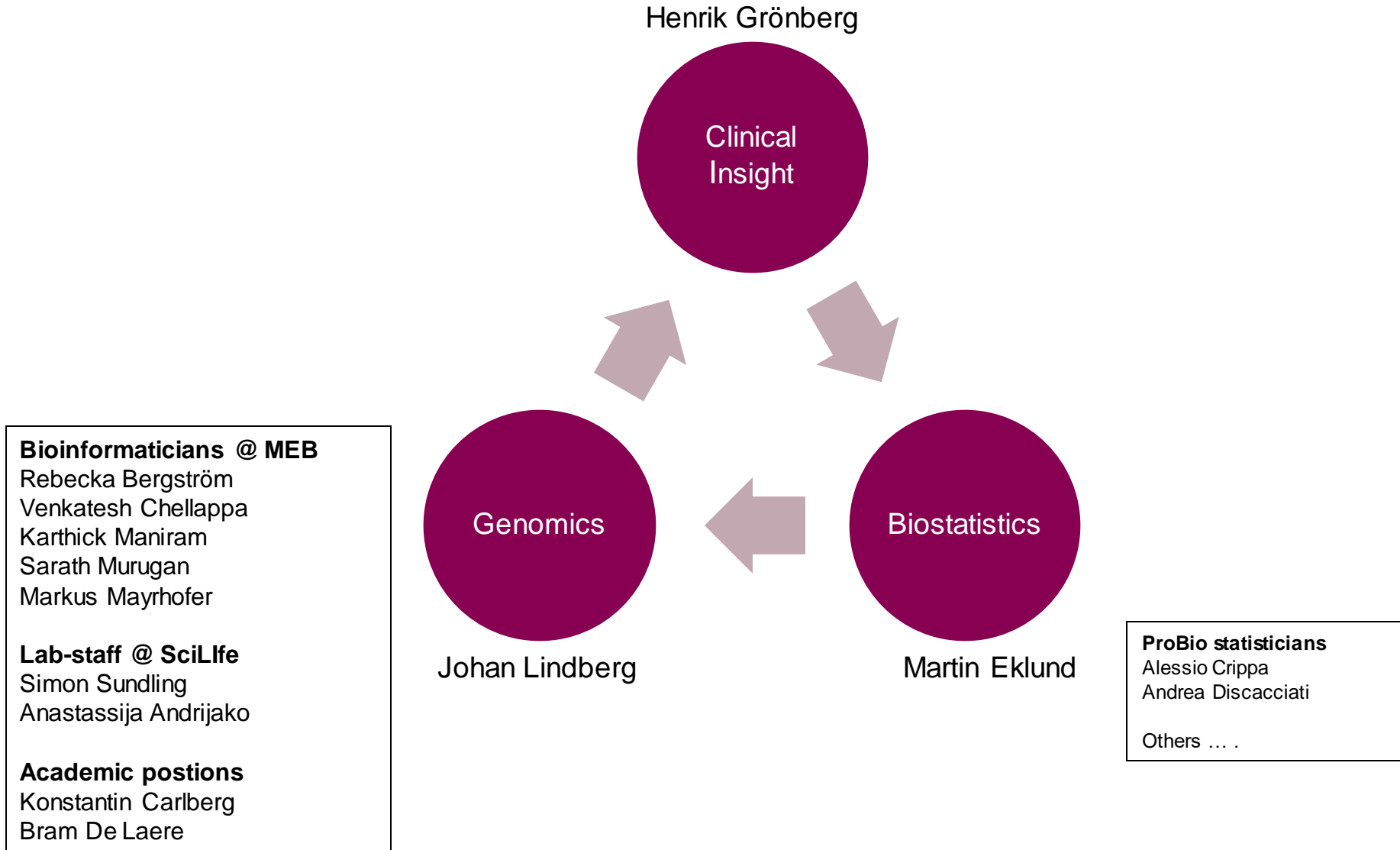


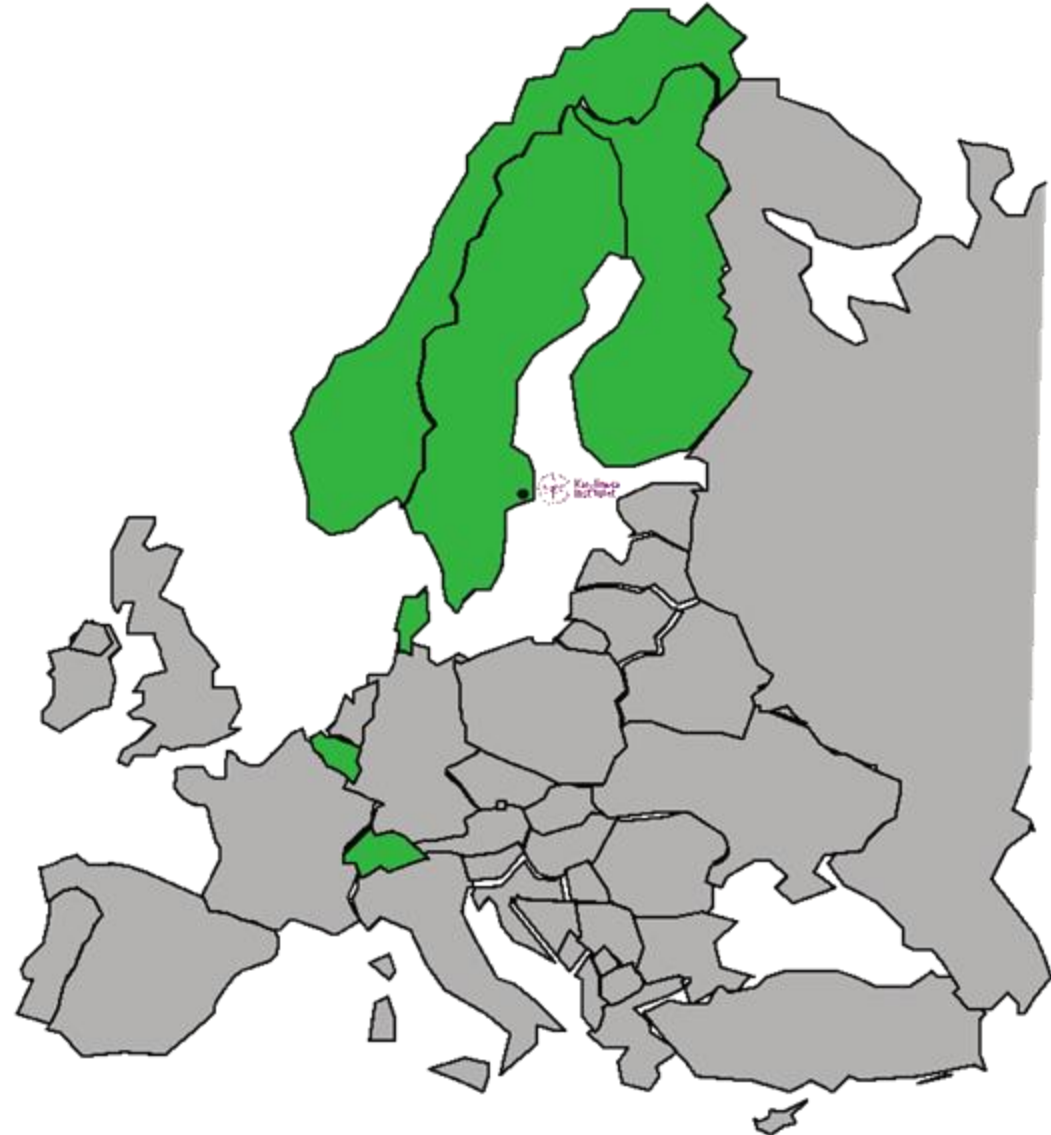
Welcome to Clinical Cancer Genomics, vt 2023

Who are we?



Prospective clinical trials using genomics

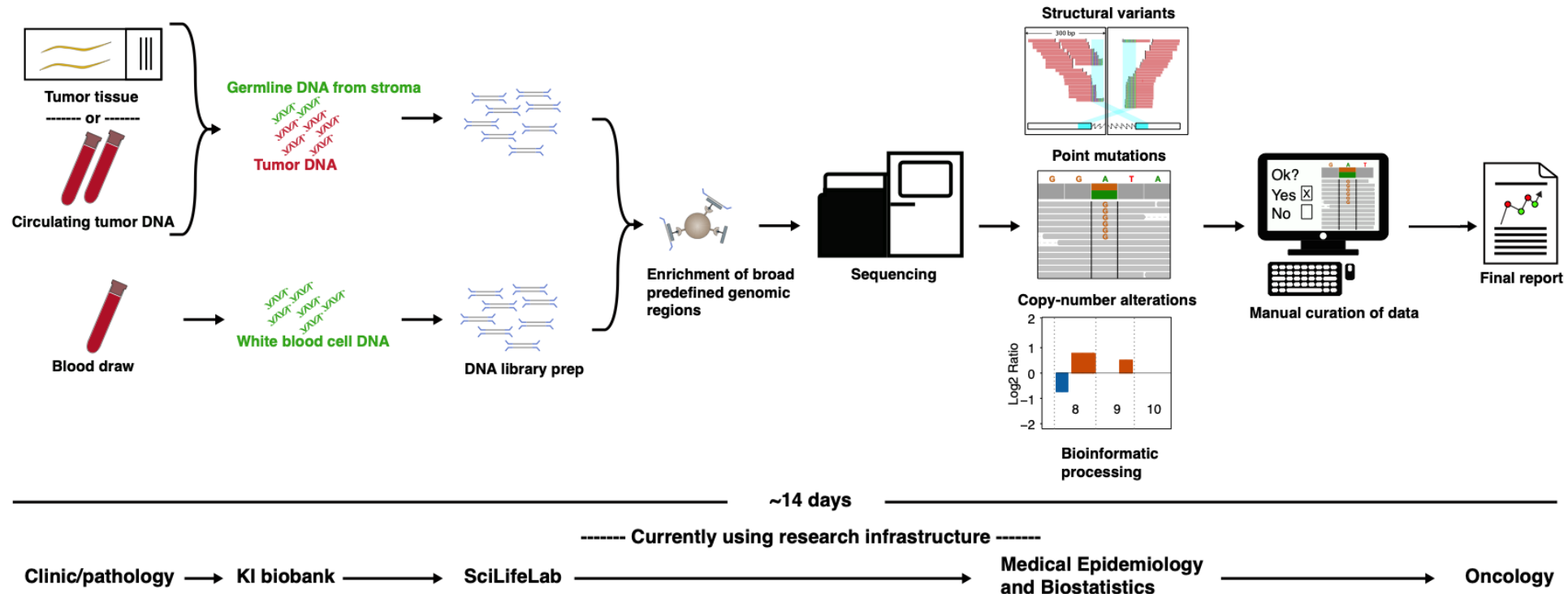
- Established infrastructure to enable the trials
- ~50 sites in 6 European countries ship biomaterial to KI



Genomics infrastructure

Experience

- > 4500 cases profiled using a validated research infrastructure/process
 - Research projects (DNA and RNA analysis)
 - Prospective clinical trials (DNA analysis)
 - Work towards accreditation and clinical implementation

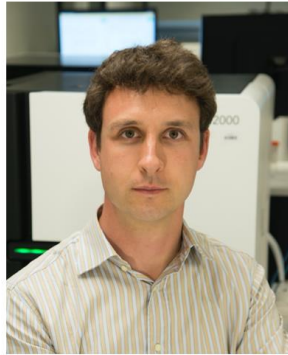


Modified version of an existing course

- <https://pmbio.org/course/>
- Course content under MIT license and Creative Commons.



Malachi Griffith, PhD
Assistant Professor of Medicine
Assistant Professor of Genetics
Assistant Director, MGI

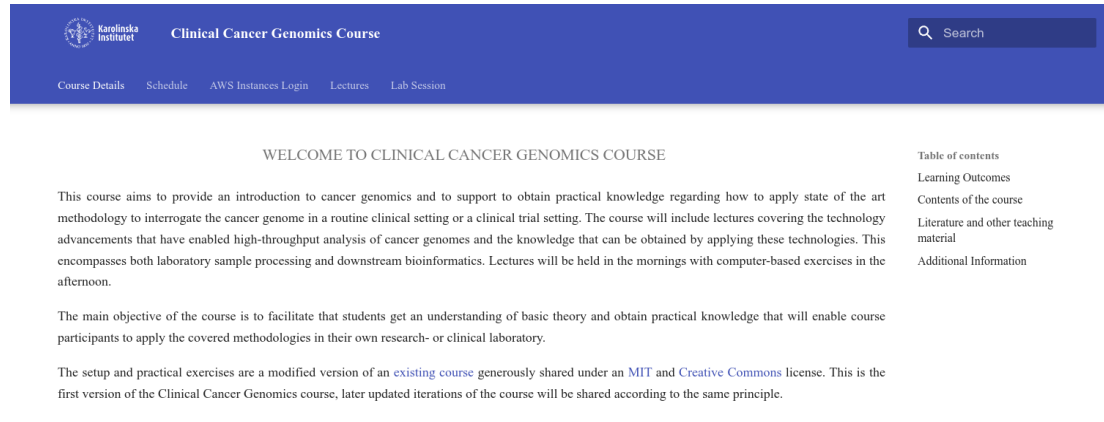


Obi Griffith, PhD
Assistant Professor of Medicine
Assistant Professor of Genetics
Assistant Director, MGI

McDonnell Genome Institute, Washington University School of Medicine

Online set-up

Course webpage, info, lectures and practical sessions.



Lectures, labwork supervision and breakout sessions via Zoom.



AWS server



Student

Hands-on exercises



Lab Session

Day 1 >

Day 2 v

[Human Genome](#)

Annotation

Processing of DNA sequencing data

Introduction to IGV

Day 3 >

Day 4 >

Day 5 >

HUMAN GENOME REFERENCE FILES

The human genome is work in progress. Different versions exist (assemblies) and is often a source of confusion in genomics. Quick use of published data is a bit more challenging if a different genome build was applied than what is used in your lab.

Also, there may even be differences in the *latest* versions, e.g. the GRCh37 and the hg19 assemblies from NCBI and USCS, respectively, had different mitochondrial genome.

In genomics, the reference genome offers a scaffold upon which new data can be mapped, which is a much more efficient way rather than building a genome from scratch.

A very nice overview of the human genome versions can be found [here](#).

During this course version GRCh38, from the Genome Reference Consortium, with modifications from the 1000 genomes consortium will be used. It includes extra decoy and HLA sequences. The files are already available on AWS but can be downloaded from [here](#).

As several of the analysis steps would take many hours we will use a smaller version of the reference genome which contains chr6 and chr17. These were chosen because the data that will be analysed contains interesting variants on these chromosomes. - Connect to your AWS instance

```
# Remember, source the settings, paths and variables file
source .bashrc
```

```
# Check the CHRS variable.
echo $CHRS
```

Table of contents

[Split the long fasta by chromosome](#)

[Index the fasta files](#)

Schedule

	Mon	Tues	Wed	Thur	Fri
AM	Lecture 1	Lectures about the cancer genome and how to interrogate it.	Bioinformatics pipelines and high-throughput computing environments	Copy number alterations	How to curate somatic- and germline data in a clinical trial setting
	Lecture 2	Guest lecture, Clinical Genomics	QC and somatic and germline (small) mutation variant callers.	Structural variation	Clinical trials
	Lecture 3	Lab introduction	Lab introduction	RNA-sequencing and lab introduction	Annotating, interpreting and reporting somatic- and germline variation.
PM	Labwork	Tutorials, basic tools for cancer genomics	Files, tools and running a basic bioinformatic pipeline.	Analysis of copy number data, structural rearrangements and RNAseq data.	Investigating databases and working with data interpretation. Finishing off any remaining labwork from previous days

- Examination – show for a course instructor each day that you have completed and understood the labwork.
 - During the exercises you get questions ...
- This is our first time – have patience!
 - Any ideas/things you want to discuss for your own projects – let us know during the practical sessions, we can have a project brainstorm.

Questions?