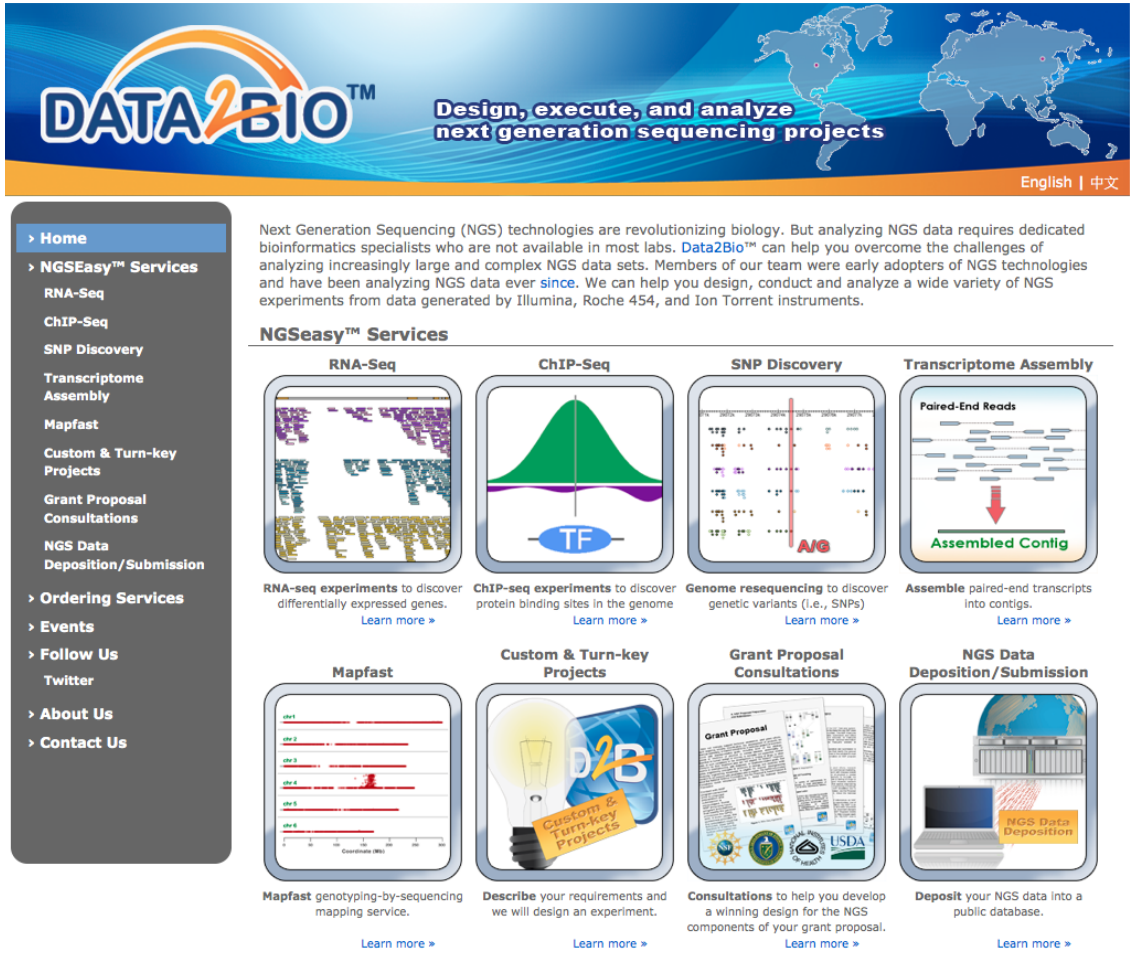
Data2Bio's tGBS Technology

Patrick S. Schnable
Data2Bio, LLC

Data2Bio, LLC



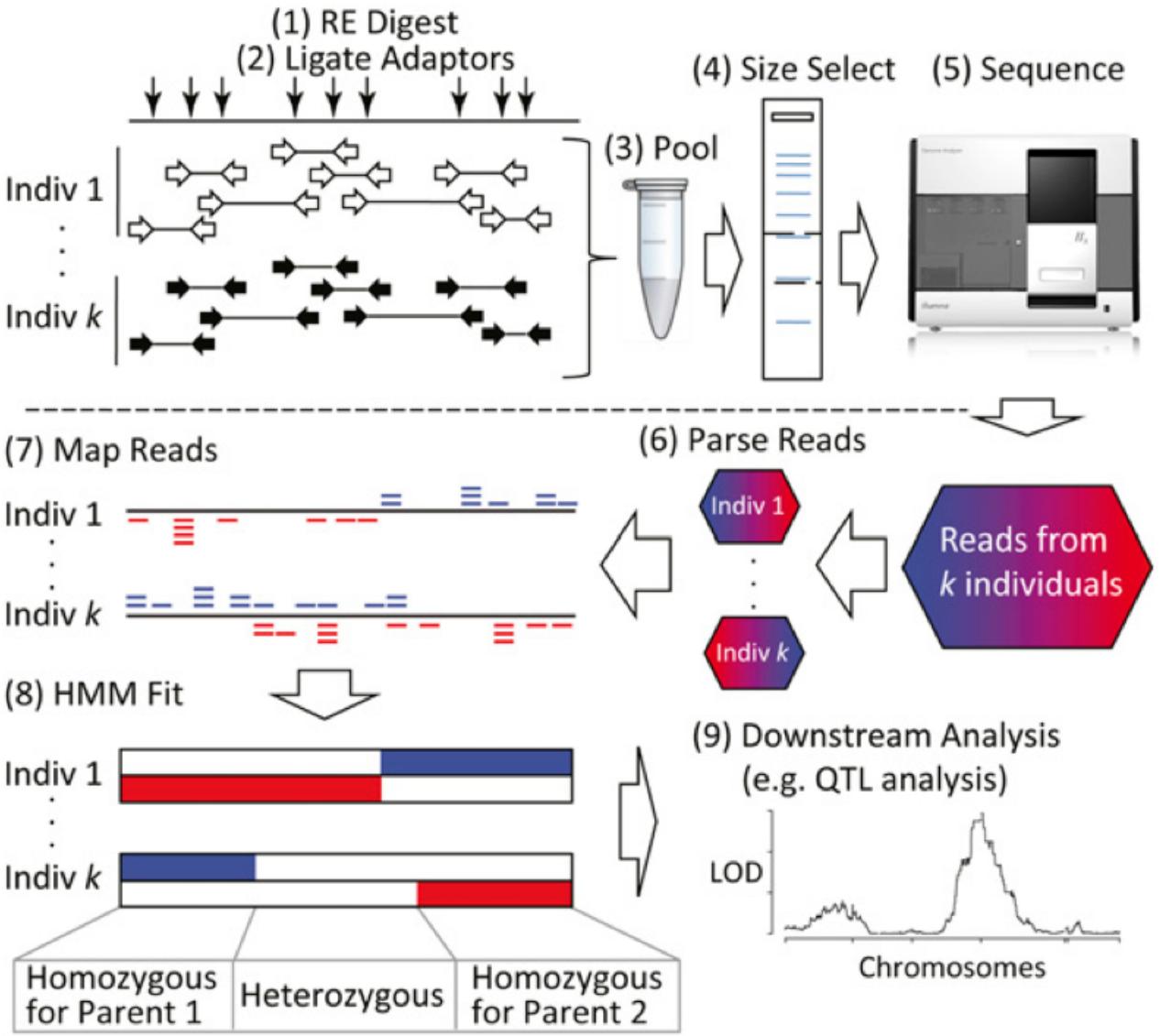
- Founded in 2010, the privately held Data2Bio designs, conducts, analyzes and interprets research projects involving next generation sequencing

- Core strengths are bioinformatics, genomics and breeding

- Offices in the US and China; Academic and private-sector customers on all continents except Africa and Antarctica

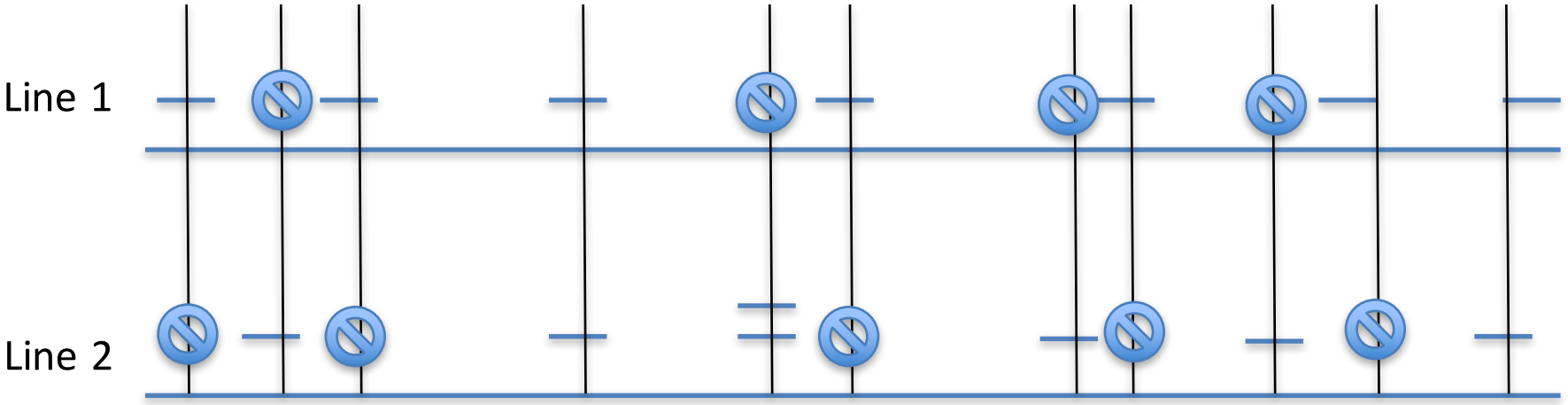
- Proprietary genomic technologies associated with DNA barcoding, genotyping-by-sequencing (tGBS) and genome sequencing (BAC-Seq) as well as proprietary bioinformatic pipelines

An Example of “Conventional” GBS



- Andolfatto et al. 2011

Conventional GBS



Up to 1M sites, i.e., more than is required for most applications

SNP calls based on few reads (increasing error rate)

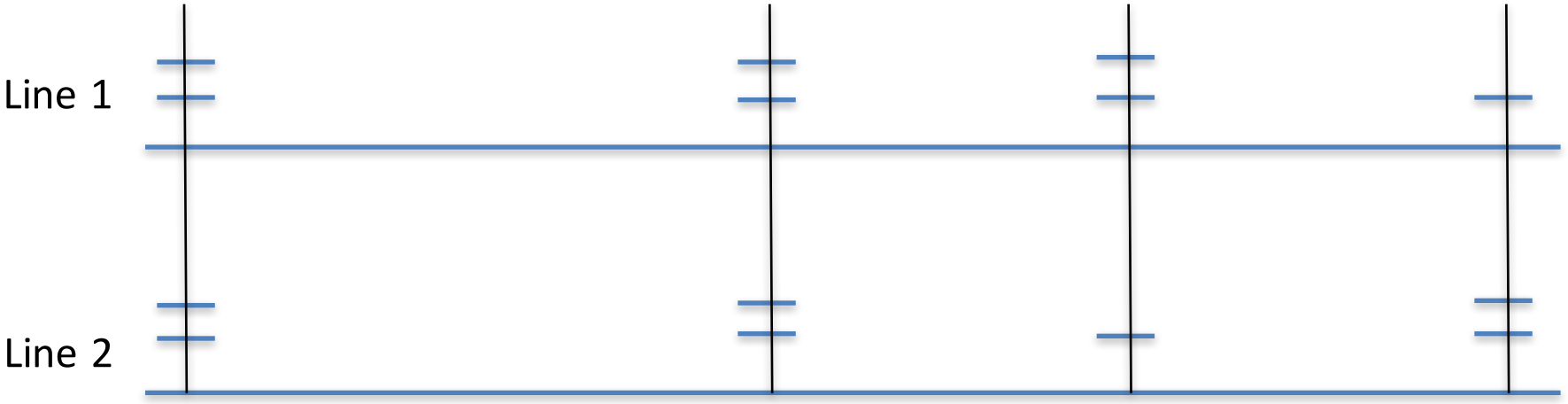
Can not call heterozygotes and rare alleles

Limited overlap of sites among lines -> missing data across lines

Requires *imputation* (heavy-duty bioinformatics)

-> difficult in species without a reference genome

Tunable GBS (tGBS)



tGBS more stringently controls fraction of genome that is sequenced and genotyped ("read selection"). Given same number of reads/sample, reads are clustered at fewer sites, increasing read depth:

- 1,000s to tens of thousands of SNPs ("tunable") -> adequate for most applications
- *Less missing data* across lines; more repeatability
- *High confidence SNP calls-even for rare or minor frequency alleles*
- *Can confidently genotype heterozygous loci*