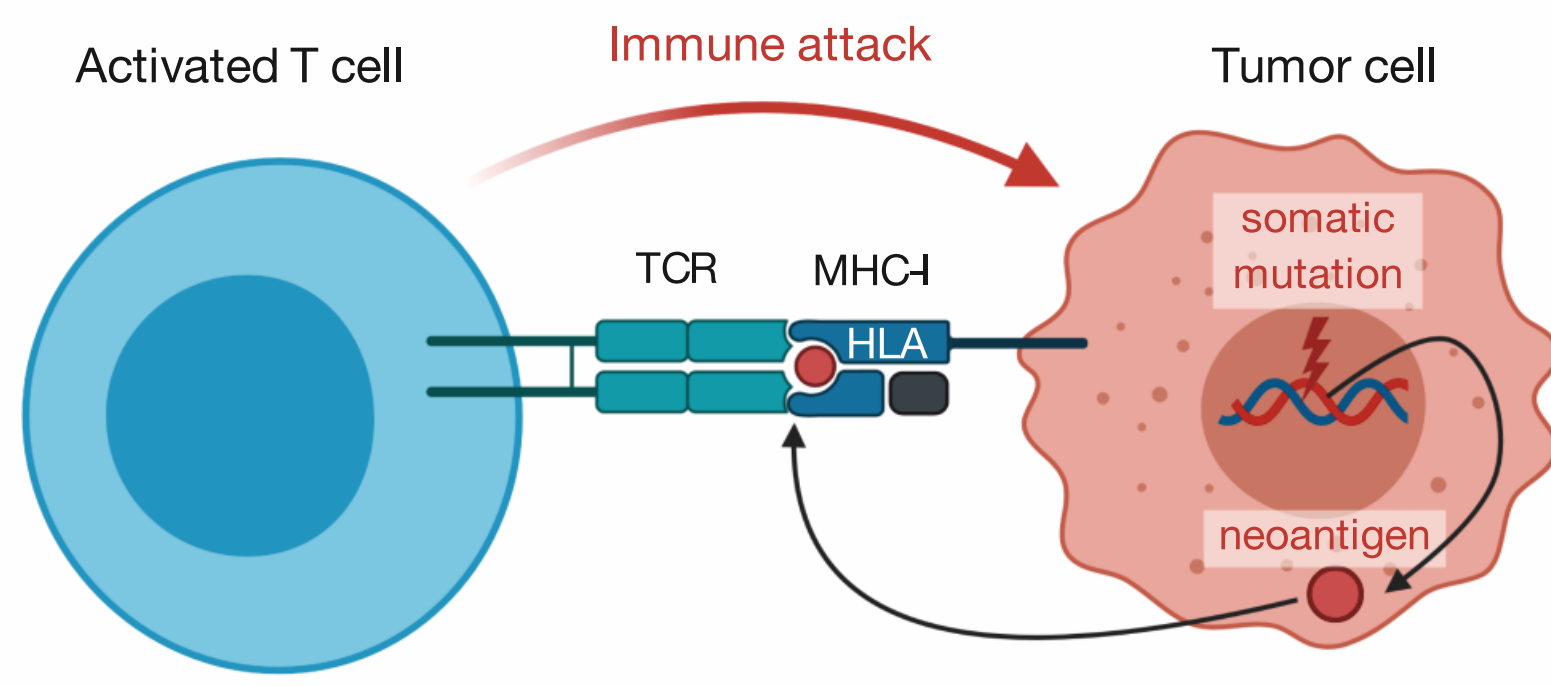


Benchmark of in silico tools for prediction of HLA genotypes from NGS data

Arne Claeys, Jasper Staut, Peter Merseburger, Kathleen Marchal and Jimmy Van den Eynden

Background



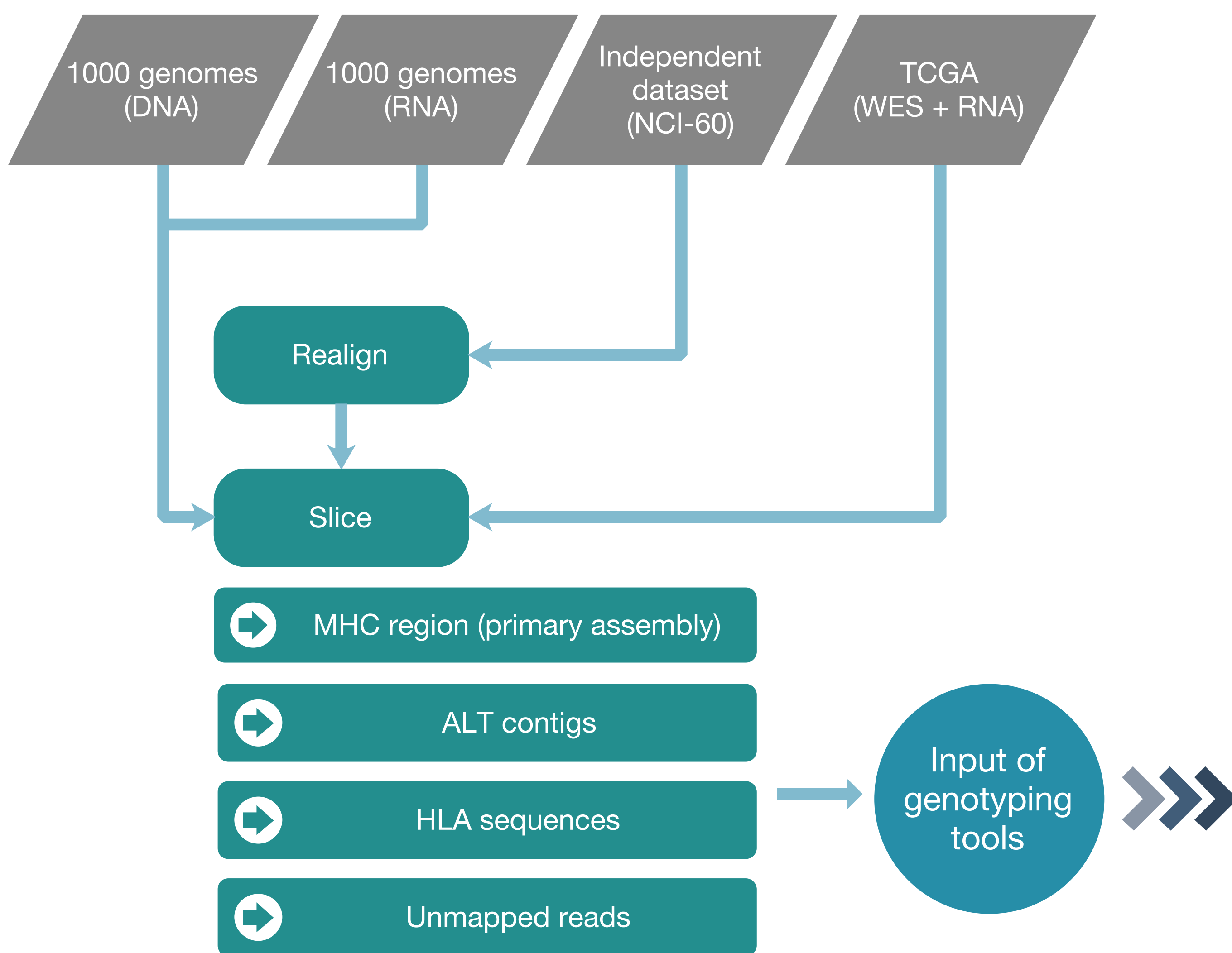
- HLA molecules of the MHC complex play a central role in **anti-tumour immunity**.
- The **presentability** of a particular antigen to the immune system depends on the **HLA genotype** of the subject.
- Genomic datasets with **matched gold standard** PCR-based HLA calls are rarely available.
- Many tools for *in silico* NGS-based HLA genotyping exist. However, there is **no consensus** on what is the best approach.

13 tools were selected for the benchmark

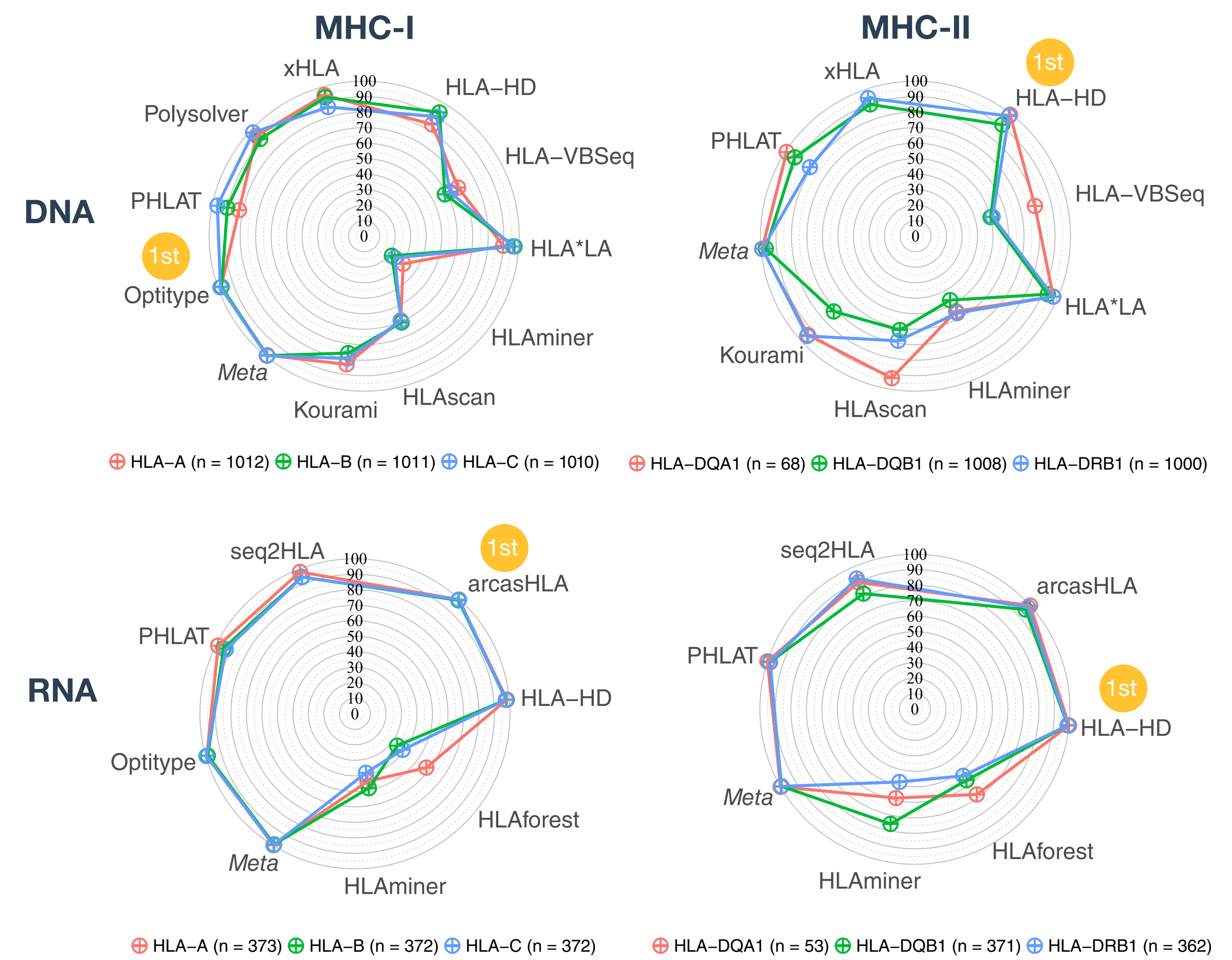
- Free for academic use
- Works on Ubuntu 20.04
- For DNA or RNA sequencing data

	Data type		Input filetype		Supported HLA loci							
	DNA	RNA	BAM	FASTQ	A	B	C	DPA1	DPB1	DQA1	DQB1	DRB1
Included	arcasHLA	X	✓	X	✓	✓	✓	✓	✓	✓	✓	✓
	HLA-HD	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓
	HLA-VBSeq	✓	X	✓	X	✓	✓	✓	✓	✓	✓	✓
	HLA*LA	✓	X	✓	X	✓	✓	✓	✓	✓	✓	✓
	HLAforest	X	✓	X	✓	✓	✓	✓	✓	✓	✓	✓
	HLAminer	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓
	HLAscan	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Kourami	✓	X	✓	X	✓	✓	✓	✓	✓	✓	✓
	Optitype	✓	✓	X	✓	✓	✓	✓	X	X	X	X
	PHLAT	✓	✓	X	✓	✓	✓	✓	X	X	✓	✓
	Polysolver	✓	X	✓	X	✓	✓	✓	X	X	X	X
	seq2HLA	X	X	X	✓	✓	✓	✓	✓	✓	✓	✓
	xHLA	✓	X	✓	X	✓	✓	✓	X	✓	X	✓
Not included	ALPHARD(-NT)	✓	X	?	?	✓	✓	✓	✓	✓	✓	✓
	ATHLATES	✓	X	X	✓	✓	✓	X	X	X	✓	✓
	HLAProfiler	X	X	X	✓	✓	✓	✓	✓	✓	✓	✓
	HLAreporter	✓	X	X	✓	✓	✓	✓	✓	✓	✓	✓
	HLAssign	✓	X	X	✓	✓	✓	✓	✓	✓	✓	✓
	OncoHLA	✓	X	X	✓	✓	✓	✓	✓	✓	✓	✓
	PolyPheMe	✓	X	X	✓	✓	✓	X	X	X	✓	✓
	SNP2HLA	X	X	X	X	✓	✓	✓	✓	✓	✓	✓
	SOAP-HLA	✓	X	✓	X	✓	✓	✓	✓	✓	✓	✓

2 Sample preprocessing

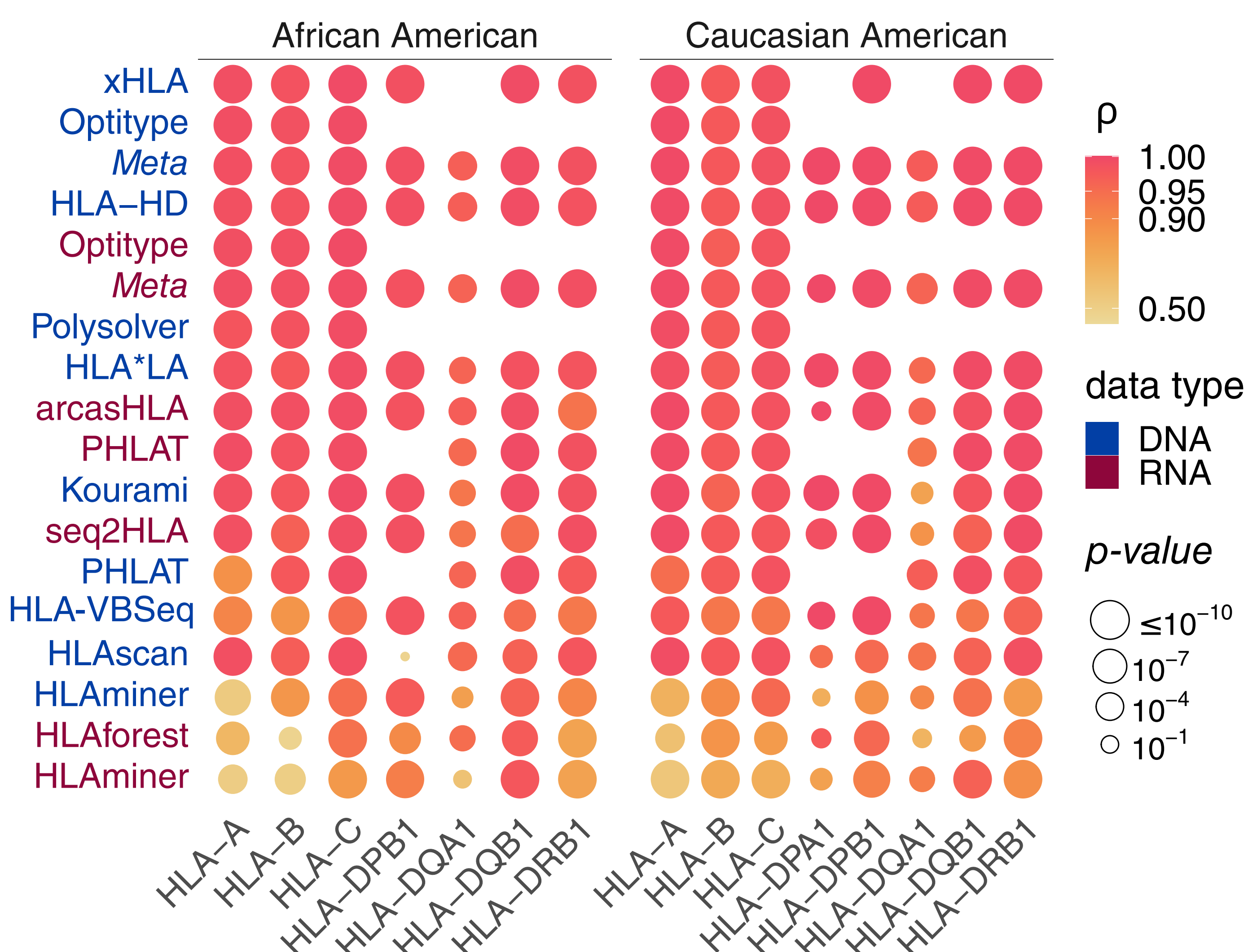


3 Benchmark on 1000 genomes data



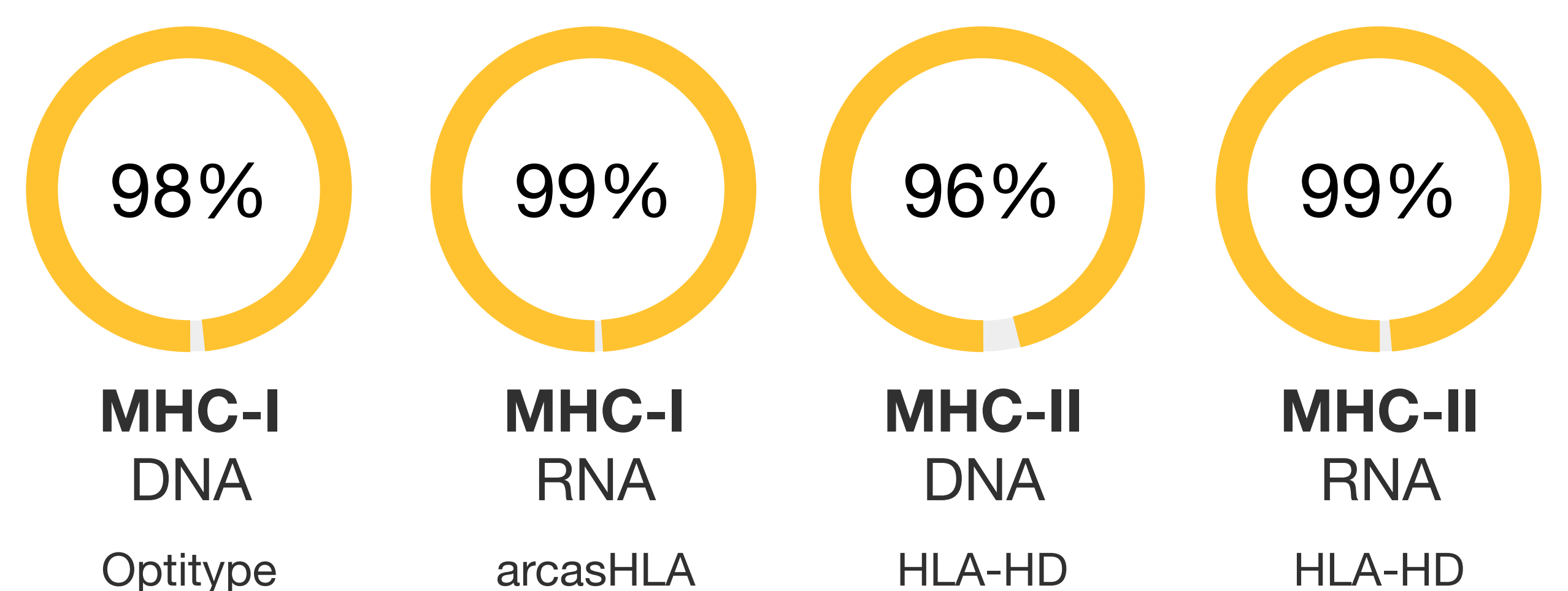
similar conclusions on the cell line data

4 Validation on TCGA data



correlation of predicted and expected population frequencies

Conclusion



For both MHC classes tools exist that **reliably predict** the HLA genotype based on DNA and RNA NGS data at **2 field resolution** on the G group (peptide binding domain) level.