BROAD INSTITUTE

iSeqTools
http://iseqtools.org/

GATK Workshop

University of Pretoria
South Africa, 22-26 June, 2015

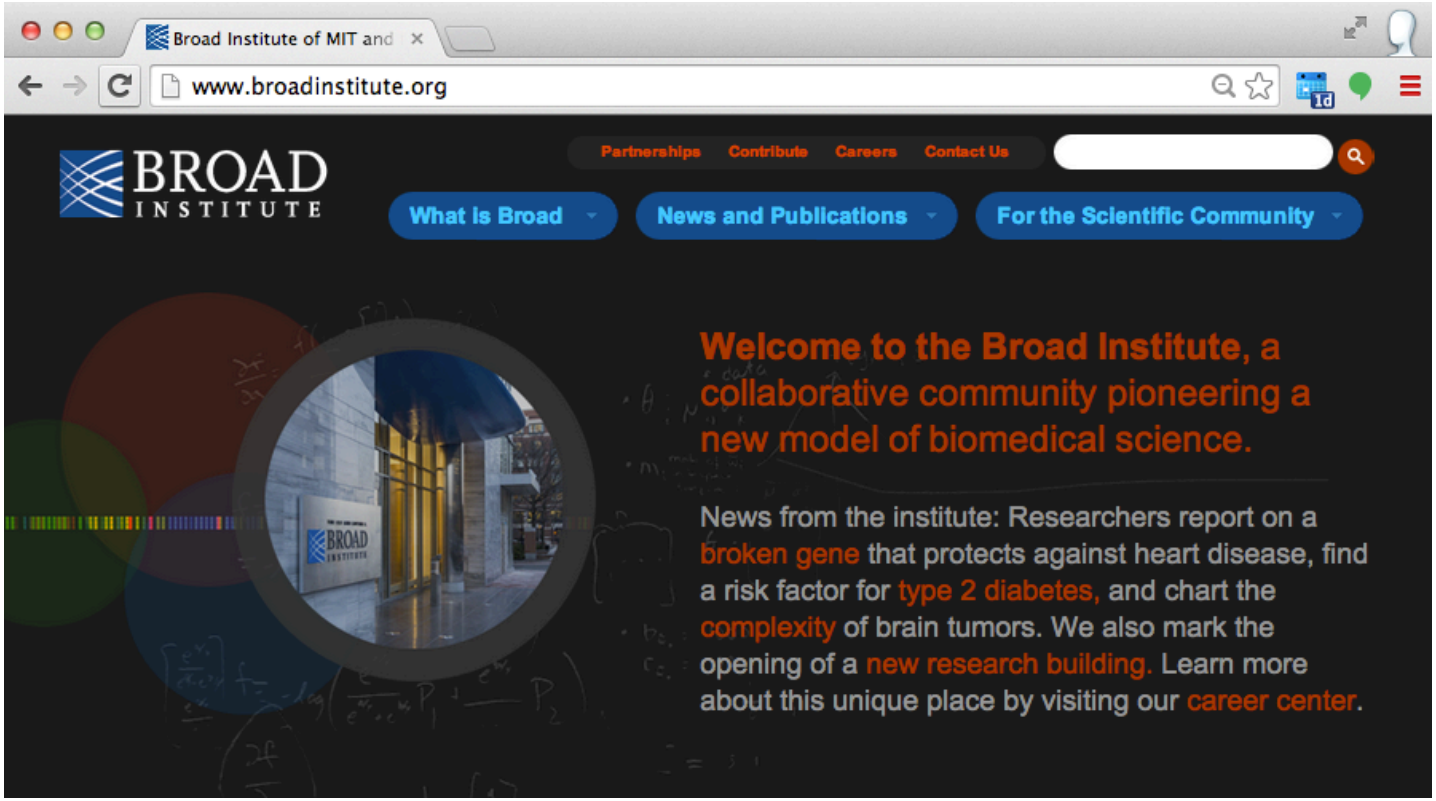
Data Sciences & Data Engineering
Broad Institute of Harvard and MIT
http://www.broadinstitute.org/gatk/

@gatk_dev

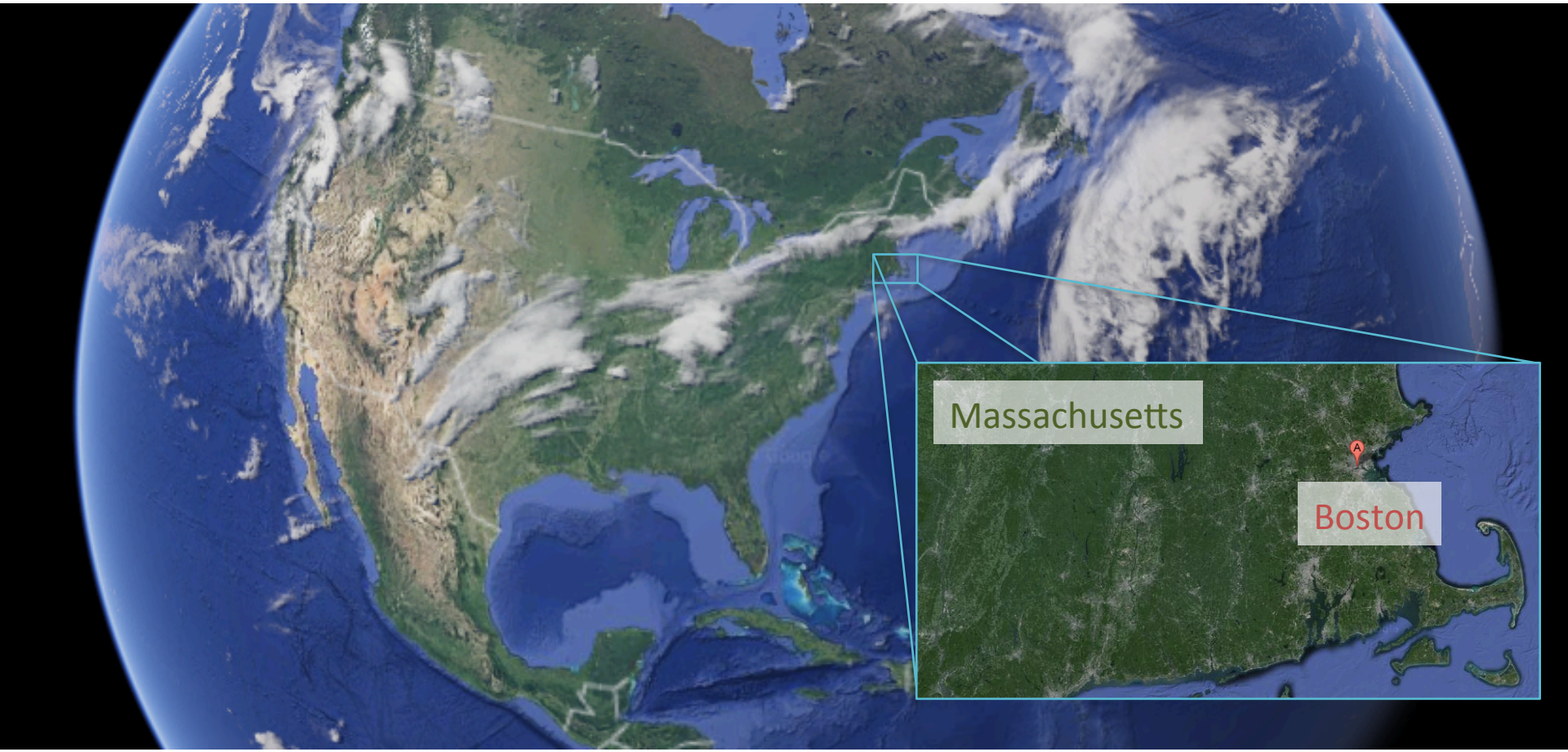
BROAD INSTITUTE

What / who is the Broad Institute?

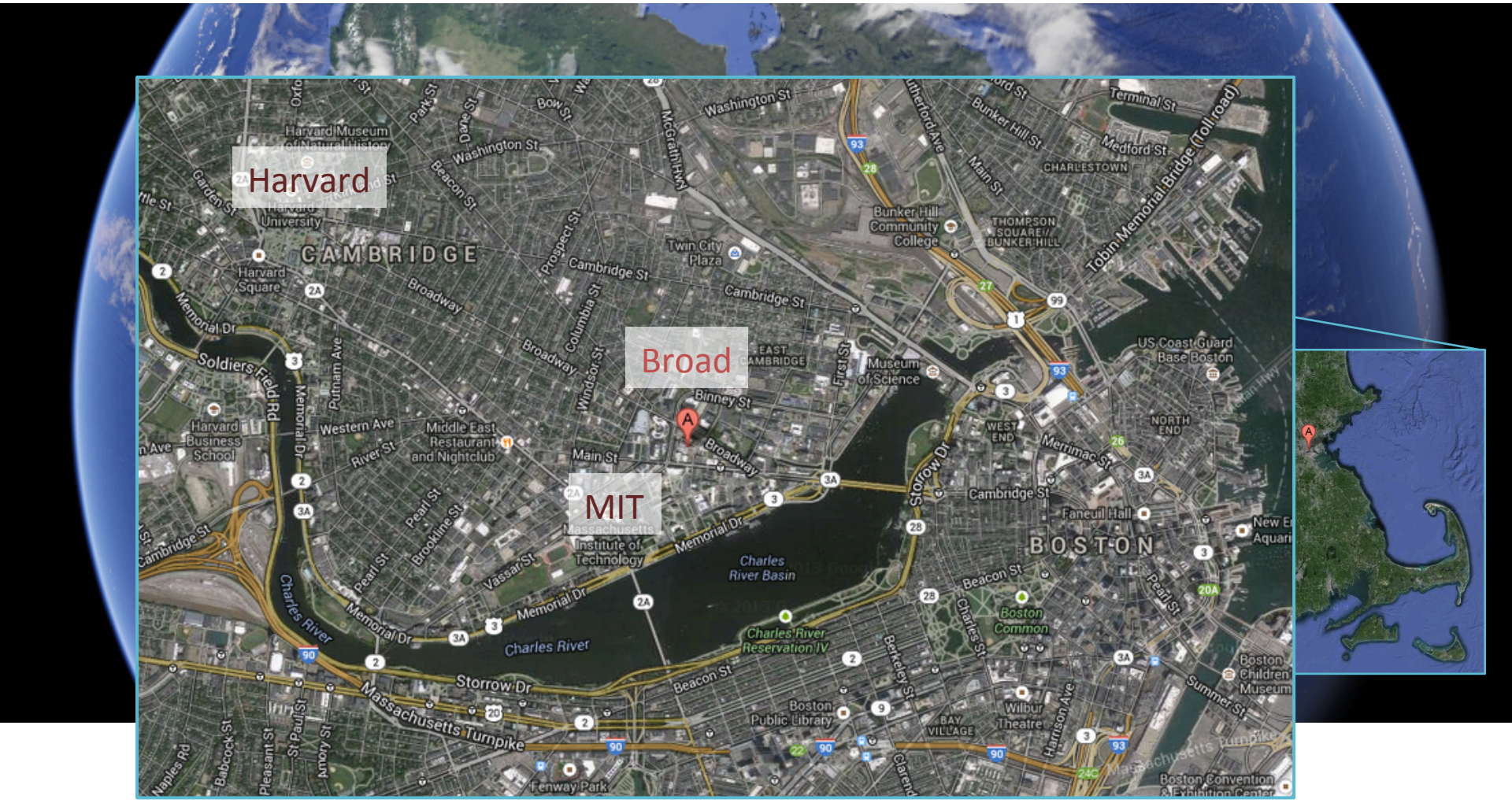
- Spinoff of Harvard & MIT -- Eric Lander and philanthropists Eli & Edyth Broad
- Use the full power of genomics to transform the understanding and treatment of disease



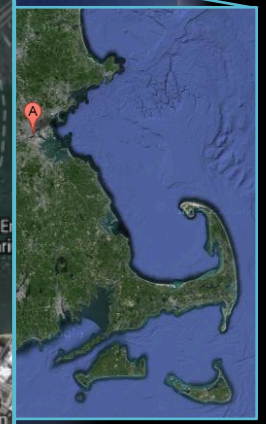
Where in the world is the Broad?



Where in the world is the Broad?

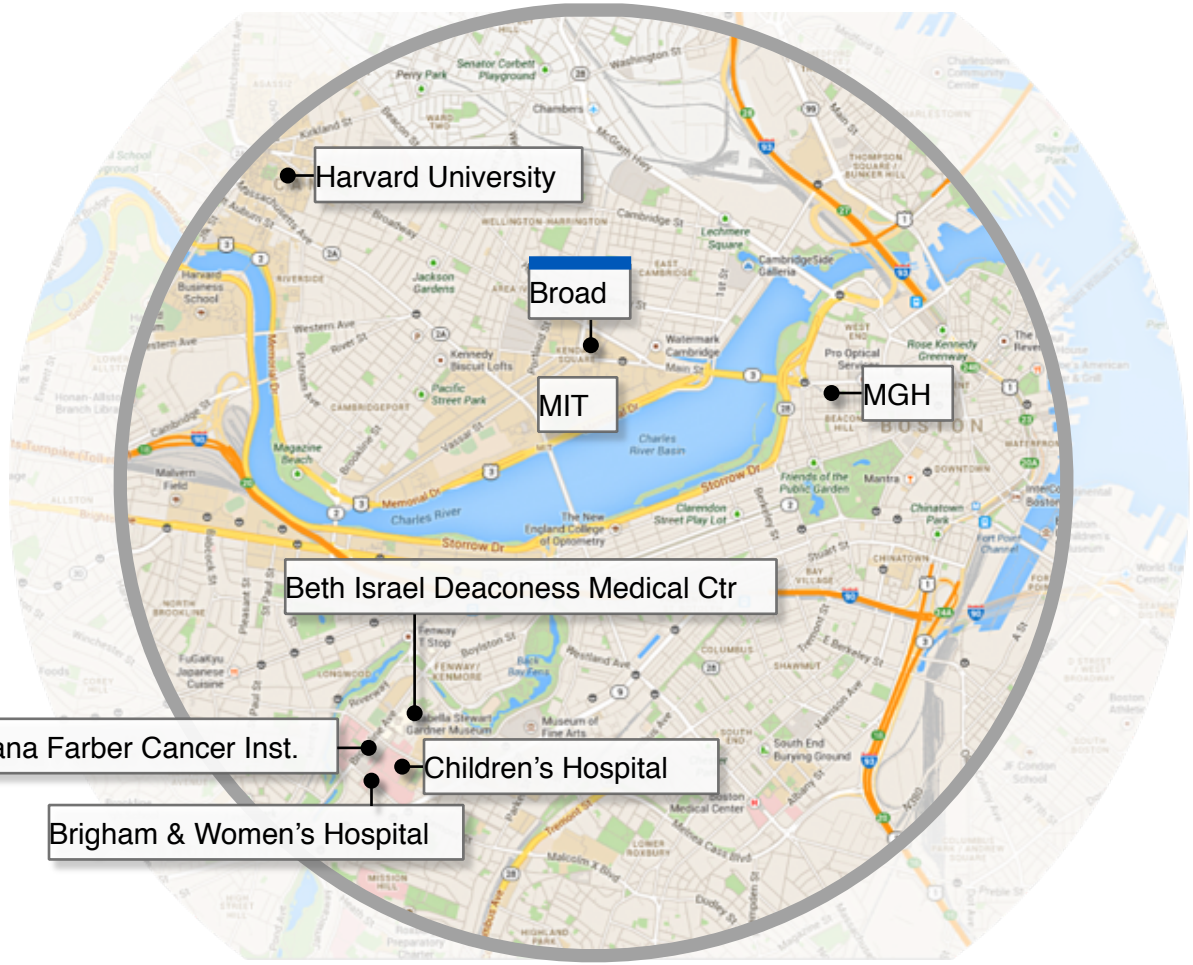


Where in the world is the Broad?

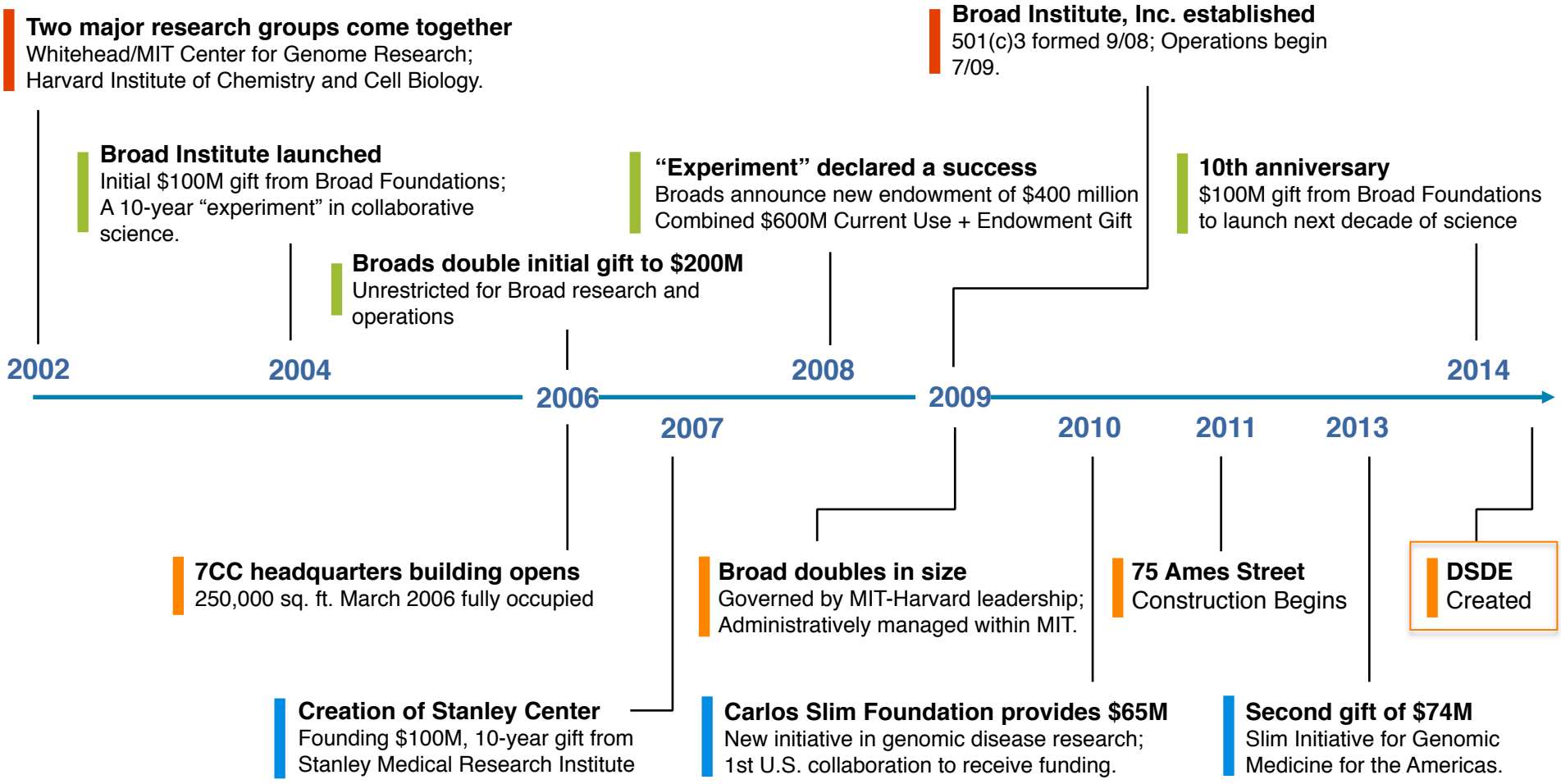


Where in the world is the Broad?

A highly collaborative and research-dense environment



A brief history of the Broad Institute



Data Science & Data Engineering @ 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/

DSDE's extended family: GP + CGA + BITS

DSDE: Methods + Engineering + Operations + Communications

Genomics Platform (GP)

Cancer Genome Analysis Group (CGA)

Broad IT (BITS)

+ many collaborators who beta-test the methods and software

General program

Day 1
Introductions, pipelines and QC

Day 2
Data pre-processing (BAM cleanup)

Day 3
Germline variant discovery

Day 4
Somatic variant discovery

Day 5
Working with variants

AM Lecture session: 9am – 12pm

PM Hands-on tutorial: 1pm – 4pm

Day 1: Sequencing for variant analysis

- 9:00 Workshop introduction
- 9:20 High-throughput sequencing for variant analysis
- 10:00 GATK Best Practices and the Broad pipelines
- *10:30 Coffee break*
- 11:00 Sequence data quality control
- 11:30 Pipelining and parallelism
- *12:00 Lunch break*
- 13:00 Hands-on session

Approximate schedule – timing may vary depending on Q&A

Day 2: Mapping and cleanup

- 9:00 Recap of Day 1 and Q&A
- 9:15 Mapping and pre-processing
- 10:00 Indel realignment
- *10:30 Coffee break*
- 11:00 Base recalibration (BQSR)
- *12:00 Lunch break*
- 13:00 Hands-on session

Approximate schedule – timing may vary depending on Q&A

Day 3: Germline variant discovery

- 9:00 Recap of Day 2 and Q&A
- 9:15 Introduction to joint variant discovery
- 9:45 Variant calling and joint genotyping
- *10:30 Coffee break*
- 11:00 Variant recalibration (VQSR)
- 11:45 Manual filtration
- *12:00 Lunch break*
- 13:00 Hands-on session

Approximate schedule – timing may vary depending on Q&A

Day 4: Somatic variant discovery

- 9:00 Recap of Day 3 and Q&A
- 9:15 Introduction to somatic variant discovery
- 10:00 Estimation of contamination with ContEst
- *10:15 Coffee break*
- 10:45 Variant calling with MuTect
- 11:30 Rescuing TiN variants and eliminating artifacts
- *12:00 Lunch break*
- 13:00 Hands-on session

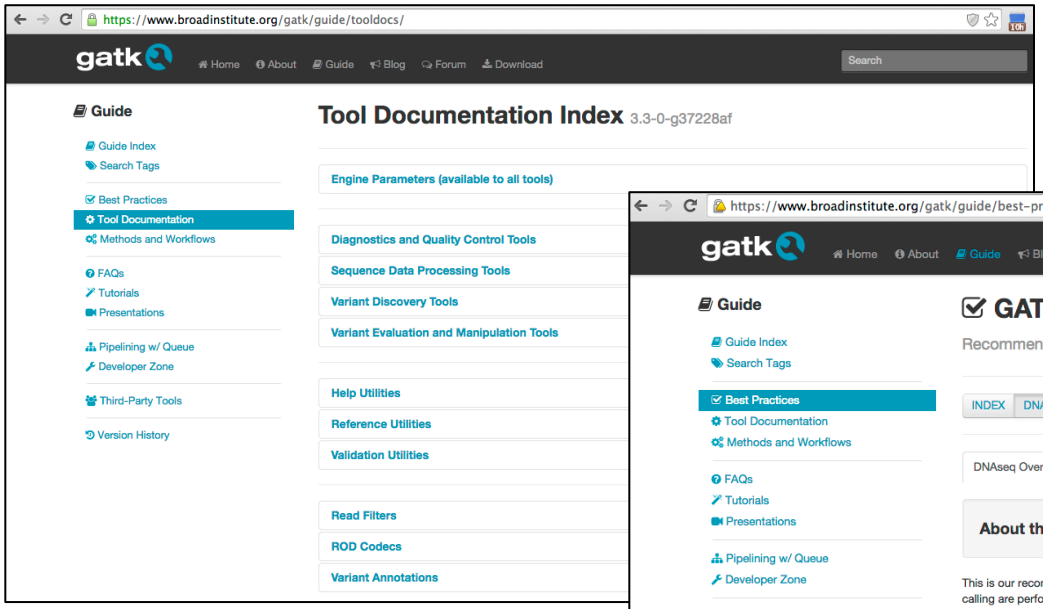
Approximate schedule – timing may vary depending on Q&A

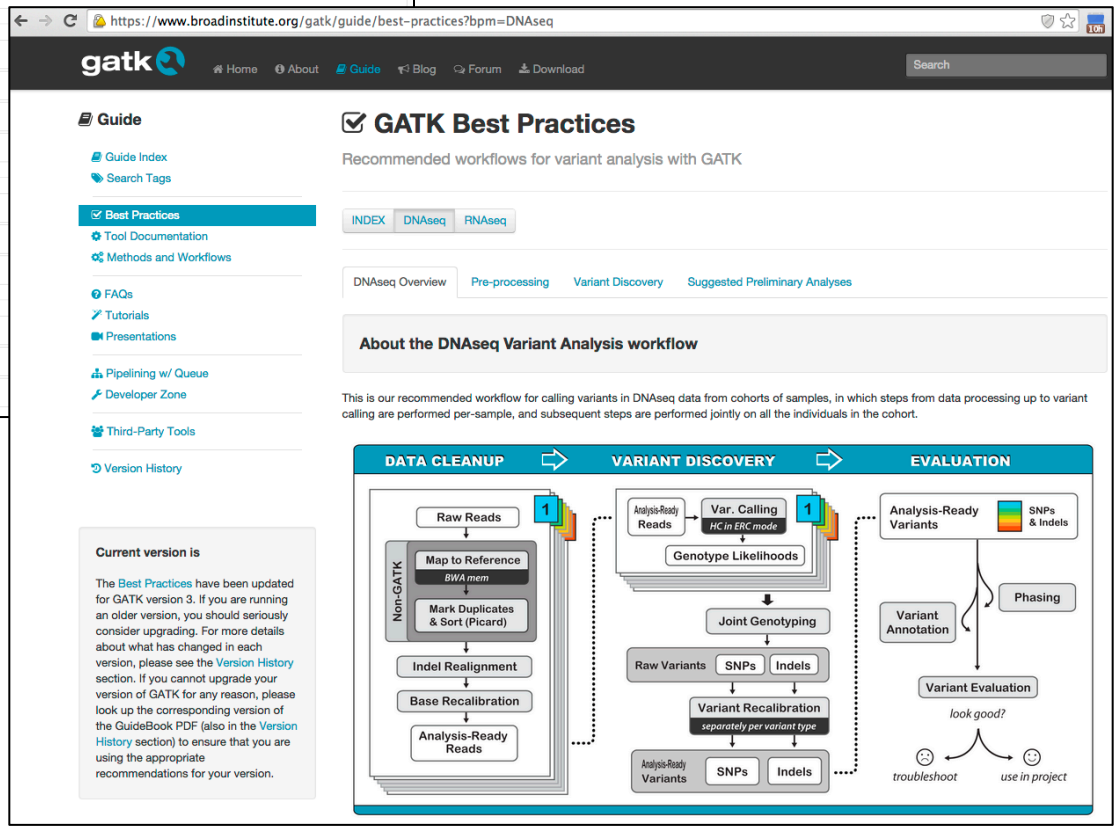
Day 5: Somatic variant discovery

- 9:00 Recap of Day 3 and Q&A
- 9:15 Introduction to working with variants
- 9:45 Genotype refinement
- *10:15 Coffee break*
- 10:45 Evaluating variants
- 11:30 Annotating variants
- *12:00 Lunch break*
- 13:00 Hands-on session

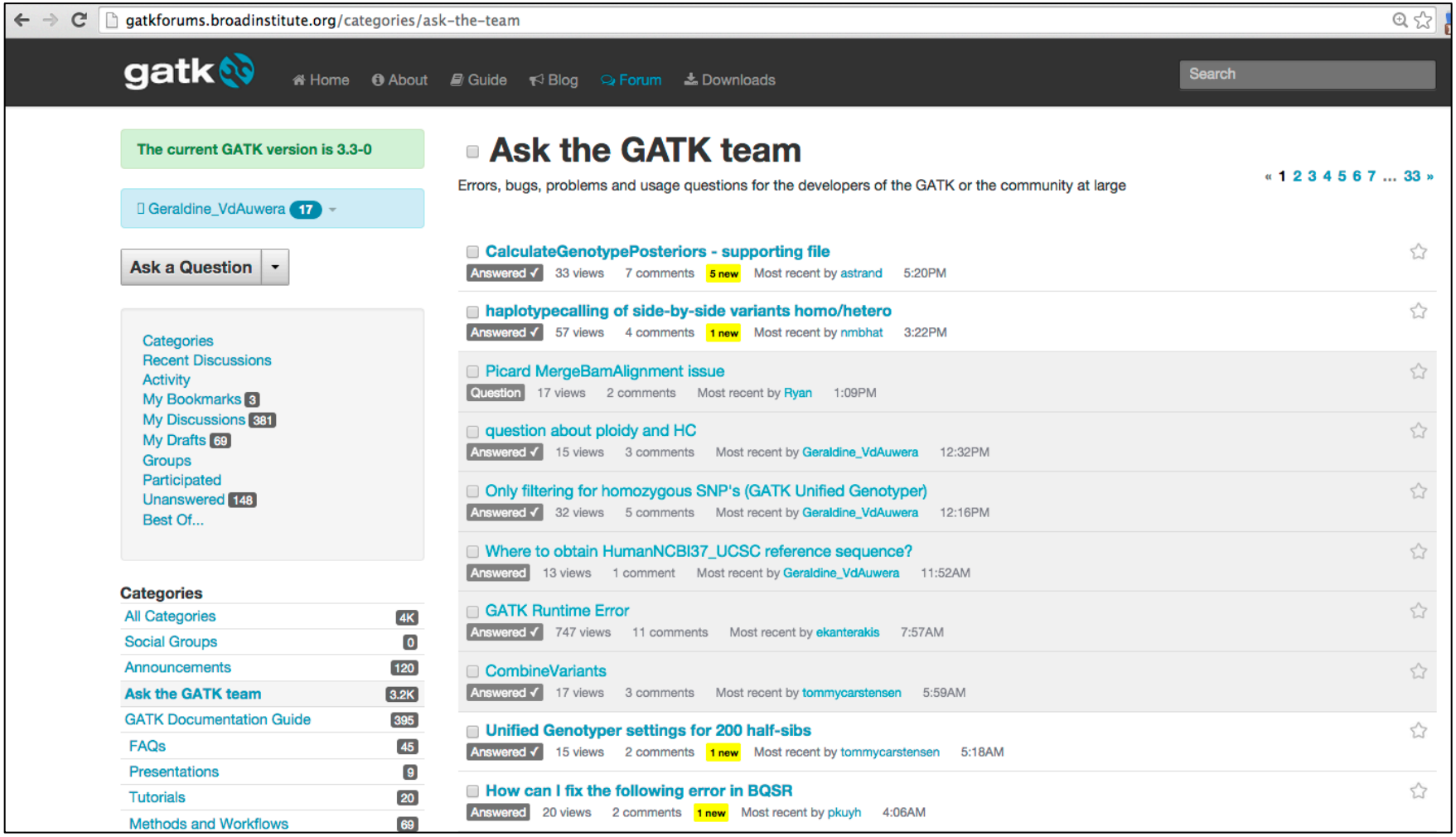
Approximate schedule – timing may vary depending on Q&A

Documentation is at https://www.broadinstitute.org/gatk/guide/





See the support forum for questions and bug reports



General program

Day 1
Introductions, pipelines and QC

Day 2
Data pre-processing (BAM cleanup)

Day 3
Germline variant discovery

Day 4
Somatic variant discovery

Day 5
Working with variants

AM Lecture session: 9am – 12pm

PM Hands-on tutorial: 1pm – 4pm