What / who is the Broad Institute?

- Spinoff of Harvard & MIT -- Eric Lander and philanthropists Eli & Edyth Broad
- Use the full power of genomics to transform the understanding and treatment of disease
Where in the world is the Broad?
Where in the world is the Broad?
Where in the world is the Broad?
A new organization bringing together software engineers, computational biologists, and computing infrastructure specialists. A vision that articulates an advanced computing infrastructure, set of data and analysis services leveraging modern cloud computing paradigms.

https://www.broadinstitute.org/dsde/
GATK = Genome Analysis Toolkit

- **Toolkit** focused on variant discovery (SNP & indel)

- **Components:**
  - Engine and infrastructure
  - Tools (walkers)

- Also a *programming framework* for developing genome analysis software
Variant discovery = identify **variants** in sequencing data
GATK Best Practices = complete reads-to-variants workflows

Data Pre-Processing → Variant Discovery → Callset Refinement

FASTQ → BAM → VCF

SAM/BAM → VCF

FASTQ
Expanding ecosystem of modular Best Practices workflows
Alpha GATK 4: cloud-friendly and more scalable (Apache Spark) + extended functionality (CNVs, Picard)

https://github.com/broadinstitute/gatk
Workshop agenda

Day 1

9 am – 10:30 am
Introduction to Variant Discovery

10:30 am – noon
Pre-processing methods

Day 2

9 am – noon
Germline variant discovery methods

Day 3

9 am – noon
Somatic variant discovery methods

Afternoons

2 pm – 5 pm
Hands-on tutorials