



Swiss Institute of
Bioinformatics

Introduction to bioinformatics: Clinical Bioinformatics

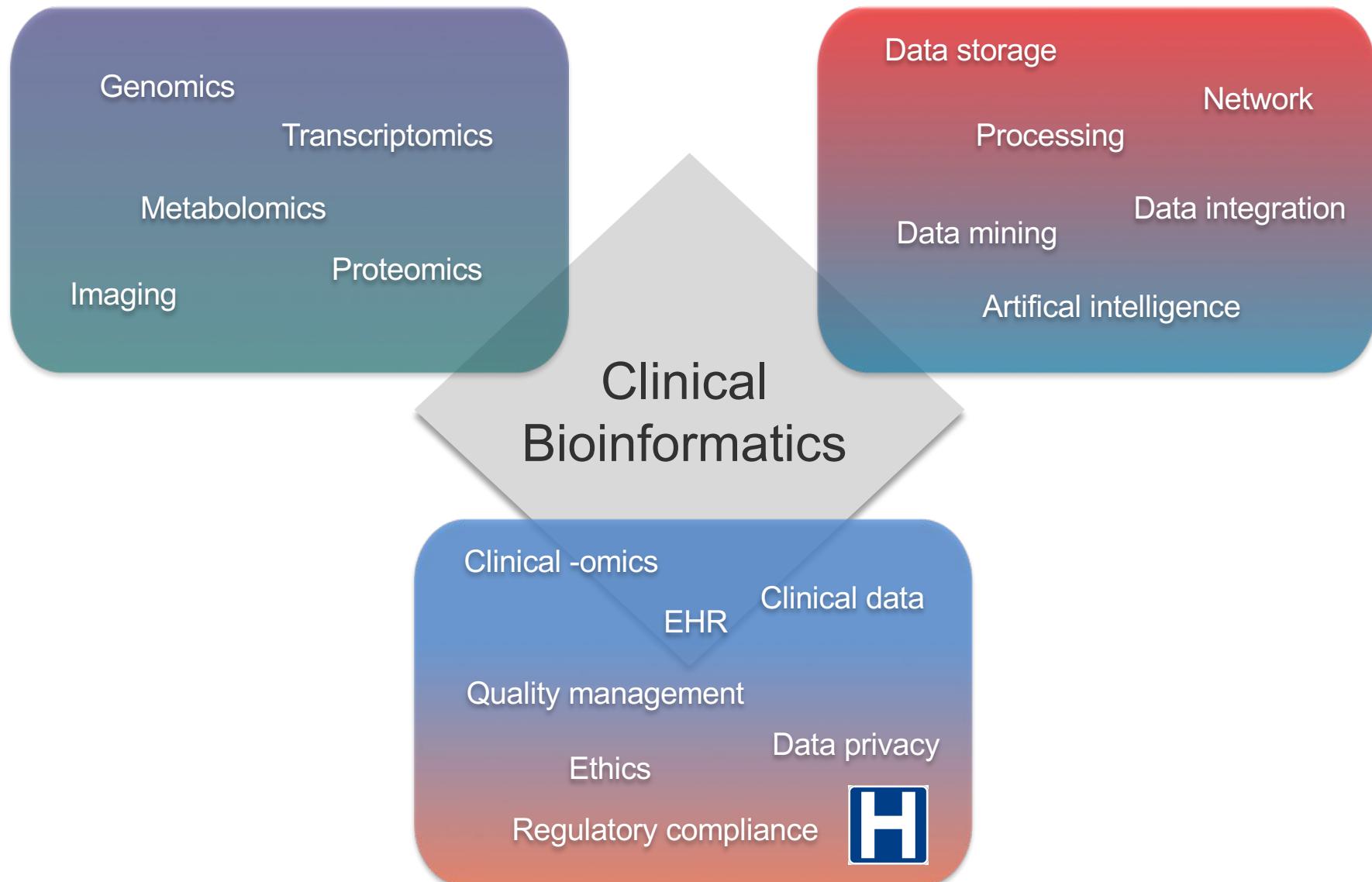
Valérie Barbié, head of SIB Clinical Bioinformatics

Zürich, 26 November 2019

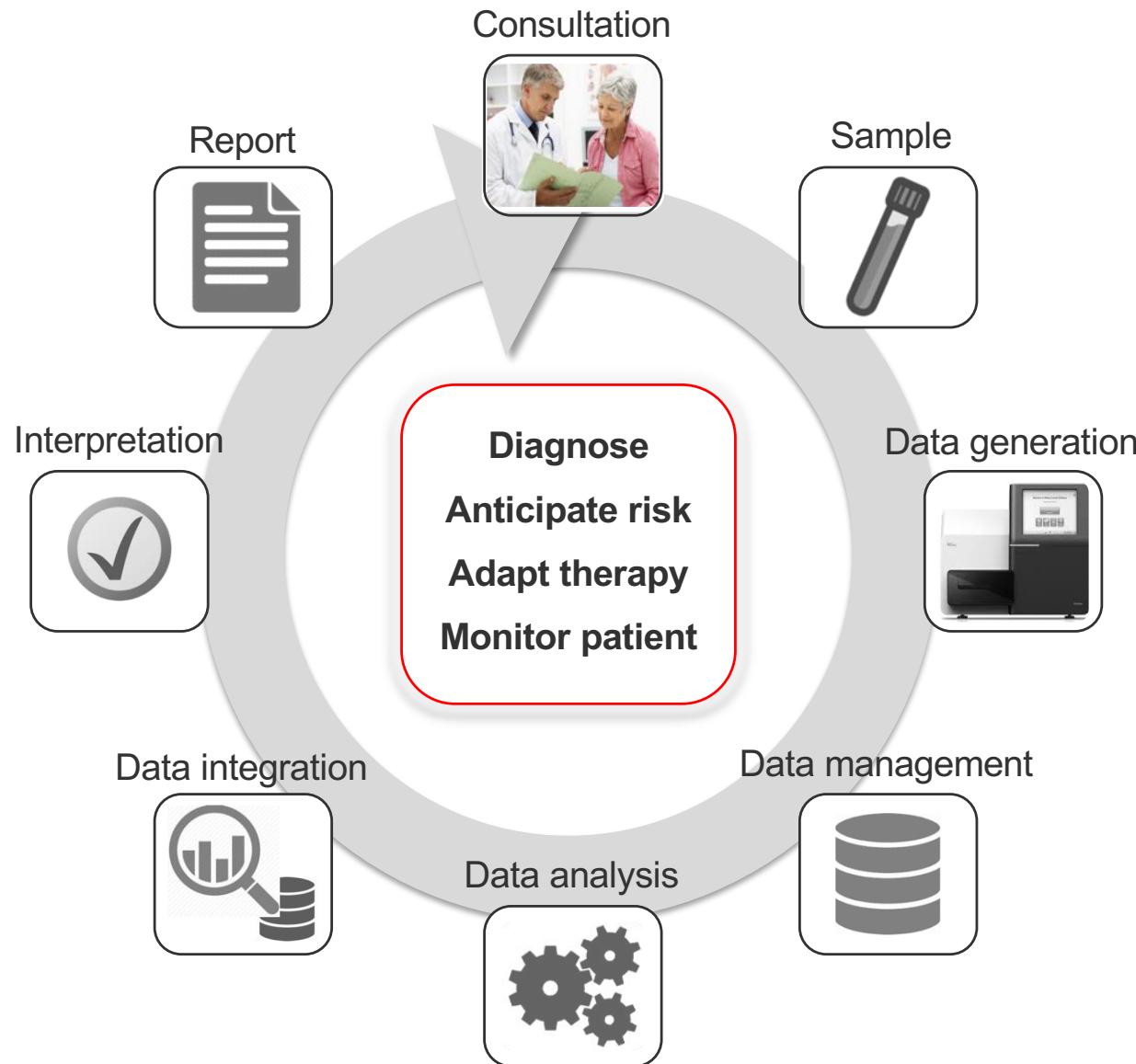


www.sib.swiss

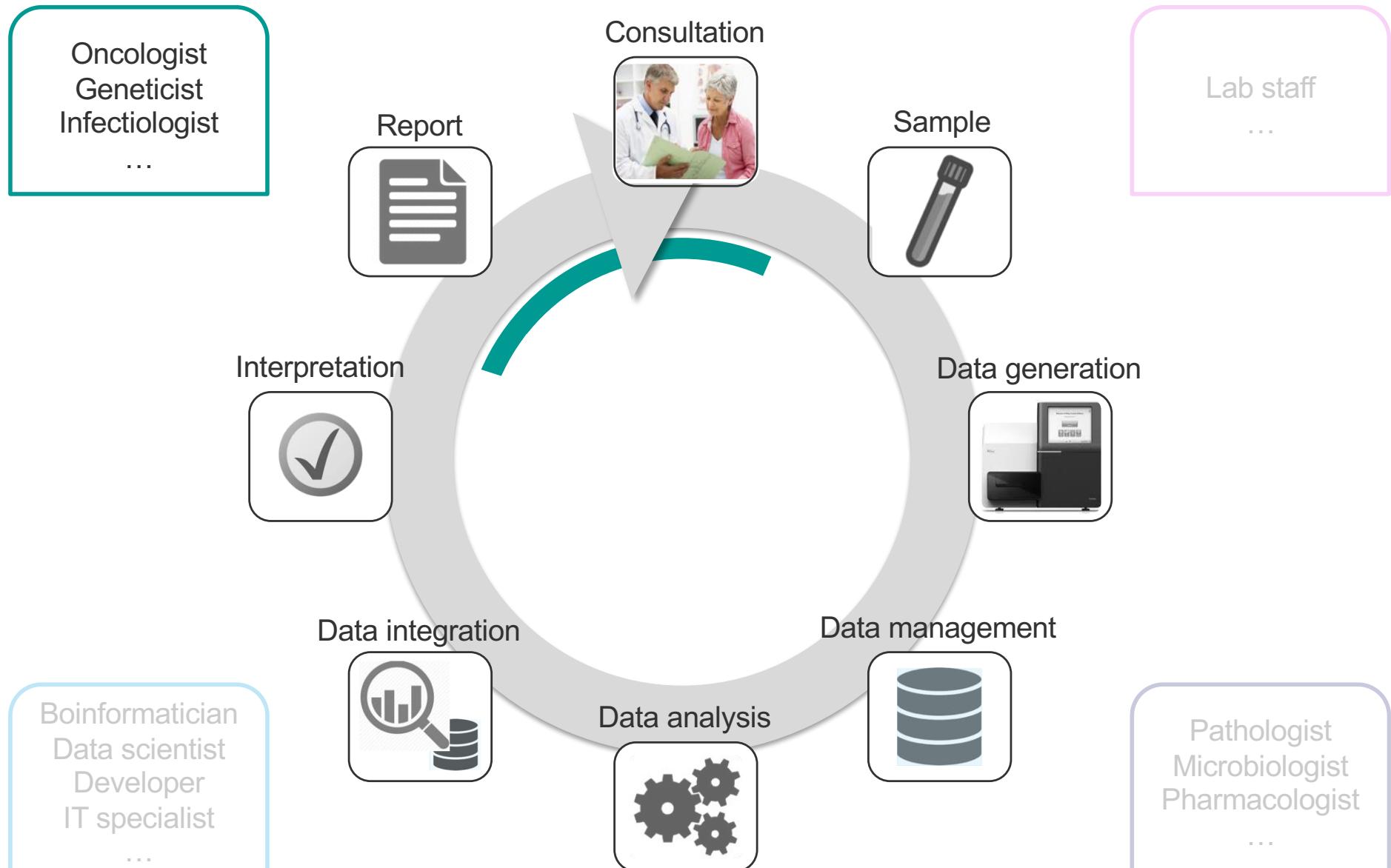
What is clinical bioinformatics?



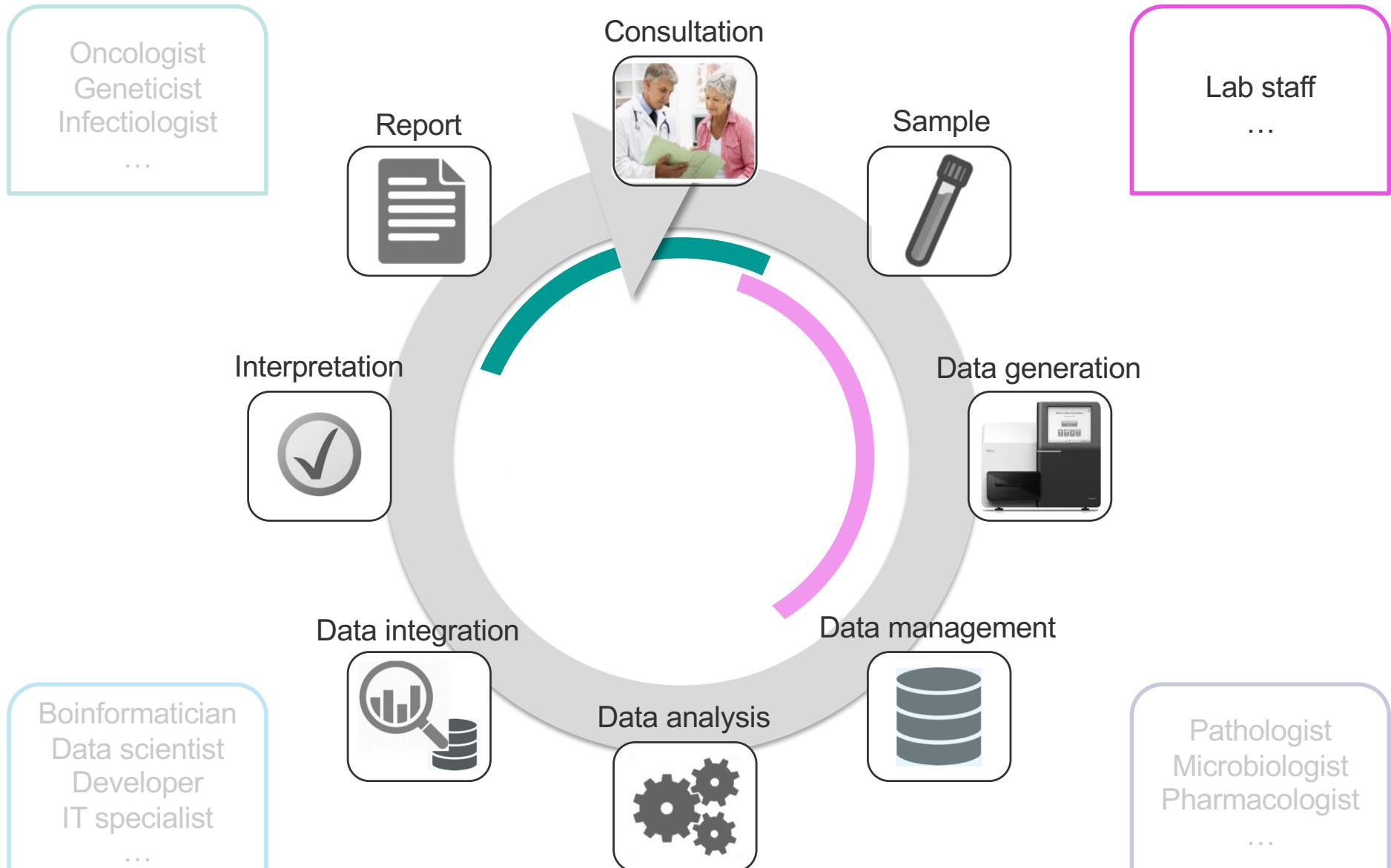
Clinical bioinformatics for medical care



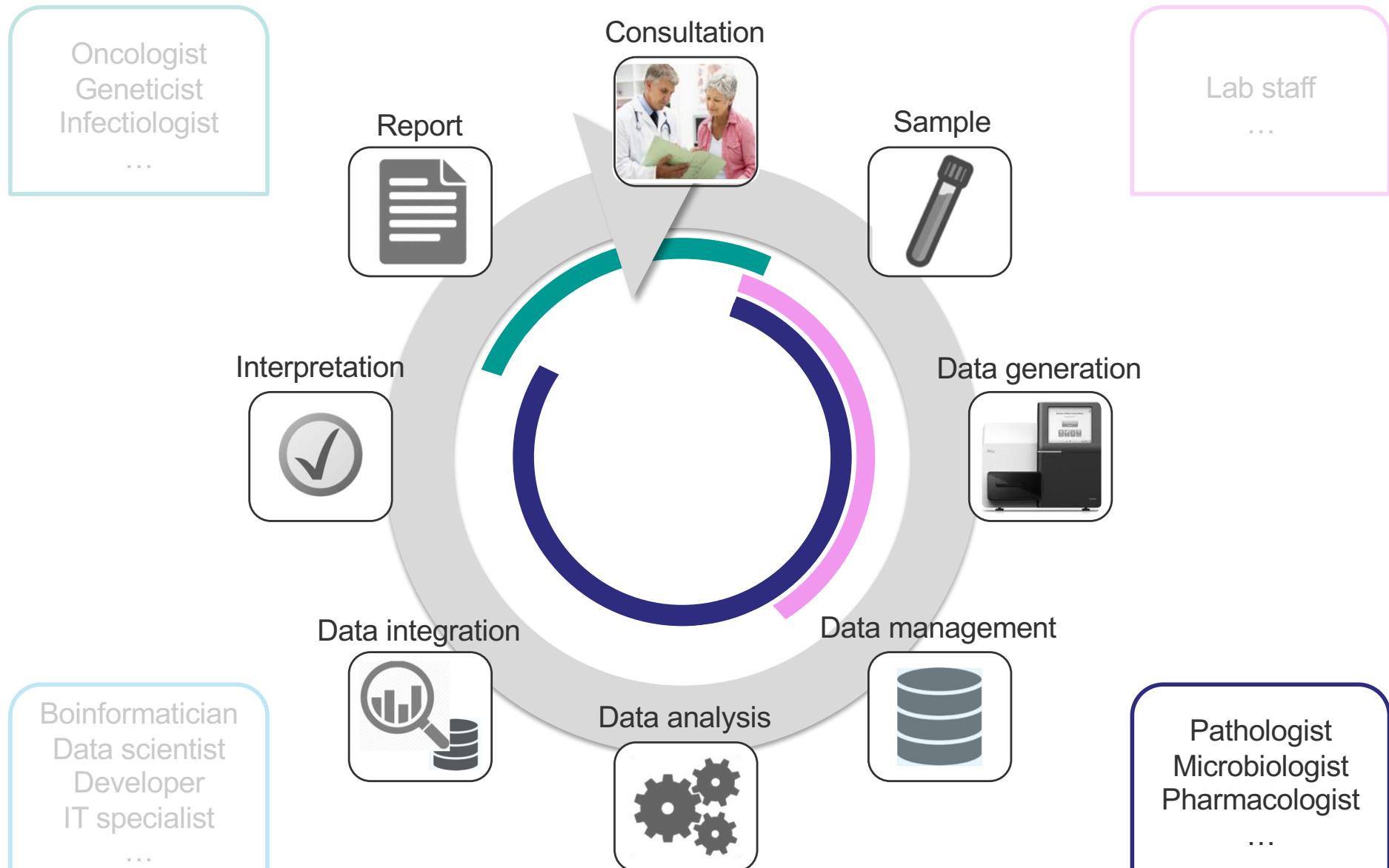
Multiple expertise



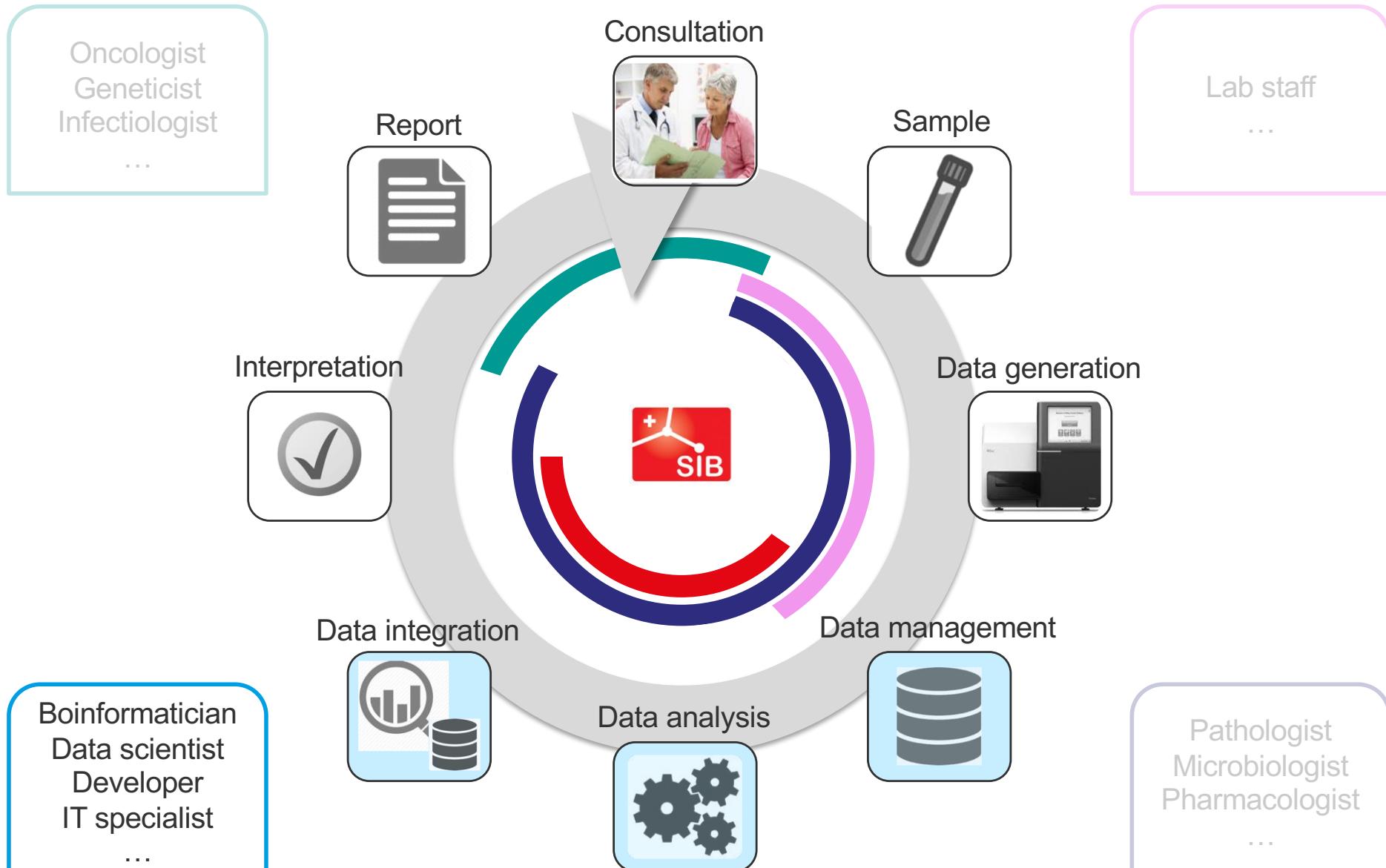
Multiple expertise

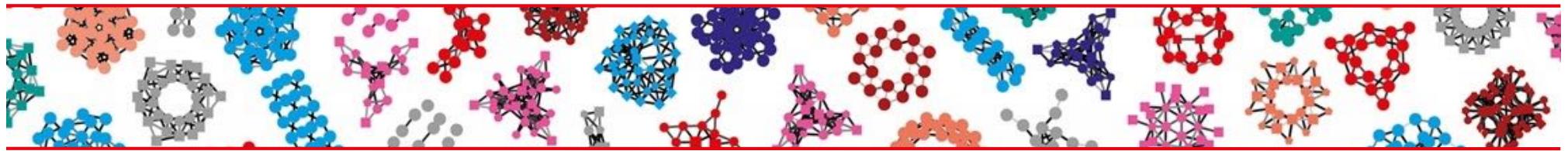


Multiple expertise



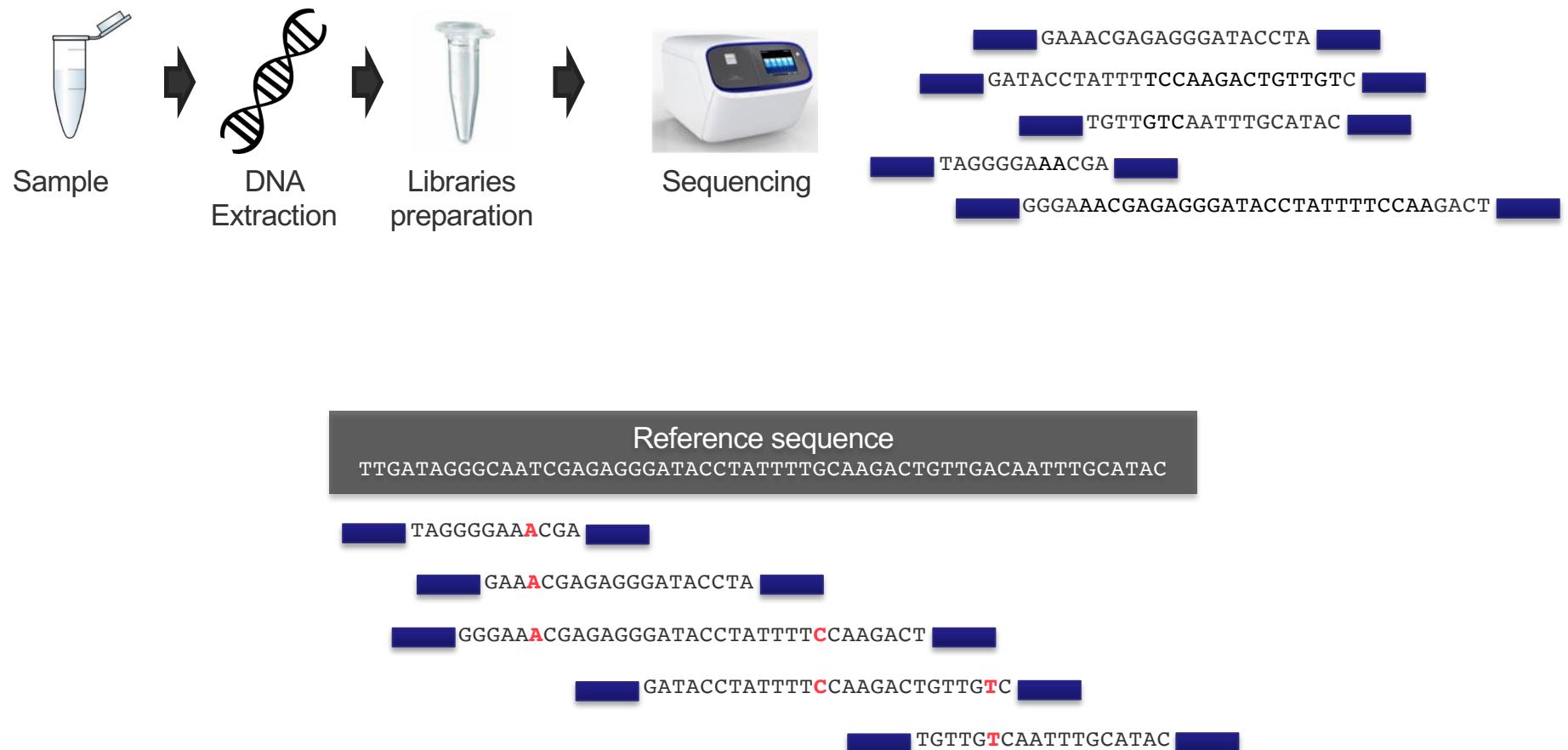
Multiple expertise



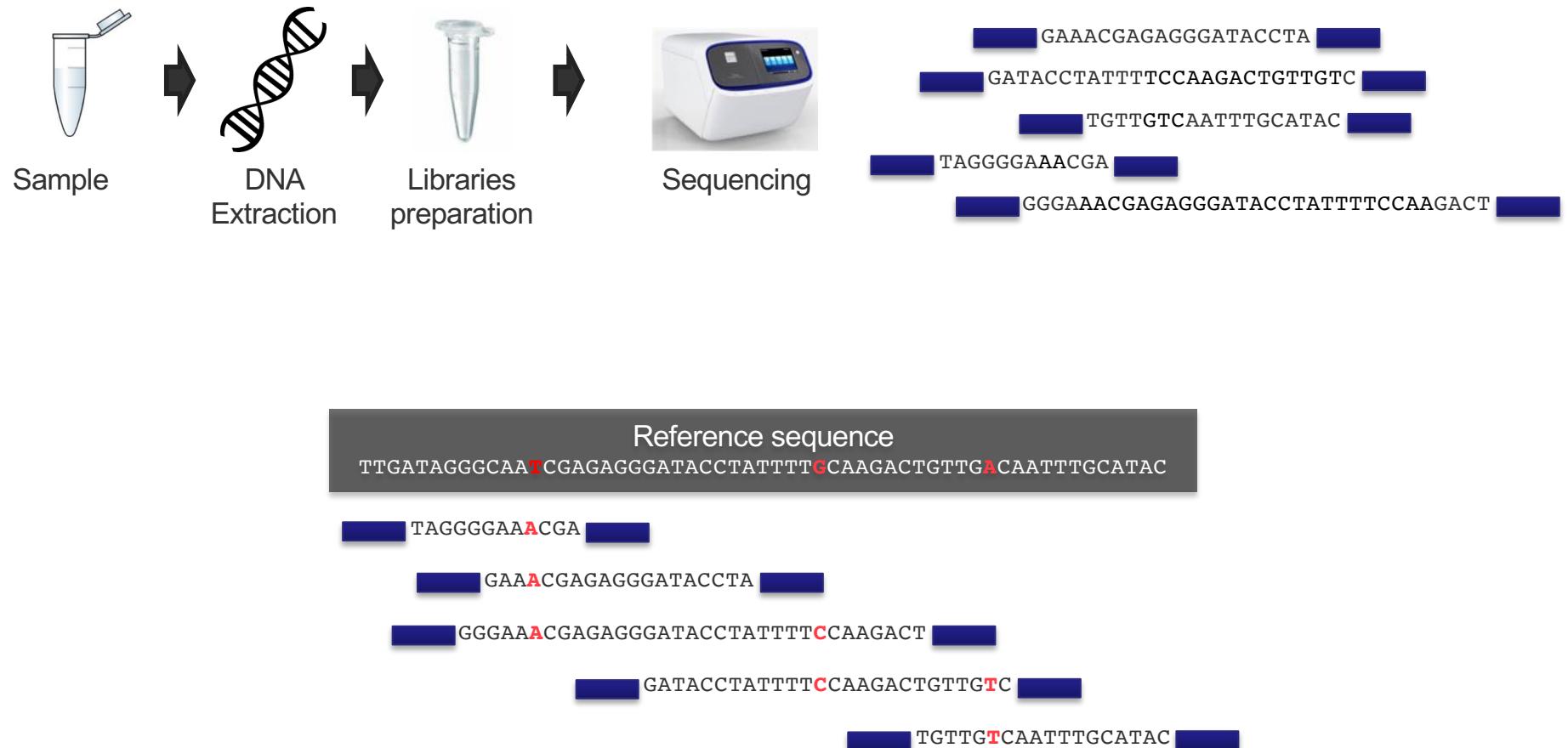


Applications of Next Generation Sequencing (NGS) in medical diagnosis

Next Generation Sequencing principle



Next Generation Sequencing principle

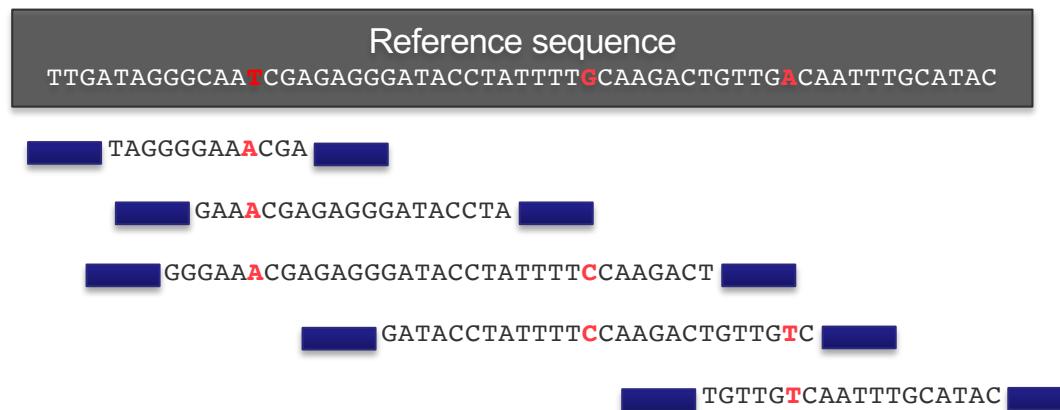


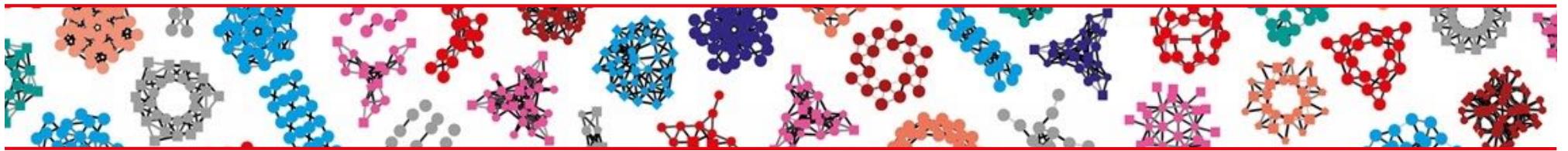
A matter of scale...



NGS clinical applications

| | Source DNA | Reference DNA |
|------------------|------------------------|---|
| Oncology | Patient tumor or blood | Consensus human genome Germline |
| Microbiology | Patient | Pathogens genomes, resistance genes |
| Medical genetics | Patient | Family members, known defects |
| Pharmacogenetics | Patient | Drug-response or -sensitivity mutations |





Applications of Next Generation Sequencing (NGS) in medical diagnosis

Focus on oncology



PART I

Overview of an NGS bioinformatics pipeline

NGS in cancer diagnosis?

- Identify single nucleotide variants (SNVs), insertions-deletions (indels) to inform clinical management

at~~t~~cgggtcatgccatagggg

Single Nucleotide Variant (SNV)

at~~g~~cgggtcatgccatagggg

Insertion

atgcgggtcat~~cgtgtcc~~gcccatagggg

Deletion

atgcgggtcatcgtgtccg....tagggg

ccca

Overview of a NGS bioinformatics pipeline



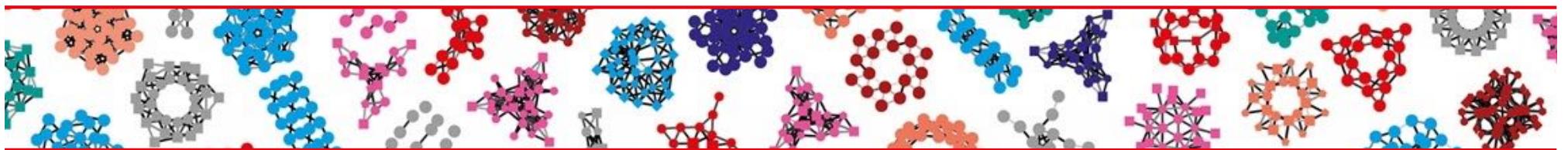
■ Gene panels analysis in clinical routine

- Identify **artifacts**: quality control
- Identify **somatic** vs. germline variants
- Variant **annotation**: is it pathogenic, actionable?



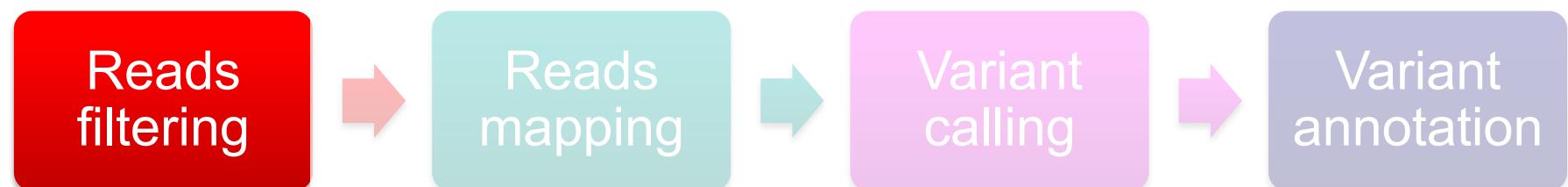
PART II

Quality control



PART II

Quality control



Out of the sequencer: FASTQ

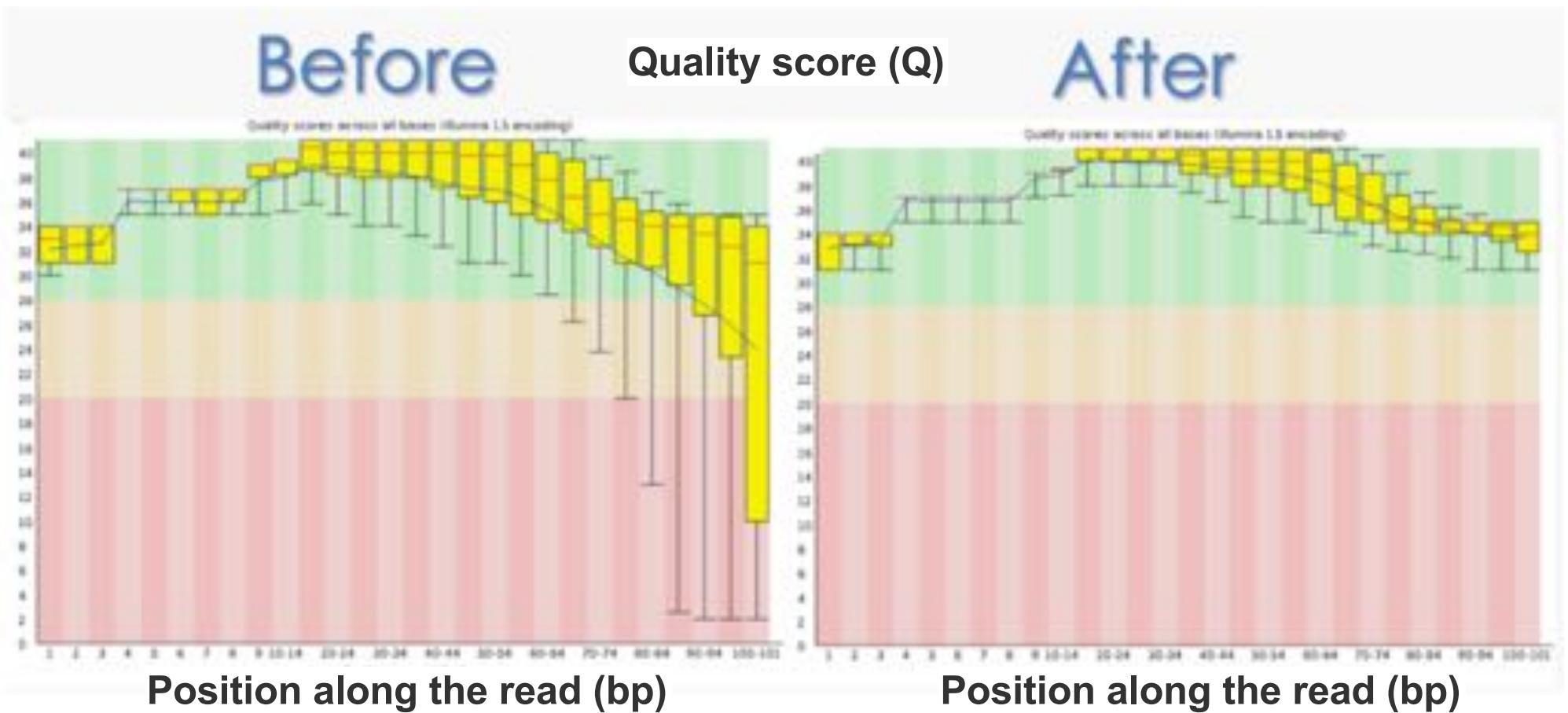
| | |
|----------------|--|
| Identifier | @SRR566546.970 HWUSI-EAS1673_11067_FC7070M:4:1:2299:1109 length=50 |
| Sequence | TTGCCTGCCTATCATTAGTGCCTGTGAGGTGGAGATGTGAGGATCAGT |
| '+' sign | + |
| Quality scores | hhhhhhhhhhghhhhhfhhhhfffffe'ee[X]b[d]ed[Y][^Y] |
| Identifier | @SRR566546.971 HWUSI-EAS1673_11067_FC7070M:4:1:2374:1108 length=50 |
| Sequence | GATTTGATGAAAGTATAACAACACTAAAAGTCAGGTGGATCAGAGTAAGTC |
| '+' sign | + |
| Quality scores | hhhhgfihcgghggfcffdhfehhhcehdchhdhahehffffde'bVd |

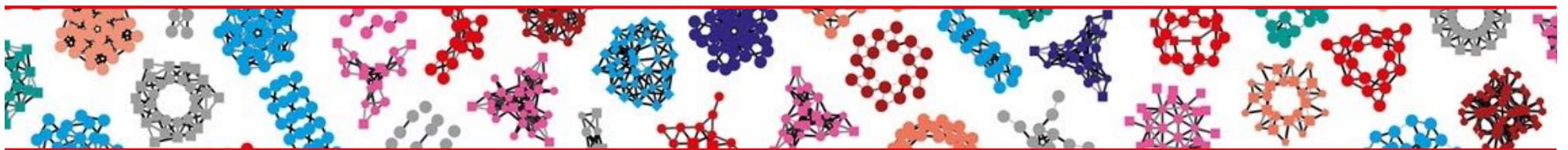
Each nucleotide has a **quality score (Phred score)** representing the probability that a base was miscalled by the sequencer

$$Q = -10 \log_{10} P$$

| Phred Score | Prob. of incorrect base call | Base call accuracy | Code |
|-------------|------------------------------|--------------------|------|
| 10 | 1 in 10 | 90% | J |
| 20 | 1 in 100 | 99% | T |
| 30 | 1 in 1'000 | 99.9% | ^ |
| 40 | 1 in 10'000 | 99.99% | h |

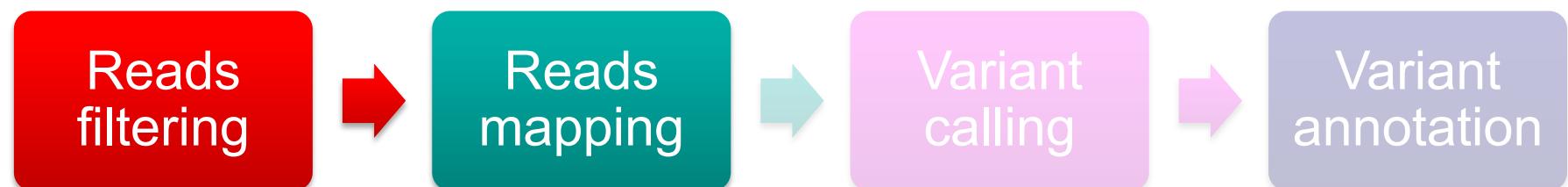
Quality-based reads trimming





PART II

Quality control



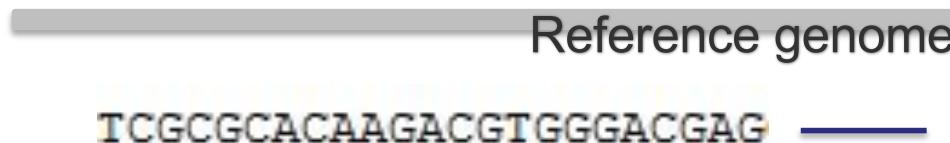
Let's align the reads



! Short reads are likely to map at several positions along the reference genome

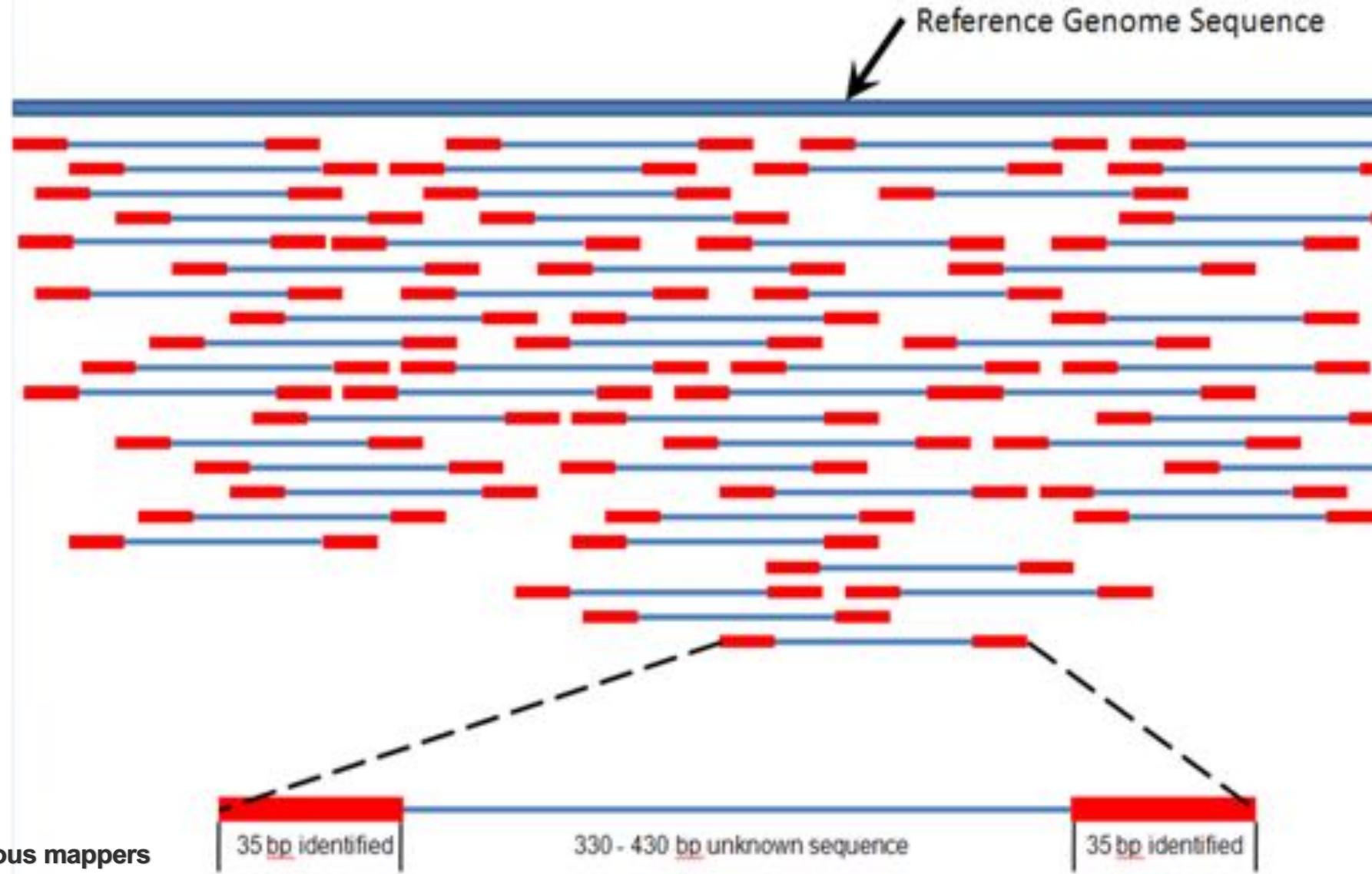


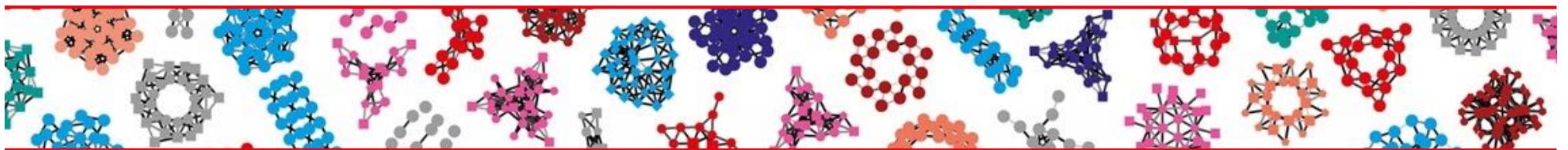
! Mismatches and gaps allowed
→ algorithms have scoring functions



! Longer reads are less ambiguous
→ but computationally more expensive

Mapping: finding the best position for each read





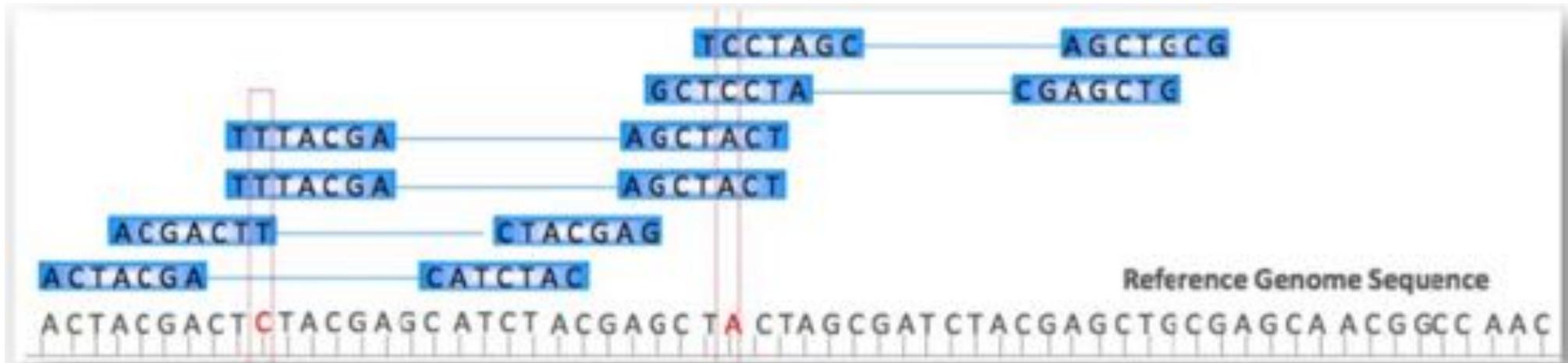
PART II

Quality control



Variant calling: putting it all together

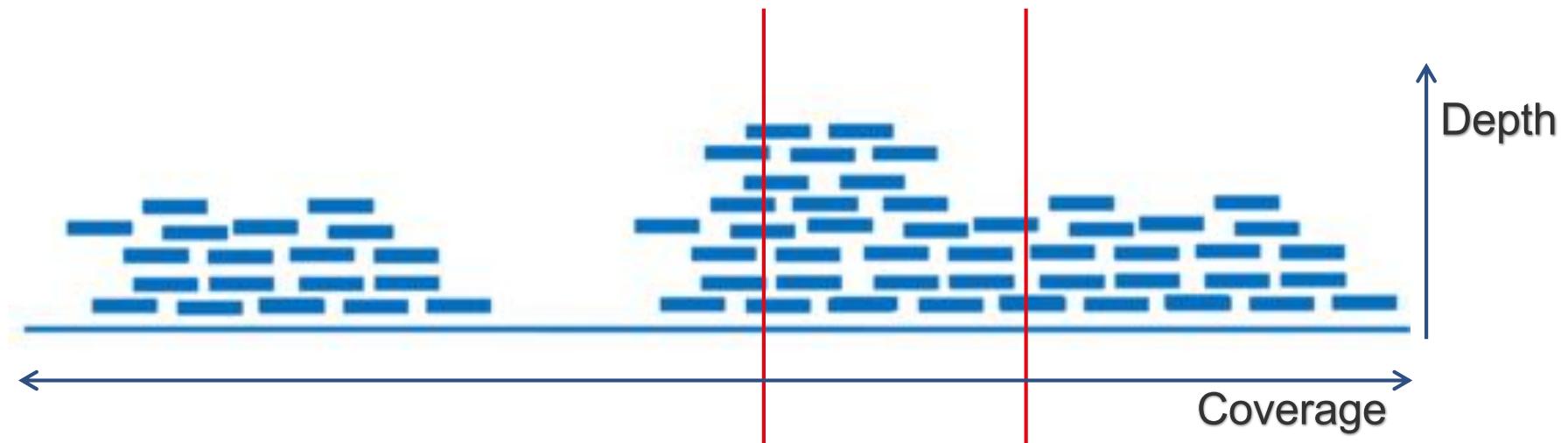
- Is it a true variant or a sequencing error?
- Variant callers generally assume that sequencing errors are independent across reads



Things to watch out for
when assessing variant quality

Depth and coverage

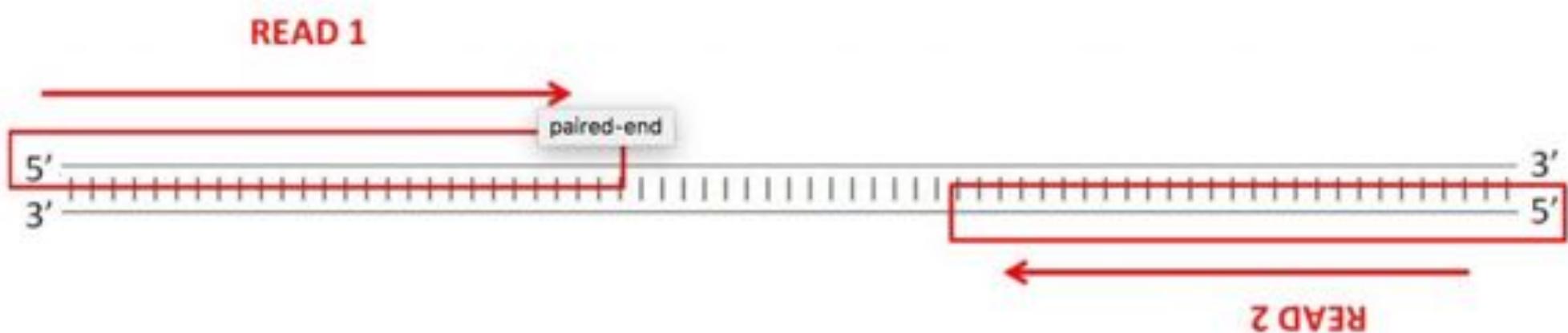
- **Depth:** nb of reads that include a given nucleotide, at a given position (e.g. 1000X)
- **Coverage:** percentage or nb of bases of a reference genome covered with a certain depth, e.g. 90% at 5X



*Many people use “coverage” for “depth”.
Watch out if % or X*

Strand bias

- Both DNA strands are sequenced
- Bias occurs when one strand is favored over the other
- Normal mutations should occur on both + and – strands with equal frequencies

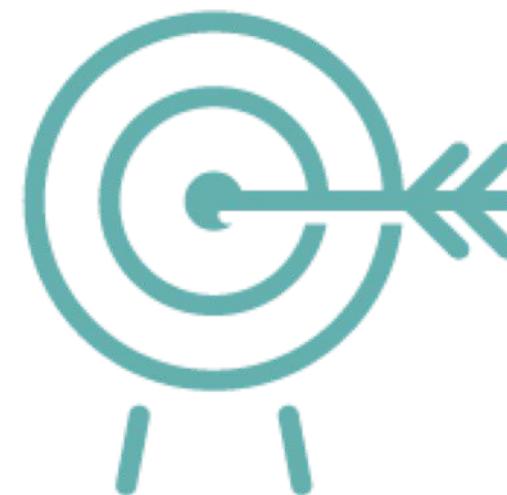
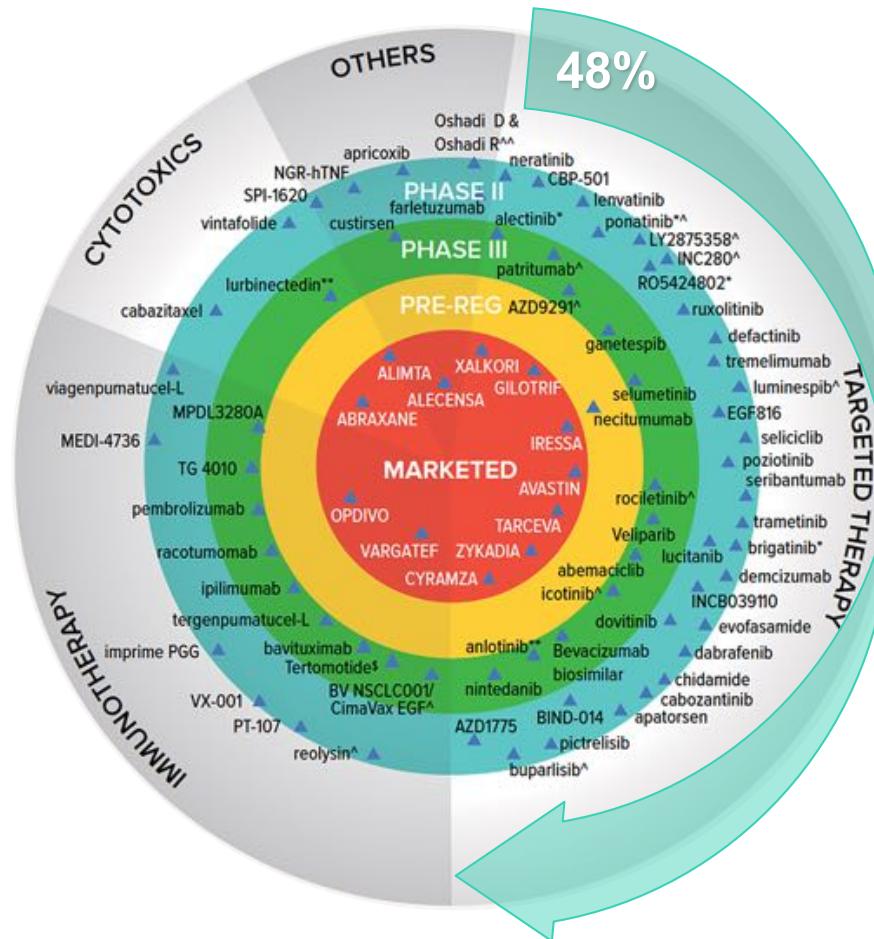




PART III

Variant annotation
and interpretation

Predictive: personalized molecular oncology

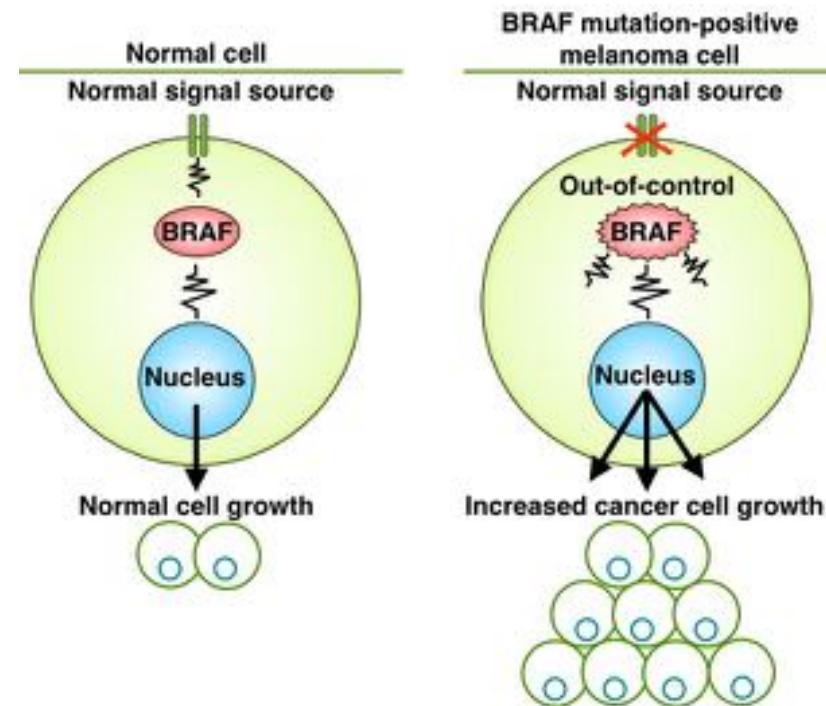


Lung cancer: about 100 drugs at different stages of development process

Medical genetics: focus on pathogenicity

- **Pathogenic variant:** genetic alteration that increases an individual's **susceptibility or predisposition** to a certain disease.

- **5 levels (guidelines)**
 - Pathogenic
 - Likely pathogenic
 - Variant of unknown significance
 - Likely benign
 - Benign



Oncology: focus on clinical significance

The Journal of Molecular Diagnostics, Vol. 19, No. 1, January 2017



SPECIAL ARTICLE

**Standards and Guidelines for the Interpretation
and Reporting of Sequence Variants in Cancer**



CrossMark

*A Joint Consensus Recommendation of the Association for
Molecular Pathology, American Society of Clinical Oncology,
and College of American Pathologists*

Finding **actionable**
variants

“Unlike interpretation of germline sequence variation, which focuses on pathogenicity (...), interpretation of somatic variants should be focused on their impact on clinical care”.

Finding actionable variants



Filtered list
of variants

Predisposition

*Indicates risk
to develop a disease*

Diagnostic

Supports disease characterization

Prognostic

Indicates disease evolution

Predictive

Supports treatment decisions

Other important questions

- Is it prevalent in the cancer subtype of interest?
- Is it known in other cancer subtypes or diseases?
- Is it present in the general population?
- Are there other known variants in the same gene?
- Is it related to an ongoing clinical trial?
- What is the evidence level? Observed vs. predicted

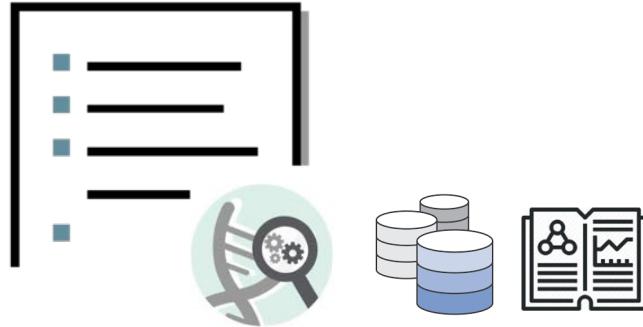


PART III

Variant annotation
and interpretation

... bioinformatics at the rescue

Bioinformatics to the rescue... for annotation



- **Genes and transcripts affected by the variant**
- **Location of the variants (e.g. coding, noncoding region...)**
- Predict variant effect (e.g. stop gained, missense...)
- Predict variant impact on protein function
- Retrieve annotations from public databases

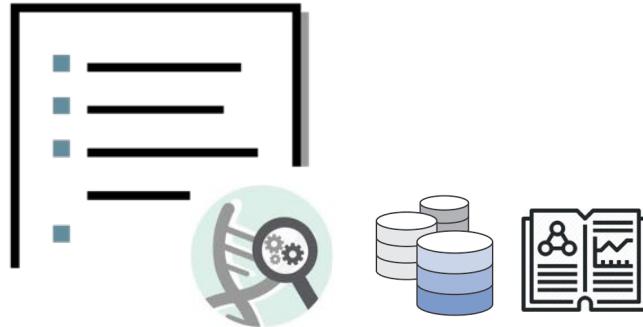
Locating variants

- Convert **genomic coordinates** (chromosome, position) to the corresponding **cDNA/amino-acid coordinate system**

- HGVS nomenclature: <http://varnomen.hgvs.org>
 - Substitution c.76A>T
 - Deletion c.76~~A~~T
 - Insertion c.76_77insG
 - Protein sequence p.Lys76Asn
 - Genomic sequence g.476A>T

- *Important to store for tracking:*
 - Version of the human genome assembly
 - Accession and version of the mRNA transcripts

Bioinformatics to the rescue... for annotation

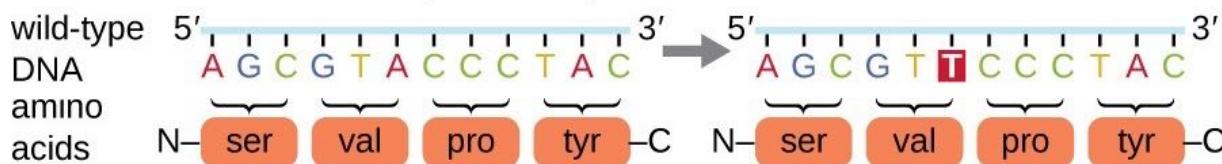


- **Genes and transcripts affected by the variant**
- **Location of the variants (e.g. coding, noncoding region...)**
- **Predict variant effect (e.g. stop gained, missense...)**
- Predict variant impact on protein function
- Retrieve annotations from public databases

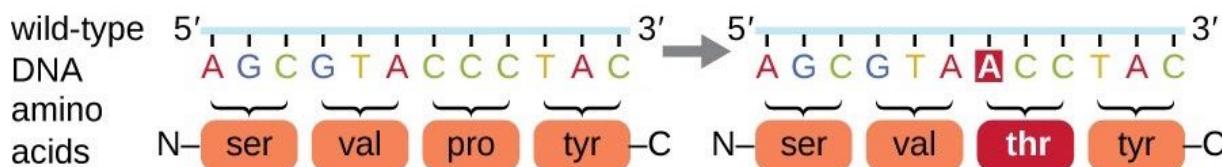
Variant effects on the protein

point mutation: substitution of a single base

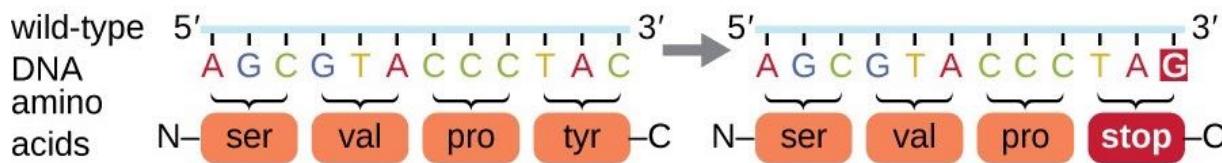
silent: has no effect on the protein sequence



missense: results in an amino acid substitution



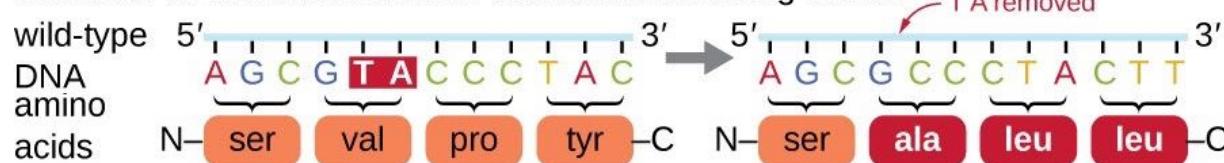
nonsense: substitutes a stop codon for an amino acid



Variant effects on the protein

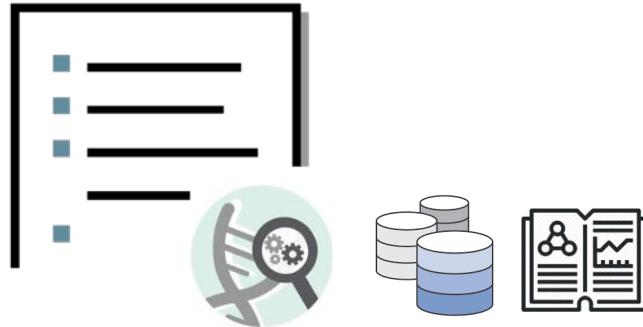
frameshift mutation: insertion or deletion of one or more bases

Insertion or deletion results in a shift in the reading frame.



→ An ontology exists to describe variant effects: sequence ontology
<http://www.sequenceontology.org/browser/obob.cgi>

Bioinformatics to the rescue... for annotation



- **Genes and transcripts affected by the variant**
- **Location of the variants (e.g. coding, noncoding region...)**
- **Predict variant effect (e.g. stop gained, missense...)**
- **Predict variant impact on protein function, splicing**
- **Retrieve annotations from public databases**

What is the impact of non-silent mutations?

- Is the mutation in a **gene, exon, regulatory region...**?
- Is the mutation in an **evolutionarily conserved** region accross species?



Predicting the impact: examples of tools

not exhaustive

| TOOLS | SnpEff (ClinEff) | VEP | SIFT | PolyPhen-2 | FATHMM |
|--|-------------------------------|-----|---------------------------------|------------------------------------|--|
| Variant effect and location (sequence ontology) | ✓ | ✓ | | | |
| Prediction of impact (score or category) | ✓ | | ↔ ↔ | ✓ | ✓ |
| >> Features used for impact prediction | rules based on variant effect | | aa conservation in related seq. | aa conservation + structural feat. | aa conservation + protein tolerance to mutations |

SnpEff impact rules

| Putative Impact | Sequence Ontology term |
|-----------------|-----------------------------|
| HIGH | start_lost |
| HIGH | stop_gained |
| HIGH | stop_lost |

http://snpeff.sourceforge.net/VCFanotationformat_v1.0.pdf

What if different tools predict different things?

■ ACMG/AMP Guidelines

- Use a combination of tools.
- Only keep variants with consensus predictions.

■ But which combination of tools to use?

- A recent study proposes combinations of tools with increased concordance for clinically relevant variants

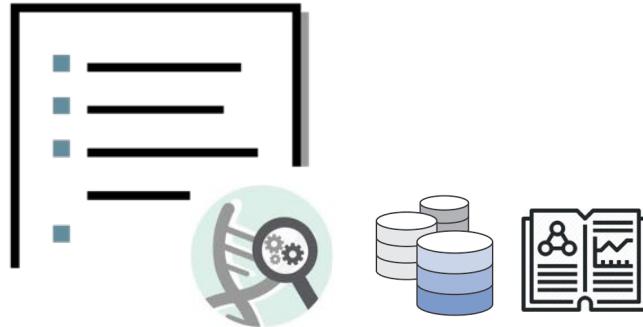
<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1353-5>

I found a damaging mutation: is it always bad?

- Keep the mutation in context: what is the gene function?
 - **Tumor suppressor genes**
→ damaging mutations are pathogenic
 - **Oncogenes**
→ activating mutations are pathogenic
(beware damaging mutation can be activating!)

→ **Keep gene function in mind
when interpreting deleteriousness**

Bioinformatics to the rescue... for annotation



- **Genes and transcripts affected by the variant**
- **Location of the variants (e.g. coding, noncoding region...)**
- **Predict variant effect (e.g. stop gained, missense...)**
- **Predict variant impact on protein function, splicing**
- **Retrieve annotations from public databases**



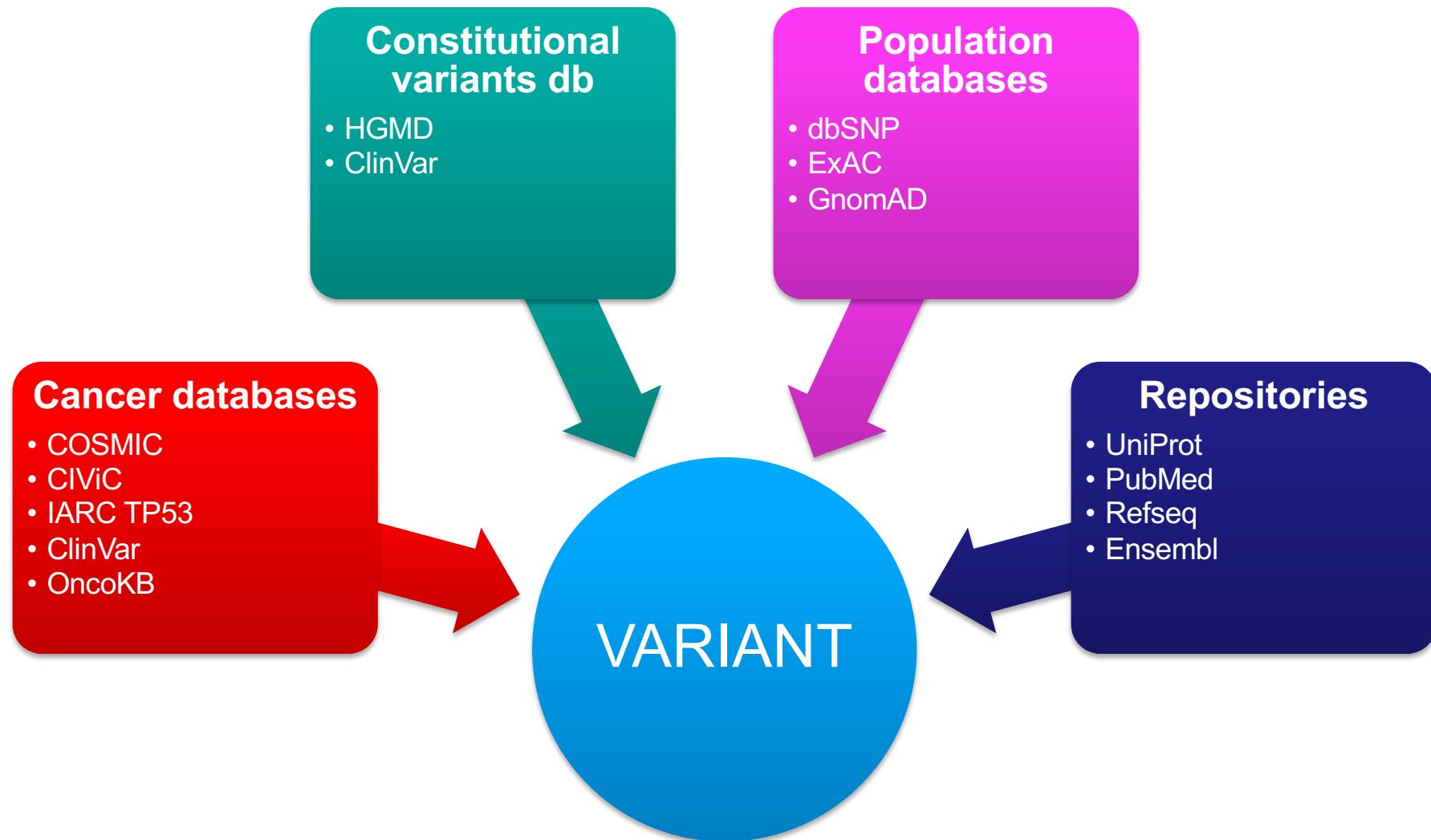
PART III

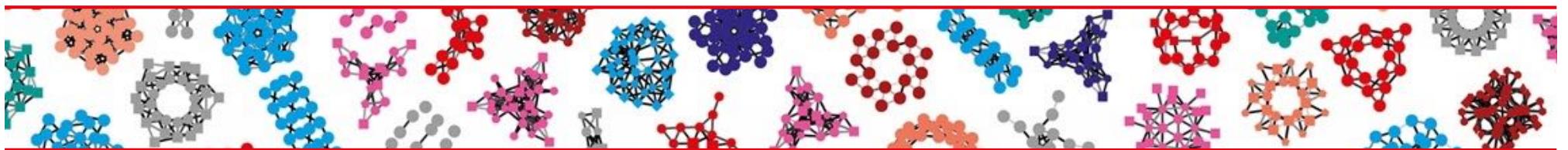
Variant annotation
and interpretation

... with knowledge-bases

Annotating a variant: knowledgebases

not exhaustive





Real-life implementations

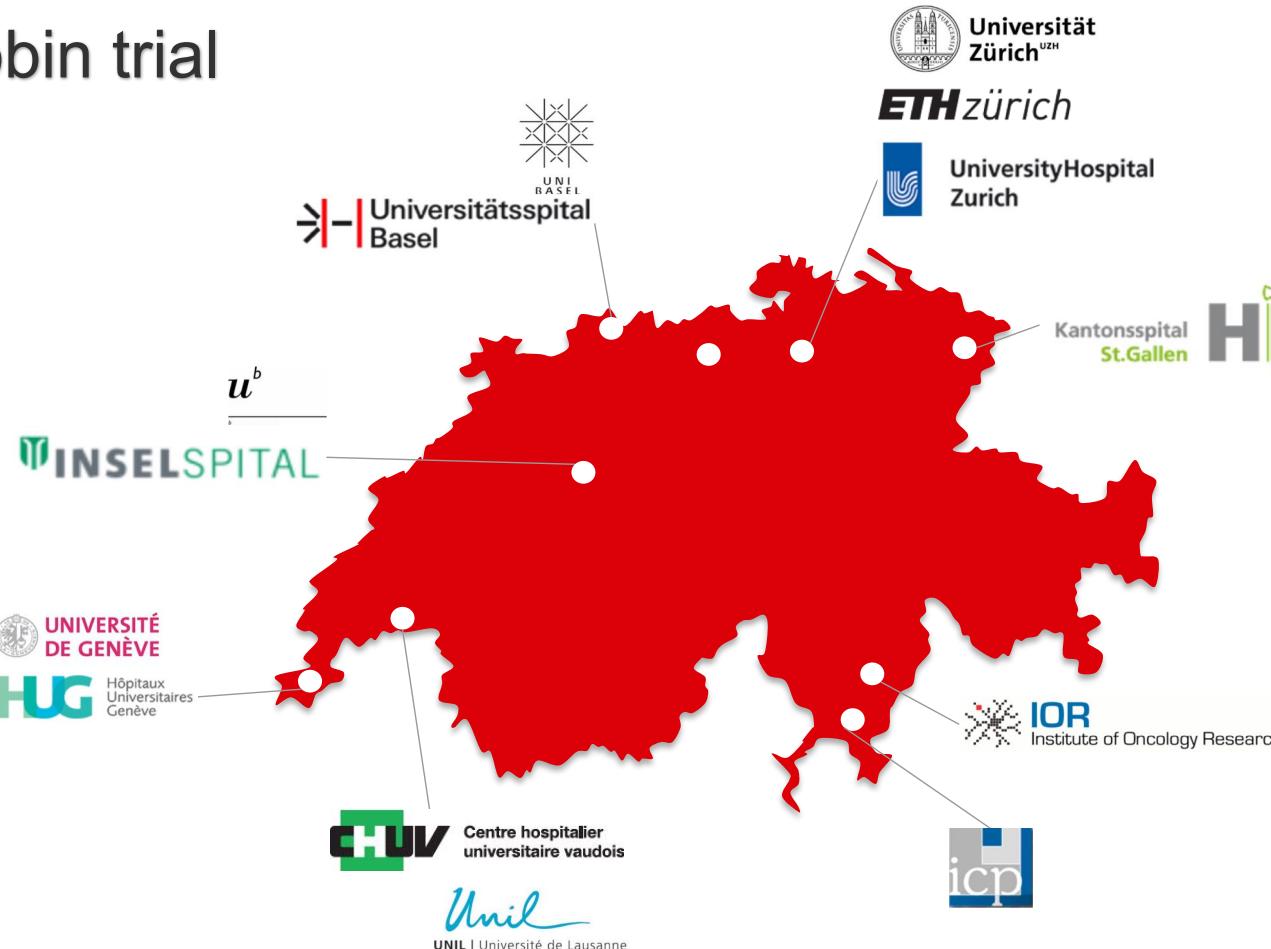
Mission of SIB Clinical Bioinformatics

Support the organization, analysis and interpretation
of -omics data for **diagnostic** purpose



SIB NGS cancer working group

- Since Nov 2015
- Round robin trial
- Training

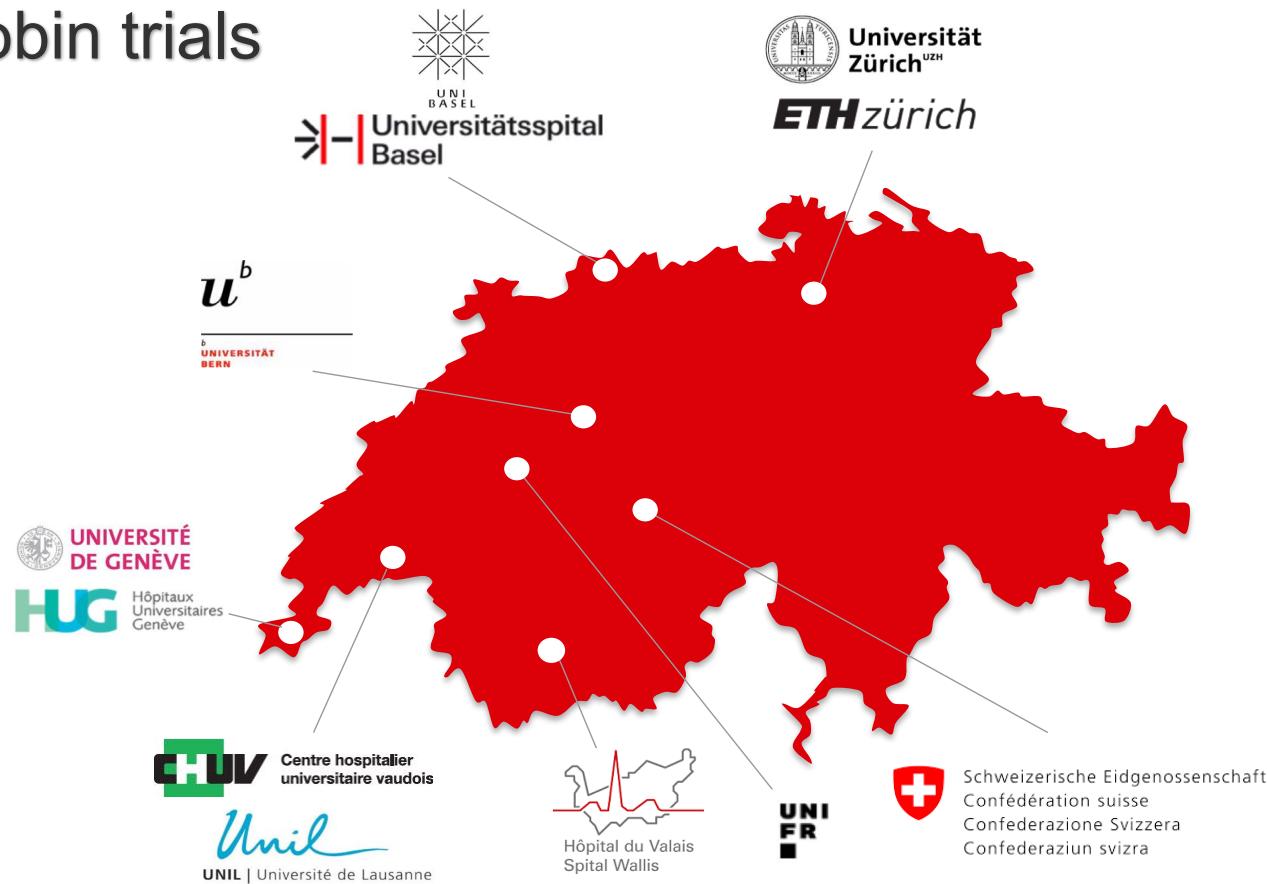


SIB NGS infectious diseases working group

- Since Sept 2016

- Round robin trials

- Training

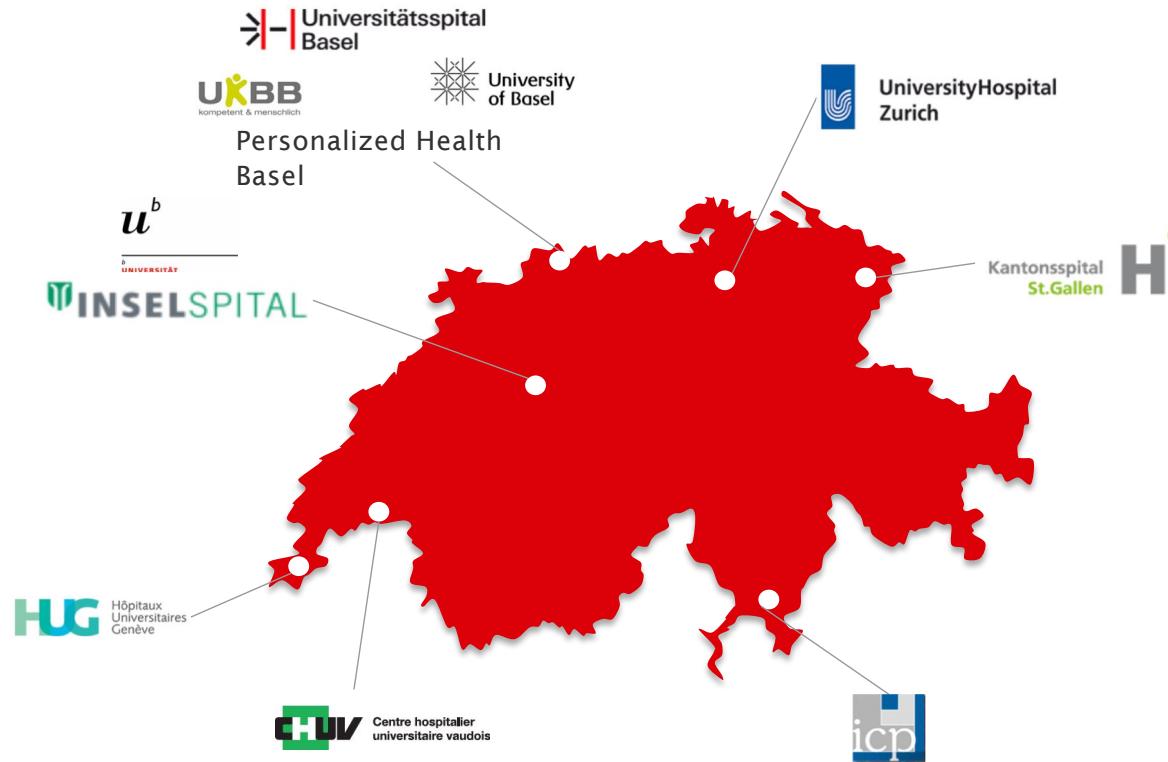


Mission of SIB Clinical Bioinformatics

Support the organization, analysis and interpretation
of -omics data for **diagnostic** purpose



Swiss Variant Interpretation Platform

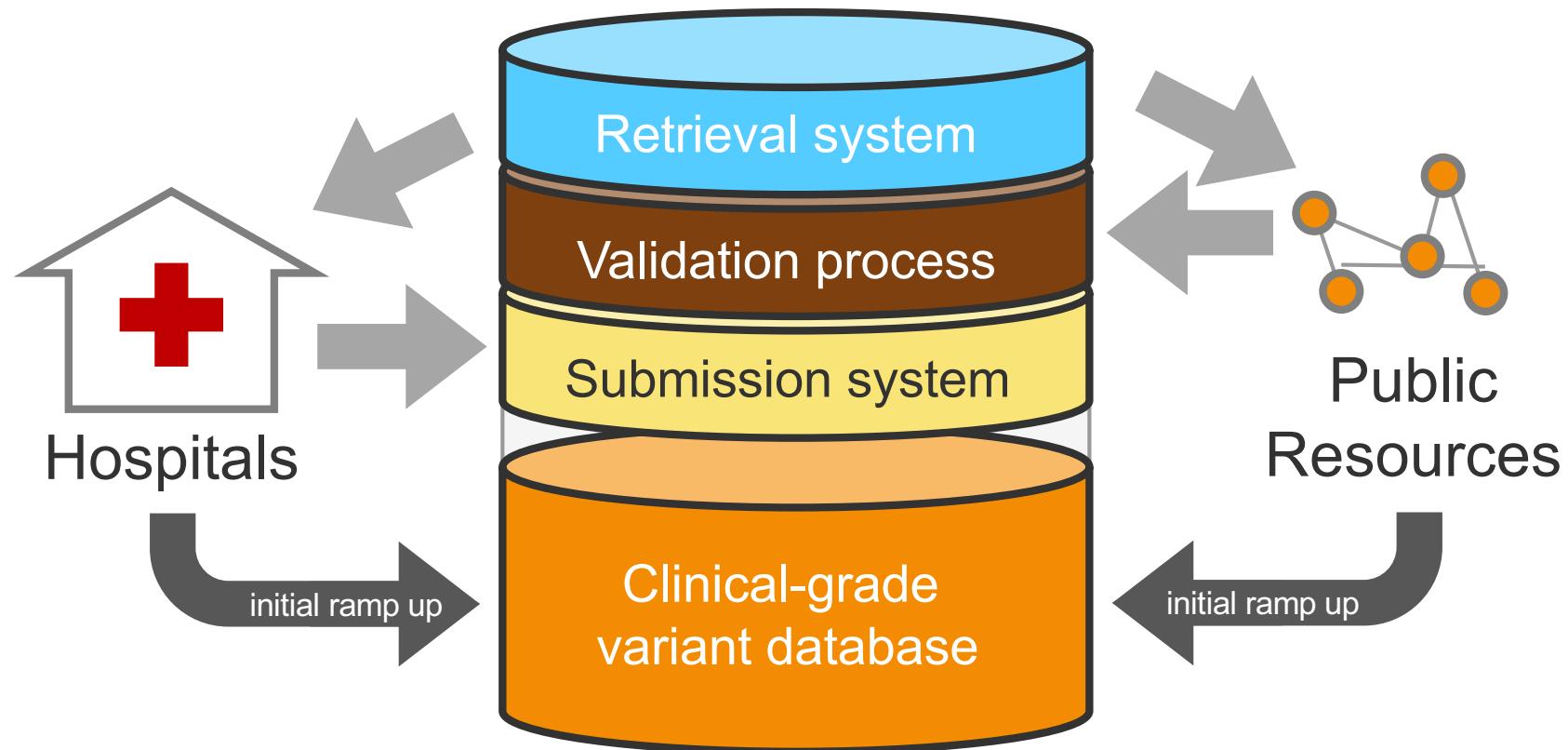


- Centralize cancer variants identified in Swiss patients in one single place, agreeing on their clinical interpretation

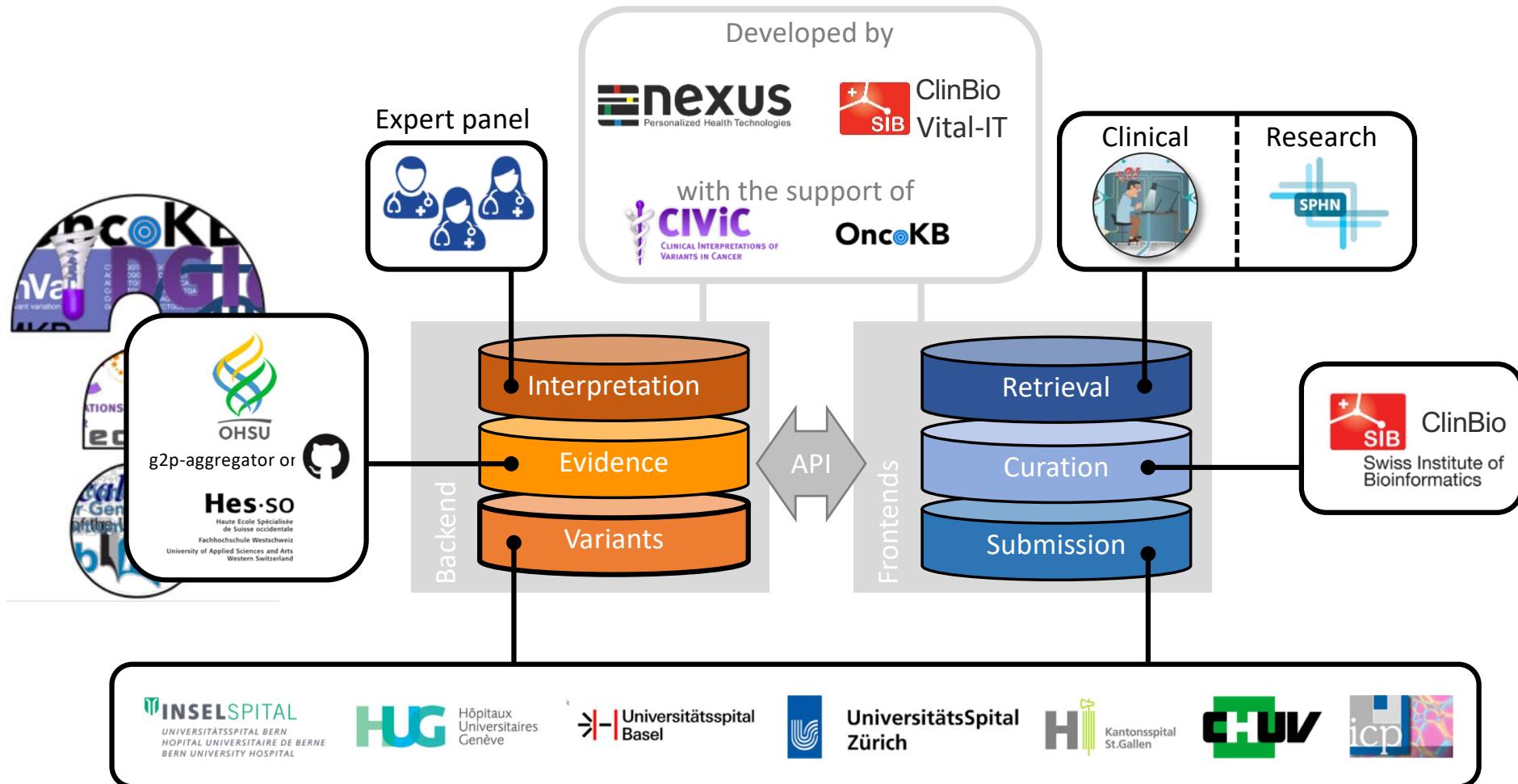
SVIP proposal designed with the SIB Somatic Mutation Calling Working Group since 2016

- One-stop shop for variants in Swiss patients
- Use SIB curation expertise
- Clinical expert panel for validation
- Harmonized clinical interpretation
- Support diagnostics and enable research

Platform workflow...

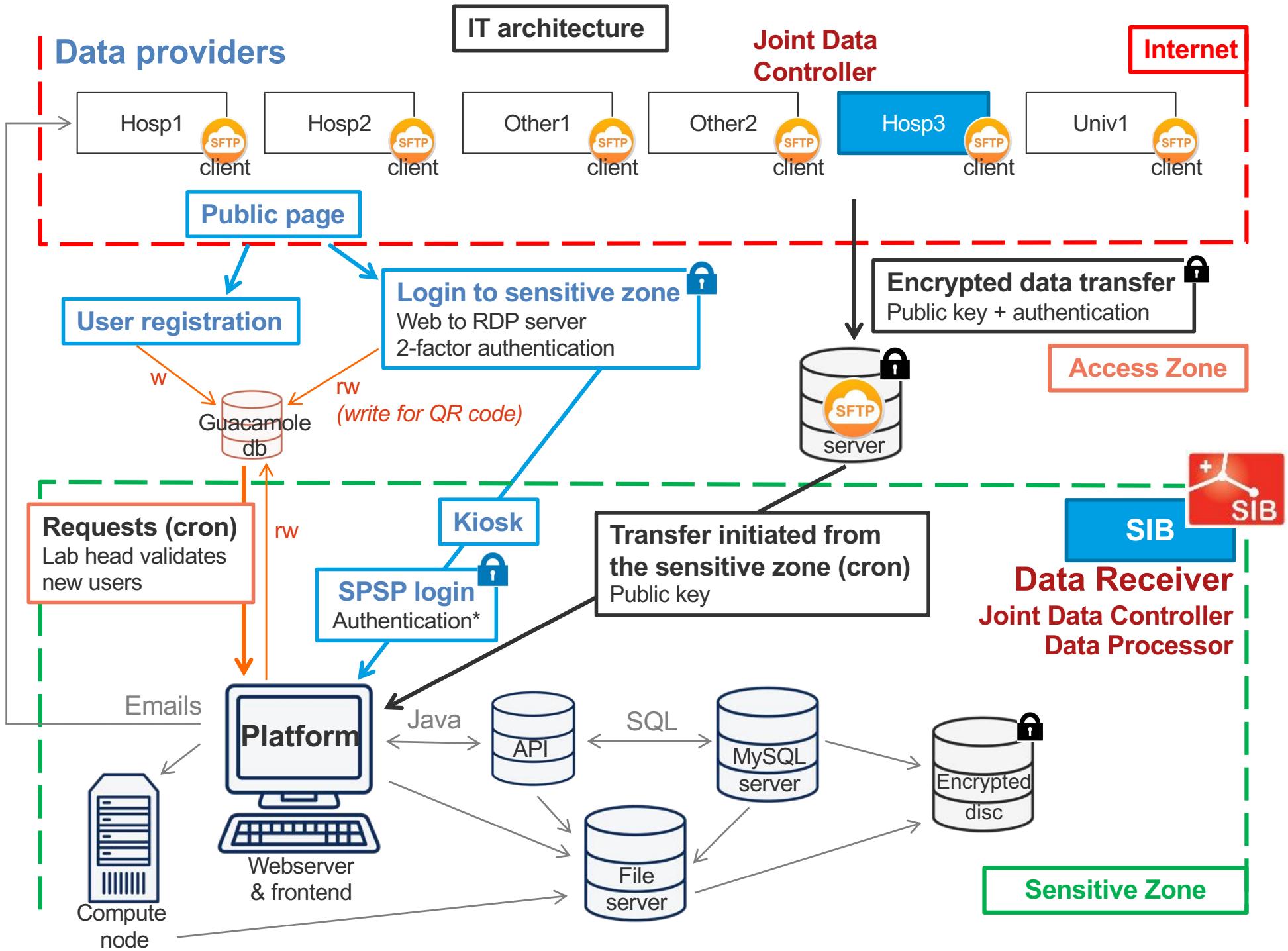


...but there's more to it



Specific requirements

- **Contain some identifiable data**
- **Two access layers**
 - Clinical layer, accessible to partner clinicians only
 - Research layer, public
- **Need for a specific IT and system architecture**
- **SPHN-wide mechanisms should be used**



Swiss Pathogen Surveillance Platform



FNSNF
SWISS NATIONAL SCIENCE FOUNDATION

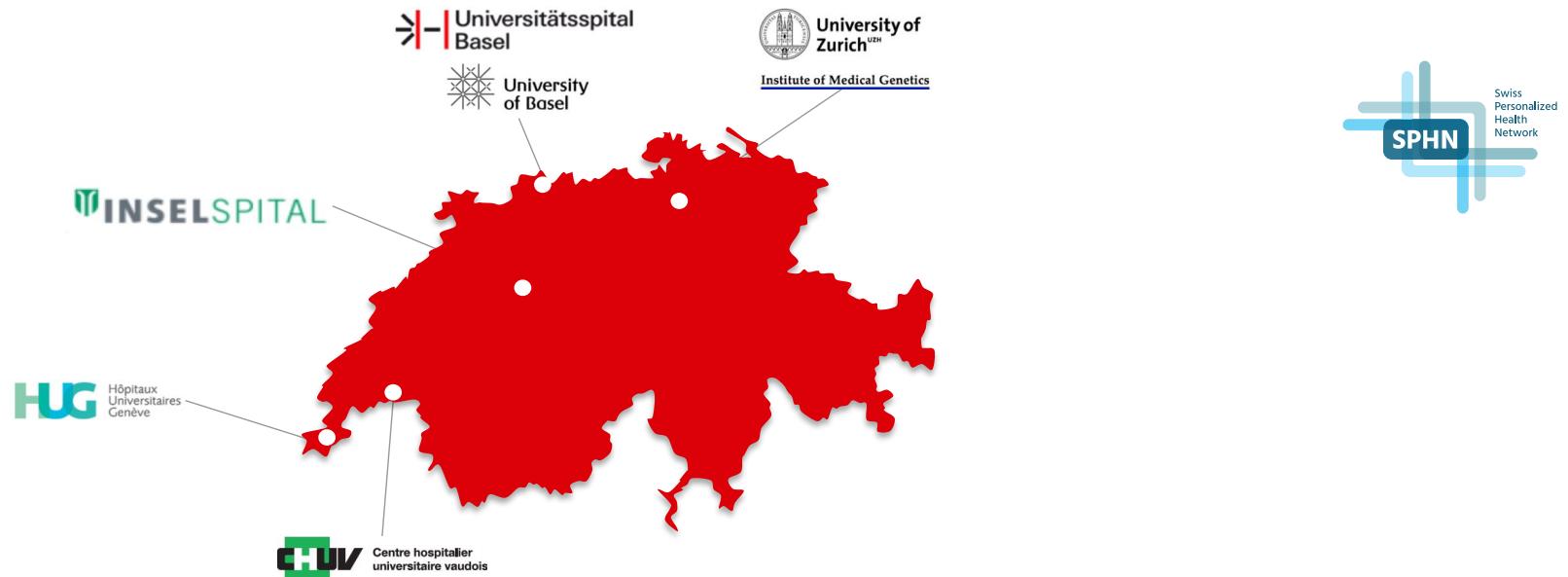
72
NRP
Antimicrobial Resistance
National Research Programme

Swiss national project
led by Dr. Egli



- Common platform for microbial WGS analyses
- Identify risks of multi-drug resistant bacterial pathogens
- Explore risks by predicting dynamics of spread
- Inter-cantonal, standardized data, harmonized methods

Now coming: SwissGenVar



- Collection of genetic variants identified in patients by Swiss clinical genetic laboratories
 - Harmonization, sharing and up-scaling of manually curated evidences and variant interpretation by clinical genetic experts
 - Accessibility of these expert-annotated variants for clinicians and researchers, integrated into SPHN and international efforts
-

Mission of SIB Clinical Bioinformatics

Support the organization, analysis and interpretation
of -omics data for **diagnostic** purpose



Training & outreach

- Certificate of Advanced Studies in Personalized Molecular Oncology (pmo.unibas.ch)



UNI
BASEL

- NGS QC and annotation for cancer diagnosis **HUG**

- MOOC on Precision Medicine **HUG**  **UNIVERSITÉ
DE GENÈVE**

- ESCMID Workshop on bioinformatics for bacterial genomics – 2019 

- Outreach events
(health exhibitions, scientific cafés, schools kids...)

Certificate of Advanced Studies (CAS) in Personalized molecular oncology

pmo.unibas.ch



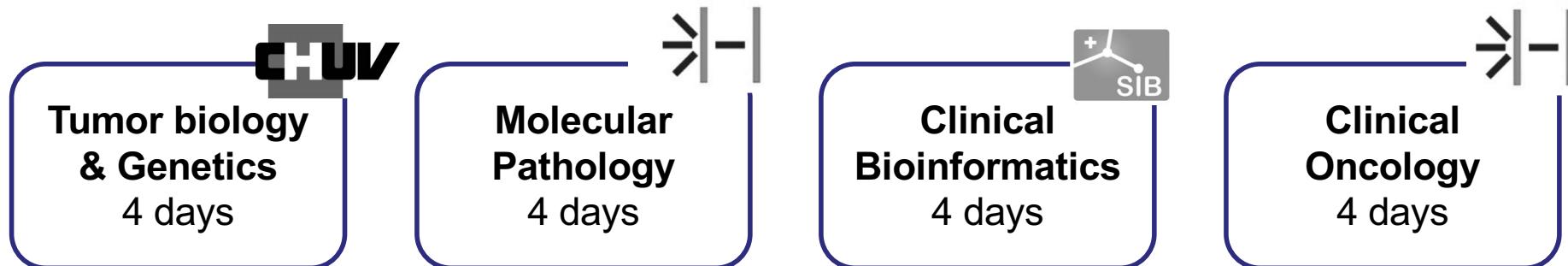
Universitätsspital
Basel



Swiss Institute of
Bioinformatics



CAS PMO: 4 modules and a mini-thesis



- | | | | |
|---|--|---|---|
| <ul style="list-style-type: none">• Cytogenetics and molecular genetics• Genetic modifications• Tumor biology: solid and hematological• Tumor genetics | <ul style="list-style-type: none">• Omics technologies• From sample to data: extraction, sequencing, panels• Quality control and accreditation• Molecular profile interpretation, reports | <ul style="list-style-type: none">• NGS data processing: mapping, calling, annotation• Data quality control• Hardware, security, privacy• Artificial intelligence applications | <ul style="list-style-type: none">• Tumor physiology & immunology• Prognostic and predictive markers• Interpretation of genetic results• Clinical trials and tumor board |
|---|--|---|---|

Mission of SIB Clinical Bioinformatics

Support the organization, analysis and interpretation
of -omics data for **diagnostic** purpose



Medical and industrial applications

■ Geneva University Hospital

- OncoBench™: genetics diagnostic pipeline for pathology



■ Zürich Medical Genetics Institute

- GenBench™ in development: medical genetics diagnosis platform



■ Pharma industry

- Bacterial genotyping tool for drug substance manufacturing (to be GMP validated)

■ Start-up

- Collaboration on imaging technology with a start-up and an hospital, in preparation

SIB Clinical Bioinformatics



Florent
Tassy



Valérie
Barbié



Aitana
Lebrand



Steffen
Pade



Valérie
Hinard



Miriam
Tesfai



Dillenn
Terumalai



Abdullah Kahraman
(USZ, Molecular Pathology)



Yann Christinat
(HUG, Molecular Pathology)

SIB collaborations

- Vital-IT (M. Ibberson)
- Nexus (D. Stekhoven)
- Text Mining (P.Ruch)
- SIS (B. Rinn)
- Core-IT (H. Stockinger)
- SwissProt (A. Bridge)
- C. Dessimoz's group
- V. Zoete's group
- J. Fellay's group
- ...

Thank You

