1 Общая информация о пайплайне

Пайплайн для xQuest/xProphet написан на языке snakemake, являющимся расширением языка Python. Snakemake отслеживает зависимость между этапами пайплайна через граф. Если какой-либо из используемых файлов был изменен (например, конфиг), то все последующие этапы пайплайна, зависящие от этого файла, переделываются.

2 Установка

1. Клонируем репозиторий `https://github.com/noel-noel/xquest-pipeline`:

   ```
   git clone https://github.com/noel-noel/xquest-pipeline
   ```

2. В директории `xquest-pipeline` из терминала запускаем

   ```
   chmod +x *.sh
   ./install.sh
   ```

3 Анализ

1. XQuest использует конфигурационные файлы для задания параметров эксперимента и анализа. Создадим в произвольной директории шаблонные конфиги, вызвав из терминала

   ```
   /usr/local/share/xquest/deffiles/create_configs.sh
   ```

2. В любом текстовом редакторе изменяем конфиги под себя. На сервере можно воспользоваться `nano` или `vim`. В разделе 4 приведены наиболее важные параметры.

3. В любом текстовом редакторе открываем Snakefile. Необходимо задать **все** параметры из секции CONFIGURATION:

   - `proj_name` – название проекта.
     Например, `proj_name = "ph-analysis-sample-1"`.
   - `xquest_def_path` – абсолютный путь файла `xquest.def` (см. п.1.).
     Например, `xquest_def_path = "/home/noel/xquest.def"`.
   - `xmm_def_path` – абсолютный путь файла `xmm.def` (см. п.1.).
     Например, `xmm_def_path = "/home/noel/xmm.def"` .
   - `xprophet_def_path` – абсолютный путь файла `xproph.def` (см. п.1.).
     Например, `xprophet_def_path = "/home/noel/xproph.def"`.
   - `raw_path` – абсолютный путь директории, в которой хранятся сырые файлы (.raw).
     Например, `raw_path = "/home/noel/Documents/RAW/"`.
     Комментарий 1: **запрещено** использовать системные директории в качестве `raw_path`.
     Можно: `"/home/user/", "/home/user/dir/", "/home/user/dir1/.../dirn/"`.
     Нельзя: `"/usr/lib/", "/mnt/data/", "/home/"`.
     Комментарий 2: Все файлы .raw, содержащиеся в директории `raw_path` будут использоваться для анализа.
   - Запускаем пайплайн анализа, вызвав:

     ```
     snakemake all
     ```

     Комментарий 1: можно посмотреть количество запланированных snakemake задач:

     ```
     snakemake -n all
     ```

     Комментарий 2: можно распараллелить пайплайн, указав число потоков (e.g. 10):

     ```
     snakemake -j 10 all
     ```

4 Конфигурационные файлы

4.1 xmm.def

TABLE 1 | Important parameters of xmm.def when used with isotopically labeled cross-linkers or unlabeled cross-linkers.

Parameter	Value type	Example value	Description
Parameters for isotopically labeled cross-linkers			
Isopair_Mr_tolerance_measure	String	p.p.m.	a.m.u.|p.p.m., tolerance measure a.m.u. corresponds to atomic mass units (Da)
Isopair_Mr_tolerance	Number in unit	15	Allowed tolerance for scan pairs in tolerance measure units
Isopair_Tr_tolerance	Number in unit	3.0	Maximum relative retention time tolerance of light and heavy scans (in minutes)
Isotopeshift	Mass	12.075321	Isotope shift or the cross-linker (always in Da)
Printisotopicscanpairs	Integer	1	Use this to print the isotopic scan pairs, which are then used by the subsequent programs
Parameters that need to be adjusted if an unlabeled cross-linker is used (light-only search)			
Isopair_Mr_tolerance	Mass	0	
Isopair_Tr_tolerance	Number in unit	0.02	
Isotopeshift	Mass	0	
Printisotopicscanpairs	Integer	0	
Printlightonlypairs	Integer	1	Prints only the light scans

4.2 xquest.def

TABLE 2 | Important xQuest parameters defined in xquest.def.

Parameter	Value type	Example value	Description
Spectrum comparison parameters (compare_peaks3.pl)			
cp_threshold	Integer	1	Minimum peak intensity
cp_tolerancemeasure	String	Da	(Da|p.p.m.) Tolerance measure in Da or p.p.m.
cp_tolerance	Mass	0.2	Allowed tolerance in Da or p.p.m. for peak matching of common ions
cp_tolerancexl	Mass	0.3	Allowed tolerance in Da or p.p.m. for peak matching of cross-linker ions
cp_peakratio	Ratio	0.3	Required peak-intensity ratio lower peak/higher peak to be considered the same peak
cp_isotopediff	Mass	12.075321	Isotope difference of light and heavy cross-linker in Da
cp_nhighest	Integer	100	Use only *n* highest peaks after spectrum comparison
cp_minpeaknumber	Integer	25	Minimum number of common ions to be considered as a spectrum for the search
xQuest search modes and database parameters			
Iontagmode	Integer	1	1|0: If 1, then the ion-tag mode is used
Enumerate	Integer	0	1|0: If 1, then the enumeration mode is used
RuntimeDecoys	Integer	1	1|0: If 1, search in target-decoy mode used in combination with the ion-tag mode

xQuest database definitions			
Database	Path	/path/…	Path to the target database (please do not use whitespaces in the paths)
database_dc	Path	/path/…	Path to the decoy database **don't change!**
General xQuest search parameters			
enzyme_num	Integer	1	Enzyme to be used for the *in silico* digest of the database. See enzyme reference at the end of the definition file
missed_cleavages	Integer	2	Maximum number of allowed missed cleavages. (recommended minimum 2)
Mindigestlength	Integer	5	Minimal digest length for a peptide to be considered for the search
Maxdigestlength	Integer	50	Maximal digest length for a peptide to be considered for the search
Nocutatxlink	Integer	1	1|0: If 1, then positions that correspond to the enzyme cleavage site are excluded as a potential cross-linking site
Ionseries	Sequence	010010	1|0 corresponding to the sequence abcxyz. Ion series that should be considered for the search. e.g., 010010 # abcxyz, considers b- and y-ions
ioncharge_common	Integers separated by ','	1,2,3	Charge states to be considered for common fragment ions. e.g., 1,2,3
ioncharge_xlink	Integers separated by ','	2,3,4,5	Charge states to be considered for cross-linker containing fragment ions. e.g., 2,3,4,5
Cross-linker–specific search parameters			
AArequired	Amino acid letter separated by ','	K	Amino acid that is cross-linked. For more than one cross-linkable amino acid separated by comma, e.g., K,S
Xlinkermw	Mass	138.0680796	Net mass shift for cross-linked peptides caused by the cross-linker in Da
Monolinkmw	Mass	156.0786442, 155.0964278	Net mass shift for mono-links caused by the cross-linker in Da. Multiple masses can be separated by comma
xQuest search tolerances			
Tolerancemeasure	String	ppm	(Da|p.p.m.) MS1 tolerance measure in Da or p.p.m.
ms1tolerance	Number in unit	10	Tolerance for parent mass matching MS1
tolerancemeasure_ms2	String	Da	(Da|p.p.m.) MS2 tolerance measure in Da or p.p.m.
ms2tolerance	Number in unit	0.2	Tolerance for peak matching on MS2 for common fragment ions
xlink_ms2tolerance	Number in unit	0.3	Tolerance for MS2 matching for cross-linker containing fragment ions
CID_match2ndisotope	Integer	1	1|0, 1 allows matching of 2nd isotope
Minionsize	*m/z*	200	Minimum ion size in MS2 mode to be considered in *m/z*
Maxionsize	*m/z*	2000	Maximum ion size in MS2 mode to be considered in *m/z*
xQuest modification settings			
variable_mod	Amino acid, mass	M,15.99491	0|mod: Defines a variable modification for a particular amino acid, separated by comma
nvariable_mod	Integer	1	Maximum number of variable modifications per peptide
modifications fixed	Amino acid, mass	C, 57.02146	Defines fixed modification for a particular amino acid, separated by tab
Parameters that need to be adjusted if an unlabeled cross-linker is used (light-only search)			
cp_isotopediff	Mass	0	

4.3 xproph.def

TABLE 3 | Important xProphet parameters defined in xproph.def.

Parameter	Value type	Example value	Description
Minborder	Number in unit	−10	Filter for minimum precursor mass error in precursor mass tolerance unit used for the xQuest search (usually p.p.m.)
Maxborder	Number in unit	10	Filter for maximum precursor mass error in precursor mass tolerance unit used for the xQuest search (usually p.p.m.)
Mindeltas	Number in unit	0.95	0 for not using this filter; filter for delta score, minimum delta score required, hits are filtered if larger or equal than the given value
Uniquexl	Integer	1	1|0: 1 calculate statistics based on unique IDs
Qtransform	Integer	1	1|0: 1 transform simple FDR to q-FDR values as described in ref. 48
Nidx	Integer	1	1|0: 1 reparse xml file