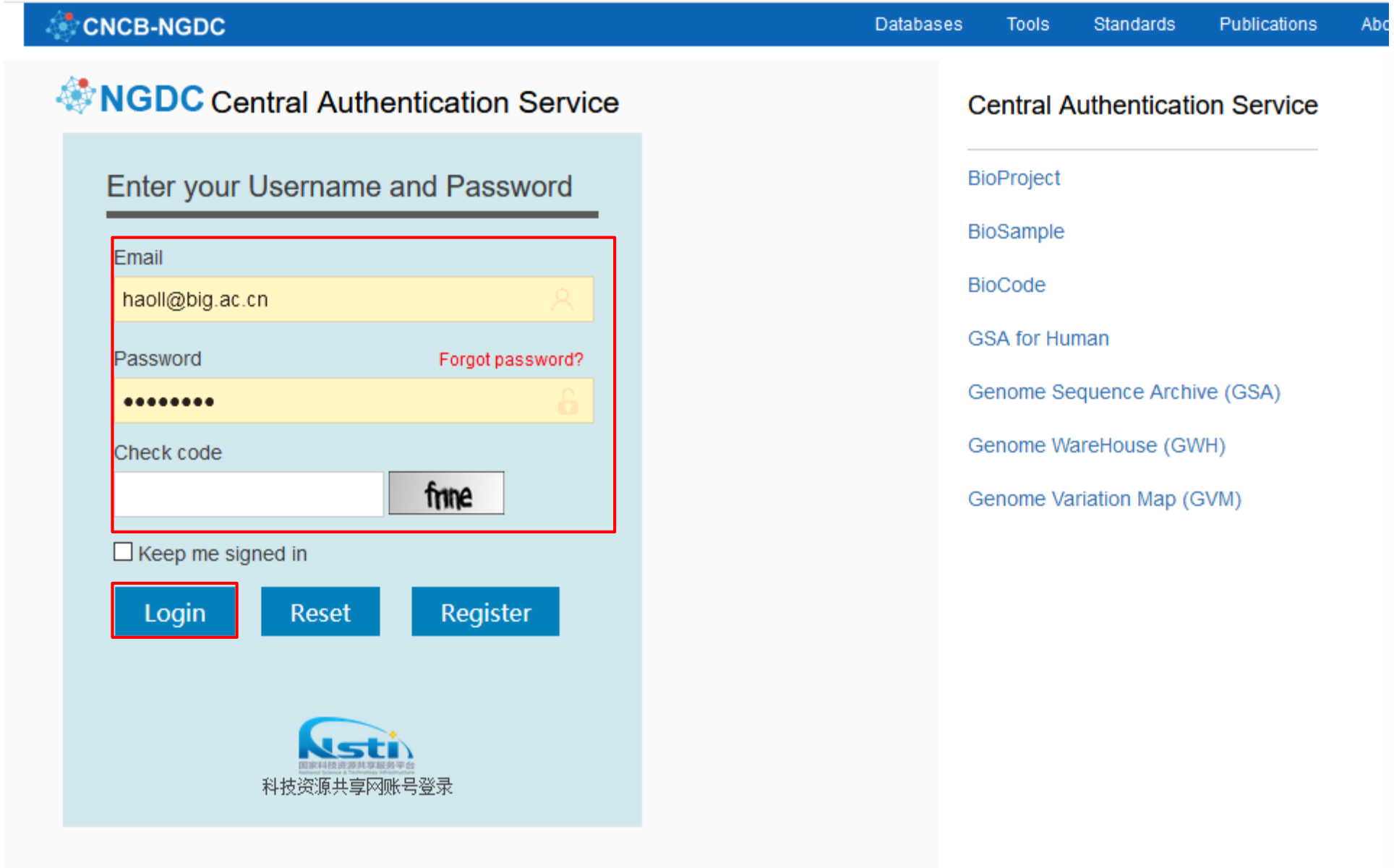
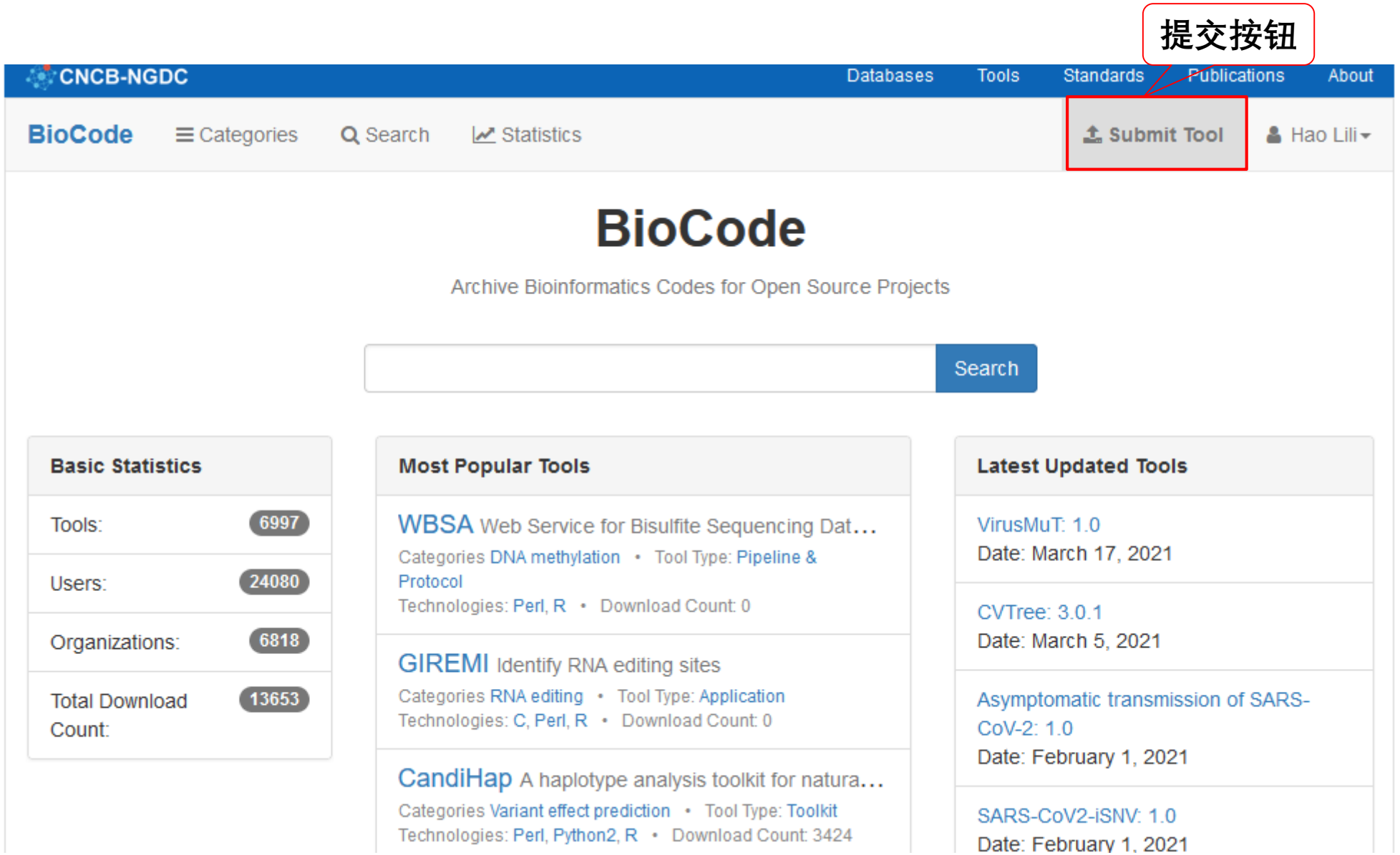
BioCode数据库

科学数据辅助工具软件提交入口

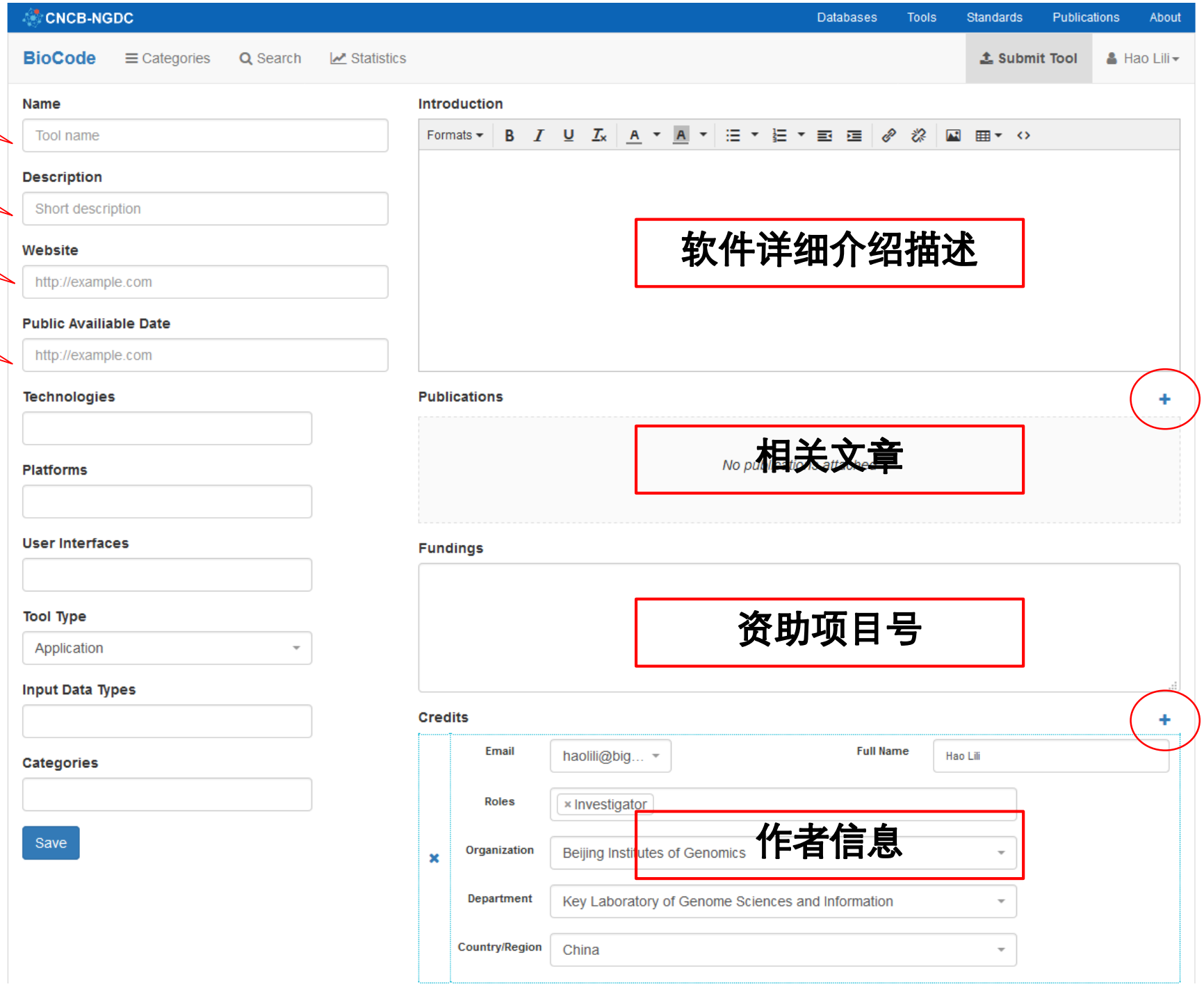
- 需要工具提交者提前准备的内容
  - 项目资助号信息
  - 软件工具基本描述性元信息
  - 作者信息、相关文章
  - 软件版本及源代码文件
  - 软件工具说明文档
  - 软件信息公开时间



https://bigd.big.ac.cn/sso/login?service=https://bigd.big.ac.cn/biocode/login


提交按钮
CNCB-NGDC
Databases
Tools
Standards
Publications
About
BioCode
Categories
Search
Statistics
Submit Tool
Hao Lili
BioCode
Archive Bioinformatics Codes for Open Source Projects
Search
Basic Statistics
Tools: 6997
Users: 24080
Organizations: 6818
Total Download Count: 13653
Most Popular Tools
WBSA Web Service for Bisulfite Sequencing Dat...
Categories DNA methylation • Tool Type: Pipeline & Protocol
Technologies: Perl, R • Download Count: 0
GIREMI Identify RNA editing sites
Categories RNA editing • Tool Type: Application
Technologies: C, Perl, R • Download Count: 0
CandiHap A haplotype analysis toolkit for natura...
Categories Variant effect prediction • Tool Type: Toolkit
Technologies: Perl, Python2, R • Download Count: 3424
Latest Updated Tools
VirusMuT: 1.0
Date: March 17, 2021
CVTree: 3.0.1
Date: March 5, 2021
Asymptomatic transmission of SARS-CoV-2: 1.0
Date: February 1, 2021
SARS-CoV2-iSNV: 1.0
Date: February 1, 2021

信息提交界面



软件名称

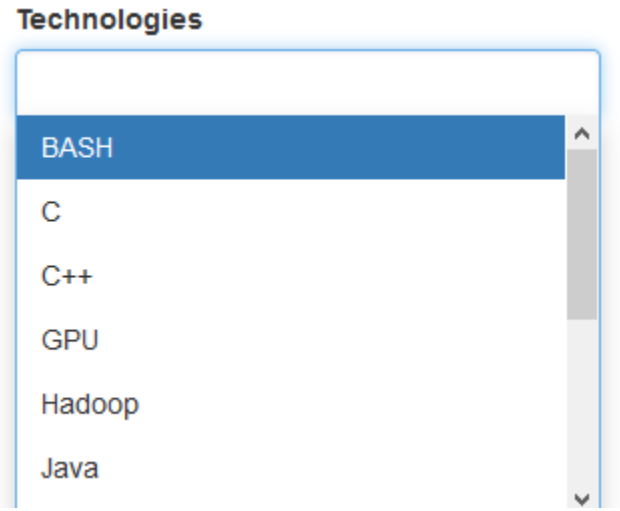
简要描述

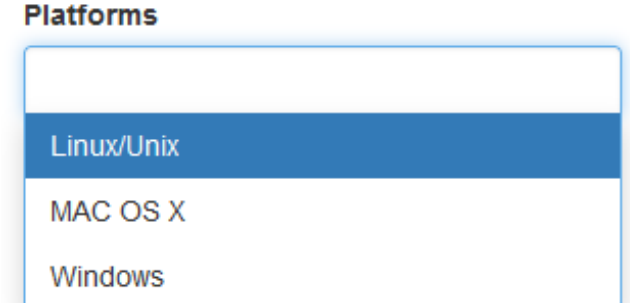
网站URL

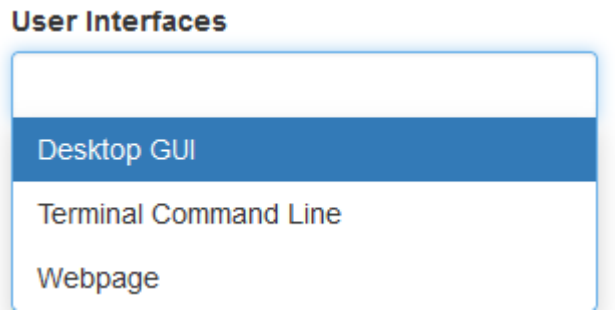
公开时间

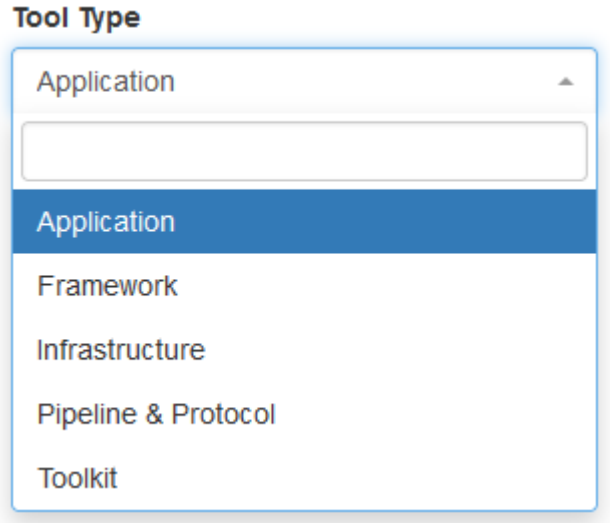
受控可选项

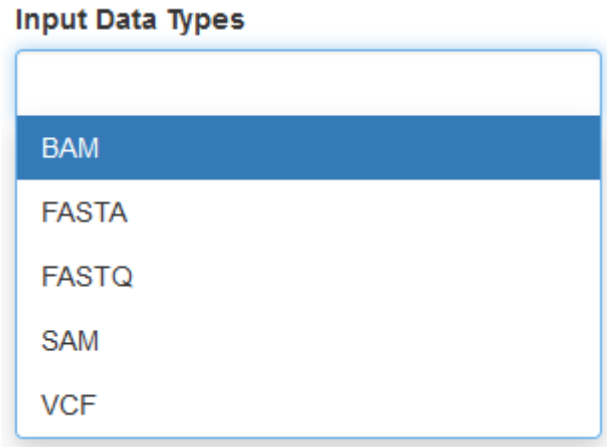
受控可选项内容

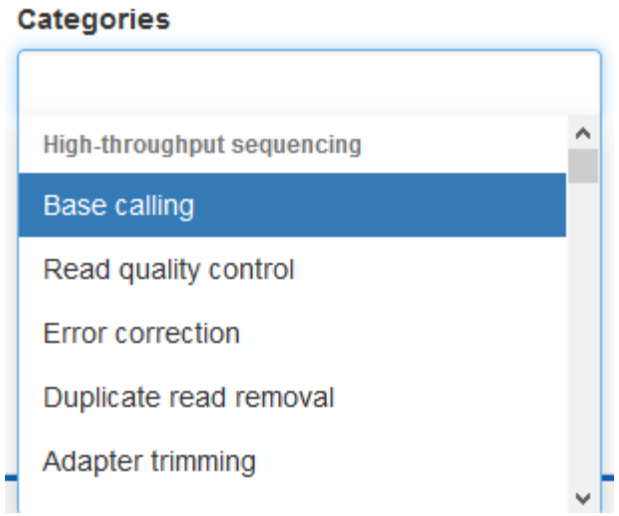

Technologies
BASH
C
C++
GPU
Hadoop
Java

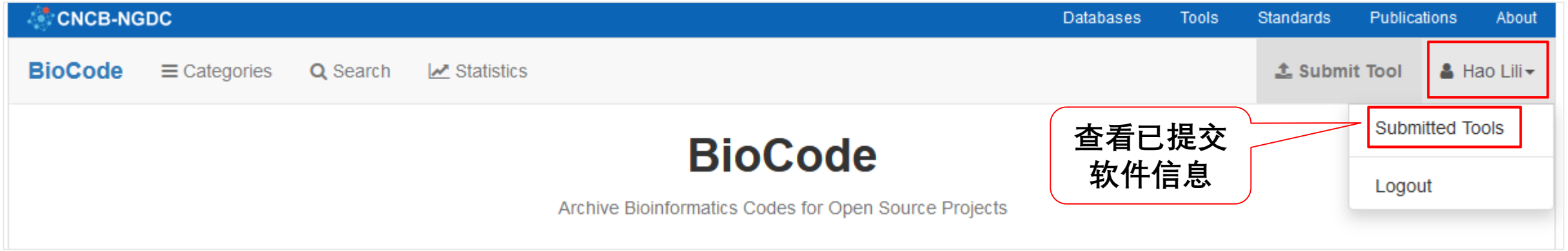

Platforms
Linux/Unix
MAC OS X
Windows

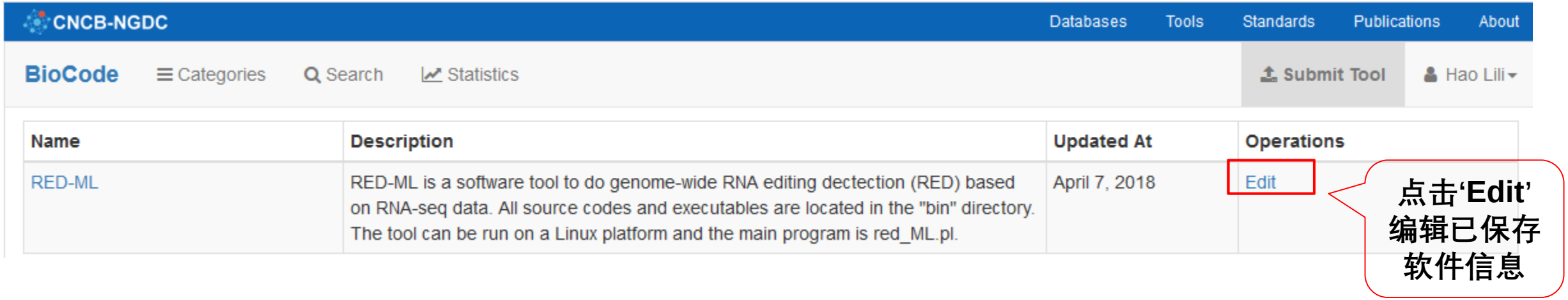

User Interfaces
Desktop GUI
Terminal Command Line
Webpage


Tool Type
Application
Application
Framework
Infrastructure
Pipeline & Protocol
Toolkit

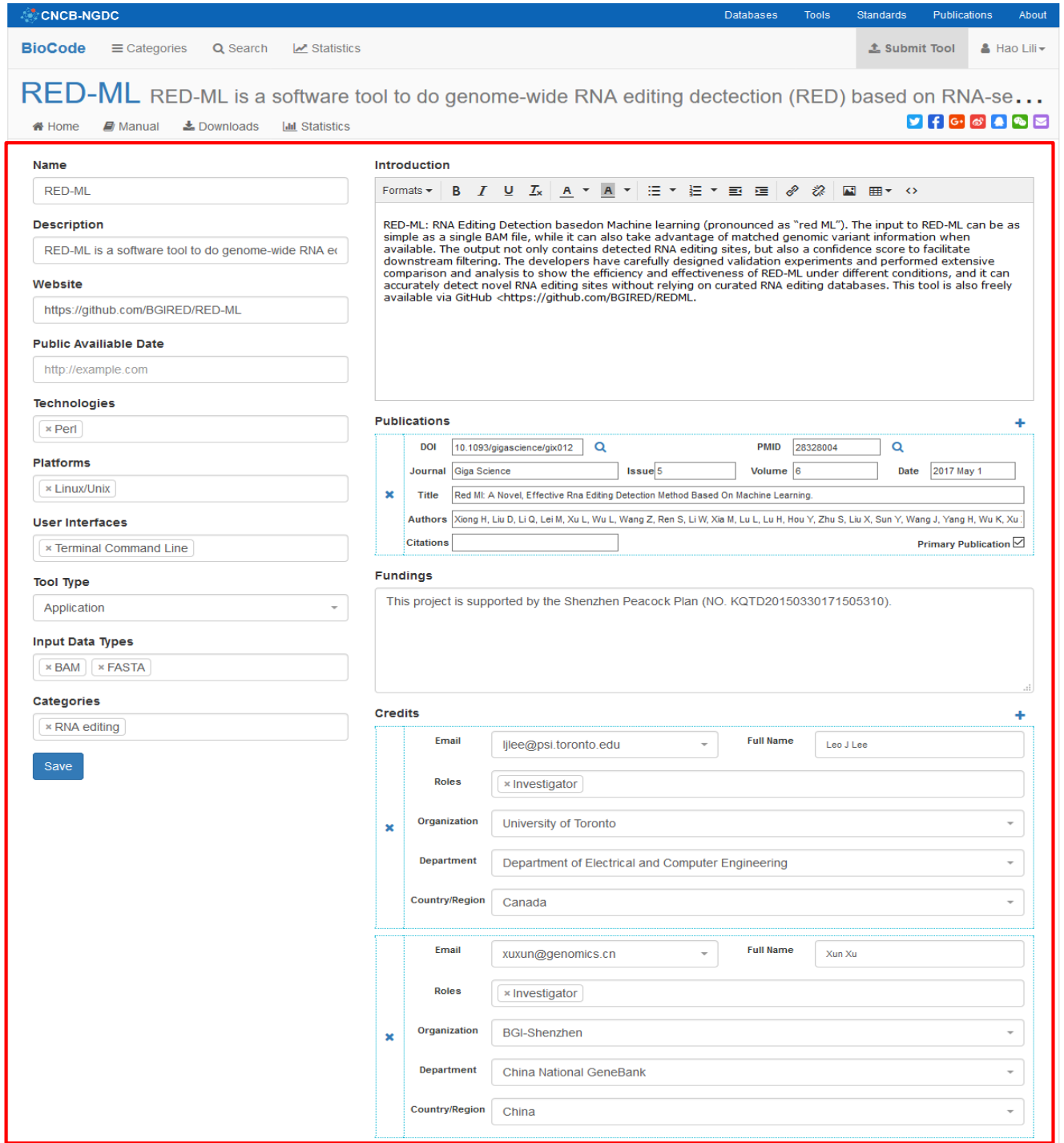

Input Data Types
BAM
FASTA
FASTQ
SAM
VCF


Categories
High-throughput sequencing
Base calling
Read quality control
Error correction
Duplicate read removal
Adapter trimming

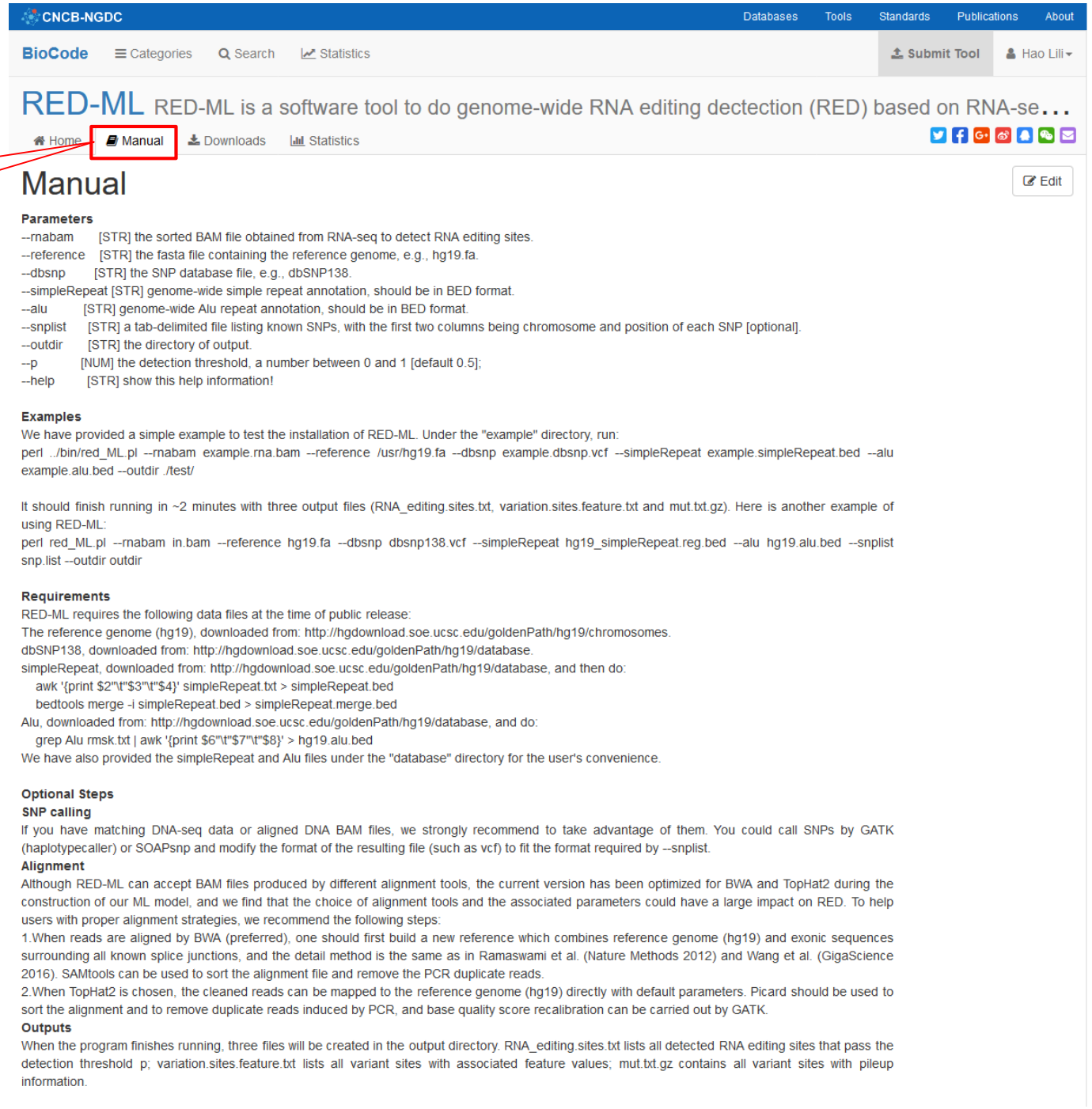

CNCB-NGDC
Databases
Tools
Standards
Publications
About
BioCode
Categories
Search
Statistics
Submit Tool
Hao Lili
Submitted Tools
Logout
BioCode
Archive Bioinformatics Codes for Open Source Projects
查看已提交
软件信息


CNCB-NGDC
Databases
Tools
Standards
Publications
About
BioCode
Categories
Search
Statistics
Submit Tool
Hao Lili
Name
Description
Updated At
Operations
RED-ML
RED-ML is a software tool to do genome-wide RNA editing dectection (RED) based on RNA-seq data. All source codes and executables are located in the "bin" directory. The tool can be run on a Linux platform and the main program is red_ML.pl.
April 7, 2018
Edit
点击‘Edit’
编辑已保存
软件信息

已保存基本信息条目
皆可编辑修改保存



点击‘Manual’可填写和修改软件说明文档



Manual

Parameters

--rnabam [STR] the sorted BAM file obtained from RNA-seq to detect RNA editing sites.
--reference [STR] the fasta file containing the reference genome, e.g., hg19.fa.
--dbsnp [STR] the SNP database file, e.g., dbSNP138.
--simpleRepeat [STR] genome-wide simple repeat annotation, should be in BED format.
--alu [STR] genome-wide Alu repeat annotation, should be in BED format.
--snplist [STR] a tab-delimited file listing known SNPs, with the first two columns being chromosome and position of each SNP [optional].
--outdir [STR] the directory of output.
--p [NUM] the detection threshold, a number between 0 and 1 [default 0.5];
--help [STR] show this help information!

Examples

We have provided a simple example to test the installation of RED-ML. Under the "example" directory, run:
perl ../bin/red_ML.pl --rnabam example.rna.bam --reference /usr/hg19.fa --dbsnp example.dbsnp.vcf --simpleRepeat example.simpleRepeat.bed --alu example.alu.bed --outdir ./test/

It should finish running in ~2 minutes with three output files (RNA_editing.sites.txt, variation.sites.feature.txt and mut.txt.gz). Here is another example of using RED-ML:
perl red_ML.pl --rnabam in.bam --reference hg19.fa --dbsnp dbsnp138.vcf --simpleRepeat hg19_simpleRepeat.reg.bed --alu hg19.alu.bed --snplist snp.list --outdir outdir

Requirements

RED-ML requires the following data files at the time of public release:
The reference genome (hg19), downloaded from: http://hgdownload.soe.ucsc.edu/goldenPath/hg19/chromosomes.
dbSNP138, downloaded from: http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database.
simpleRepeat, downloaded from: http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database, and then do:
awk '{print $2"\t"$3"\t"$4}' simpleRepeat.txt > simpleRepeat.bed
bedtools merge -i simpleRepeat.bed > simpleRepeat.merge.bed
Alu, downloaded from: http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database, and do:
grep Alu rmsk.txt | awk '{print $6"\t"$7"\t"$8}' > hg19.alu.bed
We have also provided the simpleRepeat and Alu files under the "database" directory for the user's convenience.

Optional Steps

SNP calling

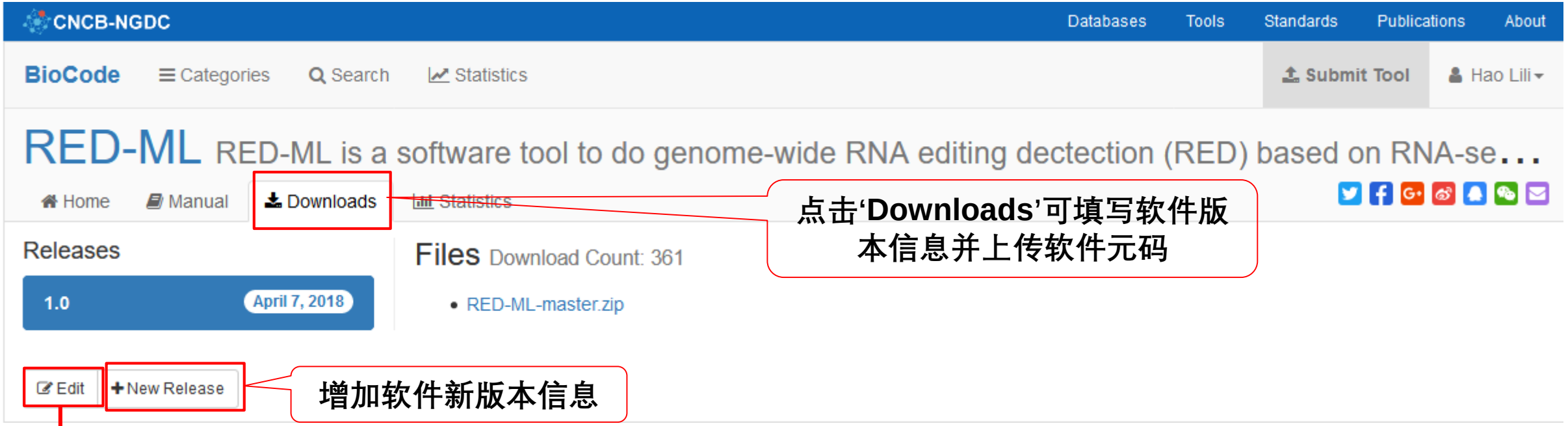
If you have matching DNA-seq data or aligned DNA BAM files, we strongly recommend to take advantage of them. You could call SNPs by GATK (haplotypecaller) or SOAPsnp and modify the format of the resulting file (such as vcf) to fit the format required by --snplist.

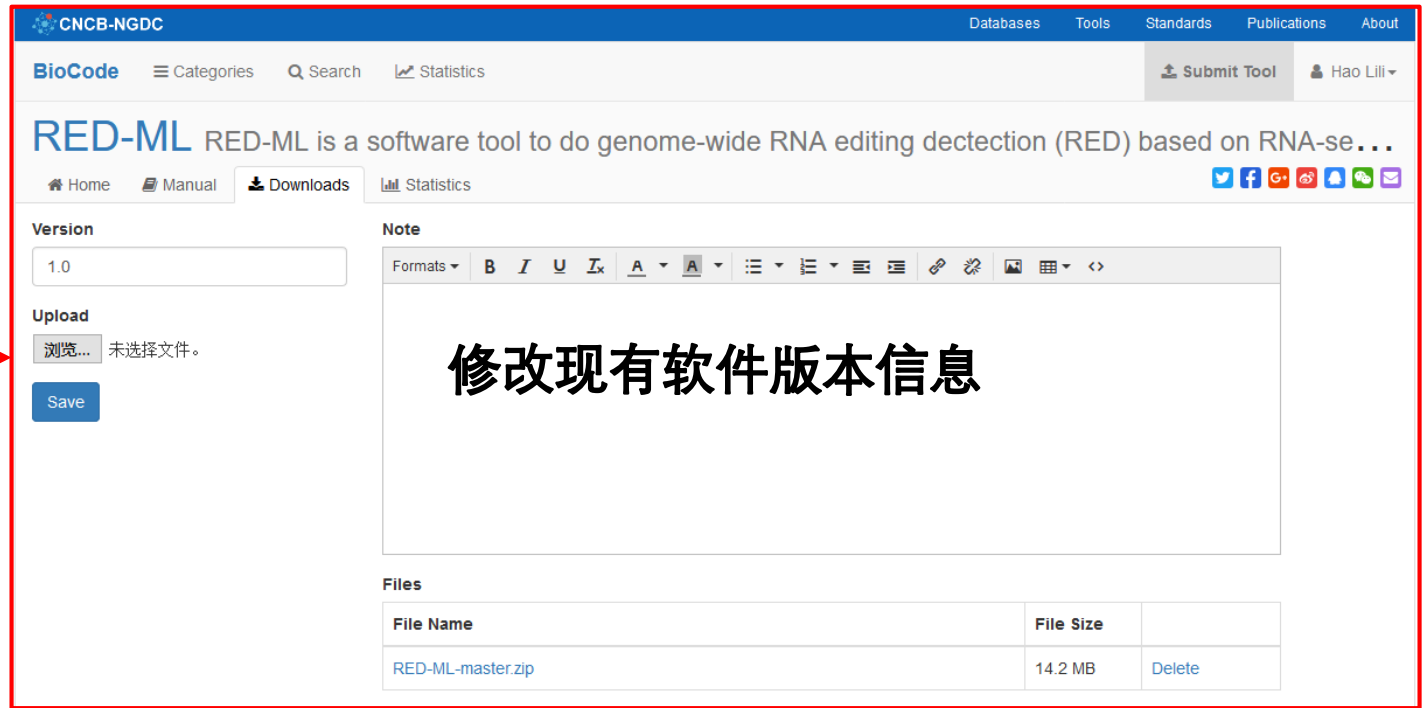
Alignment

Although RED-ML can accept BAM files produced by different alignment tools, the current version has been optimized for BWA and TopHat2 during the construction of our ML model, and we find that the choice of alignment tools and the associated parameters could have a large impact on RED. To help users with proper alignment strategies, we recommend the following steps:
1.When reads are aligned by BWA (preferred), one should first build a new reference which combines reference genome (hg19) and exonic sequences surrounding all known splice junctions, and the detail method is the same as in Ramaswami et al. (Nature Methods 2012) and Wang et al. (GigaScience 2016). SAMtools can be used to sort the alignment file and remove the PCR duplicate reads.
2.When TopHat2 is chosen, the cleaned reads can be mapped to the reference genome (hg19) directly with default parameters. Picard should be used to sort the alignment and to remove duplicate reads induced by PCR, and base quality score recalibration can be carried out by GATK.

Outputs

When the program finishes running, three files will be created in the output directory. RNA_editing.sites.txt lists all detected RNA editing sites that pass the detection threshold p; variation.sites.feature.txt lists all variant sites with associated feature values; mut.txt.gz contains all variant sites with pileup information.


CNCB-NGDC
Databases
Tools
Standards
Publications
About
BioCode
Categories
Search
Statistics
Submit Tool
Hao Lili
RED-ML RED-ML is a software tool to do genome-wide RNA editing dectection (RED) based on RNA-se...
Home
Manual
Downloads
Statistics
点击‘Downloads’可填写软件版本信息并上传软件元码
Releases
1.0
April 7, 2018
Files Download Count: 361
RED-ML-master.zip
Edit
New Release
增加软件新版本信息


CNCB-NGDC
Databases
Tools
Standards
Publications
About
BioCode
Categories
Search
Statistics
Submit Tool
Hao Lili
RED-ML RED-ML is a software tool to do genome-wide RNA editing dectection (RED) based on RNA-se...
Home
Manual
Downloads
Statistics
Version
1.0
Upload
浏览...
未选择文件。
Save
Note
Formats
修改现有软件版本信息
Files
File Name
File Size
RED-ML-master.zip
14.2 MB
Delete

RED-ML RED-ML is a software tool to do genome-wide RNA editing dectection (RED) based on RNA-se...

Introduction

RED-ML: RNA Editing Detection basedon Machine learning (pronounced as “red ML”). The input to RED-ML can be as simple as a single BAM file, while it can also take advantage of matched genomic variant information when available. The output not only contains detected RNA editing sites, but also a confidence score to facilitate downstream filtering. The developers have carefully designed validation experiments and performed extensive comparison and analysis to show the efficiency and effectiveness of RED-ML under different conditions, and it can accurately detect novel RNA editing sites without relying on curated RNA editing databases. This tool is also freely available via GitHub <https://github.com/BGIRED/REDML.

Publications

1. Red Ml: A Novel, Effective Rna Editing Detection Method Based On Machine Learning. *Cite this*
   Xiong H, Liu D, Li Q, Lei M, Xu L, Wu L, Wang Z, Ren S, Li W, Xia M, Lu L, Lu H, Hou Y, Zhu S, Liu X, Sun Y, Wang J, Yang H, Wu K, Xu X, Lee LJ, 2017 May 1 - *Giga Science*

Credits

1. Leo J Lee ljlee@psi.toronto.edu Investigator
   Department of Electrical and Computer Engineering, University of Toronto, Canada
2. Xun Xu xuxun@genomics.cn Investigator
   China National GeneBank, BGI-Shenzhen, China

Community Ratings

Usability	Efficiency	Reliability	Rated By
★★★★★	☆☆☆☆☆	☆☆☆☆☆	1 users
★★★★★ ⊖	☆☆☆☆☆ ⊖	☆☆☆☆☆ ⊖	hao***i@big.ac.cn (May 7, 2021)

Edit Delete

Summary

Accession	BT007072
Tool Type	Application
Category	RNA editing
Platforms	Linux/Unix
Technologies	Perl
User Interface	Terminal Command Line
Input Data	BAM, FASTA
Latest Release	1.0 (April 7, 2018)
Download Count	361
Country/Region	Canada
Submitted By	Hao Lili

Fundings

This project is supported by the Shenzhen Peacock Plan (NO. KQTD20150330171505310).

关联项目资助号