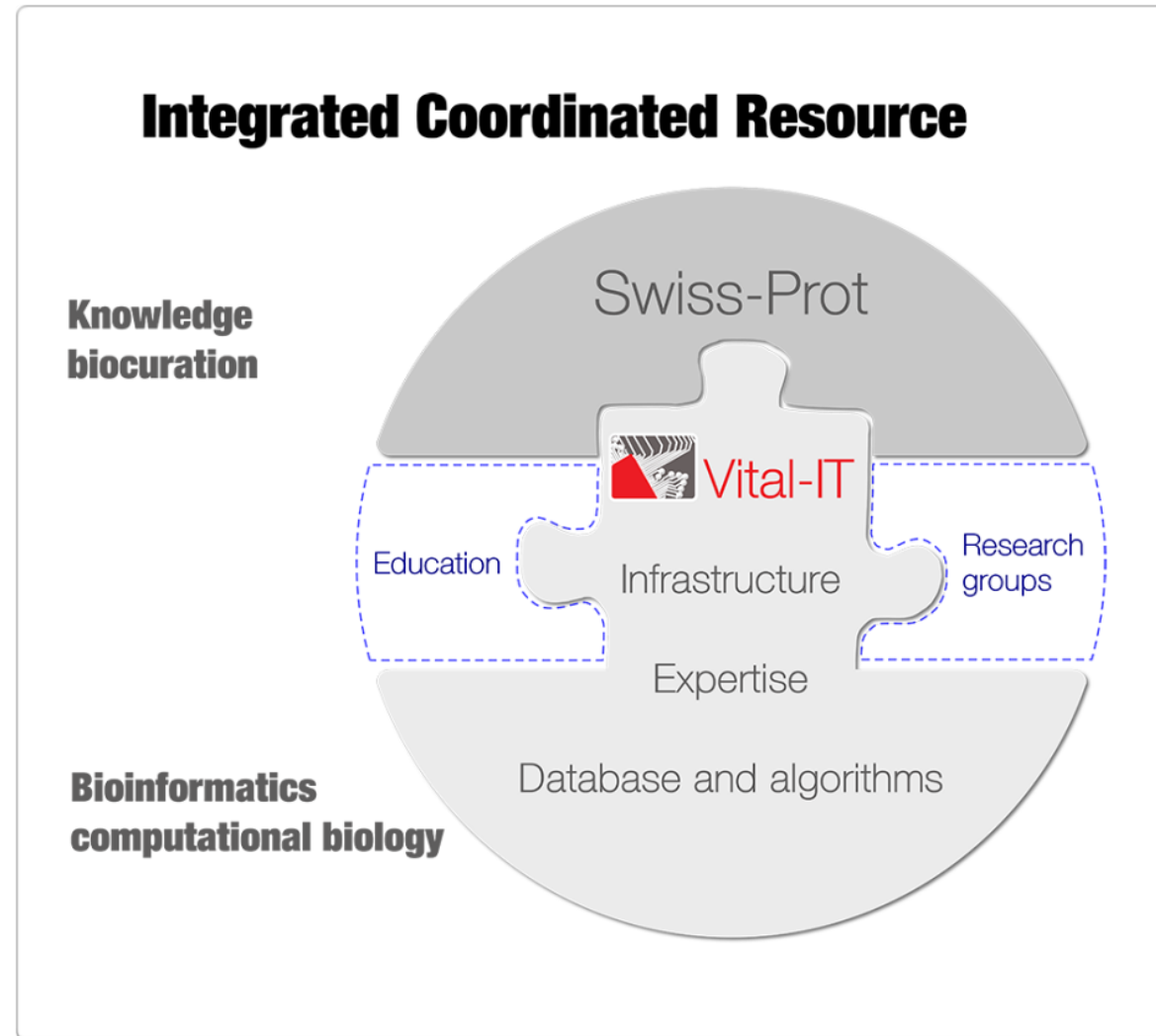
Marco Mattavelli
EPFL - (MPEG)

Yann Thoma
HES (FPGA)

SIB
Swiss Institute of Bioinformatics

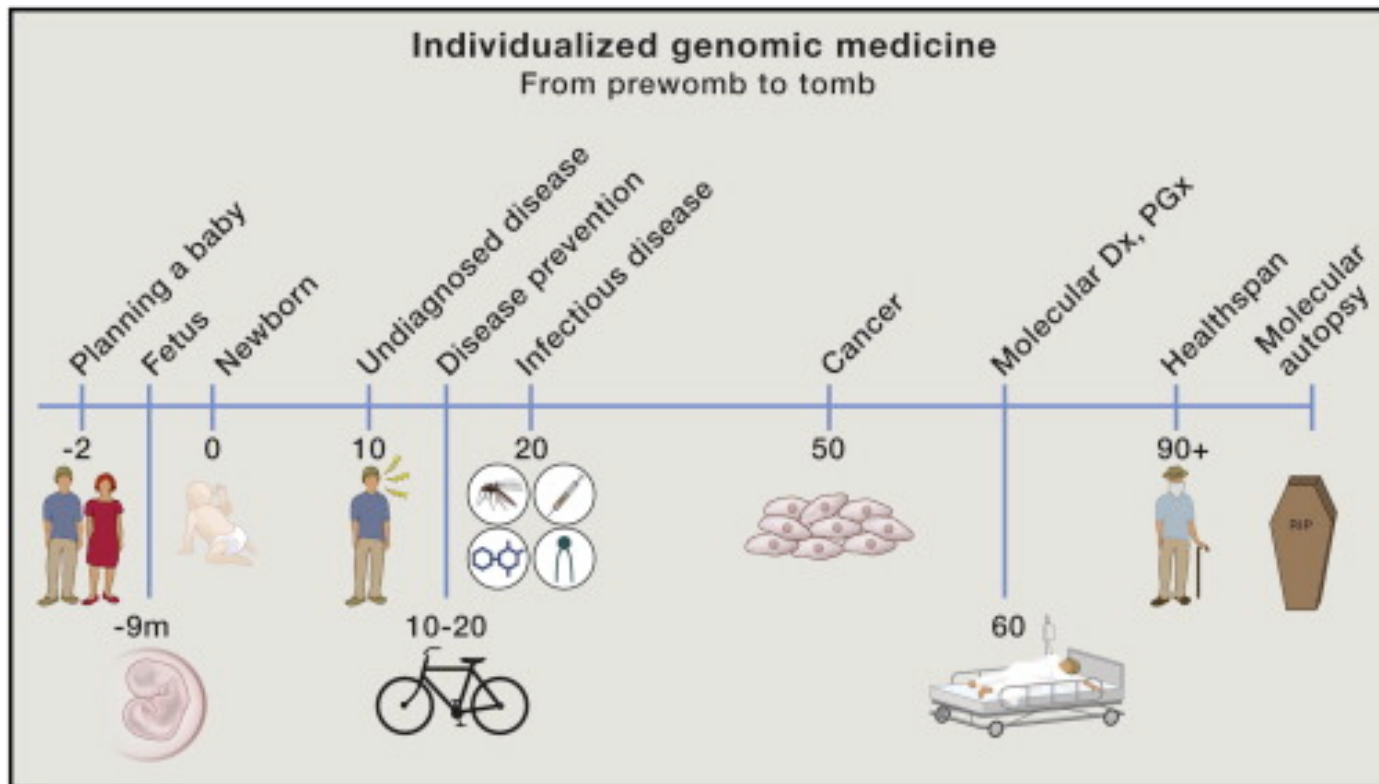# The pillars from technolog(ies) to interpretation(s)
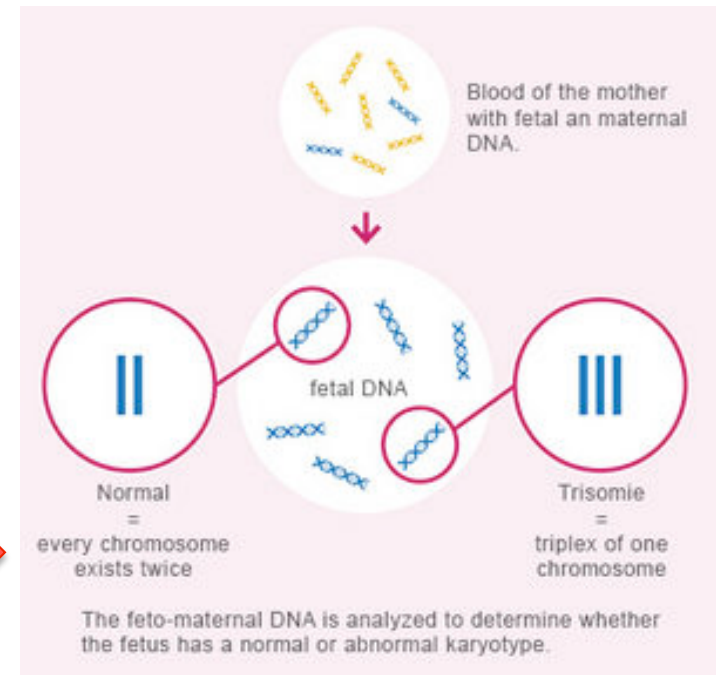
SIB competency centres - Swiss-Prot and Vital-IT

It all started with this…Eric Topol

Everybody will be sequenced at least once **in his/her** lifetime
(from womb to tomb – Eric Topol Cell **2014**)

Non-invasive prenatal test : detection *in 1 slide*

Blood of the mother with fetal an maternal DNA.

fetal DNA

Normal
=
every chromosome exists twice

Trisomie
=
triplex of one chromosome

The feto-maternal DNA is analyzed to determine whether the fetus has a normal or abnormal karyotype.

Invasive detection of Trisomy (e.g. T21)

1. Embryo
2. Amniotic cavity
3. Chorion cavity
4. Uterine cavity
5. Chorion frondosum
A. Amniocentesis
B. Chorion biopsy
C. Umbilical blood sampling
D. Transvaginal chorion biopsy
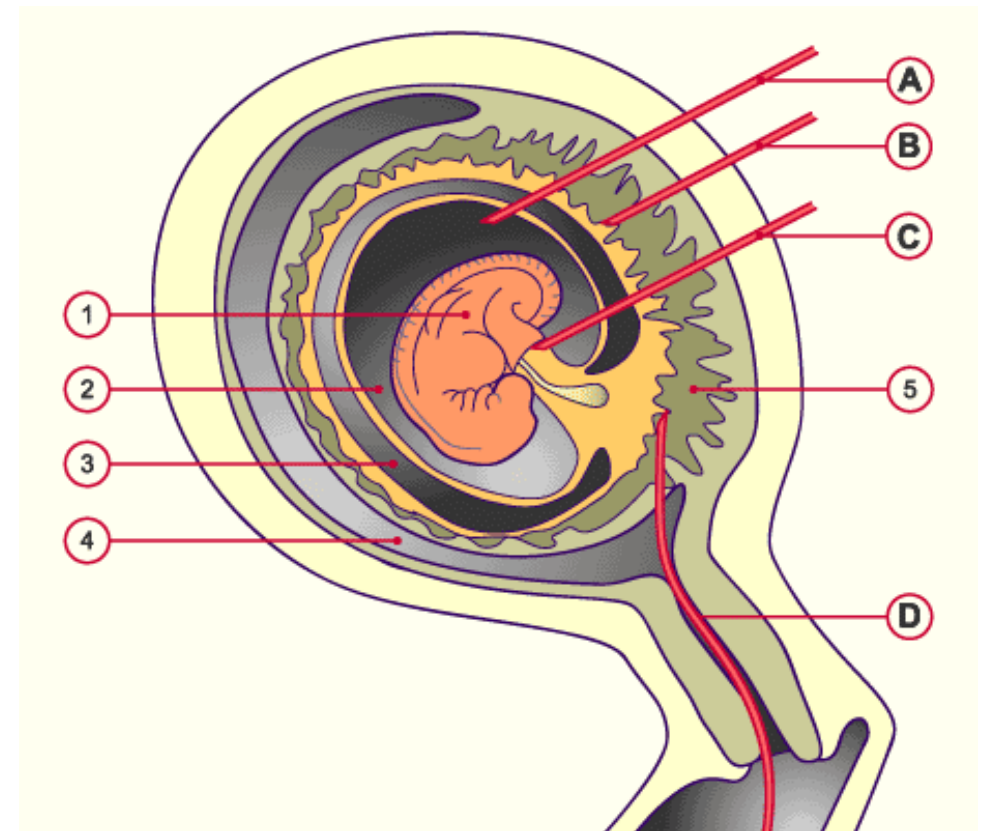
Due diligence: evaluation of the state of the art

Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood

H. Christina Fan*, Yair J. Blumenfeld[†], Usha Chitkara[†], Louanne Hudgins[‡], and Stephen R. Quake*[§]

T21

T13
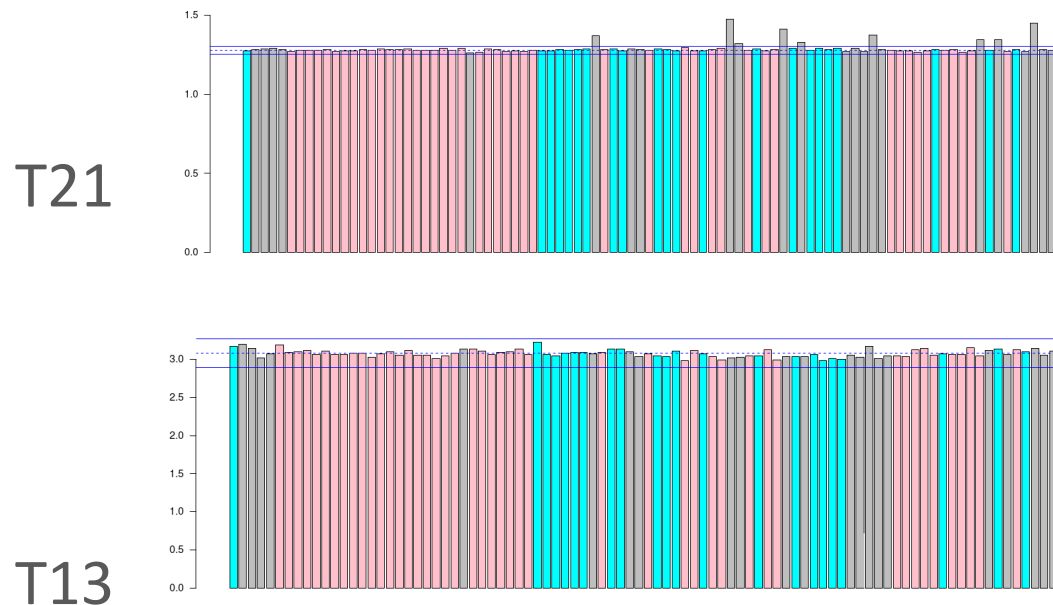
Our findings showed that although T21 is relatively easy to detect with published algorithms, other aneuploidies are more difficult to detect in a robust manner

http://www.pnas.org/content/pnas/105/42/16266.full.pdf

© 2011 SIB

SIB
Swiss Institute of Bioinformatics

Double Blind -Clinical Trial Results

(Guex N. et al, *Prenat Diagn*. 2013 Jul;33(7):707-10)

Aneuploidy	Sensitivity	Specificity
Trisomy 21 (n = 39)	39 (100%, 95% CI 88.8–100)	237/237 (100%, 95%CI 98.0–100)
Trisomy 18 (n = 24)	23 (95.8%, 95%CI 76.8–99.7)	252/252 (100%, 95%CI 97.0–100)
Trisomy 13 (n = 15)	15 (100%, 95%CI 74.6–100)	261/261 (100%, 95%CI 98.1–100)
Trisomy 16 (n = 1)	1 (100%, 95%CI 5.4–100)	275/275 (100%, 95%CI 98.2–100)
Trisomy 22 (n = 2)	2 (100%, 95%CI 19.7–100)	274/274 (100%, 95%CI 98.2–100)
45,X (n = 15)	15 (100%, 95%CI 74.6–100)	261/261 (100%, 95%CI 98.1–100)
47,XXX (n = 5)	5 (100%, 95%CI 46.2–100)	271/271 (100%, 95%CI 98.2–100)

Clinical Follow up on 6,388 consecutive cases

Table 2 CNVs overlapping or not with genomic disorders

CNVs not overlapping with genomic disorders		Data set (n)		CNVs overlapping with genomic disorders		Data set (n)	
Chromosomal region	**Breakpoints (hg19)**	**1st**	**2nd**	**Chromosomal region**	**Breakpoints (hg19)**	**1st**	**2nd**
HSA1:dup 1p31	(69,975,000–74,425,000)	1		HSA15:dup 15q11.2	(20,175,000–26,575,000)	1	
				HSA15:dup 15q11.2	(20,175,000–29,925,000)	1	
HSA2:dup 2q24.3	(16,500,001–20,300,001)	1					
				HSA16:del 16p13.12	(13,175,000–19,475,000) (13,925,000–19,925,000)	1	1
HSA3:del 3p26.1	(0–3,200,000)	1		HSA16:dup 16p13.12 (14,025,000–19,825,000)			2
HSA3:dup 3p14	(64,850,000–68,700,000)	1					
				HSA17:dup 17p12	(12,625,000–17,275,000)	1	
HSA4:dup 4q13	(59,150,000–63,500,000)	1					
HSA4:dup 4q21-22	(68,300,000–73,050,000)	1		HSA22:del 22q11.2	(17,175,000–22,100,000) (17,175,000–25,375,000)	1	1
HSA4:dup 4q?3	(149,500,000–154,200,000)	1		HSA22:dup 22q11.2	(17,175,000–23,125,000) (17,175,000–23,175,000)	1	1
HSA4:dup 4q35	(186,550,000–190,875,000)	1					
HSA4:del 4q33	(187,150,000–190,875,000)	1					
HSA5:dup 5p15.2	(20,300,000–25,050,000)	1					
HSA5:del 5q14.3	(90,250,000–126,750,000)		1				
HSA7:dup 7q21	(86,975,000–90,975,000)	1					
HSA8:dup 8p23.3	(3,025,000–6,675,000)	1					
HSA8:del 8p23	(14,175,000–18,775,000)		1				
HSA9:dup 9p21.1	(28,250,000–32,450,000)	1					
HSA9:del 9q31	(102,550,000–107,150,000)	1					
HSA9:dup 9q22	(103,750,000–107,450,000)	1					
HSA10:dup 10q?11	(43,900,000–52,225,000)	1					
HSA10:del 10q22	(63,950,000–68,600,000)	1					
HSA10:dup 10q25.3	(118,200,000–129,650,000)		1				
HSA13:dup 13q12	(19,500,000–26,450,000)	1					
HSA15:dup 15q13.1	(28,175,000–34,425,000)		1				
HSA17:del 17p12	(12,775,000–17,375,000)		1				
HSA17:dup 17q24.3	(70,225,000–74,725,000)		1				
HSA20:dup 20p12.3	(7,025,000–11,325,000)	1					

Nonrecurrent CNVs are listed according to individual chromosome (HSA) and breakpoints given in parentheses (hg19). dup, duplications; del, deletions. Recurrent CNVs are listed according to individual chromosome (HSA) and breakpoints given in parentheses (hg19).

CNV, copy-number variations.

Acknowledgments

Fasteris:
Laurent Farinelli
Et al

Medisupport:
Bernard Conrad
Et al

Vital-IT:
Nicolas Guex
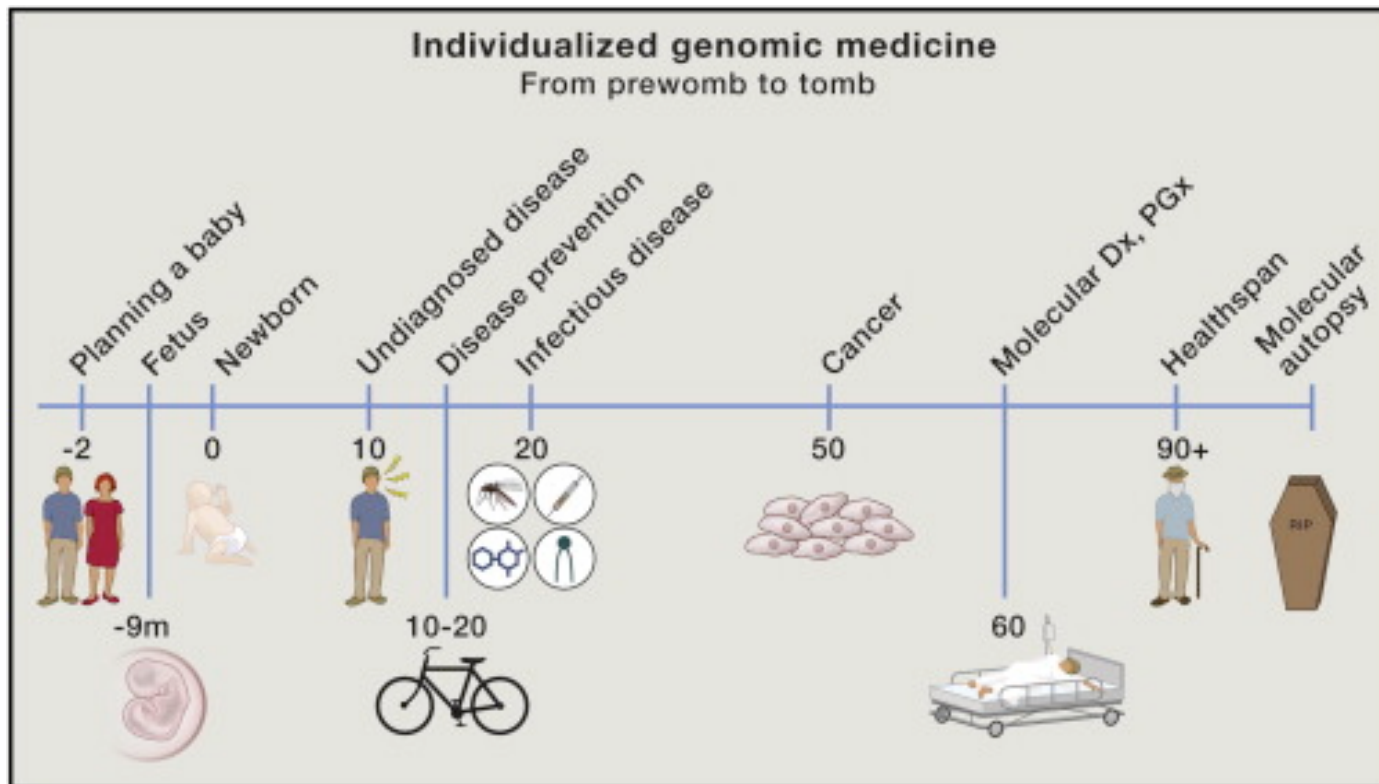Christian Iseli
Ioannis Xenarios

King's college, London
K. H. Nicolaides

It all started with this…Eric Topol

Everybody will be sequenced at least once **in his/her** lifetime
(from womb to tomb – Eric Topol Cell **2014**)

Science Translational Medicine

Personalized cancer vaccine effectively mobilizes antitumor T cell immunity in ovarian cancer

Janos L. Tanyi, Sara Bobisse, Eran Ophir, Sandra Tuyaerts, Annalisa Roberti, Raphael Genolet, Petra Baumgartner, Brian J. Stevenson, Christian Iseli, Denarda Dangaj, Brian Czerniecki, Aikaterini Semilietof, Julien Racle, Alexandra Michel, Ioannis Xenarios, Cheryl Chiang, Dimitri S. Monos, Drew A. Torigian, Harvey L. Nisenbaum, Olivier Michielin, Carl H. June, Bruce L. Levine, Daniel J. Powel, Jr., David Gfeller, Rosemarie Mick, Urania Dafni, Vincent Zoete, Alexandre Harari, George Coukos and Lana E. Kandalaft

ARTICLE

Sensitive and frequent identification of high avidity neo-epitope specific CD8[+] T cells in immunotherapy-naive ovarian cancer

Sara Bobisse[1], Raphael Genolet[1], Annalisa Roberti[2], Janos L. Tanyi[2], Julien Racle[1,3], Brian J. Stevenson[3], Christian Iseli[3], Alexandra Michel[1], Marie-Aude Le Bitoux[1], Philippe Guillaume[1], Julien Schmidt[1], Valentina Bianchi[1], Denarda Dangaj[1], Craig Fenwick[4], Laurent Derré[5], Ioannis Xenarios[3], Olivier Michielin[1,3], Pedro Romero[1], Dimitri S. Monos[6], Vincent Zoete[1,3], David Gfeller[1,3], Lana E. Kandalaft[1,2], George Coukos[1] & Alexandre Harari[1]

Current Challenges

The number of projects that use NGS is exploding.
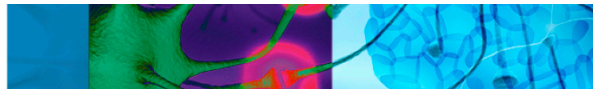
MRI
PET scans
Cognitive Testing
Longitudinal data
SNP chips
WGS

Sequencers deliver output as fastq files

```
LP6005115-DNA_G11_R1.fq   (150GB)
LP6005115-DNA_G11_R2.fq   (150GB)
                          =======
                           300GB
```

paired-end 2 x 656,935,771 **= 1.31 billion reads**

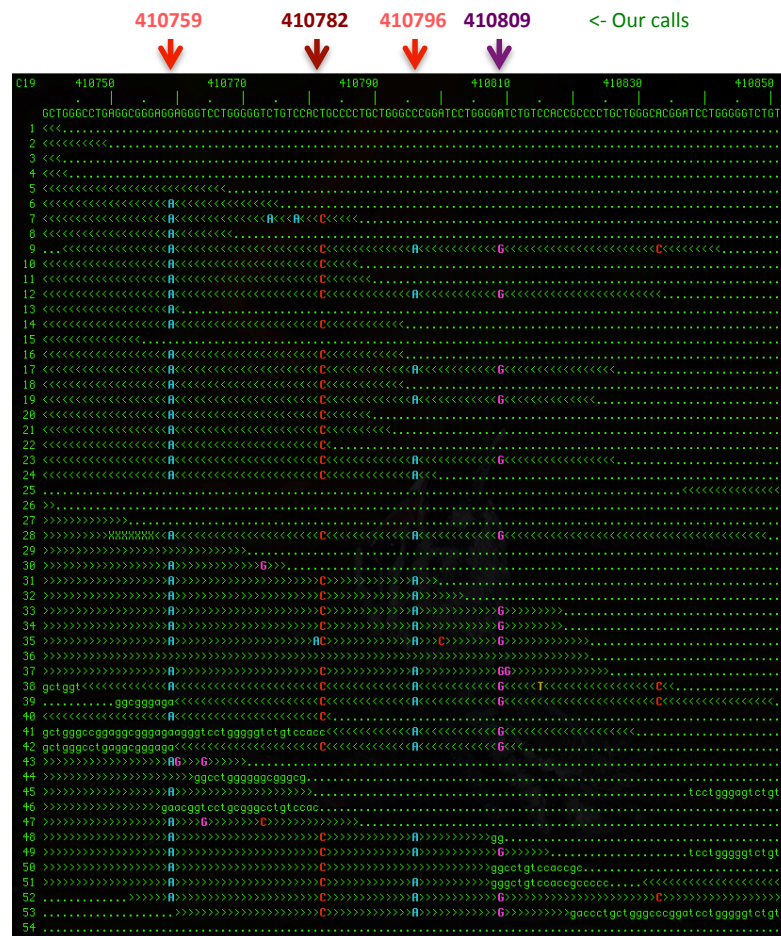For the ADNI project alone   > **800 patients.**
300Gb x 800 = 240000 GB **= 240 TB**
(BAM file– 84Tb)

```
variant calling from a single sample (LP6005115-DNA_G11)
```
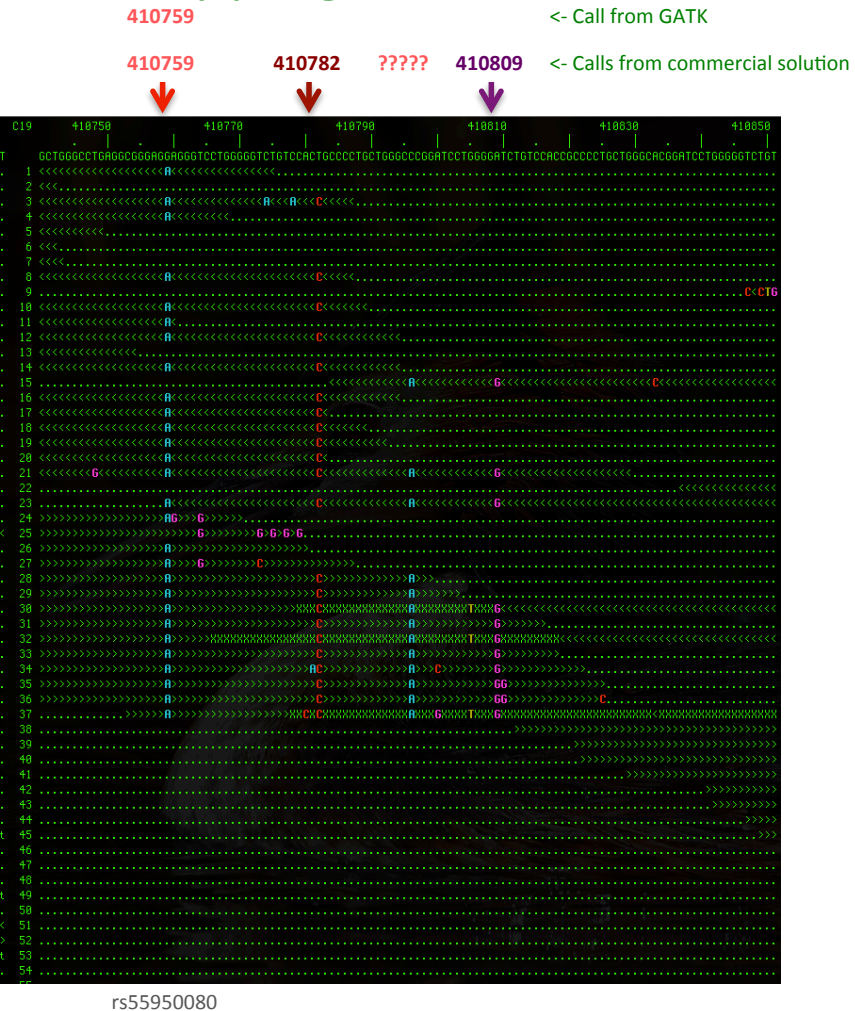**~7 days** to complete using bwa/Picard/GATKv3.2 pipeline

Swiss Institute of
Bioinformatics

Mapping / Calling galore

Our mapping /calling

Mapping we received for ADNI

rs55950080

SIB
Swiss Institute of
Bioinformatics

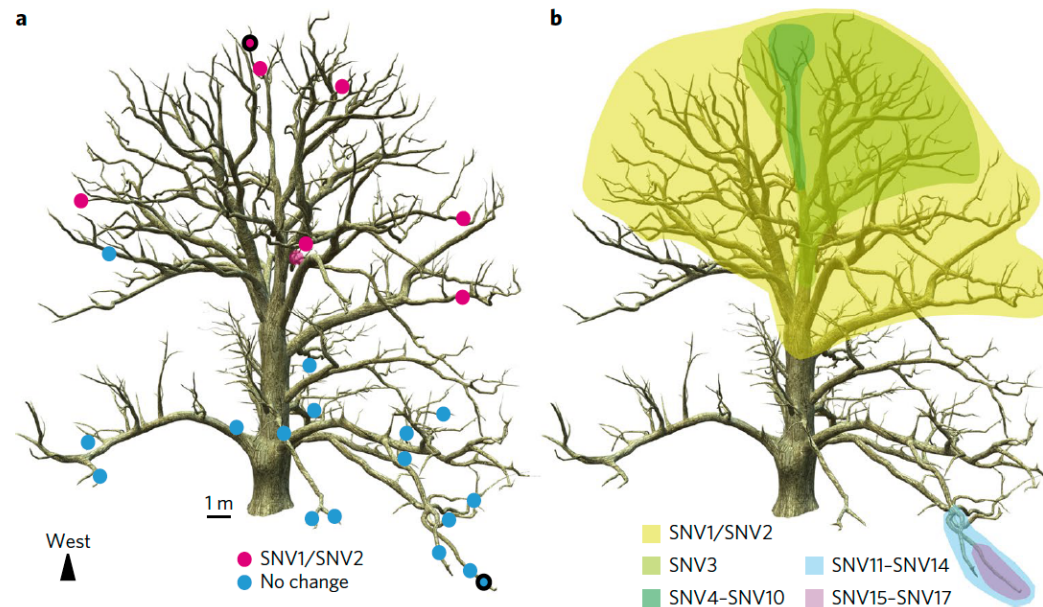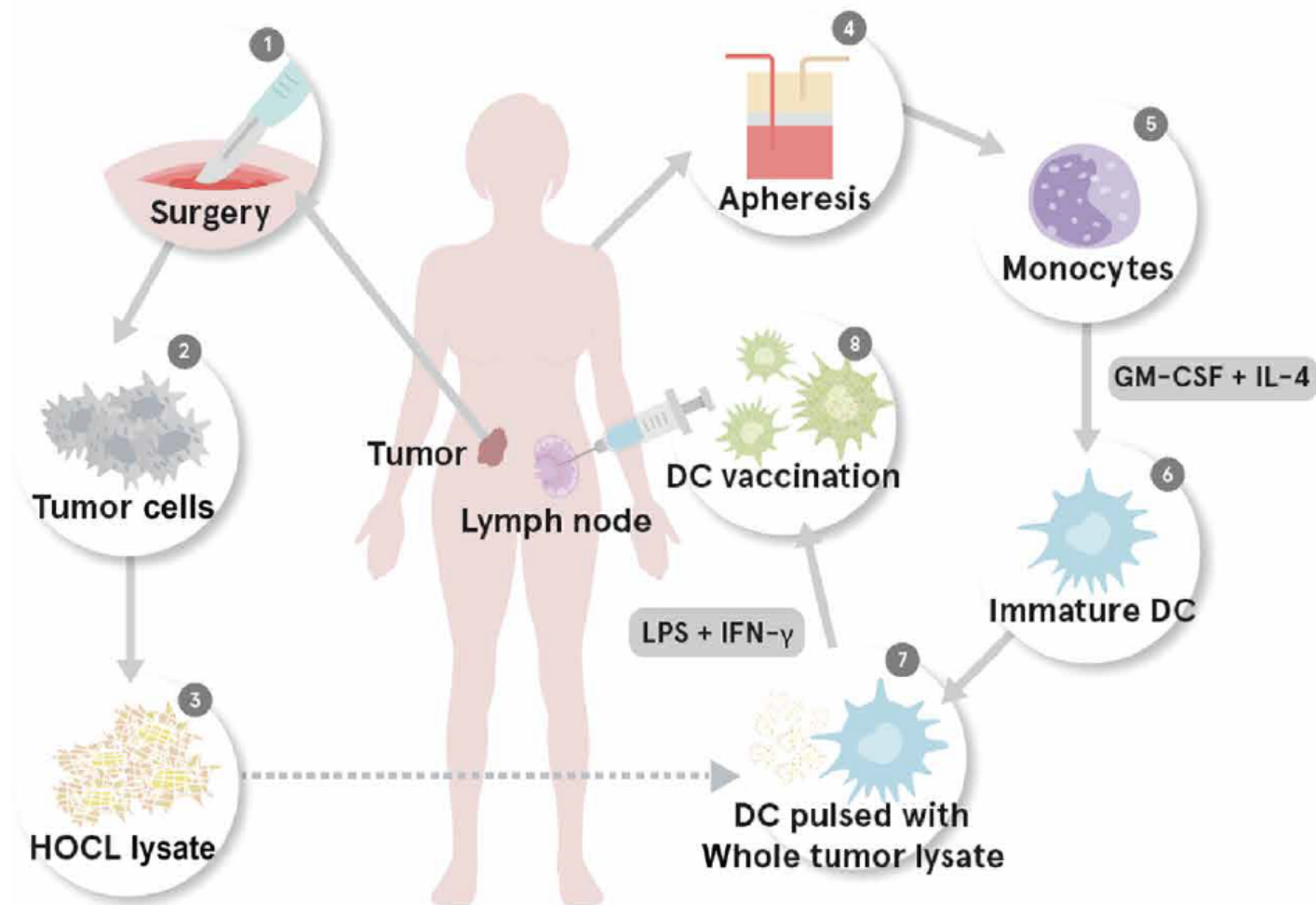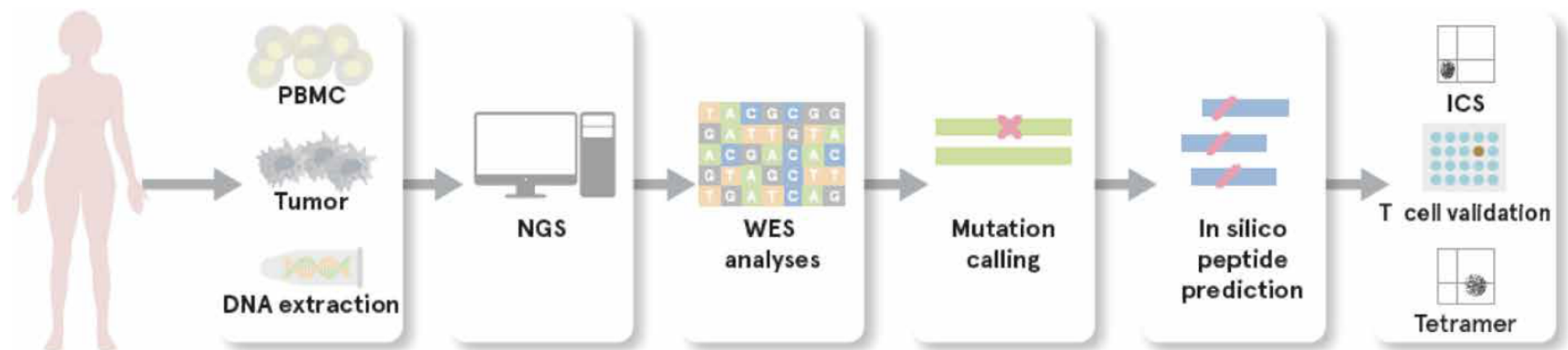# Same approaches applied to an oak tree with very *small* number of mutations– 17 on 750Mb of genome

Fig. 1 | Distribution of fixed somatic mutations in the Napoleon Oak. a, The genome of two leaf samples (outlined dots) was sequenced to identify SNVs. Seventeen SNVs were confirmed and analysed in 26 other leaf samples to map their origin. A reconstructed image of the Napoleon Oak shows the similar location of two SNVs (magenta dots) on the tree. Blue dots represent genotypes that are non-mutant for these SNVs. Three non-mutant samples are not visible on this projection. The location of other SNVs can be found in Supplementary Fig. 3. **b,** The location of all identified SNVs. Sectors of the tree containing each group of SNVs are represented by different colours.
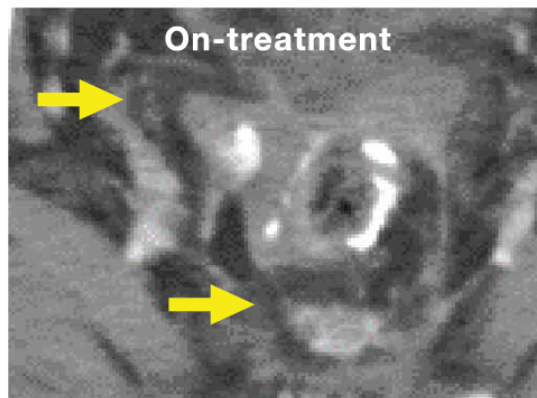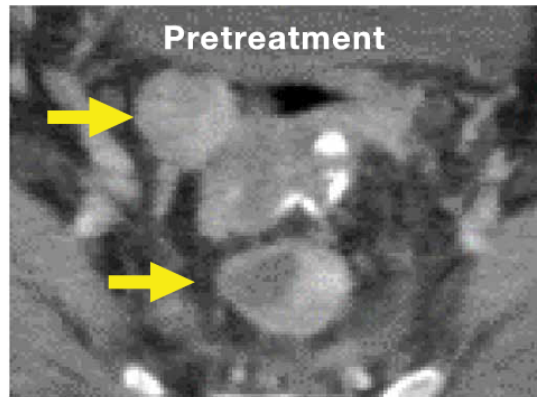
Complex workflow
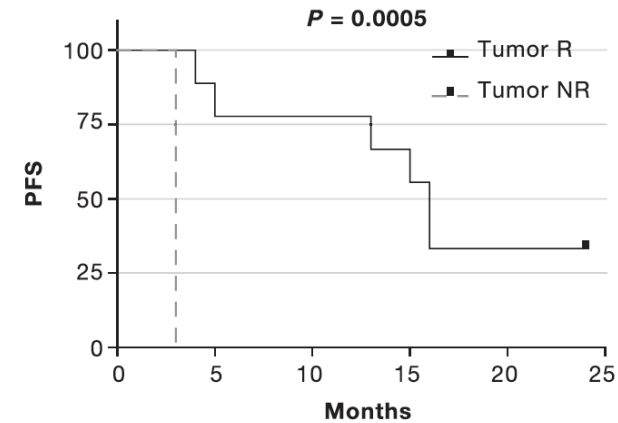## to vaccinate patient with their own tumor derived Dendritic cells

Swiss Institute of Bioinformatics

PBMC

Tumor

DNA extraction

NGS

WES analyses

Mutation calling

In silico peptide prediction

ICS

T cell validation

Tetramer

SIB
Swiss Institute of Bioinformatics

Vaccination with the neoantigen increase life span
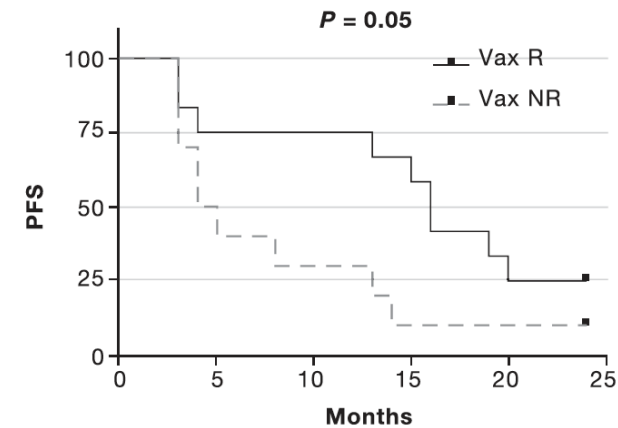
Conclusions

- High throughput sequencing has changed completely the face of biology and medecine in the last 10 years

- Efficient standards to represent genomic information are critical, MPEG-G can contribute positively with its years of experience in *developing industry* standards and creating an economy around it.

- Sequencing will become more pervasive and used more and more in diagnostic environment, the ability to preserve also part of the data with the Topol vision (from womb to tomb is critical)

SIB
Swiss Institute of Bioinformatics