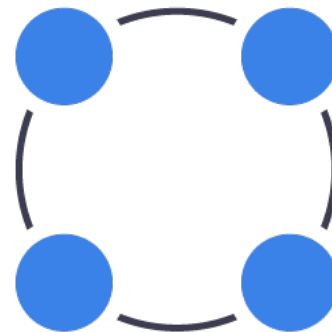
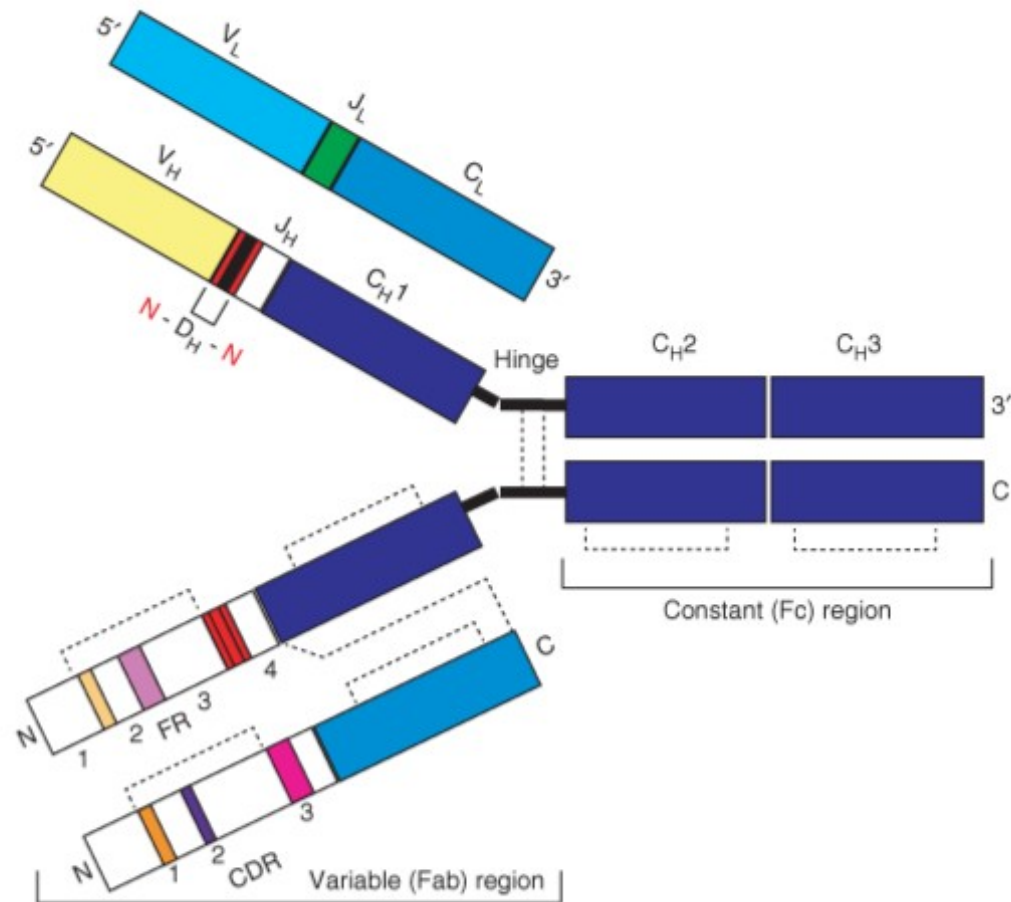


Finding novel variations of germline Immunoglobulin genes using WGS data

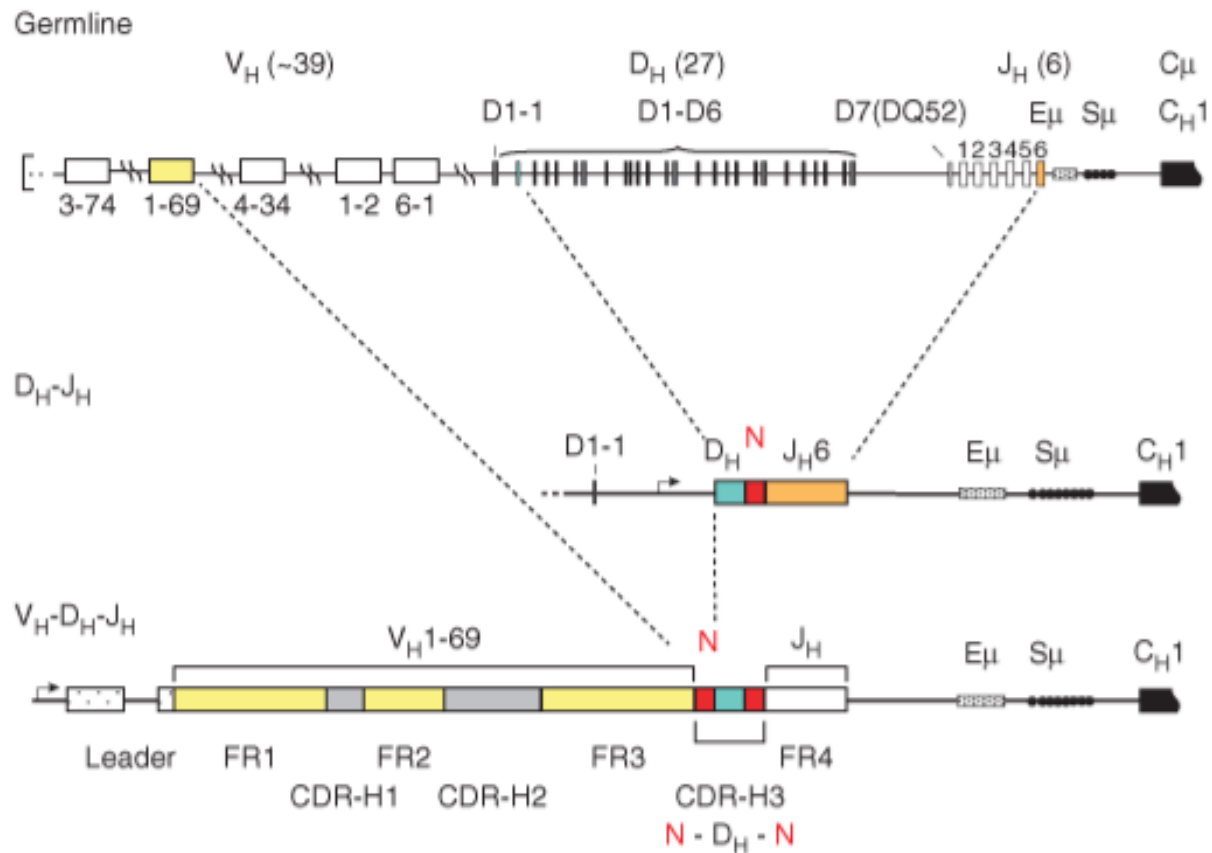
Alexander Ilin
Supervisor - Yana Safonova
UCSD



Antibodies



Origin of antibodies



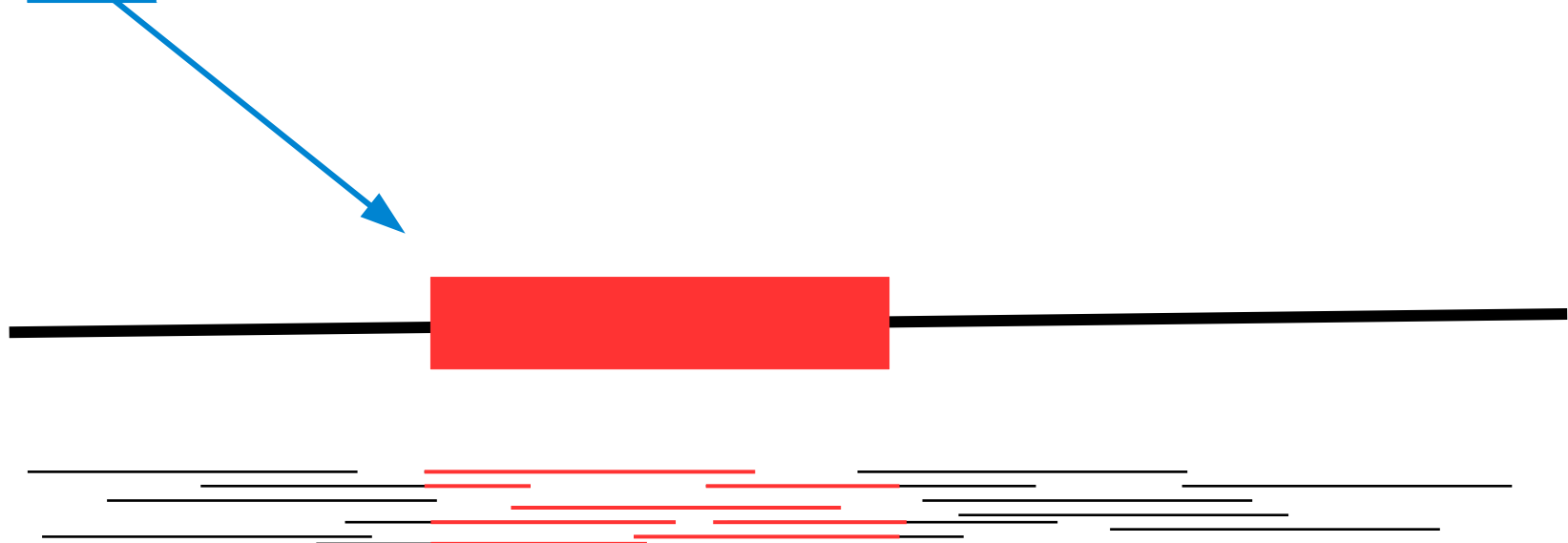
Tasks

Find new variations of Immunoglobulin genes using Whole Genome Sequencing data

- Create tool to select appropriate fragments
- Validate it via simulation
- Analyze WGS data from different populations

Ig locus and WGS

Germline configuration



Different scoring for alignments



Needleman-Wunsch

биоинформатика

б--и--о-м-----

--ти-- --м----а



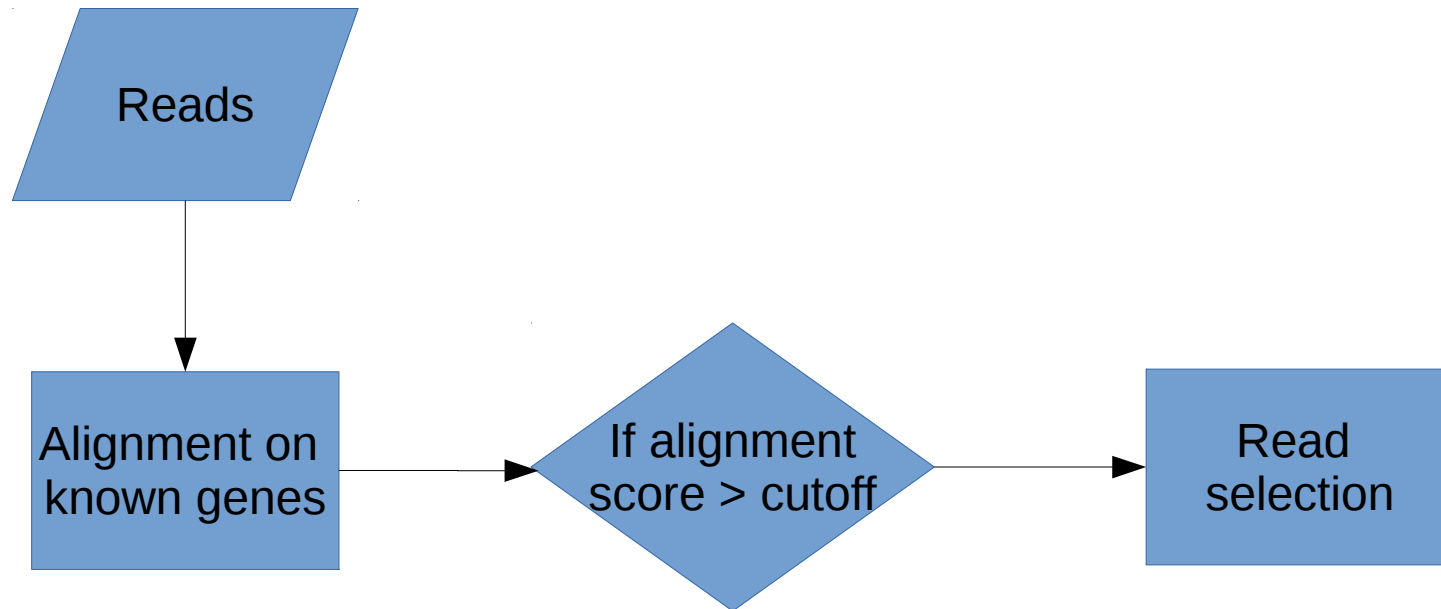
Smith-Waterman

биоинформатика

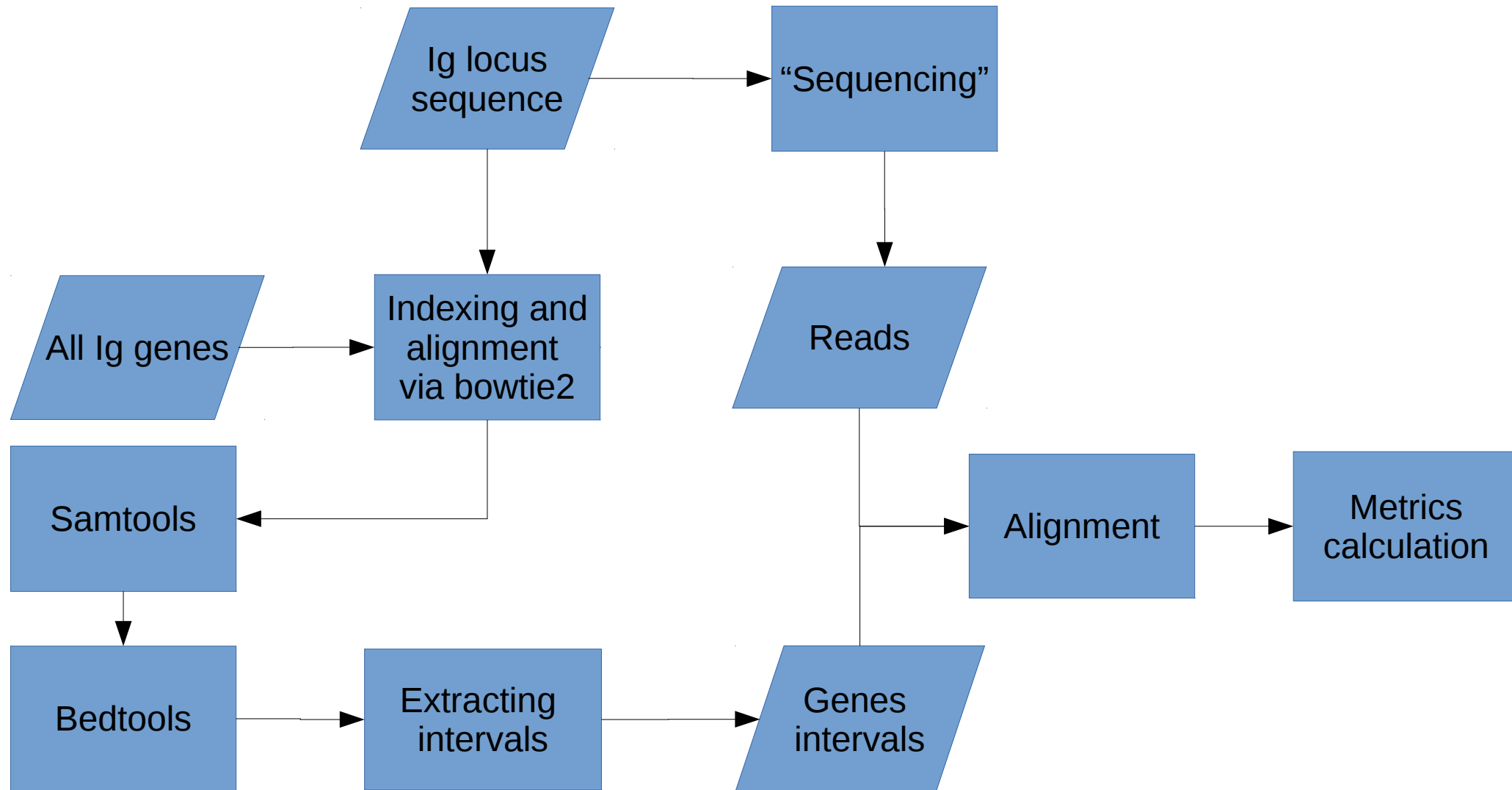
бион-----

-----тима

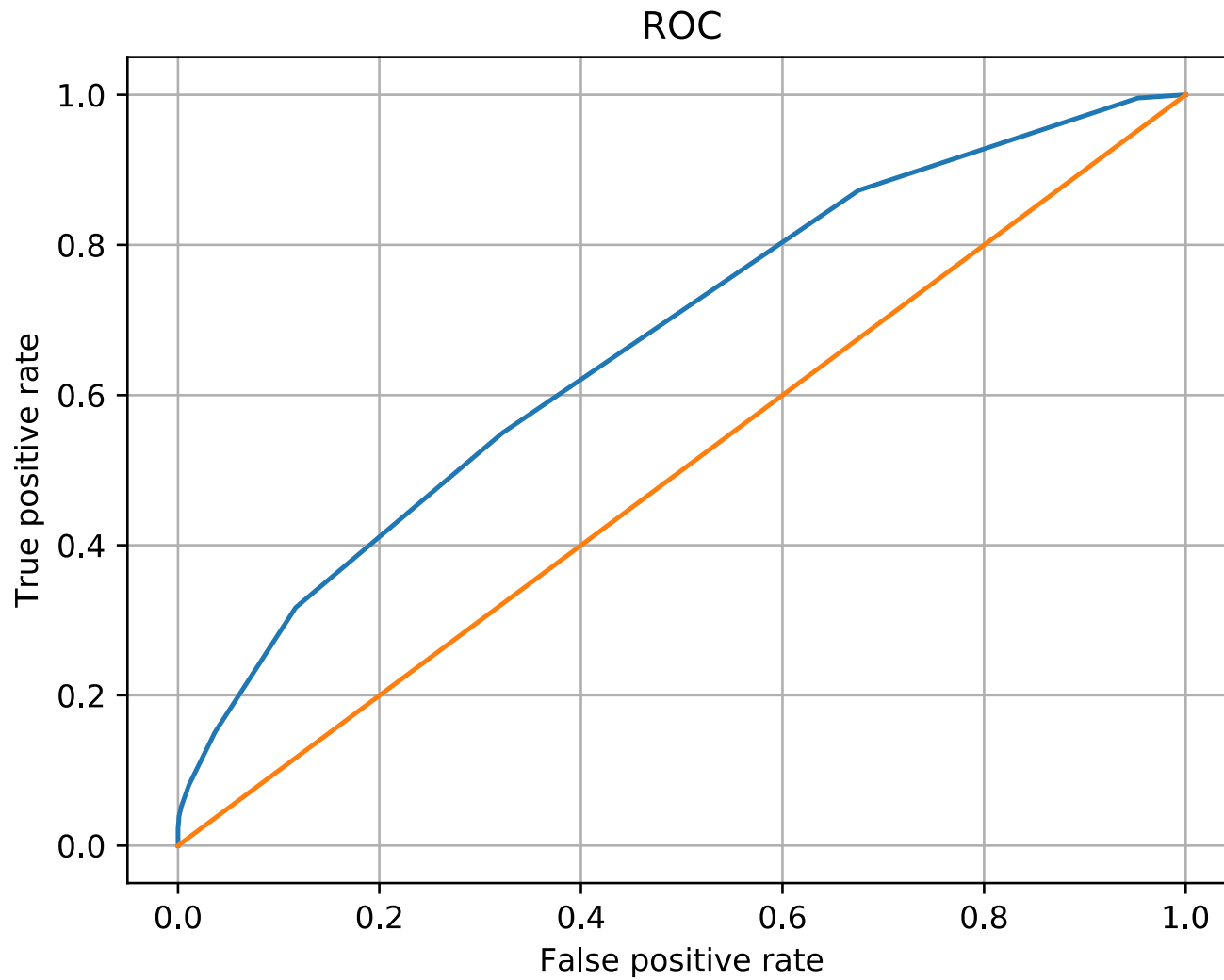
Pipeline



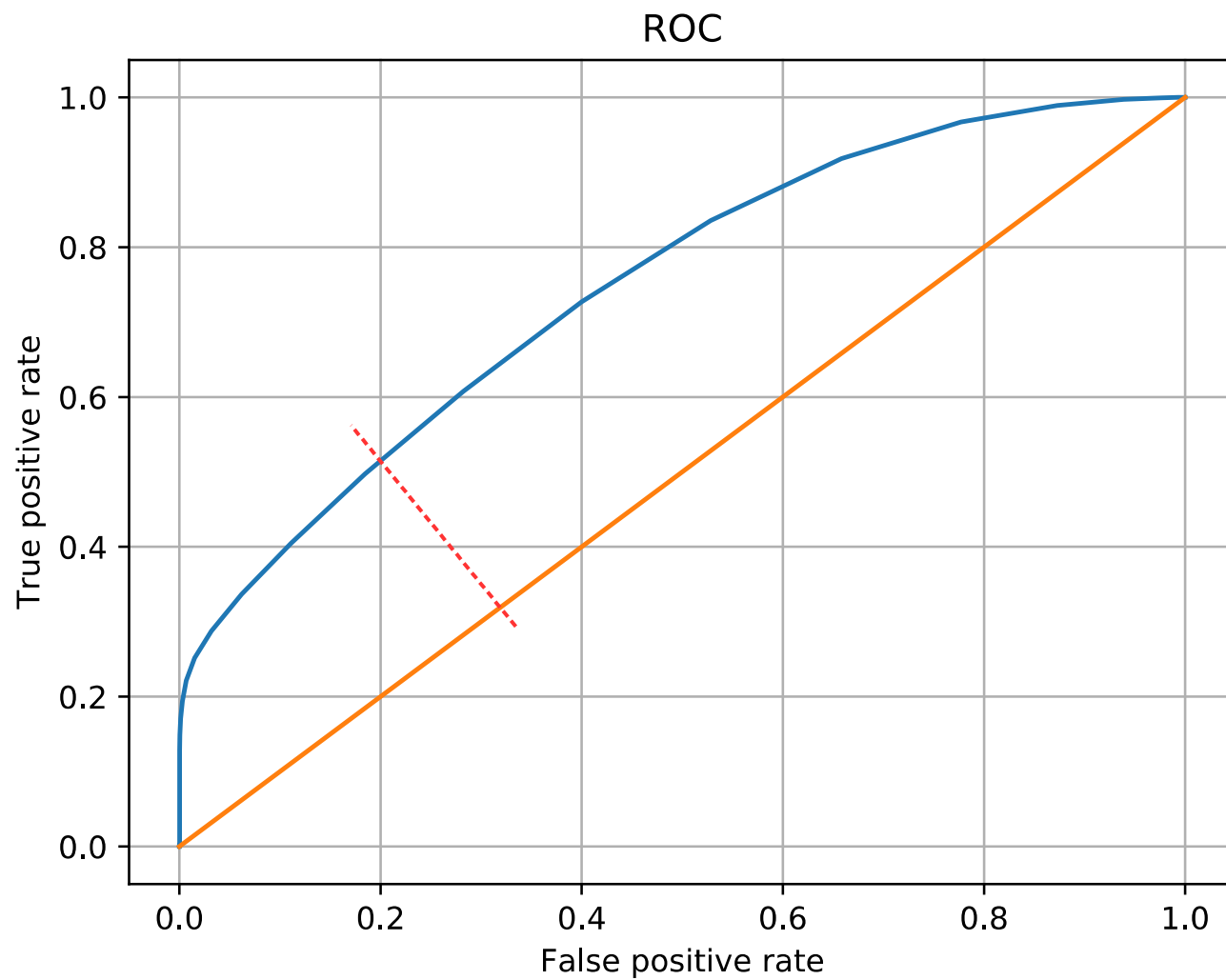
Verification via simulation



First setting



Another try



Chinese IG diversity analysis

WGS of Han Chinese populations

Accession: PRJEB11005 ID: 298320

High-coverage whole genome sequencing of Han Chinese populations (1000 Genomes Project)

The purpose of the project is to support the discovery and understanding of genetic variants that influence human disease, with particular emphasis on the Chinese population. Here we provide whole genome sequence data on 91 individuals of Chinese origin.

Accession	PRJEB11005
Scope	Monoisolate
Submission	Registration date: 9-Oct-2015 BGI
Locus Tag Prefix	BN2788

Plans

- Analyze WGS data from different populations
- Interpret results
- Optimize library

Thank you for your attention

https://github.com/Serfentum/bioinf_vsegments