NovAT

Preface

NovAT is a web application for the identification of novel HLA alleles from NGS raw data and HLA-typing results.

The software generates a reference from the HLA-alleles of HLA-typing result using IPD/IMGT-HLA database. Incomplete alleles from the database are filled via a proprietary algorithm. The generated consensus is used for alignment and variant calling. For incomplete HLA-alleles only variants identified in generic regions are reported. If the variant is detected that means that the sequenced sample carries the allele that is different from the allele reported in the HLA-typing results. This can evidence either of incorrect HLA-typing result or the presence of a novel HLA allele in the sample. It is recommended to confirm all detected novel alleles with orthogonal method.
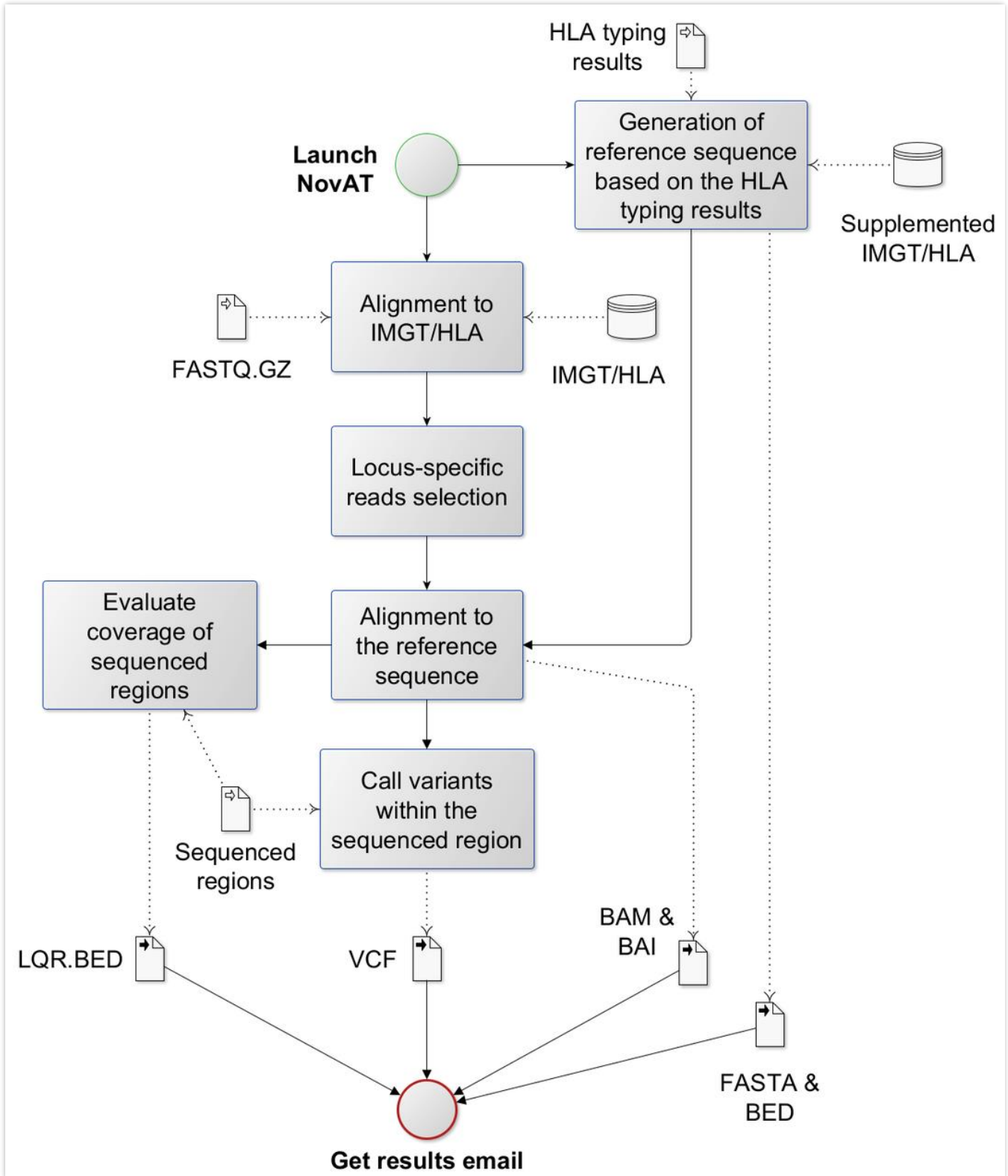
The software output contains the alignment file (BAM, BAI), variants (VCF) and regions' tracks (allele annotation and low-quality regions BED files). Regions with coverage less than 16 are designated as low-quality regions (LQR). This information is useful for data quality assessment: in general, the less LQR size the better the data.

The NovAT-used IPD/IMGT-HLA database release version is available on the website. IPD/IMGT-HLA database is updated for NovAT annually.

NovAT perform analysis only of the following 5 loci: HLA-A HLA-B HLA-C, HLA-DQB1, HLA-DRB1. Other HLA loci presented in the sequencing data won't interfere with the analysis.

NovAT software is available online: novat.parseq.pro

The algorithm of NovAT software is given on the scheme:

Getting started

Input data requirements

- NovAT is validated for paired-end Illumina sequencing data containing sequencing reads from loci: HLA-A, HLA-B, HLA-C, HLA-DQB1, HLA-DRB1. Other HLA loci presented in the sequencing data won't interfere with the analysis.
- Sequencing data format: FASTQ.GZ.
- File size restriction: 1-250 MB.
- HLA-typing results minimum resolution: 3 fields. If an allele is entered at a lower resolution, the first allele in group will be selected for the alignment. For example, if allele 02:02:02 is submitted, the alignment will be performed for allele 02:02:02:01. If submitted allele is absent in the current IPD/IMGT-HLA database version the analysis will be failed.
- It is recommended to use HLA-typing results generated with the same database version as given on the NovAT website. If you use the other version of the database, the analysis may fail if the allele was changed.
- Alleles must be written in "allele+allele" format without a prefix.

Output

- An indexed FASTA file containing the reference sequence with supplemented incomplete HLA alleles with corresponding BED files with annotated allele regions (intron, exon, UTR).
- An indexed BAM file containing the alignment.
- VCF file containing the identified variants.
- BED file containing low-covered regions.
- If analysis is completed without errors, you will receive an email with the link to download the analysis results. If the analysis is failed, you will receive an email with the error message.
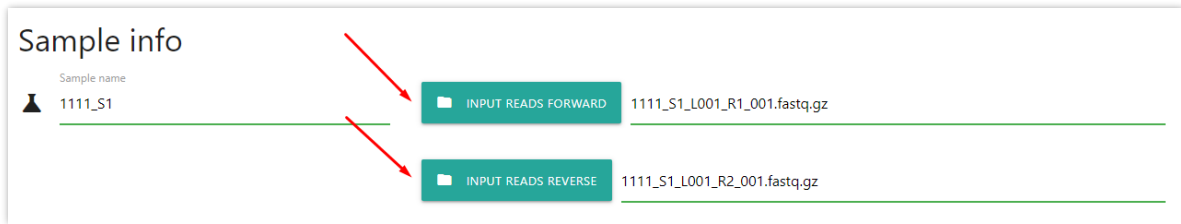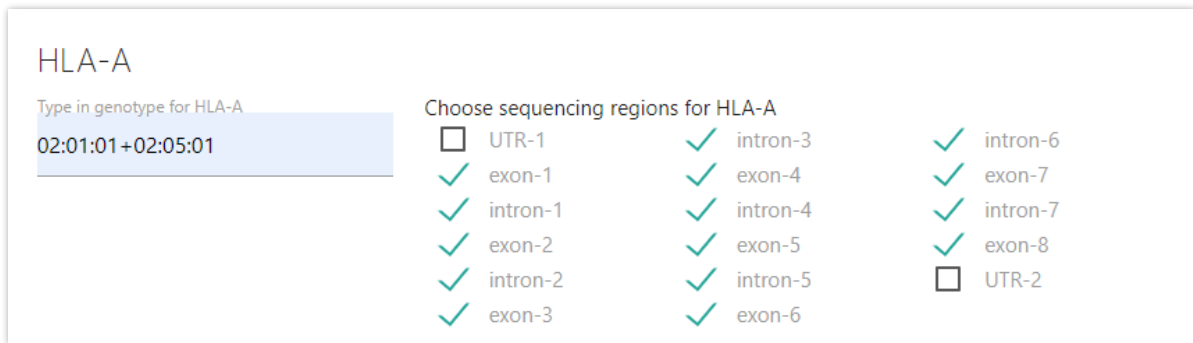
Usage

1) Go to https://novat.parseq.pro/
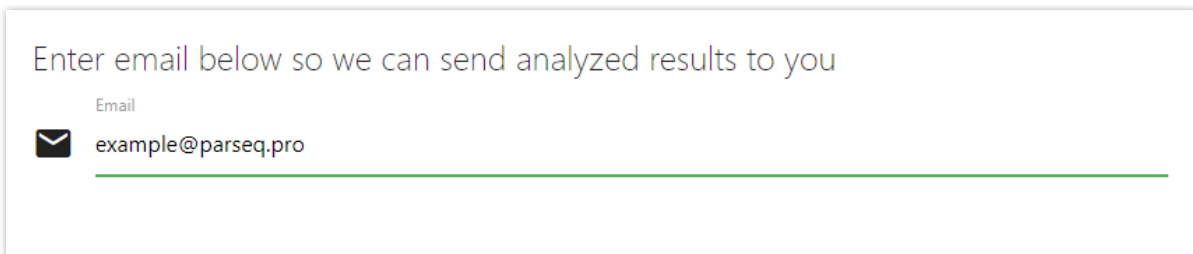2) Enter a sample name.

3) Import the forward and reverse reads using the "INPUT READS FORWARD" and "INPUT READS REVERSE" buttons. You can also drag and drop files into the input area.

4) Enter the genotype for each locus and mark the regions that have been sequenced. By default, the whole allele is analyzed.

5) Enter the email to receive the analysis result.

6) Click on the "SUBMIT" button. If data were imported correctly, you will see the message "Thank you! Your data have been successfully uploaded".