

KNIME Pipeline User Manual

Eva Freckmann

Contents

1	Before starting	1
1.1	KNIME installation	1
1.2	R Integration	1
1.3	Python Integration	2
2	Data import and pre-processing	2
3	Data summary files	2
4	Automated	2
5	User-defined	2

1 Before starting

The computer you will be running the KNIME analysis from should have either access to, or a local, copy of the folder containing your phase images, and output from CellProfiler.

1.1 KNIME installation

KNIME can be downloaded from [here](#). **Notes for KNIME workflow:**

Required KNIME version and extensions

Name	Version
KNIME Analytics Platform	1
R	2

1.2 R Integration

The KNIME workflow uses an R integration for some steps of the analysis. The workflow uses R version **version**, which can be downloaded from [here](#). Instructions for setting up the R integration can be found [here](#). **Note for Windows users:** Do not install Rserve version 1.7-3.1 as is suggested in point 1 of the “R packages installation” section of these instructions. Instead, go straight to point 2 of the section, to install Rserve v1.8.7 **check this**. More information on installing Rserve can be found [here](#).

All users: In KNIME Analytics Platform go to File → Preferences. From the list on the left, select R under KNIME. Set the “Rserve receiving buffer size limit” to 0.

The KNIME workflow should automatically download and install any missing R packages that are required for the analysis. **Check that cytofkit does this**

1.3 Python Integration

The KNIME workflow requires Python in order to run the GeoSketch algorithm for subsampling data. Instructions for setting up the Python integration can be found [here](#). First, follow the “Quickstart” and “Anaconda Setup” instructions, and download the Python environments provided on the **davebryant-lab/MethodsPaper2020** Github. Load these environments into Anaconda. Then follow the “Setting up the KNIME Python Integration” instructions, and select the provided environments within your KNIME Python preferences.

2 Data import and pre-processing

3 Data summary files

think of a better name for this section

4 Automated

5 User-defined