Installation

The raw data files received from your HT sequencing facility of choice can be processed with the Decipher Bar-Code Deconvoluter. This software is available by downloading from the Decipher web site (http://www.decipherproject.net/software/). The 32-bit version is freely distributed and can be used for raw data files up to 2 gigs in size. The 64-bit version is available commercially and includes additional features such as support for raw data files larger than 2 gigs, support for multiplexing, technical services, ability to correct two sequencing errors, and more. For the Windows version of the Enumerator, your computer also needs to have the most recent .NET framework, downloadable from Microsoft (http://www.microsoft.com/net/).

After downloading, unpack the .zip archive and run the installer package. Free registration is required to remove usage limits. Unregistered software can only be run 10 times. The opening screen will generate a computer-specific code. Forward the code to support@decipherproject.net and you will receive a one-computer license which can be entered to remove all software limitations.

Raw data conversion

High-throughput sequencing data can be returned in several formats. Illumina HT sequencers generate a set of 100 QSEQ files for every sample (text files in the format FILENAME_qseq.txt). Combined, these contain each individual sequence found in the sample. Individual QESQ files need to be merged into one file per sample before further processing. Merged files can get quite large and approach 4Gb in size. For merging files we suggest TXTCollector, a free utility available from http://bluefive.pair.com/txtcollector.htm.

Other raw data formats are sometimes used. In addition to QSEQ, Deconvoluter supports all major HT sequencing data formats, including RAW (.txt, .raw), FastA (.fasta), FastQ (.fastq), Miro (.miro), and Seq files (*_seq.txt).

Deconvoluter is able to convert raw data to TAB (.tab extension), an alternative format which presents each unique sequence and the number of times it was found in the sample. This format allows to present the same data in smaller files of ~300mb. Select your "Input file" and click the *Convert* button.
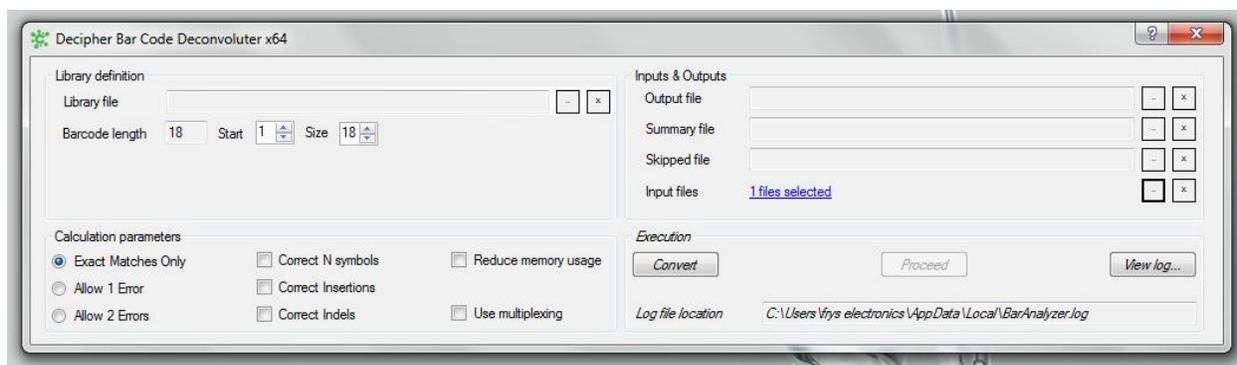
Certain HT facilities return data in TAB format. These can be used directly for the deconvolution step.

Bar-code deconvolution

Deconvolution is a process where individual bar-codes are matched to the corresponding gene names and annotations. To run bar-code deconvolution, you would need the appropriate Decipher library annotation file in a dedicated BLIB format. Currently, annotation files for Decipher Human Modules 1 and 2 and Mouse Modules 1 and 2 are available and can be downloaded from the Decipher web site (http://www.decipherproject.net/software/).
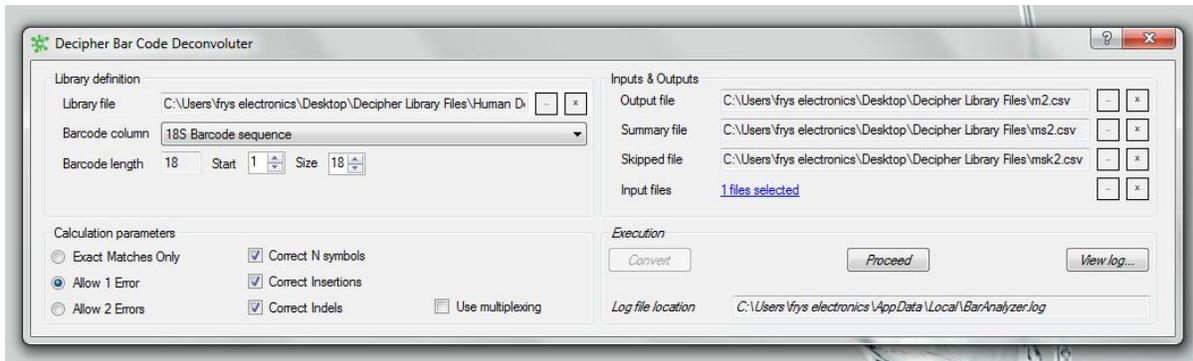
Select the appropriate annotation file in the "Library file" line on the right side of the program workspace.

Select the appropriate bar-code column in the drop-down menu directly below; for the Decipher annotation files it would be the first choice, "18S Bar-code Sequence".

Decipher bar-code design allows to detect errors in the bar-code sequence. The 32-bit program version can correct one error in the bar-code and discounts the sequence if there are more, while the 64-bit version is capable of correcting up to two errors. As a default, we suggest checking "Allow 1 Error", as well as the checkboxes "Correct N symbols", "Correct Insertions", and "Correct Indels".

Next, designate the names and locations for the output on the left side of the program workspace. Deconvoluter can save the results in .csv (comma-separated values) or .txt formats. "Output file" contains annotations and count data for each bar-code. "Summary file" keeps general statistics about the analysis, including total number of processed data points (reads), and the breakdown into the skipped ones and the counted ones; the latter include exact matches and corrected sequences by the type (one corrected error, corrected N, indels, insertions). "Skipped file" will contain individual bar-code sequences which were not included in the analysis. *Summary* and *Skipped* files are optional.

Choose the input file(s) directly below and press *Proceed*.

The deconvolution process is very memory-intensive. We recommend running Deconvoluter on systems with at least 4Gb of memory, and more if two or more files at a time are converted. Once the bar-code deconvolution is complete, one can proceed to normalization and data analysis.