Haplotype inference with Long-read sequencers

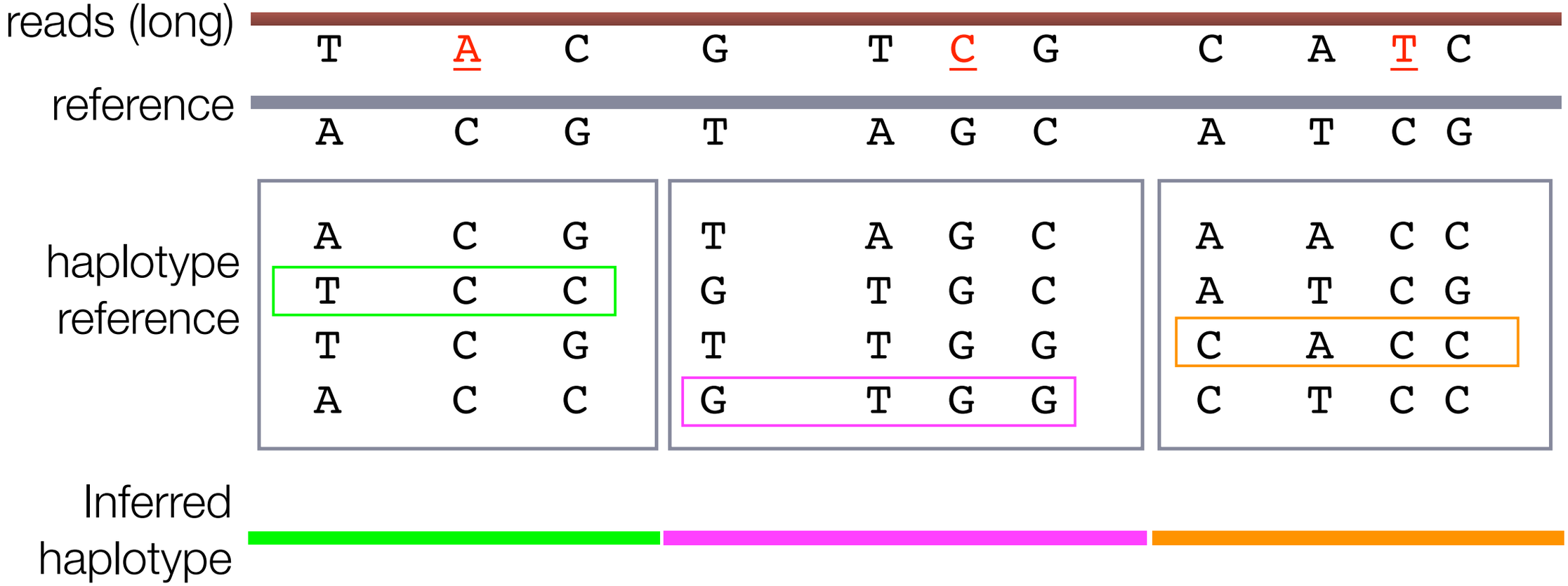
Yosuke Tanigawa (ytanigaw@stanford.edu)
M. Rivas Lab (rotation student) | Biomedical Informatics Ph.D. Program

[Background]
Long read sequencer & Haplotype reference panel

- Long read sequencers (Oxford Nanopore, PacBio)
  - Read length >= 8 kb
- Haplotype reference
  - UK BioBank (152,729 imputed haplotypes)
  - Haplotype reference consortium (not available yet?)
- Compressed data representation of haplotypes
  - PLINK2

[Rotation project]
Compressed representation of haplotypes

- How can we compress haplotype information ?



- Is it possible to infer haplotype on the fly ?
  - Bayesian approach with haplotype frequencies

[Data]
cDNA data set (by Helio)

- Technology: Oxford Nanopore MinION sequencer
  R9.3 chemistry + minKnow v1.1.14
  (current version: R9.4 & v1.1.17)

- "WGS" data set
  20161008_wgs_caucasian_48hr
  - 48hr(?) run
  - 44 M nt (in total)
  - N50: 2227
  - # reads > 20kb: 102
  - Error rates: 10-20 %

Read Length Distribution