# NextGenerationSequencing

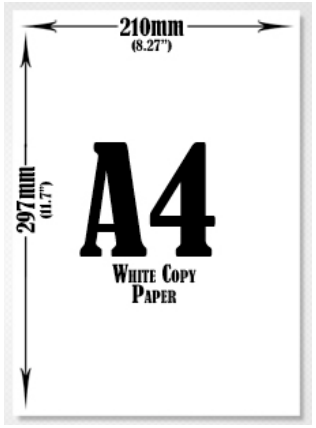
## Design

Niklaus Zemp
June 2020

Bioinformatics
Genetic Diversity Centre (GDC)
ETH Zurich

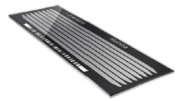

www.lustich.de

# NGS data



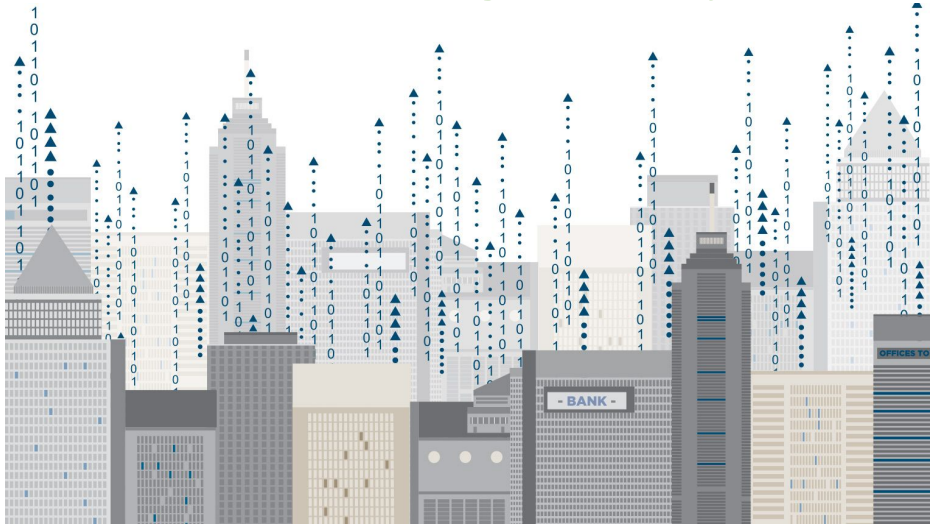

A4 => Font 10

NovaSeq
3000 Gb

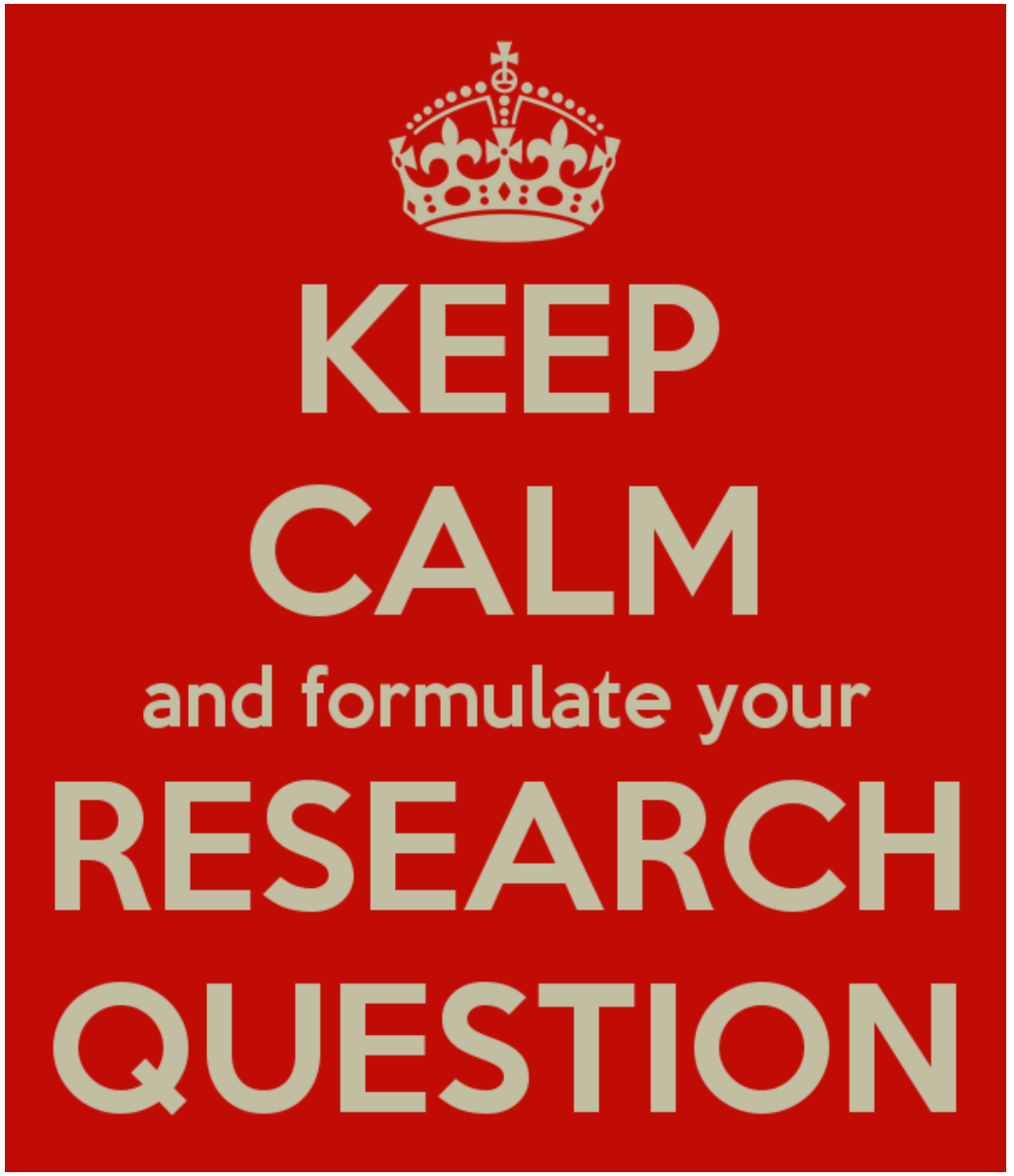


~500 Mio pages
**-> 54 km**

3Vs:
-> High volume
-> High velocity
-> High variety




Big data


KEEP
CALM
and formulate your
RESEARCH
QUESTION

# Workflow

1) Define research question and formulate hypotheses
2) Select application/methode
3) Select NGS platform
4) Define experimental design
5) Valdiate your results

# How many samples can I pool?

$$C = LN / G$$

C coverage; L read length ; N inumber of reads; G genome length

https://support.illumina.com/downloads/sequencing_coverage_calculator.html

## What theoretical coverage do I need?

Genome assembly: 60-100 X

Re-sequencing: 20 X

RAD-seq: 20 X

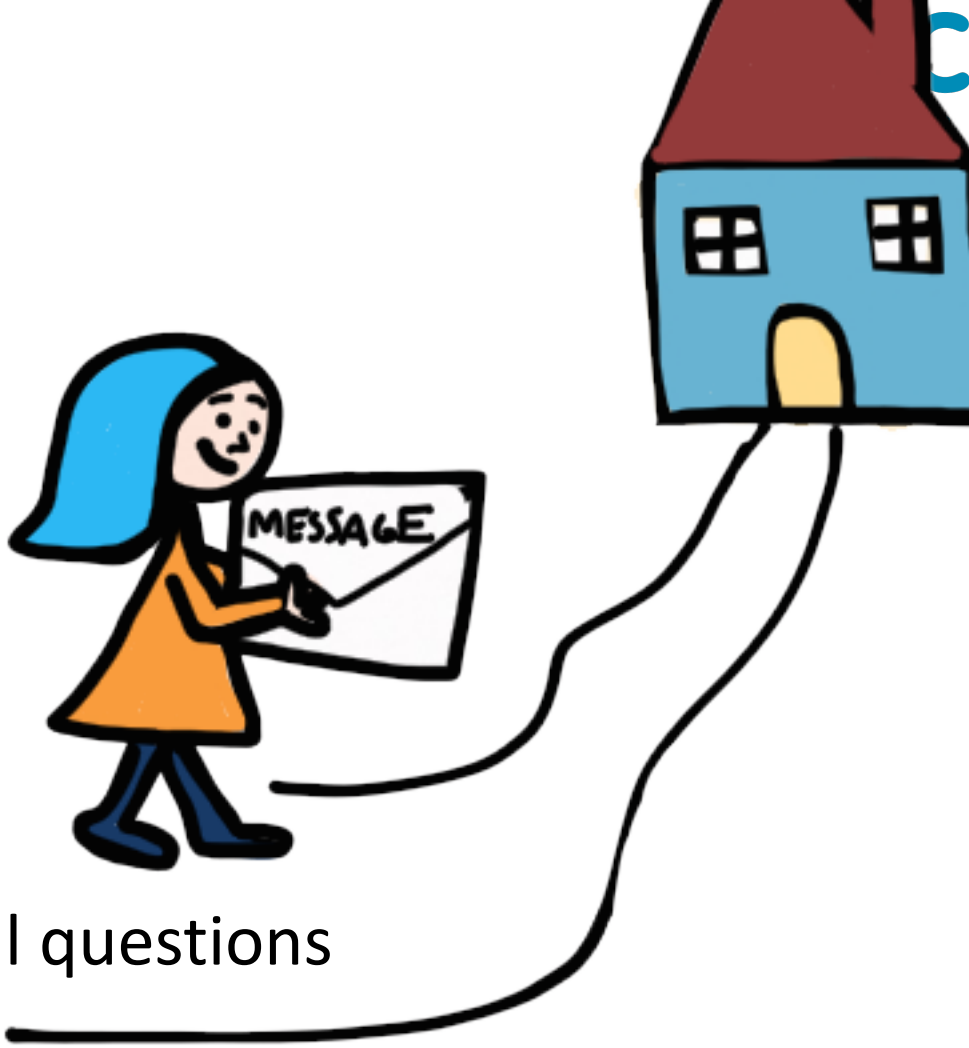
Pool-seq: 100 X

RNA-seq: ~15-20 Mio reads for Eukaryotes

Amplicon:  >10 per OUT

Low coverage re-sequencing: 0.5-2X

# Take home massage




- Formulate your research questions
- There is no ideal application/method to answer all questions