Seq2C Getting Started Guide

Version Number: 1.0

|  |  |
| --- | --- |
| Document Details | |
| Document Title | Seq2C Getting Started Guide |
| Status / Version | Final/1.0 |
| Reference Code |  |

|  |
| --- |
| Disclaimer |
| If you are reading a copy of this document (i.e. print out of controlled electronic document or an electronic copy of a wet ink signed paper document) you may be reading an uncontrolled version. Before making updates, please ensure that you are using the latest version of this document and the associated template. |

|  |  |  |  |
| --- | --- | --- | --- |
| Document Approval / Review | | | |
| Criteria | Name & Role | Signature | Date |
| Author |  |  |  |
| Approved |  |  |  |
| Approved |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Version Control | | | |
| Version | Date | Description of Changes | Author |
| 1.0 |  | Initial version | Viktor Kirst |

| Referenced Documents | | | |
| --- | --- | --- | --- |
| No. | Document Title and ID | Version | Location |
|  |  |  |  |

**Table of Contents**

1. Source Location 3

2. Build Instruction 3

1.1. Distribution Package Structure 3

1.2. Third-Party Libraries 3

3. Launch Instruction 4

# Source Location

The Seq2C source code is located at <https://github.com/AstraZeneca-NGS/Seq2CJava>.

To load the project, execute the following command:

git clone --recursive https://github.com/AstraZeneca-NGS/Seq2CJava.git

# Build Instruction

The following prerequisites are necessary to build the Seq2C project:

* JDK1.7 (or higher)
* Internet access

The project uses Gradle (<http://gradle.org/>) and already includes a gradlew script.

To build the project, in the root folder of the project, run the following command:  
./gradlew clean installDist

To build the project as a distribution archive, in the build/distributions folder, run the following command:

./gradlew clean distZip

This command builds the project and creates an archive of the distribution package in the folder.

## Distribution Package Structure

When the build command completes successfully, the build/install/Seq2C folder contains the distribution package.

The distribution package has the following structure:

bin/ - contains the launch scripts

lib/ - has the jar file that contains the compiled project code and the jar files of the third-party libraries that the project uses.

You can move the distribution package (the content of the build/install/Seq2C folder) to any convenient location.

## Third-Party Libraries

Currently, the project uses the following third-party libraries:

* **Commons CLI** (<http://commons.apache.org/proper/commons-cli>, Apache License) – a library for parsing the command line.
* **HTSJDK** (<http://smtools.github.io/htsjdk/>) is an implementation of a unified Java library for accessing common file formats, such as SAM and VCF.
* **Commons Math** (<https://commons.apache.org/proper/commons-math>) - a library of lightweight, self-contained mathematics and statistics components.

# Launch Instruction

The following prerequisites are necessary to launch the program:

* Java 1.7 (or higher)

To launch Seq2C:

* Run the command <distribution package path>/bin/Seq2C or
* Run just Seq2C if <distribution package path>/bin/ is in path

The work of the program is identical to the function of the original Perl script and thus can replace it.

For example, the original launch command (from the ***readme*** file of the perl version) is:

seq2c.sh sample2bam.txt Illumina\_TruSeq\_Exome.bed

The same command in Seq2CJava is as follows:

<distribution package path>/bin/Seq2C sample2bam.txt Illumina\_TruSeq\_Exome.bed cov.txt

The program uses the same parameters as the Perl version. The Java version contains an additional parameter: -i. This parameter manages the number of threads that do the work. If this parameter is missing, then the mode is one-thread. If you add the -i parameter, the number of threads equals to the number of processor cores. The parameter -i threads sets the number of threads explicitly.

For launching the separate parts of code an additional parameter –r is used. For launching the first part of code option –r 1 (analogue to seq2cov.pl script in Perl) is used. For launching the second part of code option –r 2 (analogue to bam2reads.pl, cov2lr.pl and lr2gene.pl scripts in Perl). When Seq2c is launched without option –r all code will run from start to end.