

Extracting barcode split data from SOLiD 5500 XSQ files

Extracting barcode split data from SOLiD 5500 XSQ files can be done by using ABI XSQ_Tools, currently installed at /home/daras/XSQ_Tools.

1. To split the XSQ files into multiple XSQ files, by barcode:

convertFromXSQ.sh -s xsqfile -o outputpath

2. To convert each of the resulting XSQ files into csfasta and QV files:

convertFromXSQ.sh -c xsqfile

The converted files are created in the following directory: ./Libraries/<LibraryName>/<TagName>/reads

Typing just the command by itself will display all possible command line options and parameters.

To extract all the barcode sequences UNASSIGNED by the LifeTech software from an .xsq file, use:

extractBCfromXSQ <input.xsq> > bc.csfasta 2> bc_QV.qual

To extract all the unassigned F3, or F5-RNA sequences from an xsq file, use:

extractF3fromXSQ <input.xsq> > f3.csfasta 2> f3_QV.qual

or

extractF5fromXSQ <input.xsq> > f5.csfasta 2> f5_QV.qual

(note that these are all just scripts that call other scripts to do all this - pretty easy to extract anything from an hdf file. You sometimes have to use hdfview to see exactly what the tables in the .xsq file are called, then use h5dump to pull it out).

<input.xsq> should be the raw, final .xsq file for the lane, not the _Unassigned.xsq residual file generated after convertFromXSQ.sh.