

iTRAQ and TMT Crosstalk Compensation with Crosstraq

Tutorial based on Crosstraq version 0.0.2, last update on March 11, 2019

Written by Brian Searle (bsearle@systemsbiology.org)

Crosstraq is an open-source Java library for removing sources of interference from iTRAQ and TMT data. Data interpretation in TMT and iTRAQ experiments is complicated by two factors: chemical contamination and isotopic contamination. Chemical contamination in proteomics measurements occurs when multiple peptides with similar precursor masses are co-isolated and co-fragmented. Isotopic contamination is caused because isobaric tagging reagents cannot be produced with 100% purity. Crosstraq is a tool to help mitigate both sources of error.

PREREQUISITES

The newest version of Crosstraq is freely available for download from:

<https://bitbucket.org/searleb/crosstraq>

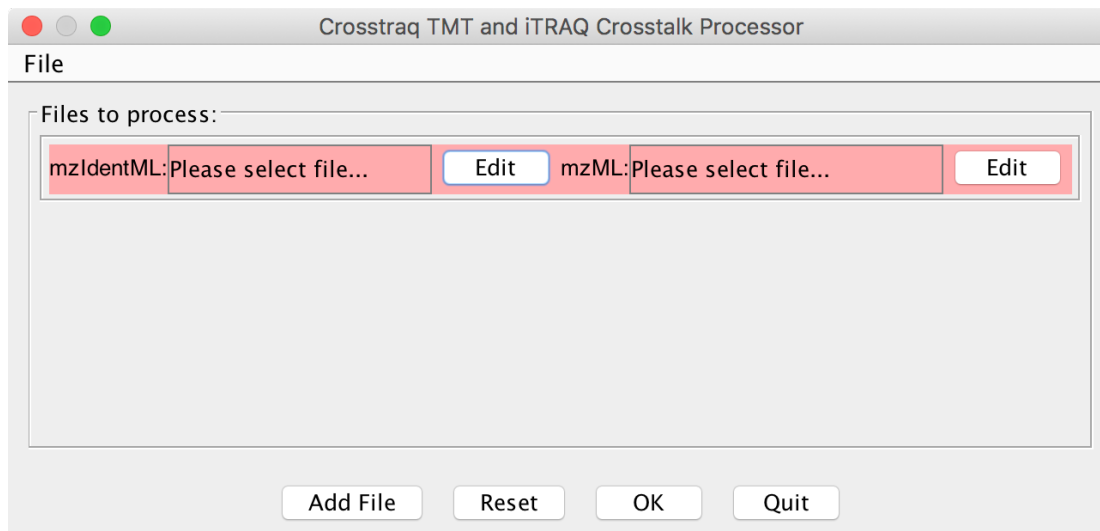
Crosstraq is a cross-platform Java application that has been tested for Windows, Macintosh, and Linux. Crosstraq requires 64-bit Java 1.8. If you don't already have it, you can download "Windows x64 Offline" from:

<http://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html>

After you have 64-bit Java 1.8, double click on the Crosstraq .JAR file to launch the GUI interface. If you are using a Macintosh, you may need to right click on the Crosstraq .JAR and select "Open" to execute it for the first time with the proper permissions.

LOADING FILES INTO CROSSTRAQ

Crosstraq has a simple graphical interface to process multiple raw files in batch. This is how Crosstraq appears after launching it for the first time:



We recommend using the msConvert tool from Proteowizard (available at <http://proteowizard.sourceforge.net/download.html>) to generate both MGFs and mzMLs from your raw files. MS1s and MS2s in mzMLs can be either centroid or profile. MS2s in MGFs are assumed to be centroid in the reporter ion region. The MGFs can be given to virtually any search engine that supports the mzIdentML standard (for example, MSGF+, available at <https://github.com/MSGFPlus/msgfplus>). Crosstraq analysis begins after searching your MGF files.

Each raw file needs an mzML raw file and an mzIdentML database search engine result file. To process a search, first queue the database search engine mzIdentML file by clicking the “Edit” button next to that field. Then queue the mzML raw file using the second “Edit” button. If you want to process multiple raw files sequentially, press the “Add File” button to add additional mzIdentML/mzML paired files. Once you have set up your jobs, press the “OK” button.

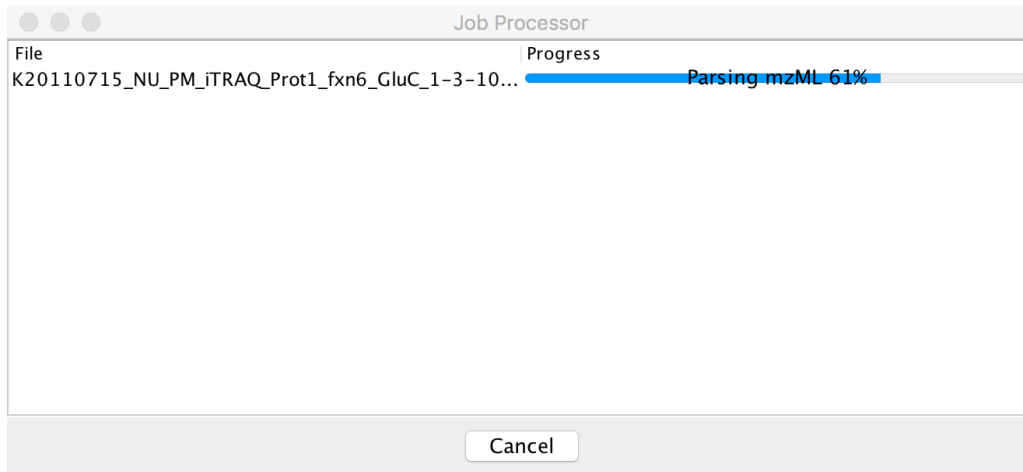
SETTING UP A CROSSTALK MATRIX

After queueing files for analysis, Crosstraq asks for a crosstalk matrix. While this matrix was specified with your iTRAQ or TMT reagents, the vendors for those reagents have mostly settled on reagent synthesis procedures so these matrices rarely change. Crosstraq has default matrices for iTRAQ-4plex, iTRAQ-8plex, TMT-6plex, TMT-10plex, and TMT-11plex based on previous batches of reagents. In addition to crosstalk compensation, Crosstraq also provides precursor impurity correction, which is useful if your iTRAQ or TMT data were not collected using MS3-capable workflows.

	-2	-1	+1	+2
iTr-114	0	0.01	0.059	0.002
iTr-115	0	0.02	0.056	0.001
iTr-116	0	0.03	0.045	0.001
iTr-117	0.001	0.04	0.035	0.001

Also apply precursor impurity correction (not for MS3)

After hitting “OK”, Crosstraq will immediately start processing your jobs in sequence. For example:



After Crosstraq is finished, it will create two files in the directory where the mzML was. First, Crosstraq will create "FILENAME_corrected.mgf", which can be used in place of the search engine MGF for loading into downstream quantitative analysis software, such as Scaffold or IDPicker. Crosstraq will also create "FILENAME.txt", which will contain tab-delimited statistics on every peptide-spectrum-match in the search, including the top peptide/protein match, the precursor intensity in the nearest precursor MS1 scan, the calculated precursor purity, and the normalized intensities for each reporter-ion channel. This file looks like this in Excel:

title	peptide	accessions	log10Intensity	purityRatio	iTr-114	iTr-115	iTr-116	iTr-117
controllerType=0 controllerNumber=1 scan=15290	RLGDPKE	sp Q9NP73	7.145914621	0.6724002	0.28278908	0.2775821	0.4381753	0.00145353
controllerType=0 controllerNumber=1 scan=11720	LRRVVAALPE	sp A4D2H0	7.142265856	0.45260495	0.21110964	0.17066613	0.6165946	0.00162968
controllerType=0 controllerNumber=1 scan=15198	FIMKYKRG	sp Q96M93	6.743581253	0.37925556	0.2197777	0.24041669	0.5377044	0.00210122
controllerType=0 controllerNumber=1 scan=15141	GLEKPOSKLE	sp Q8WXI8	6.873533511	0.47757322	0.40660006	0.21415412	0.3774998	0.00174609
controllerType=0 controllerNumber=1 scan=8035	MVAFYQE	sp Q8IV45	7.690481982	0.7623881	0.2965673	0.31098598	0.39113212	0.00131464
controllerType=0 controllerNumber=1 scan=26008	GTLNIRE	sp Q5VIR6	4.392639119	0.25	0.29543766	0.32949913	0.3712975	0.0037658
controllerType=0 controllerNumber=1 scan=4029	RGLVRSVTGE	sp Q60I27	5.790589882	0.53034073	0.03764013	0.13696489	0.8243023	0.00109273
controllerType=0 controllerNumber=1 scan=8590	LMCQIE	sp Q8IWJ2	8.180243767	0.573185	0.3205576	0.25610992	0.4217218	0.00161067
controllerType=0 controllerNumber=1 scan=19567	SGEKKVE	sp Q8NFI5	6.295768011	0.56568986	0.29038206	0.2882208	0.41969386	0.00170335
controllerType=0 controllerNumber=1 scan=7482	LLLTLE	sp O94915	7.077342864	0.38912132	0.36155352	0.29560786	0.34037223	0.00246641
controllerType=0 controllerNumber=1 scan=18717	VLKQRIAE	sp Q9Y247	6.664073355	0.53591377	0.1375796	0.26585093	0.5951904	0.00137907