CLC Genomics Workbench Premiumを利用したメタトランスクリプトーム解析

フィルジェン株式会社　バイオインフォマティクス部
(biosupport@filgen.jp)

CLC Genomics Workbench について

次世代シークエンス解析用ソフトウェア

◆リシークエンシング解析
- リファレンスゲノムへのマッピング
- 変異検出

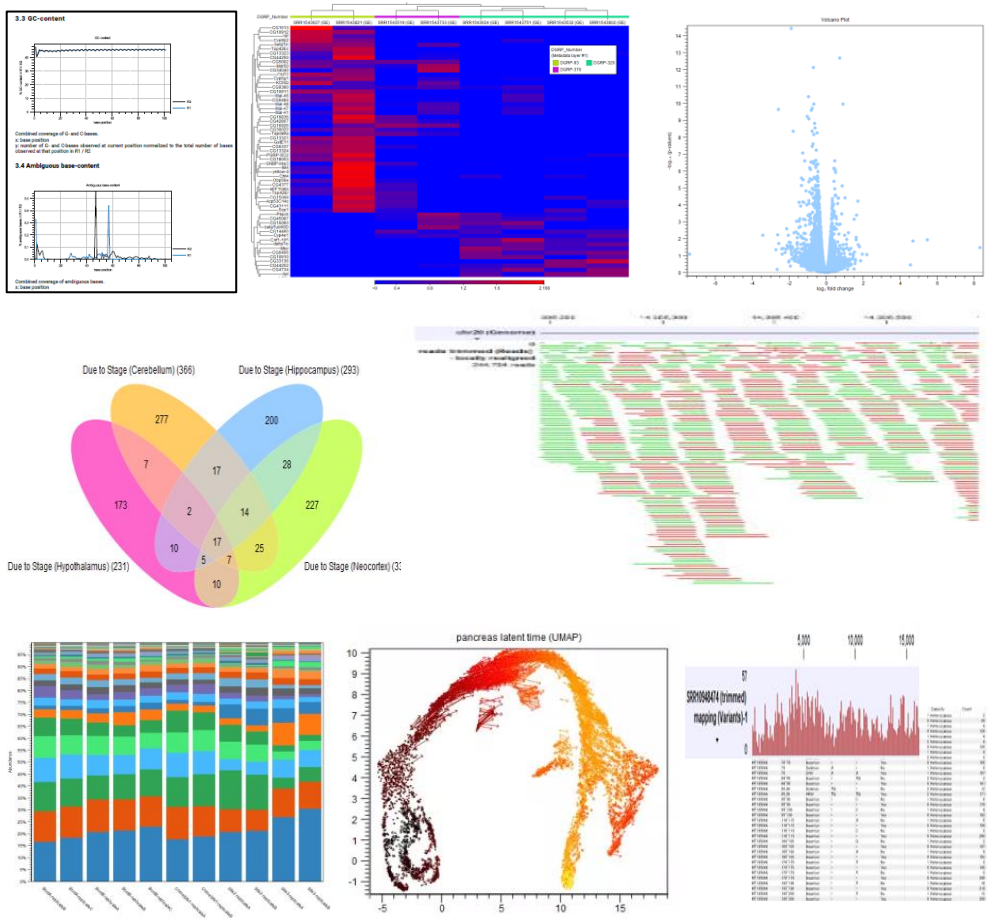
◆トランスクリプトミクス解析
- RNA-seq解析
- small RNA解析

◆エピゲノミクス解析
- ChIP-seq解析
- バイサルファイトシークエンス解析

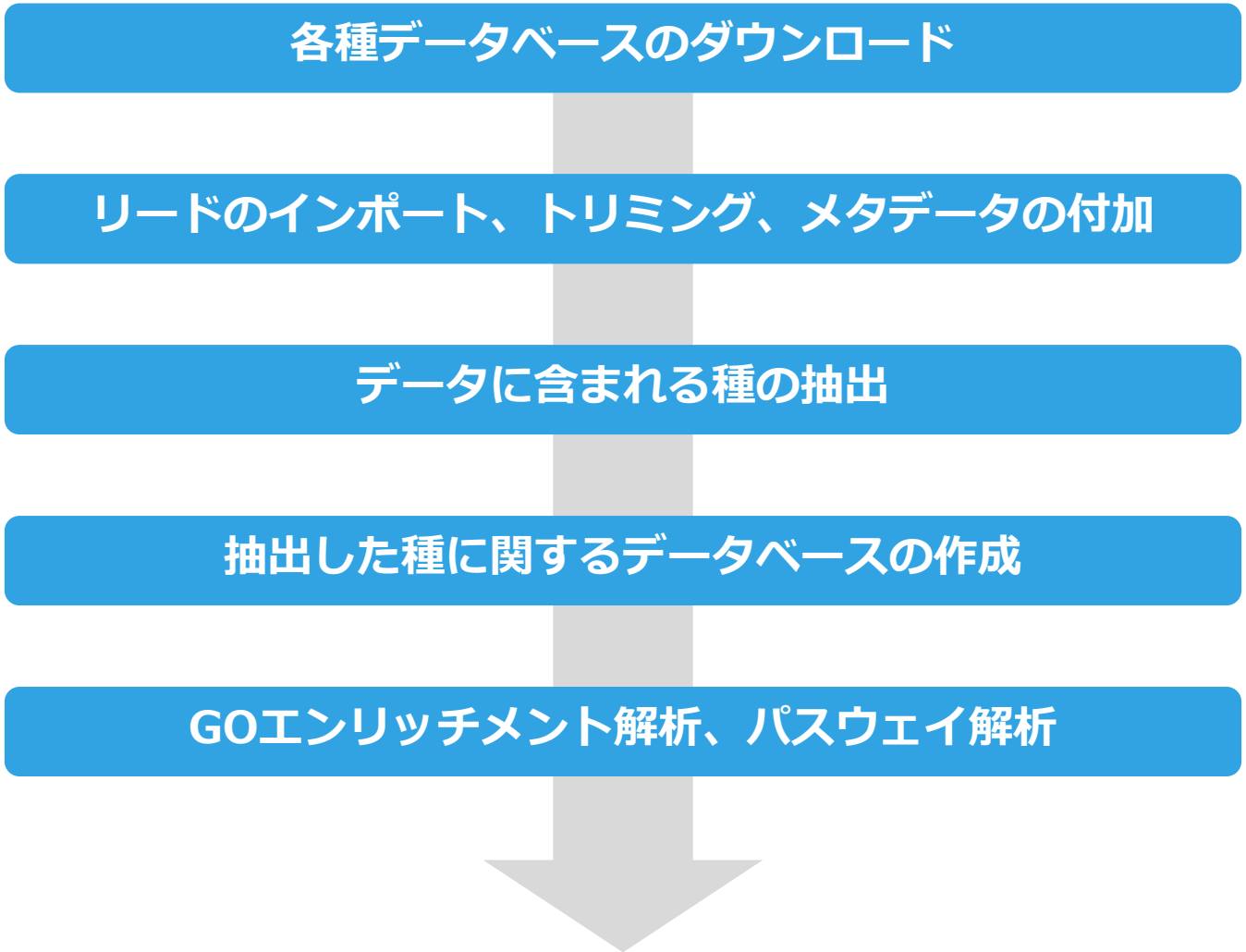
◆De Novo シークエンス解析
- De Novo Assembly
- BLAST解析

◆菌叢解析
◆ゲノムフィニッシング解析
◆シングルセル解析
◆超高速変異解析

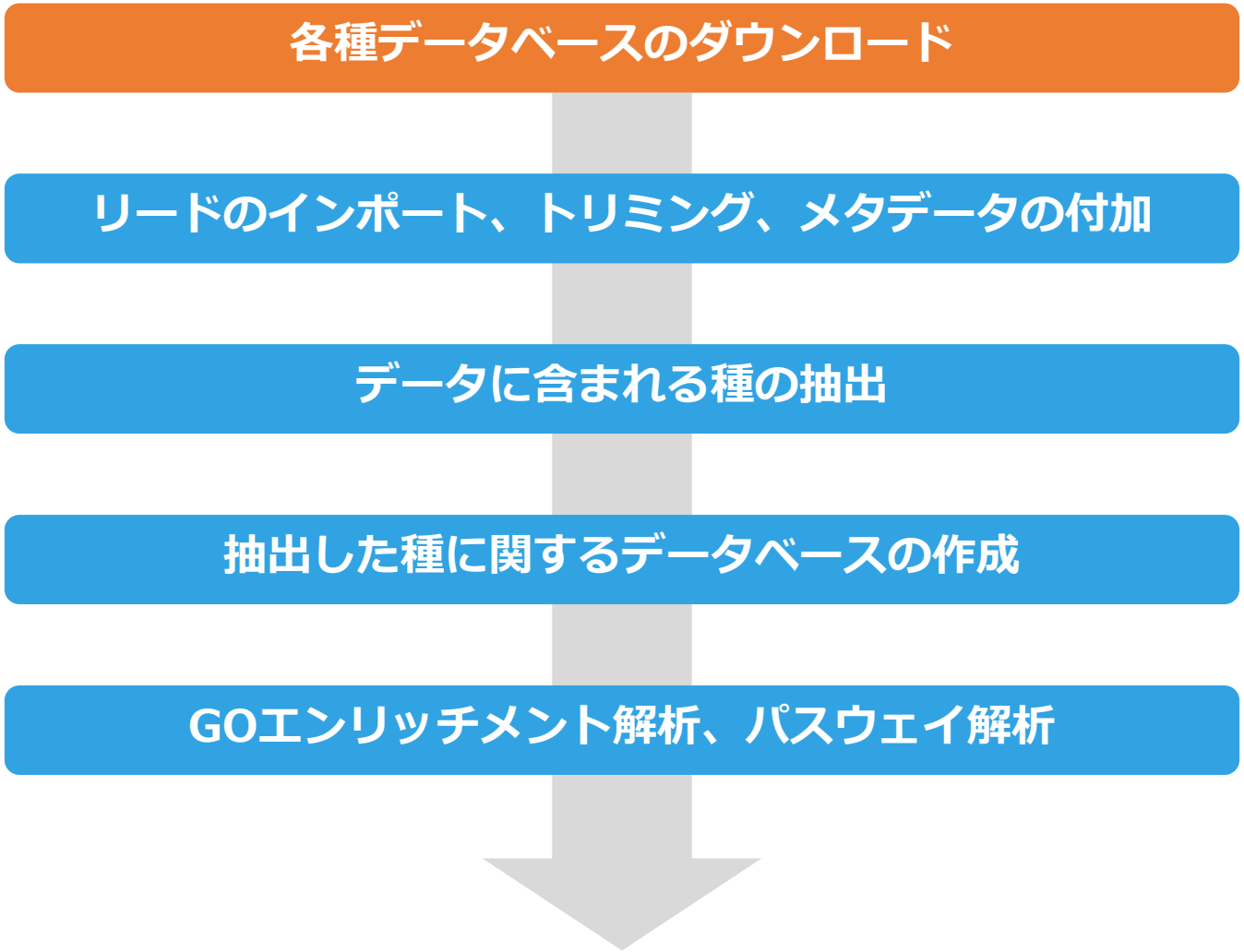
Premium版 限定機能



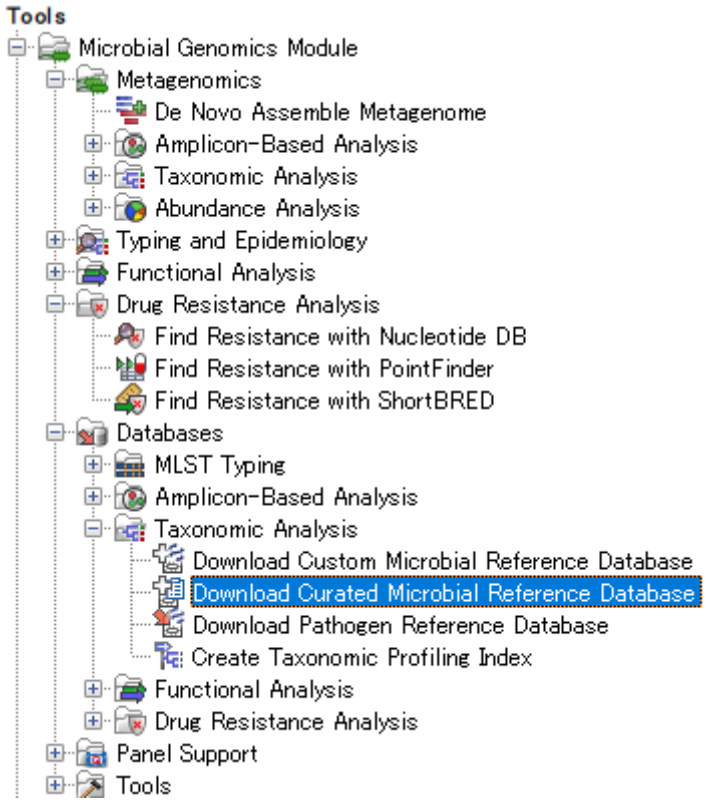
解析フロー

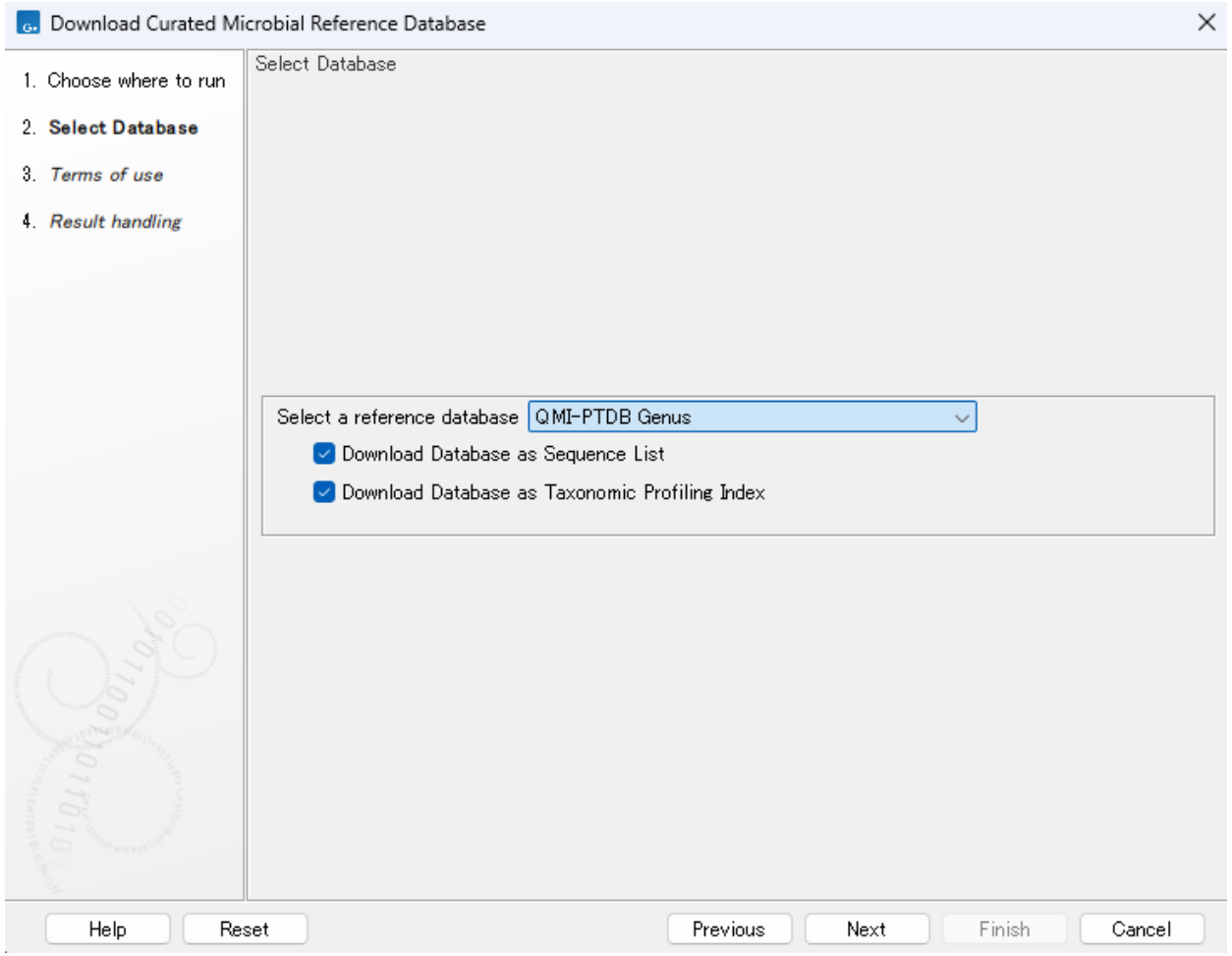

各種データベースのダウンロード
リードのインポート、トリミング、メタデータの付加
データに含まれる種の抽出
抽出した種に関するデータベースの作成
GOエンリッチメント解析、パスウェイ解析

解析フロー


各種データベースのダウンロード
リードのインポート、トリミング、メタデータの付加
データに含まれる種の抽出
抽出した種に関するデータベースの作成
GOエンリッチメント解析、パスウェイ解析

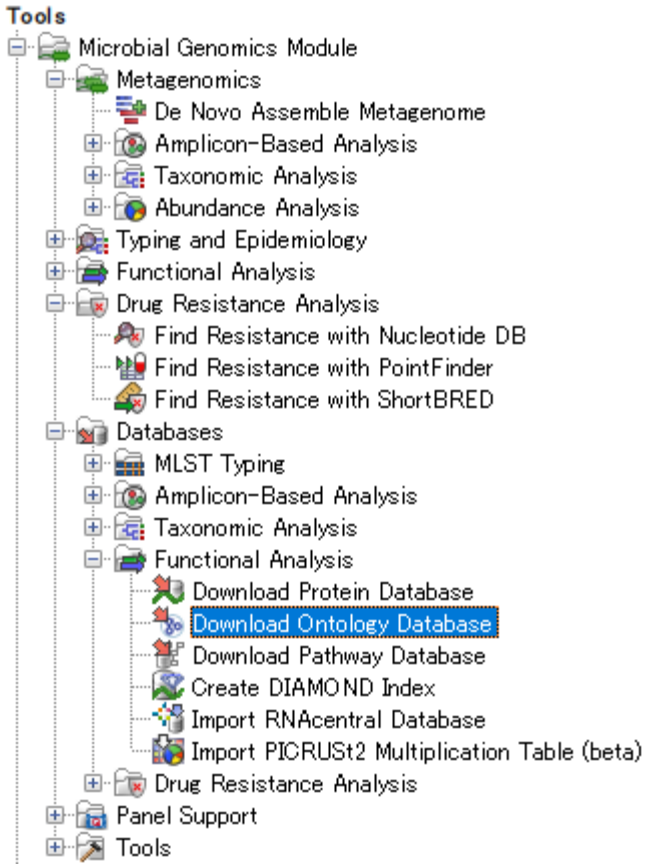
データベースのダウンロード（1/5）

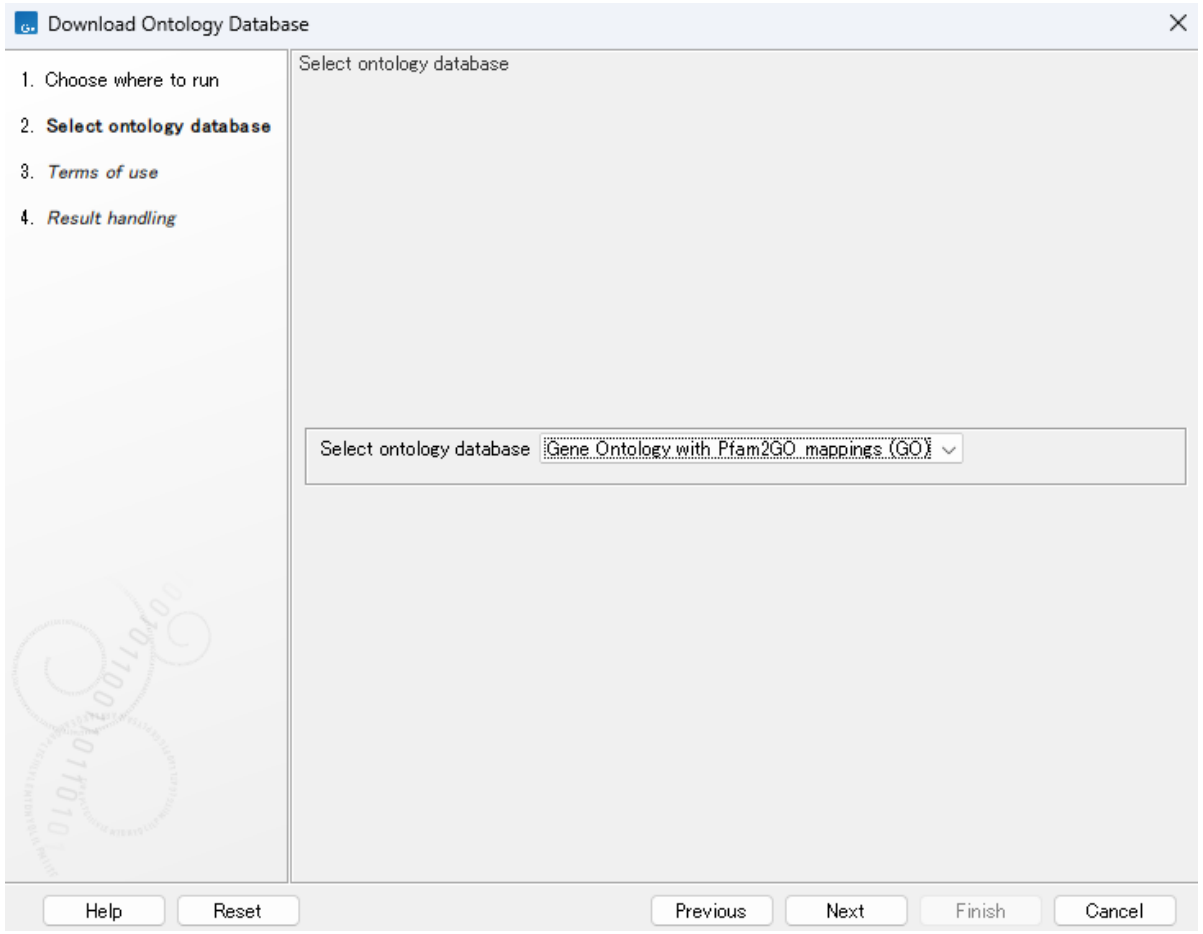




Download Curated Microbial Reference Databaseから、QMI-PTDBをダウンロードします。
※”Download Database as Taxonomic Profiling Index”にチェックをいれます。

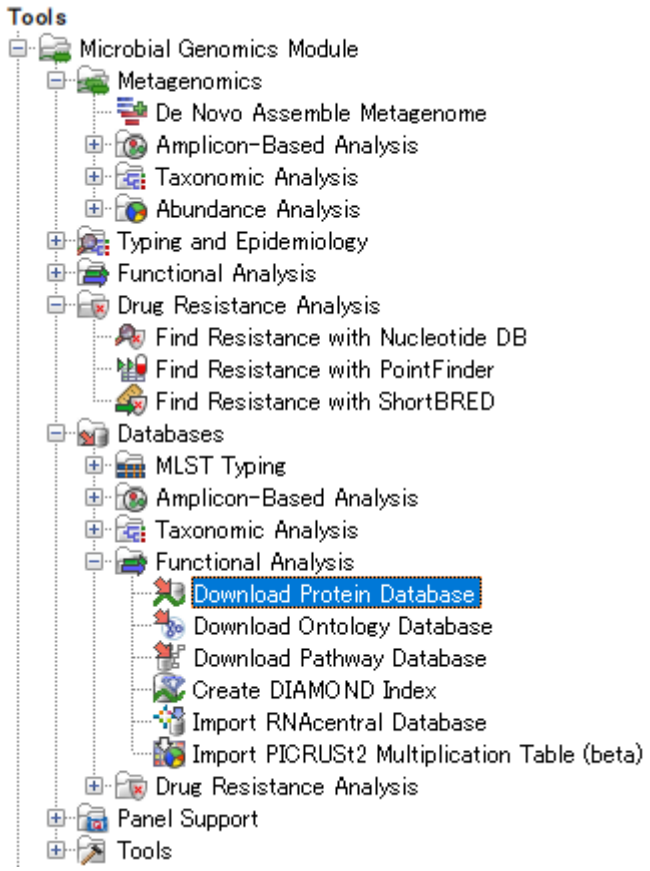
データベースのダウンロード（2/5）

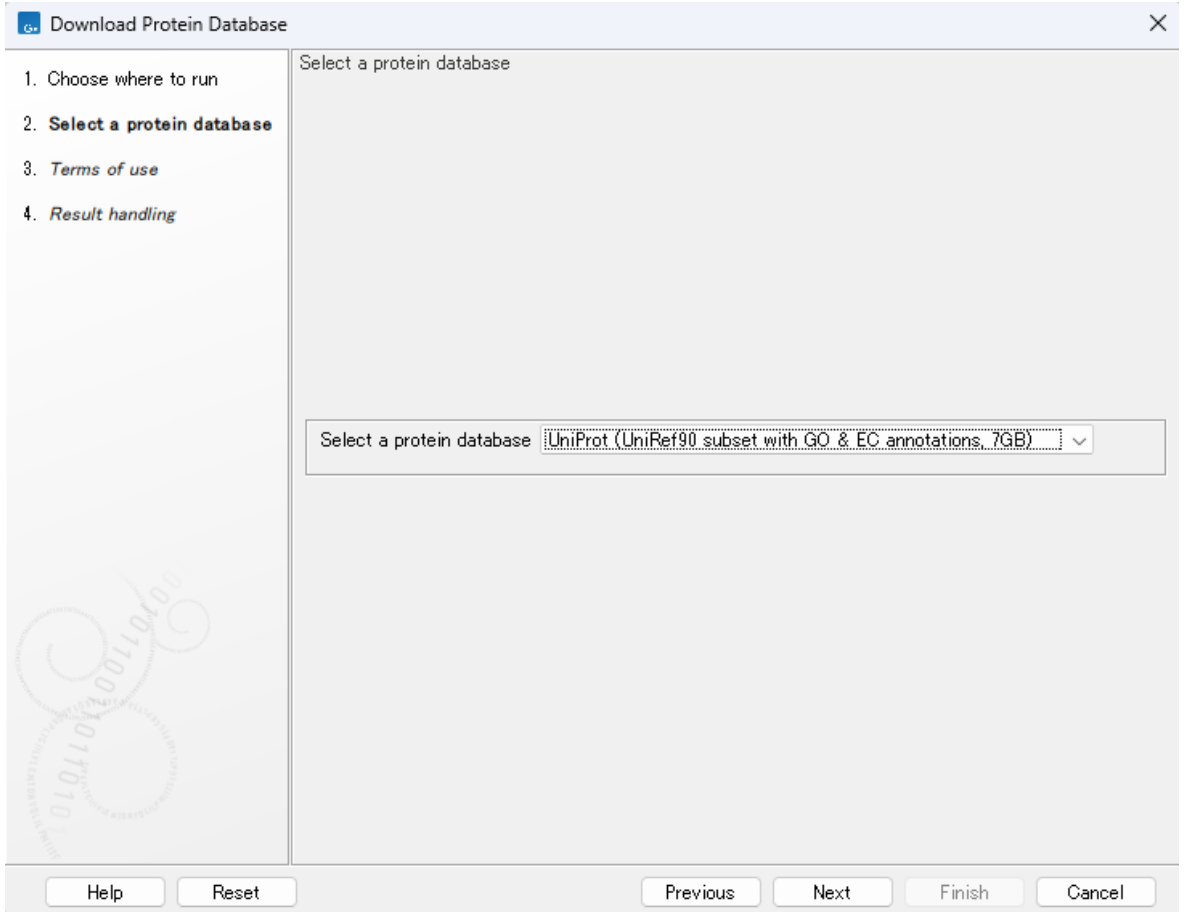




Download Ontology Databaseから、Gene Ontology with Pfam2GO mappingsをダウンロードします。

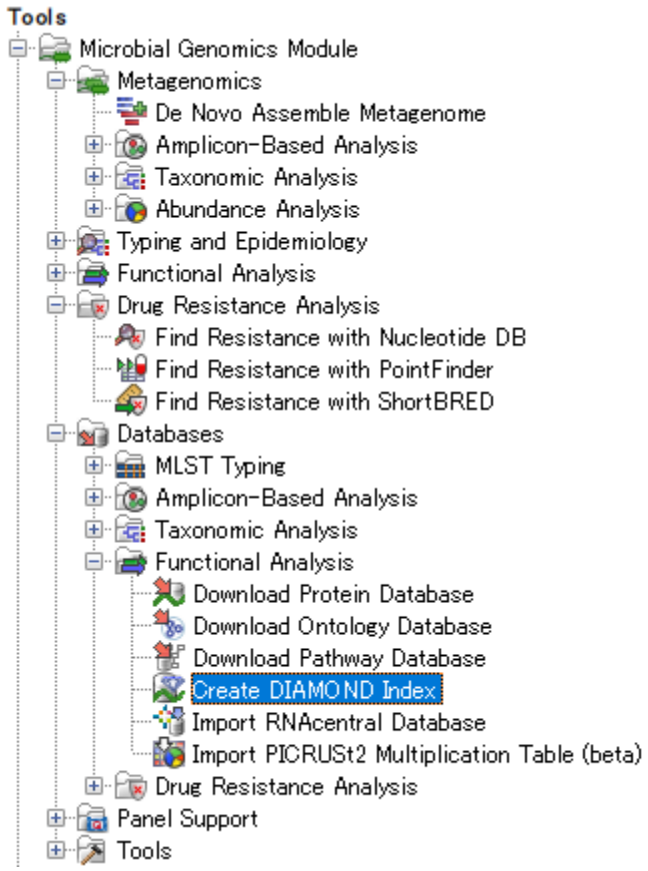
データベースのダウンロード（3/5）

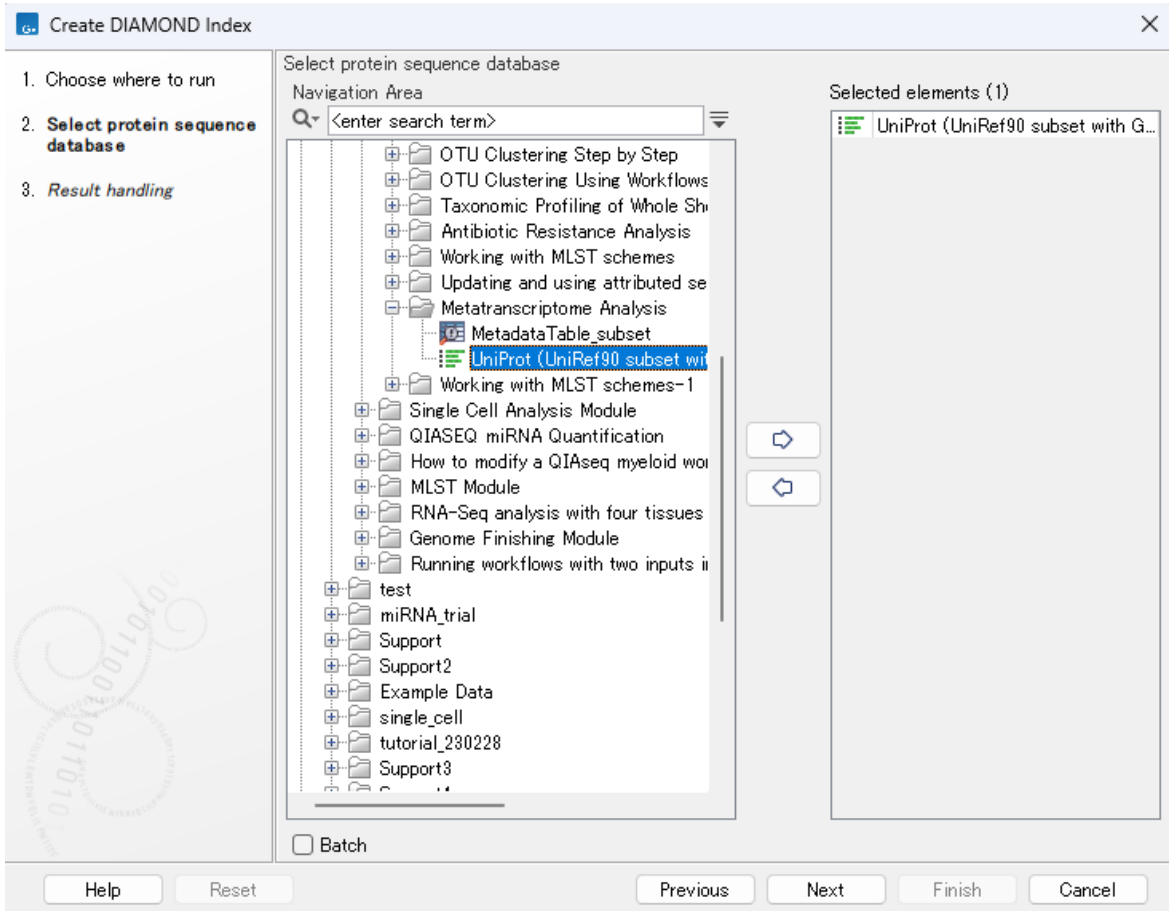




Download Protein Databaseから、UniProt90をダウンロードします。

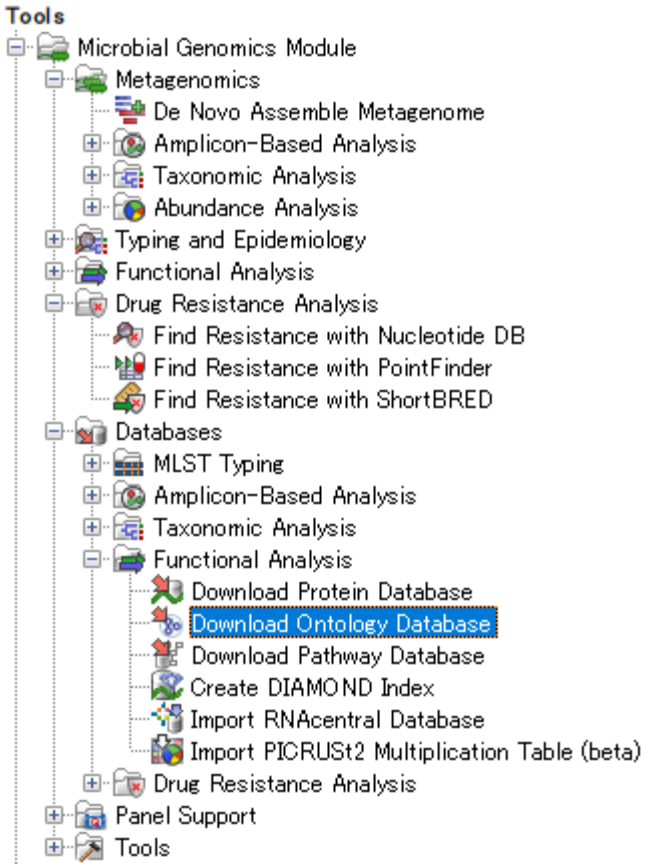
データベースのダウンロード（3/5）

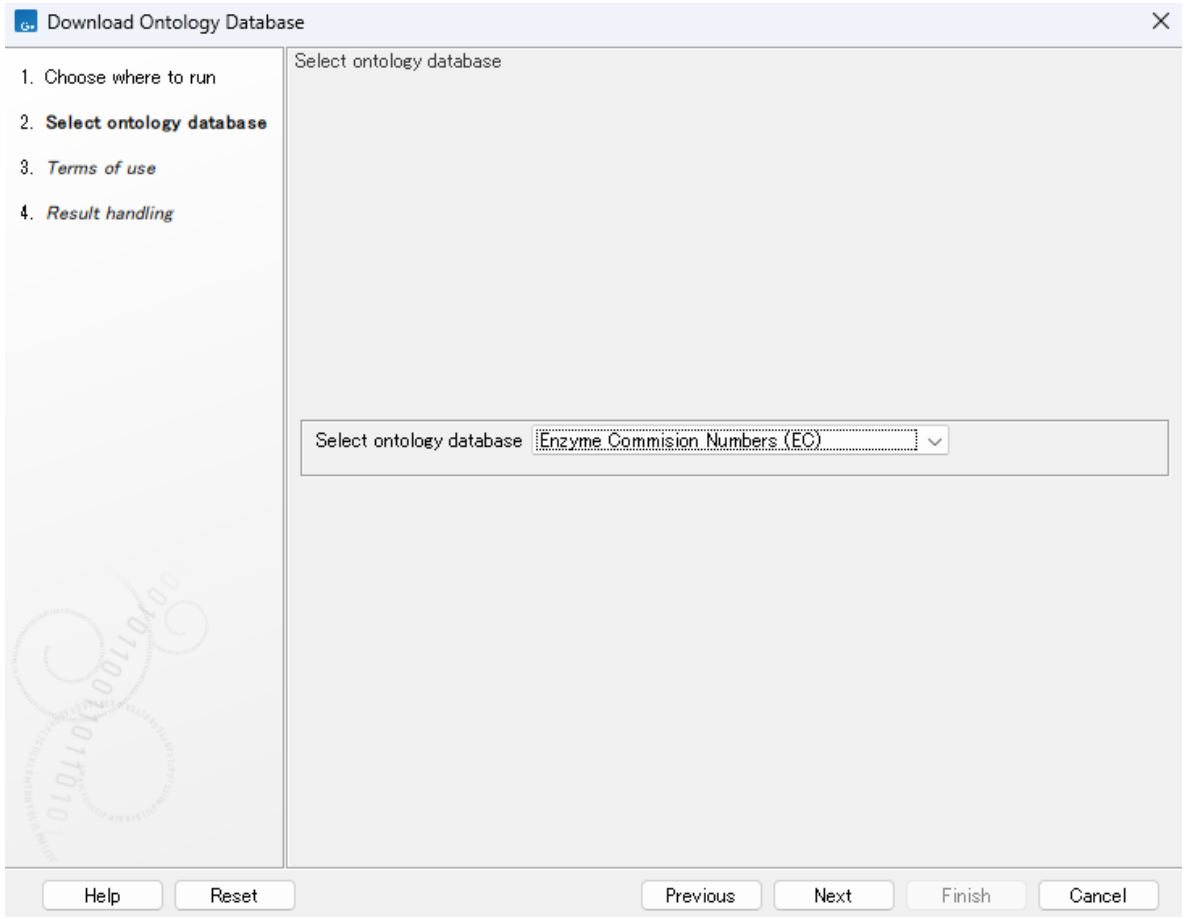




Create DIAMOND Indexで、今ダウンロードしたUniProt90を選択し、DAIMOND Indexを作成します。

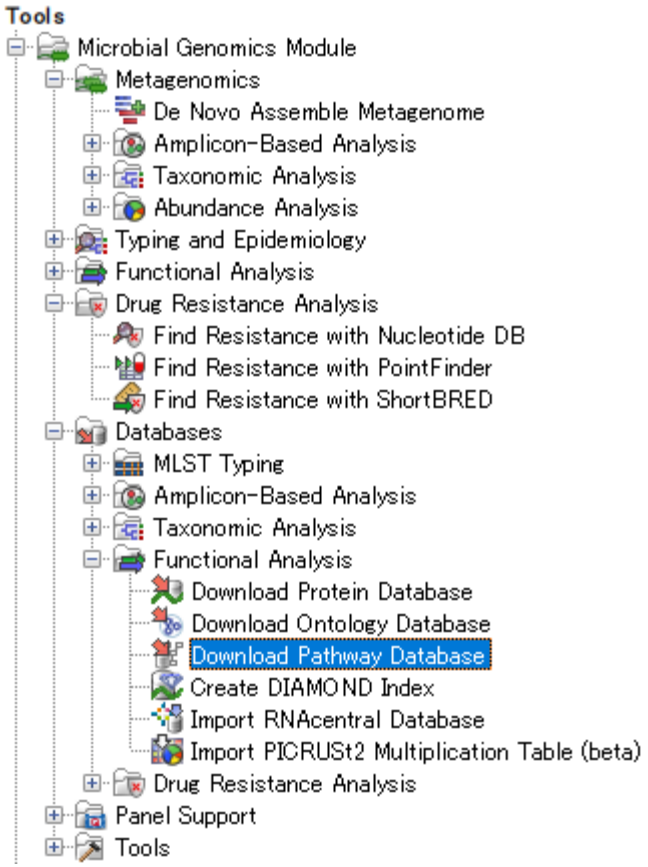
データベースのダウンロード（4/5）

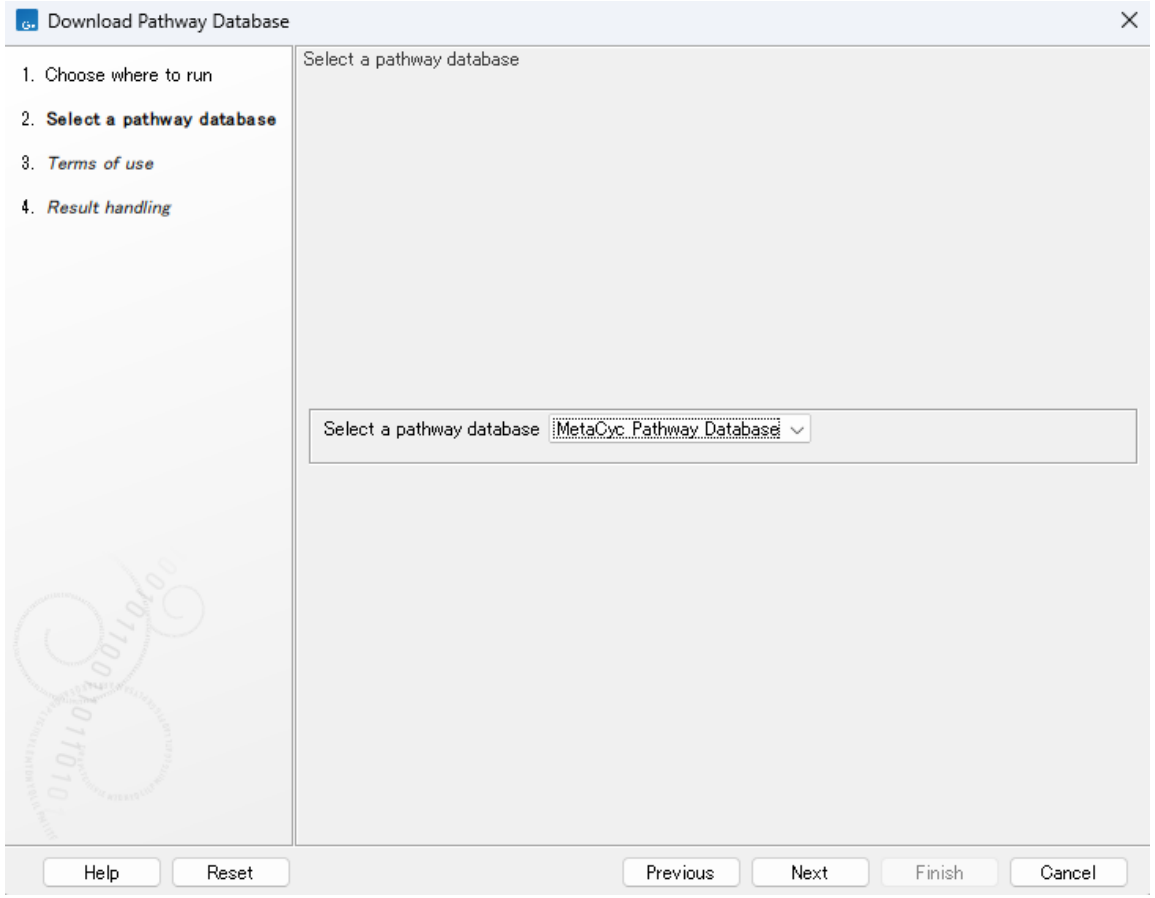




Download Ontology Databaseから、Enzyme Commission Numbersをダウンロードします。

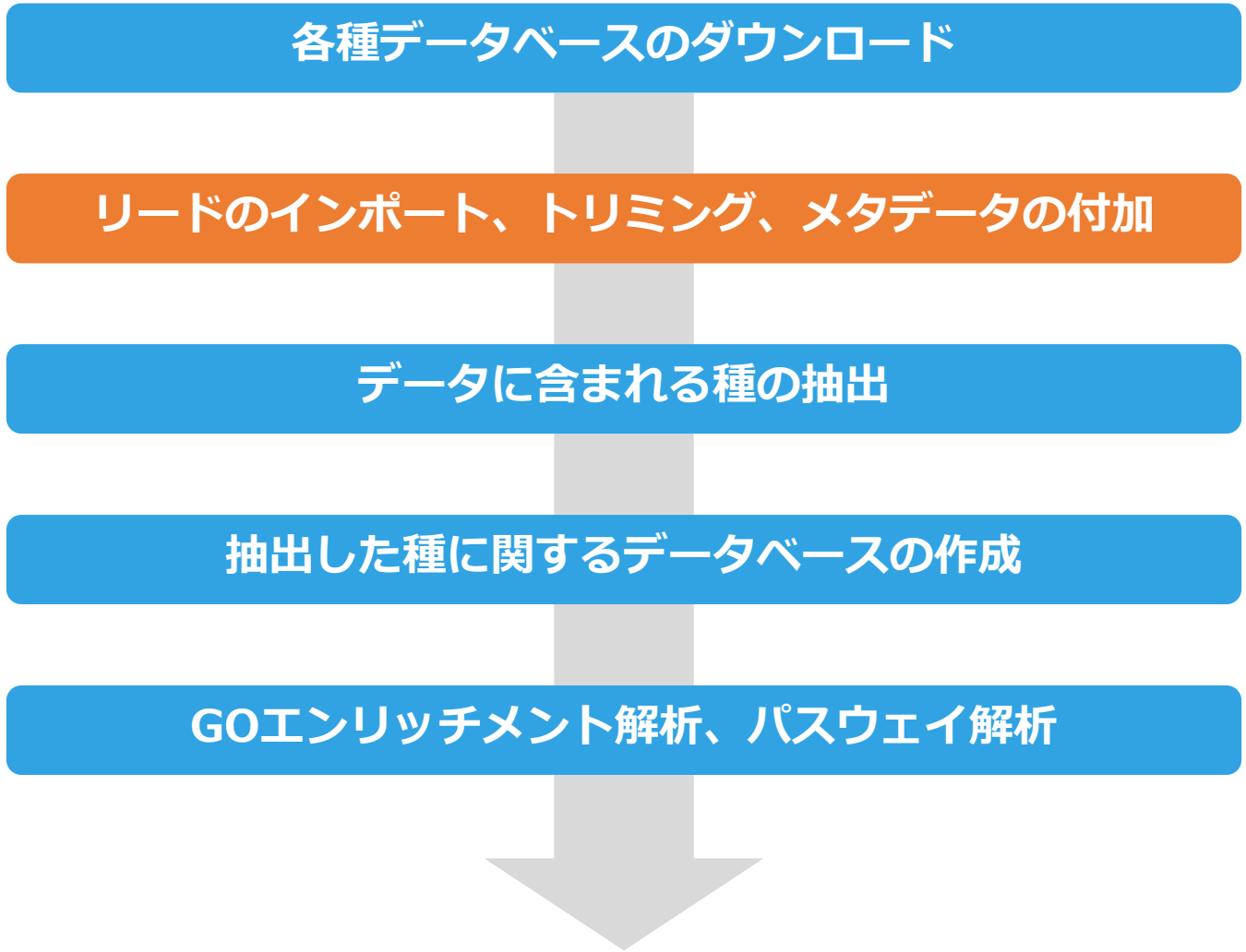
データベースのダウンロード（5/5）



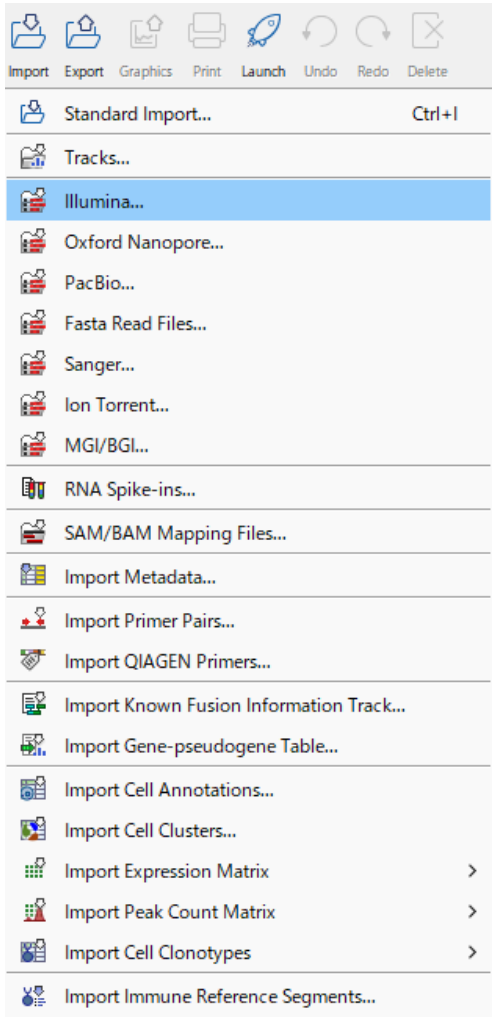


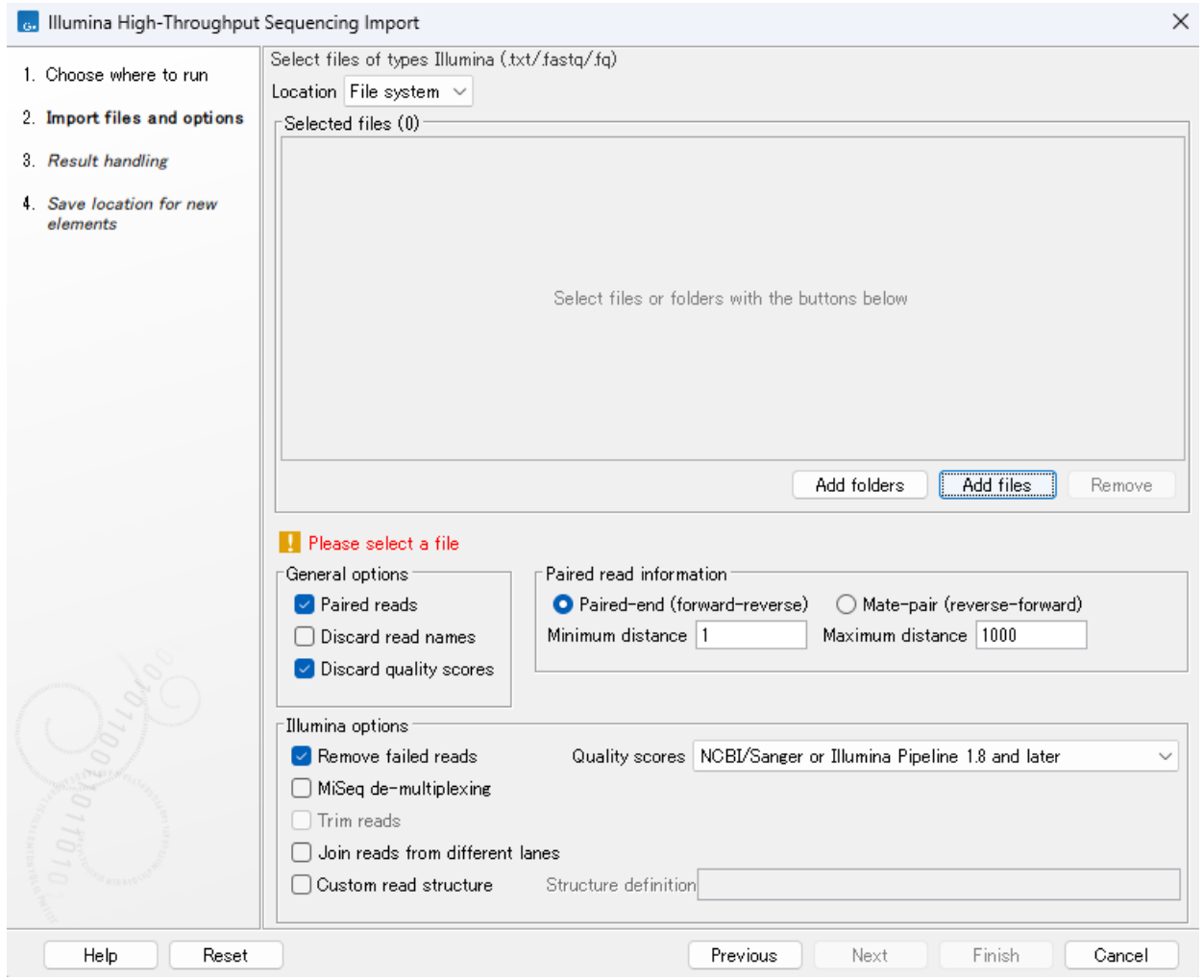
Download Pathway Databaseから、MetaCyc Pathway Databaseをダウンロードします。

解析フロー


各種データベースのダウンロード
リードのインポート、トリミング、メタデータの付加
データに含まれる種の抽出
抽出した種に関するデータベースの作成
GOエンリッチメント解析、パスウェイ解析

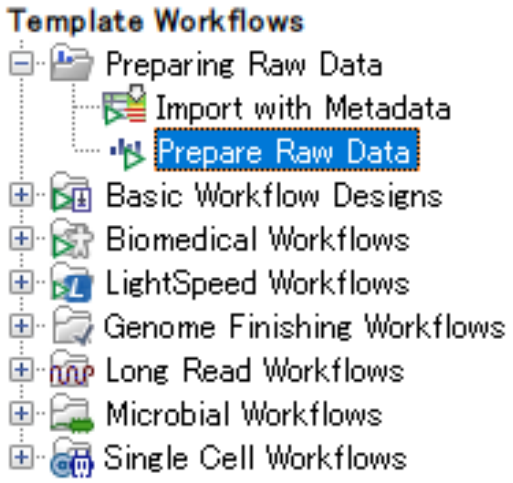
リードのインポート

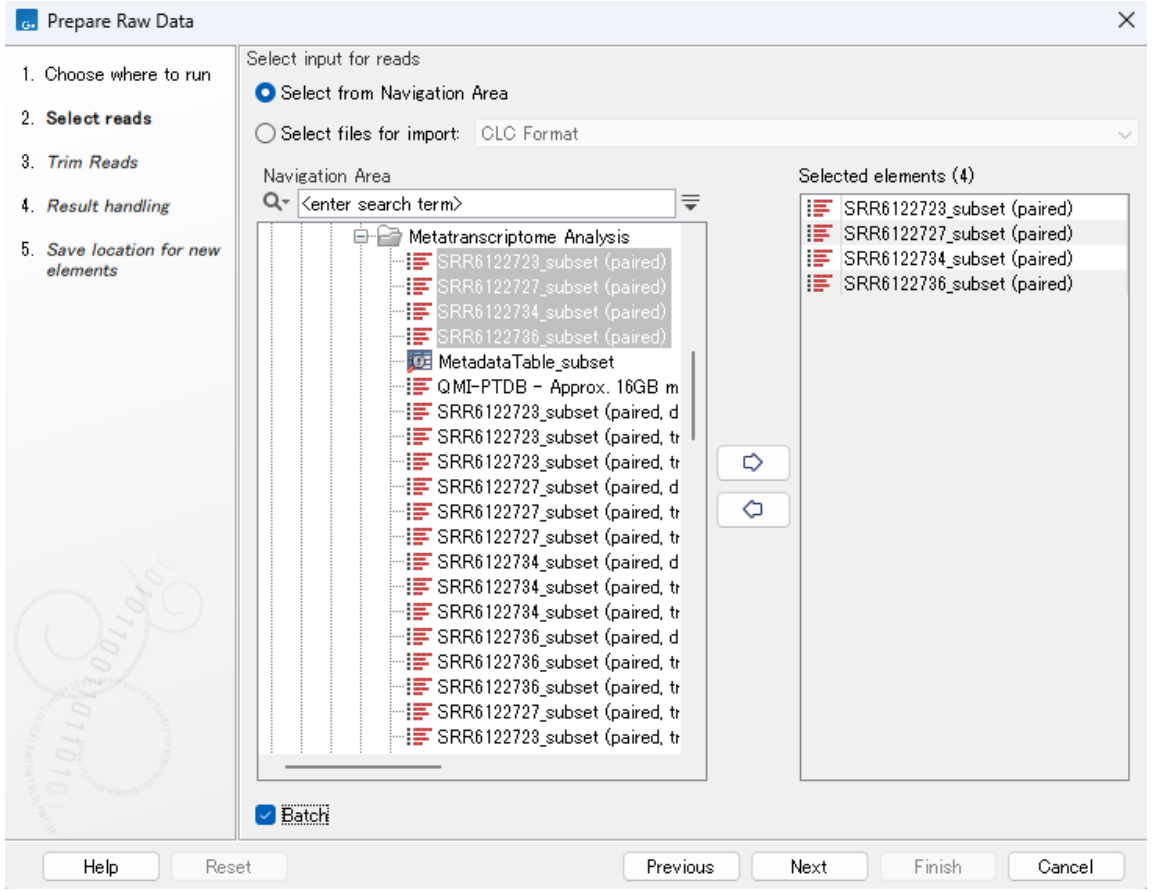




画面左上のImportメニューからシーケンスプラットフォームを選択し、Fastqファイルを選択します。

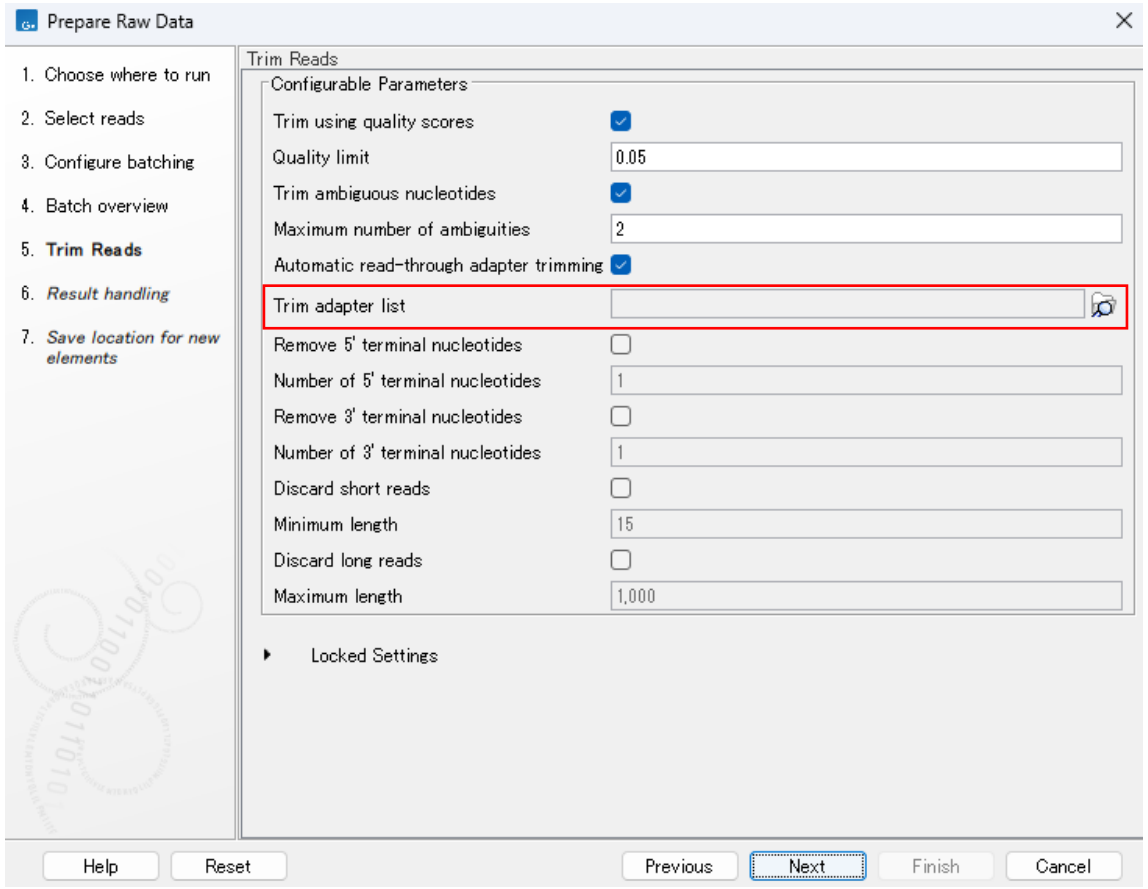
リードの前処理





Prepare Raw Dataワークフローを利用し、リードの前処理を行います。

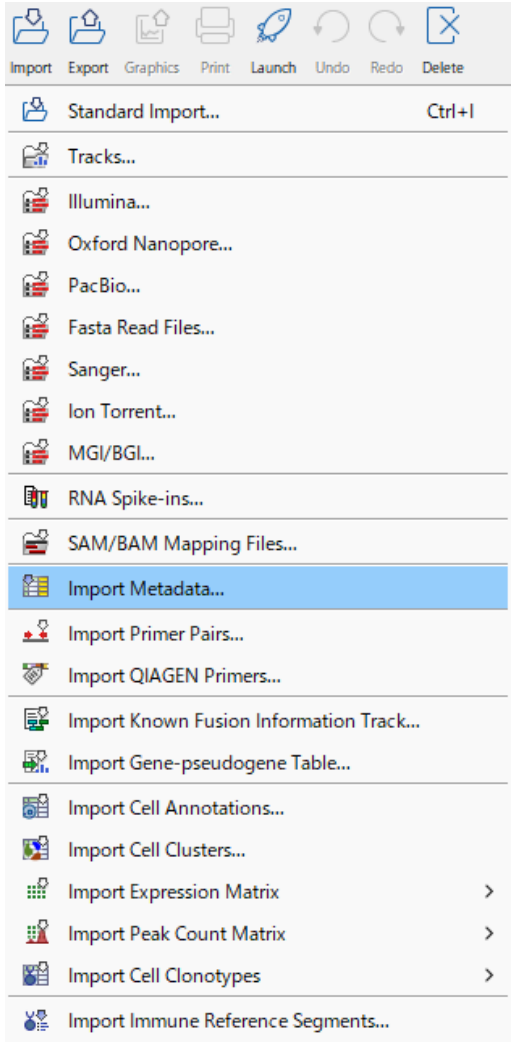
リードの前処理



アダプターのトリミングを行う場合、Trim adapter listでアダプターリストを指定します。
アダプターリストは画面左上のNew > Trim Adapter listで別途作成しておく必要があります。

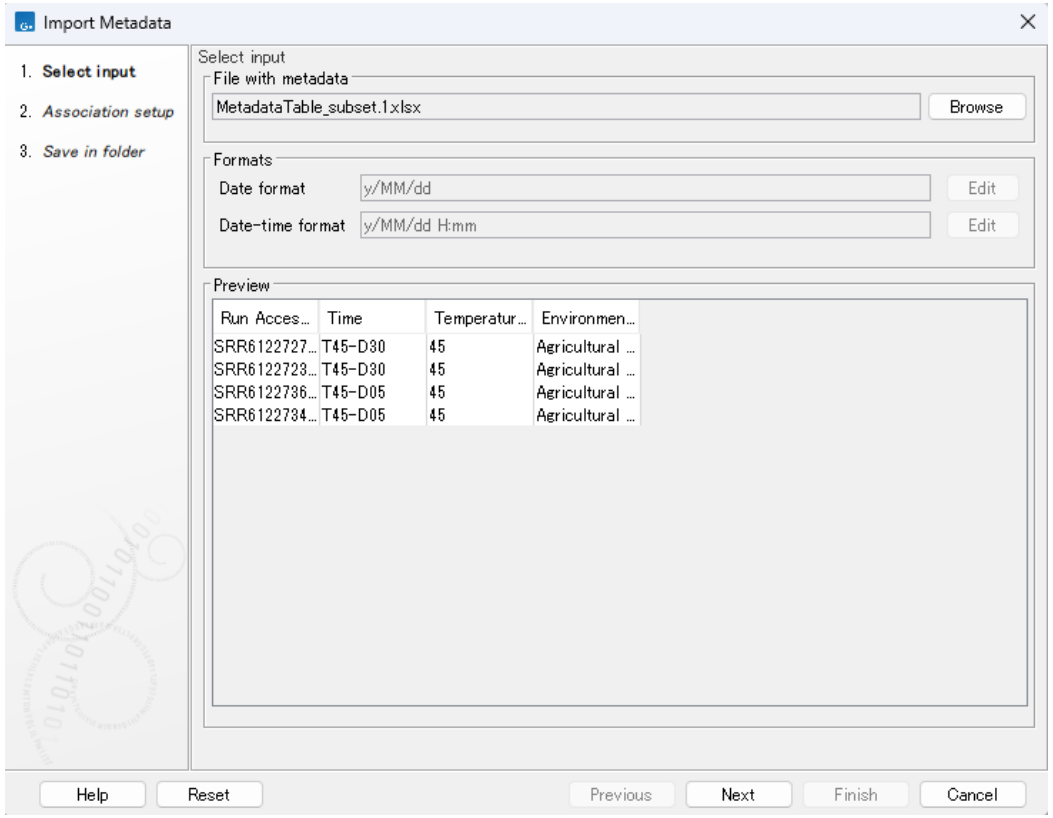
メタデータのインポート

	A	B	C	D	E	F
1	Run Accession	Time	Temperature	Environmental		
2	SRR6122727_subset	T45-D30	45	Agricultural paddy soil		
3	SRR6122723_subset	T45-D30	45	Agricultural paddy soil		
4	SRR6122736_subset	T45-D05	45	Agricultural paddy soil		
5	SRR6122734_subset	T45-D05	45	Agricultural paddy soil		

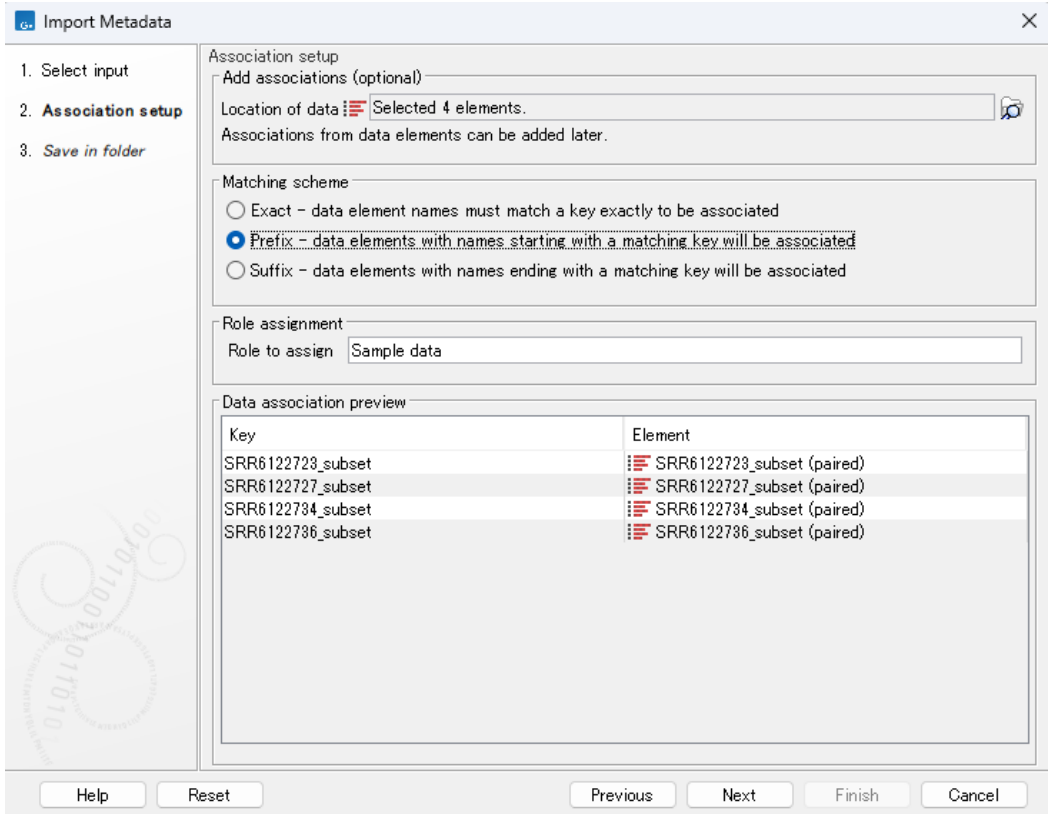


Excelなどで左のようなメタデータを用意し、ImportメニューのImport Metadataからインポートします。

メタデータのインポート

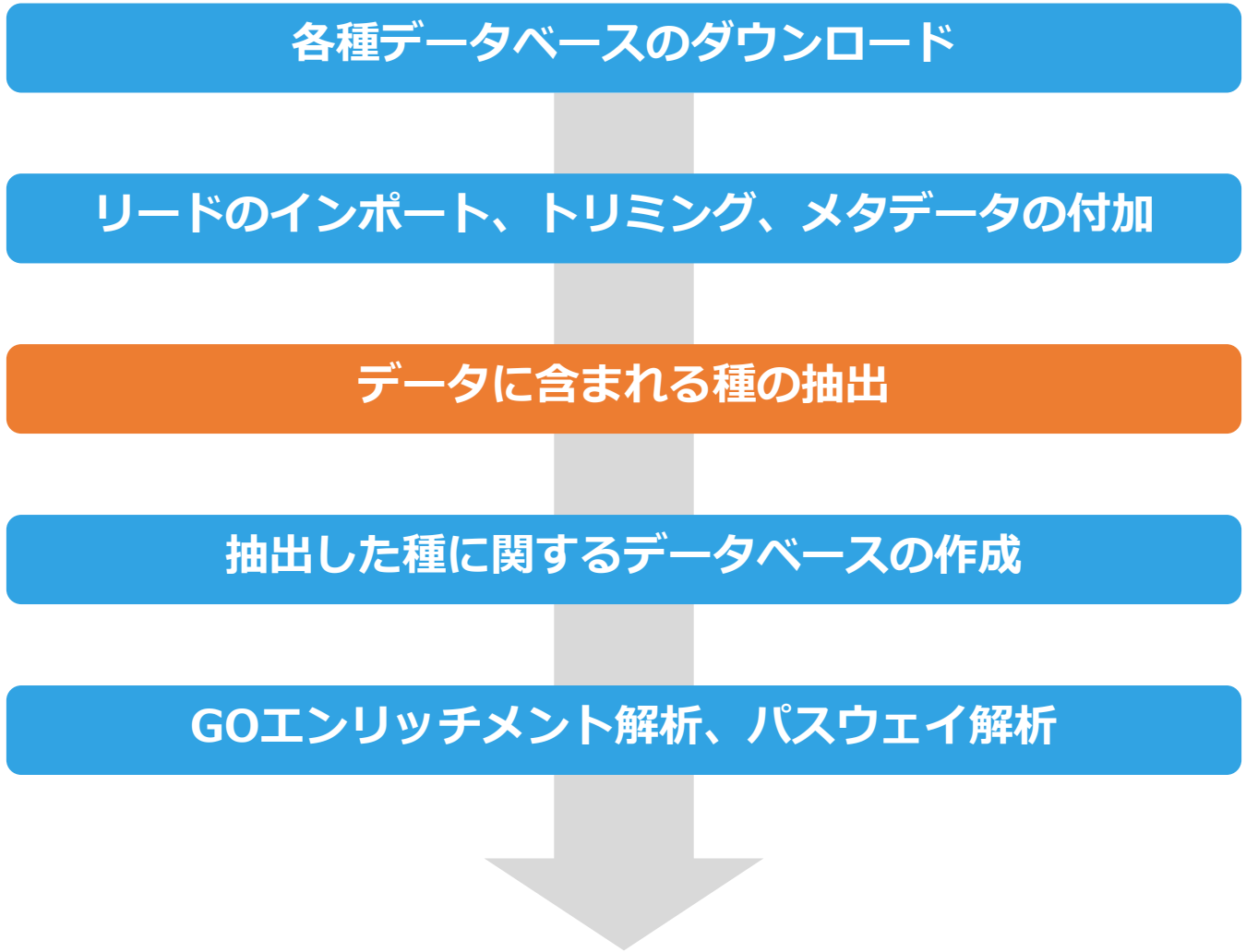


メタデータのファイルを指定します。

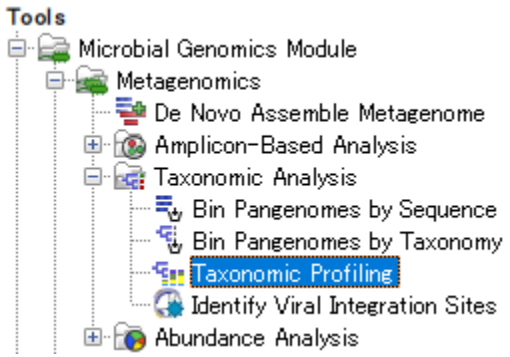


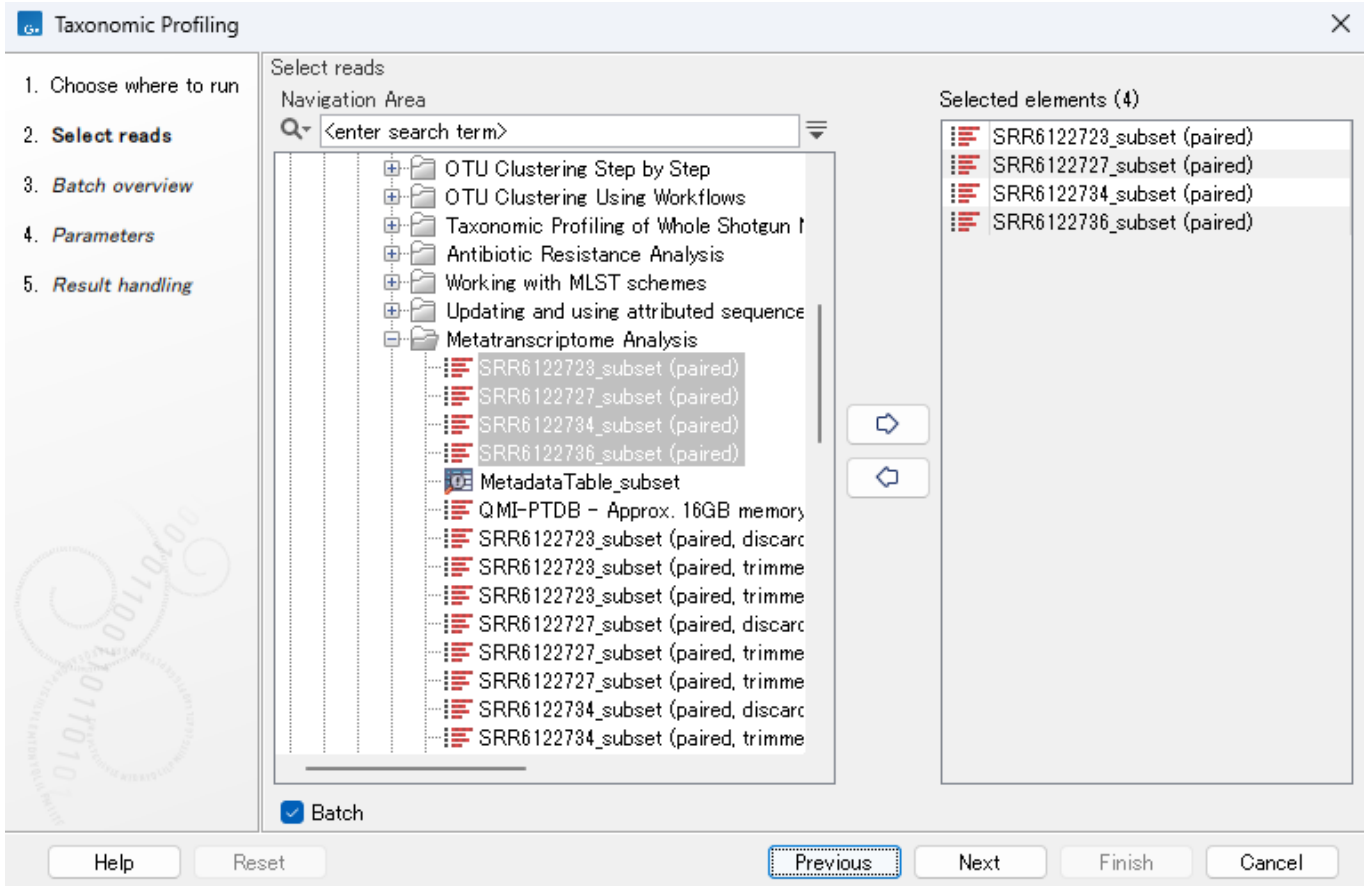
メタデータとリードを名前で紐づけます。
完全一致のほか、前方一致や後方一致で紐づけることが可能です。

解析フロー


各種データベースのダウンロード
リードのインポート、トリミング、メタデータの付加
データに含まれる種の抽出
抽出した種に関するデータベースの作成
GOエンリッチメント解析、パスウェイ解析

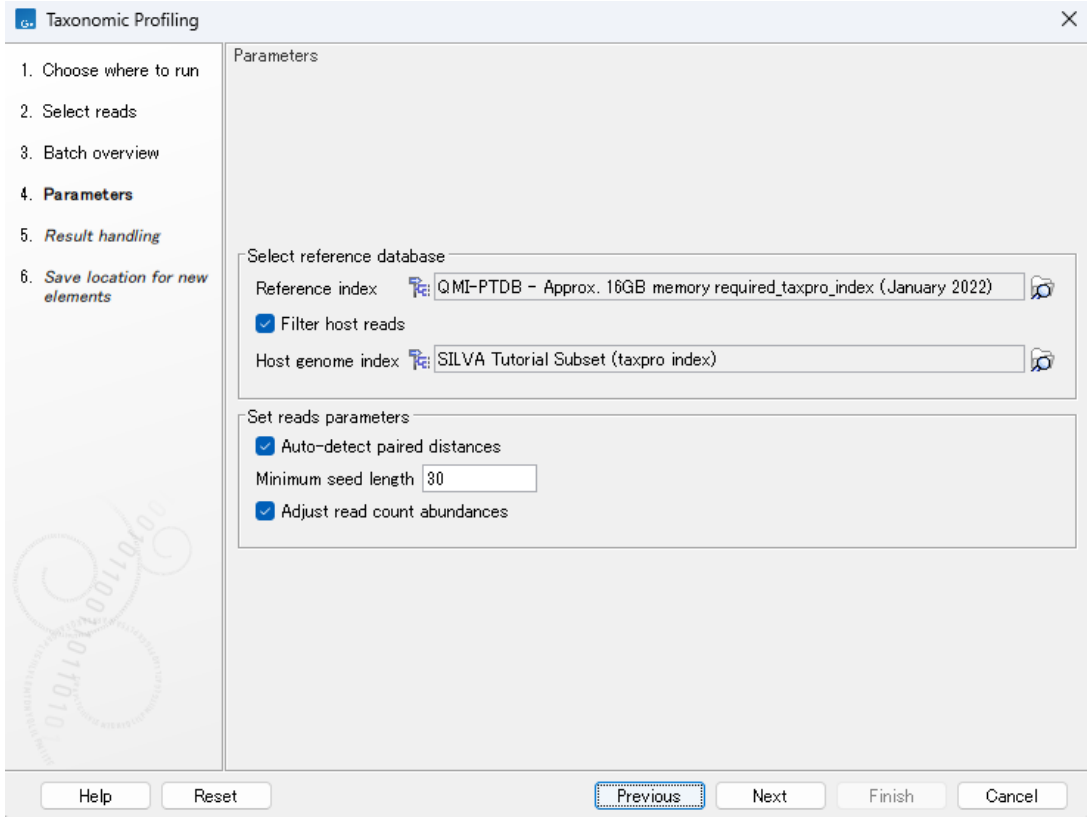
Taxonomic Profiling



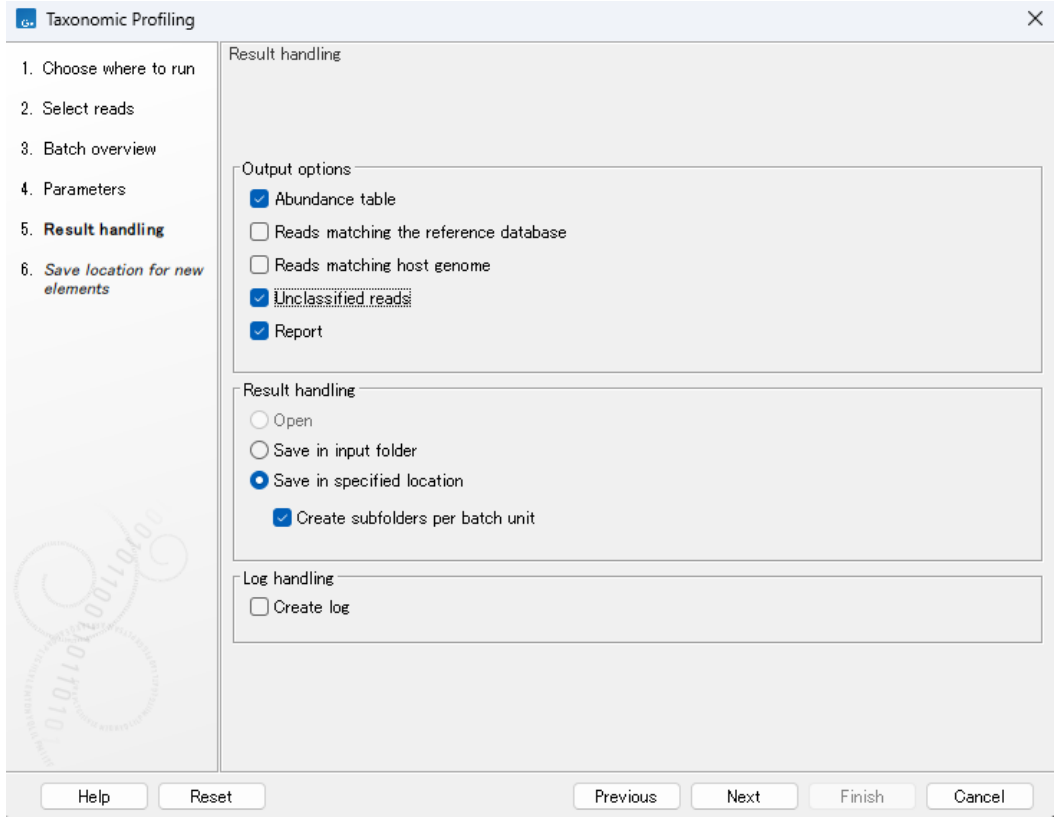


Taxonomic Profilingを起動し、Batchにチェックを入れ、全リードデータを選択します。

Taxonomic Profiling

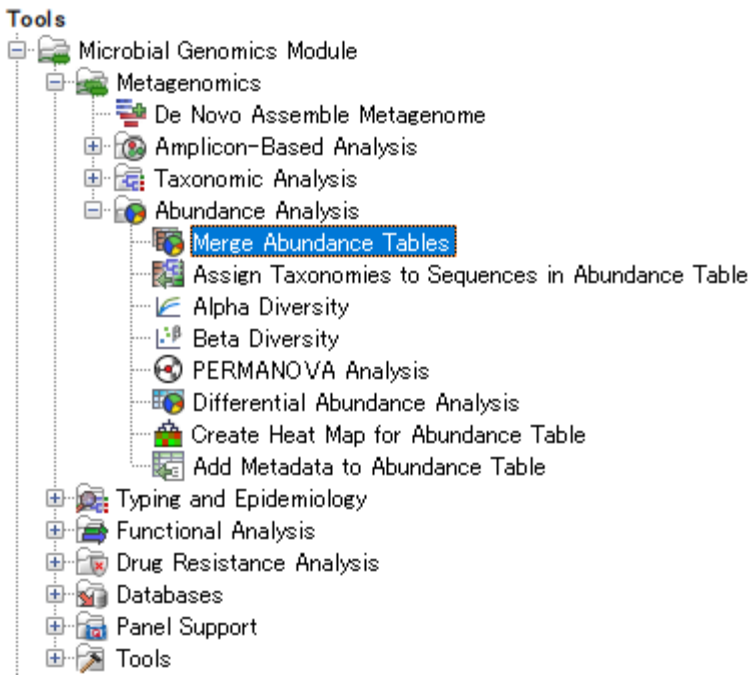


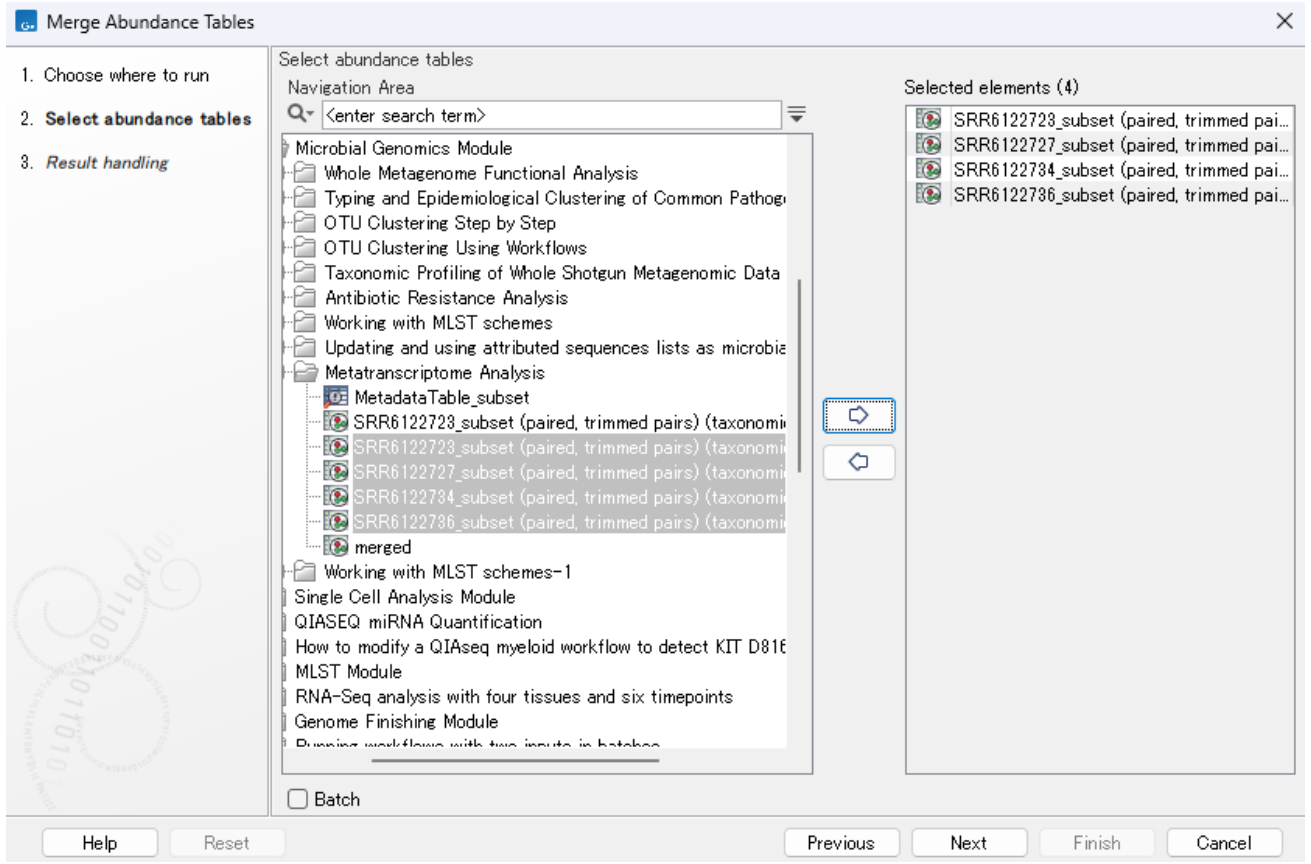
Reference IndexおよびHost genome indexを指定します。



Abundance TableおよびUnclassified Readsを出力します。

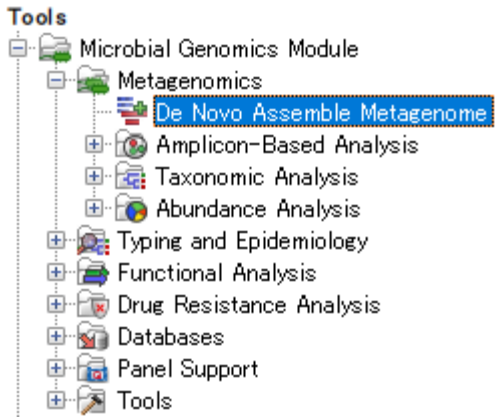
Taxonomic Profiling

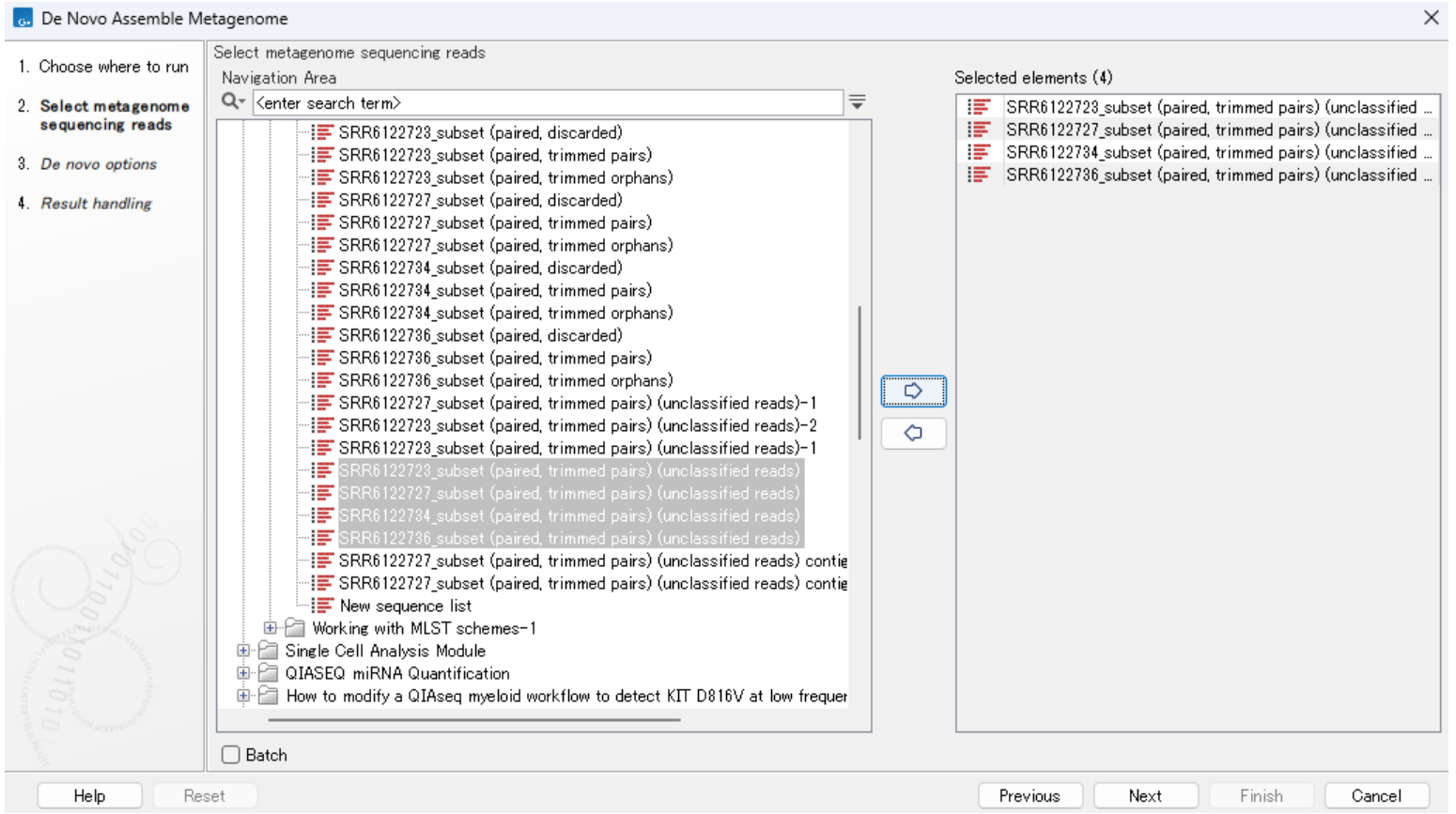




作成されたAbundance TableをMerge Abundance Tableで一つにまとめます。

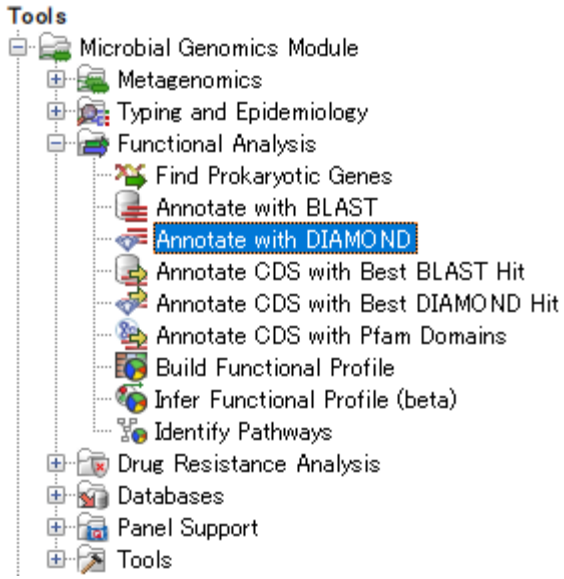
unclassified readのアセンブリ

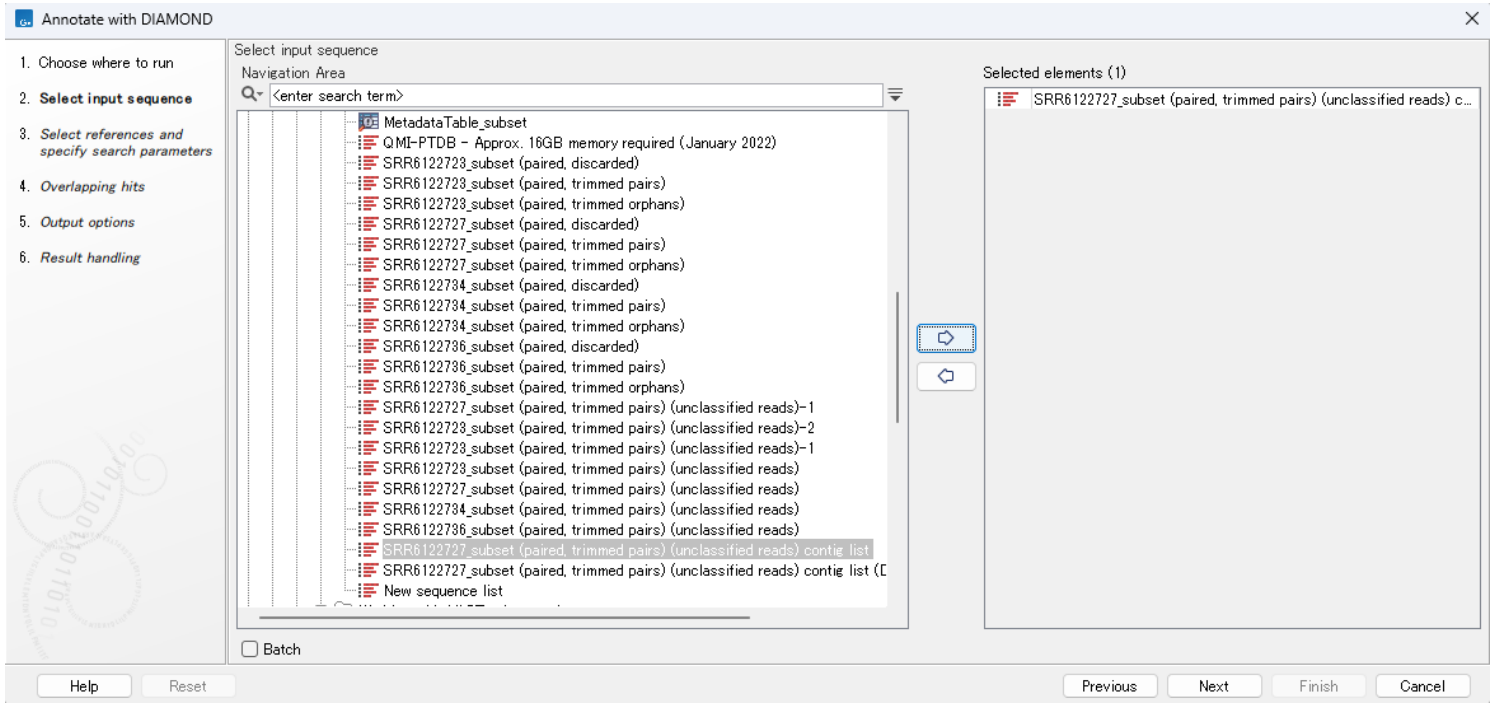




De Novo Assemble Metagenomeを起動し、
Taxonomic profilingツールで作成されたunclassified readsを選択します。

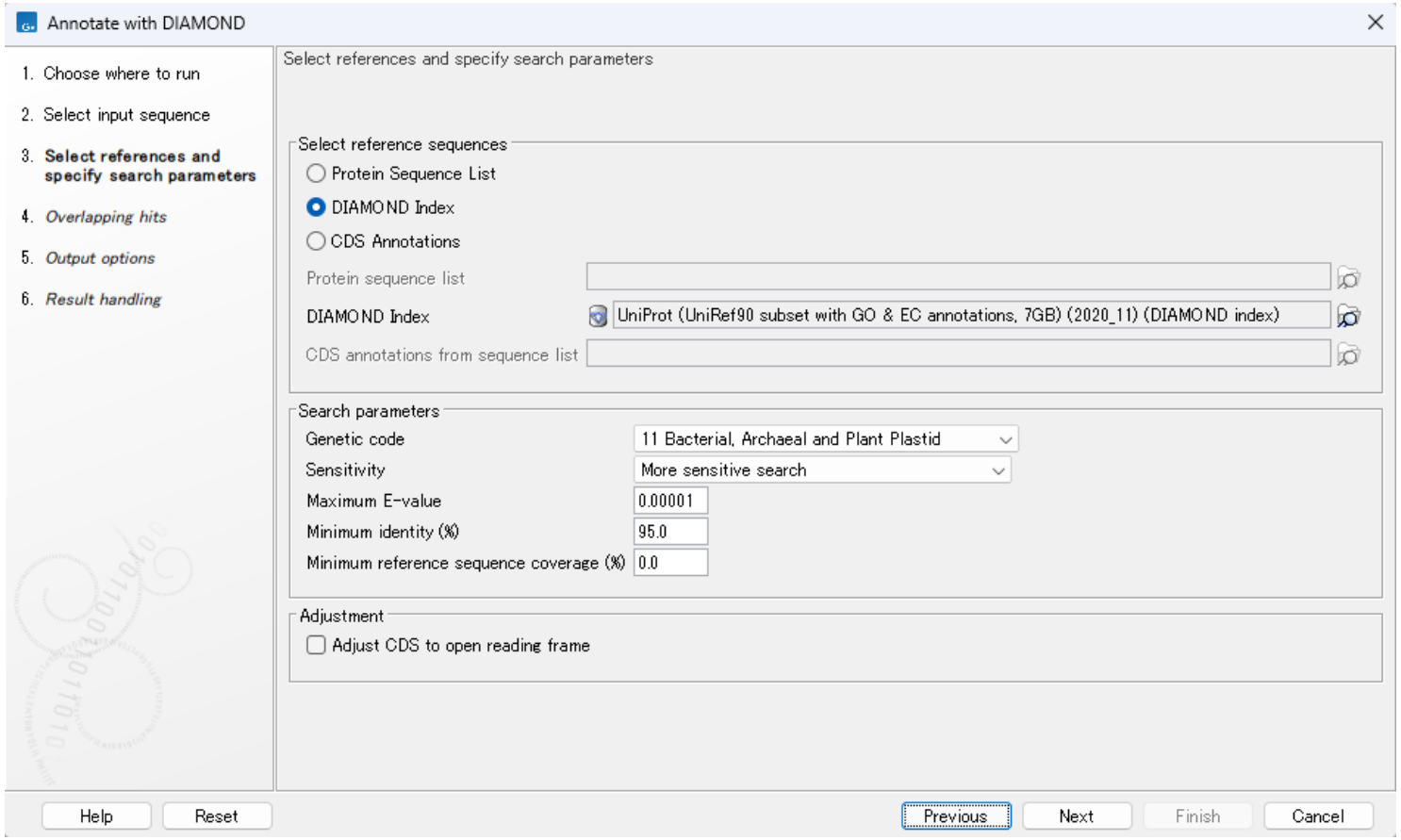
unclassified readのアノテーション





De Novo Assemblyで作成したcontig listに対し、DIAMONDによるアノテーションを行います。

unclassified readのアノテーション



DIAMOND Indexにチェックを入れ、ダウンロードしたUniProtを指定します。

unclassified readのアノテーション

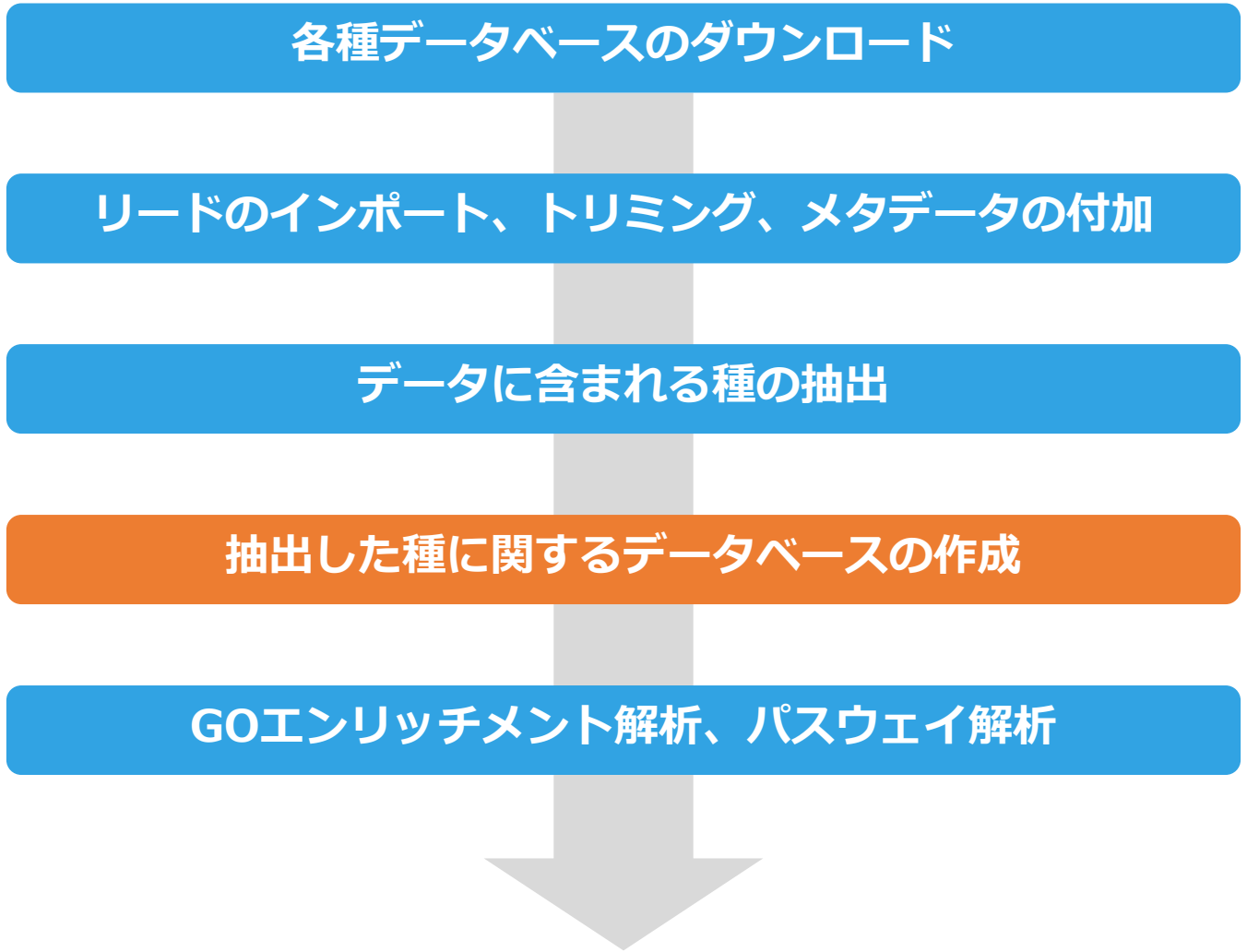
SRR6122727_subset (paired, trimmed pairs) (unclassified reads) contig list (DIAMOND annotations)

Rows: 29 Filter: All

Sequence	Name	Type	Region	Qualifiers
SRR6122727_subset_(paired,_tri...	UniRef90_Q9C4C0	CDS	complement(228..944)	/Description=Coenzyme-B sulfoethylthiotransferase (Fragment) n=4 Tax=environmental samples TaxID=68359 RepID=Q9C4C0_9EURY /EC numbers=2.8.4.1 /GO-terms=GO:0015948 /GO-terms=GO:0046872 /GO-terms=GO:0050524 /db_xref=UniProtKB/Swiss-Prot=Q9C4C0 /DIAMOND Hit=UniRef90_Q9C4C0 /E-value=4.35E-157 /Sequence identity (%)=100.0 /Reference sequence coverage (%)=100.0 /Alignment length=239 /Reference sequence length=239
SRR6122727_subset_(paired,_tri...	UniRef90_H8I4B1	CDS	3442..5112	/Description=Methyl-coenzyme M reductase subunit alpha n=9 Tax=Archaea TaxID=2157 RepID=H8I4B1_METCZ /EC numbers=2.8.4.1 /GO-terms=GO:0015948 /GO-terms=GO:0046872 /GO-terms=GO:0050524 /db_xref=UniProtKB/Swiss-Prot=H8I4B1 /DIAMOND Hit=UniRef90_H8I4B1 /E-value=0.0 /Sequence identity (%)=96.1 /Reference sequence coverage (%)=100.0 /Alignment length=557 /Reference sequence length=555

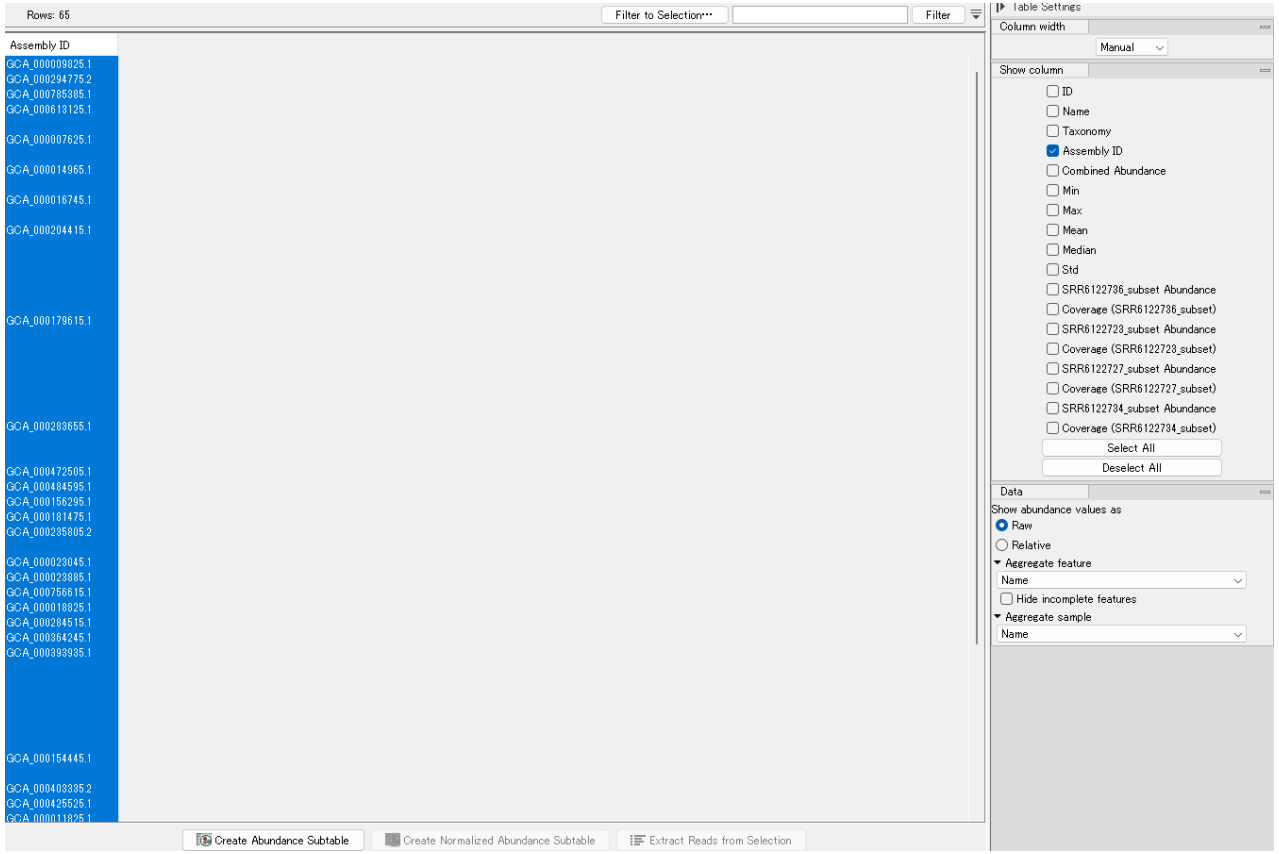
各contigにアノテーションが付与されたデータが作成されます。

解析フロー


各種データベースのダウンロード
リードのインポート、トリミング、メタデータの付加
データに含まれる種の抽出
抽出した種に関するデータベースの作成
GOエンリッチメント解析、パスウェイ解析

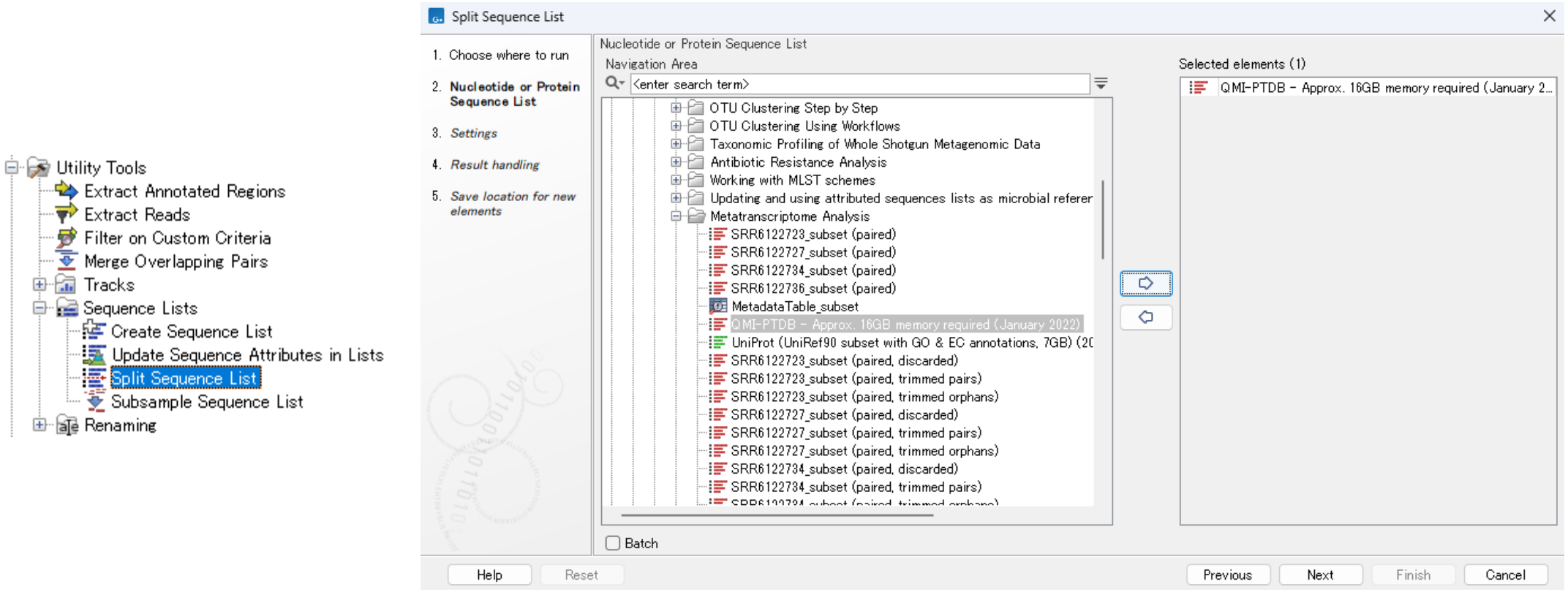
データに含まれる種を抽出

merged



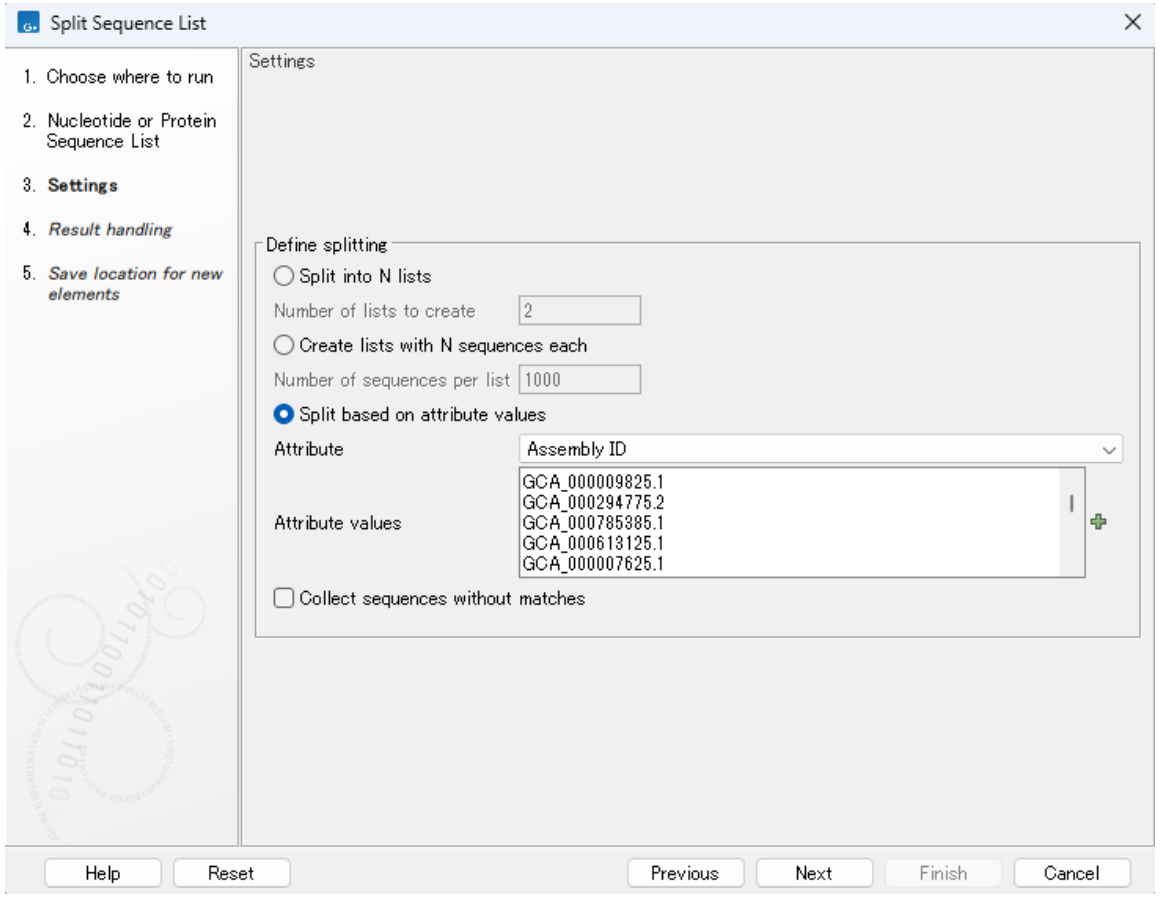
mergeしたAbundance Tableを開き、右側のパネルで”Assembly ID”以外のチェックを外します。
Assembly ID列のみの表示になるので、すべての行を選択し、クリップボードにコピーします。

リファレンスから関連する種の情報を抽出



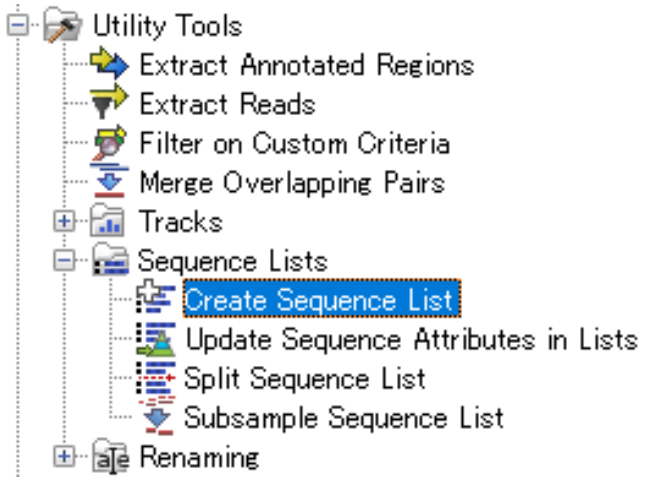
Split Sequence Listツールを起動し、ダウンロードしたQMI-PTDBを指定します。

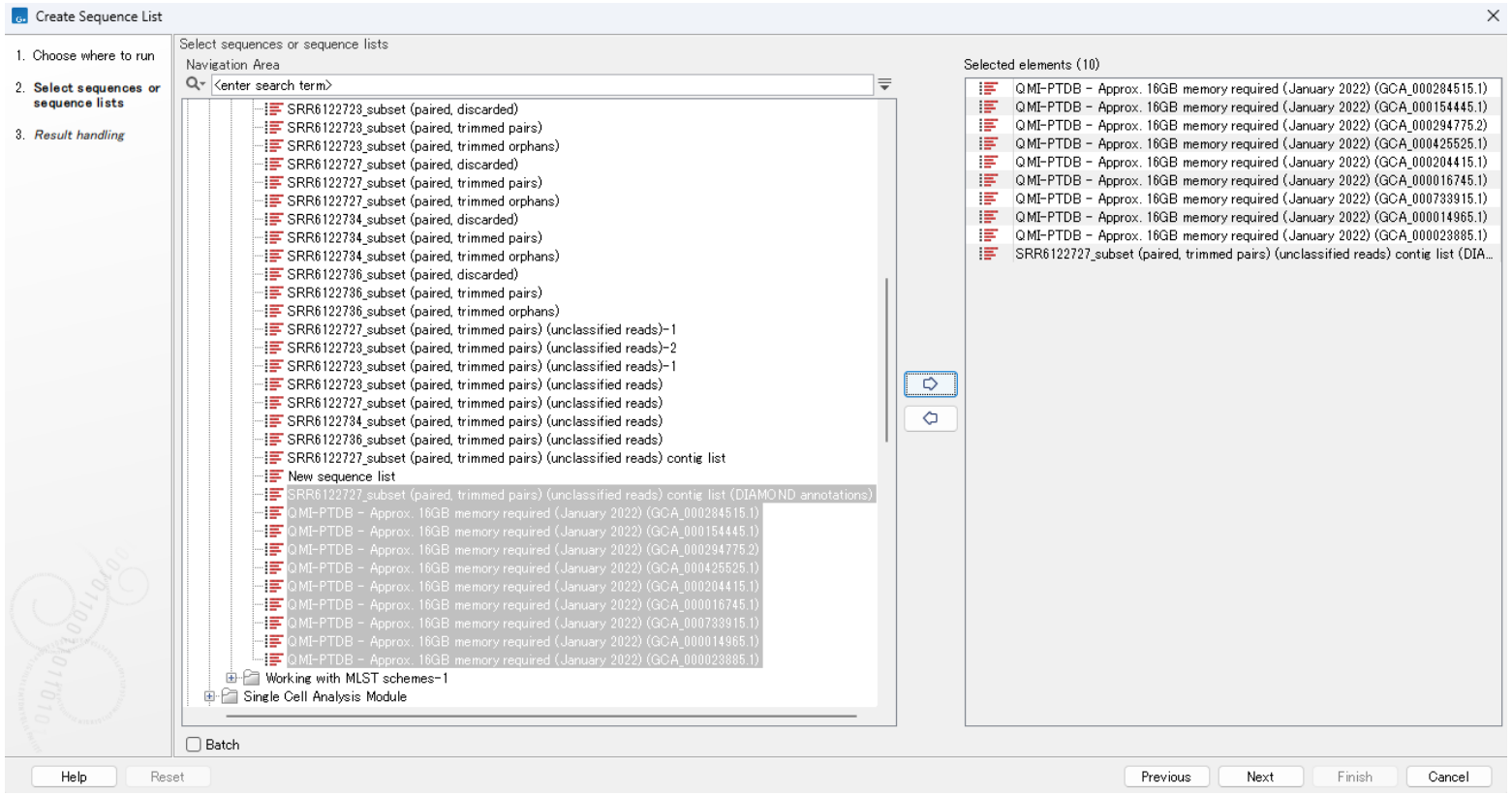
リファレンスから関連する種の情報を抽出



Split based on attribute valuesにチェックを入れ、AttributeをAssembly IDとし、
下の欄にクリップボードの内容をペーストします。

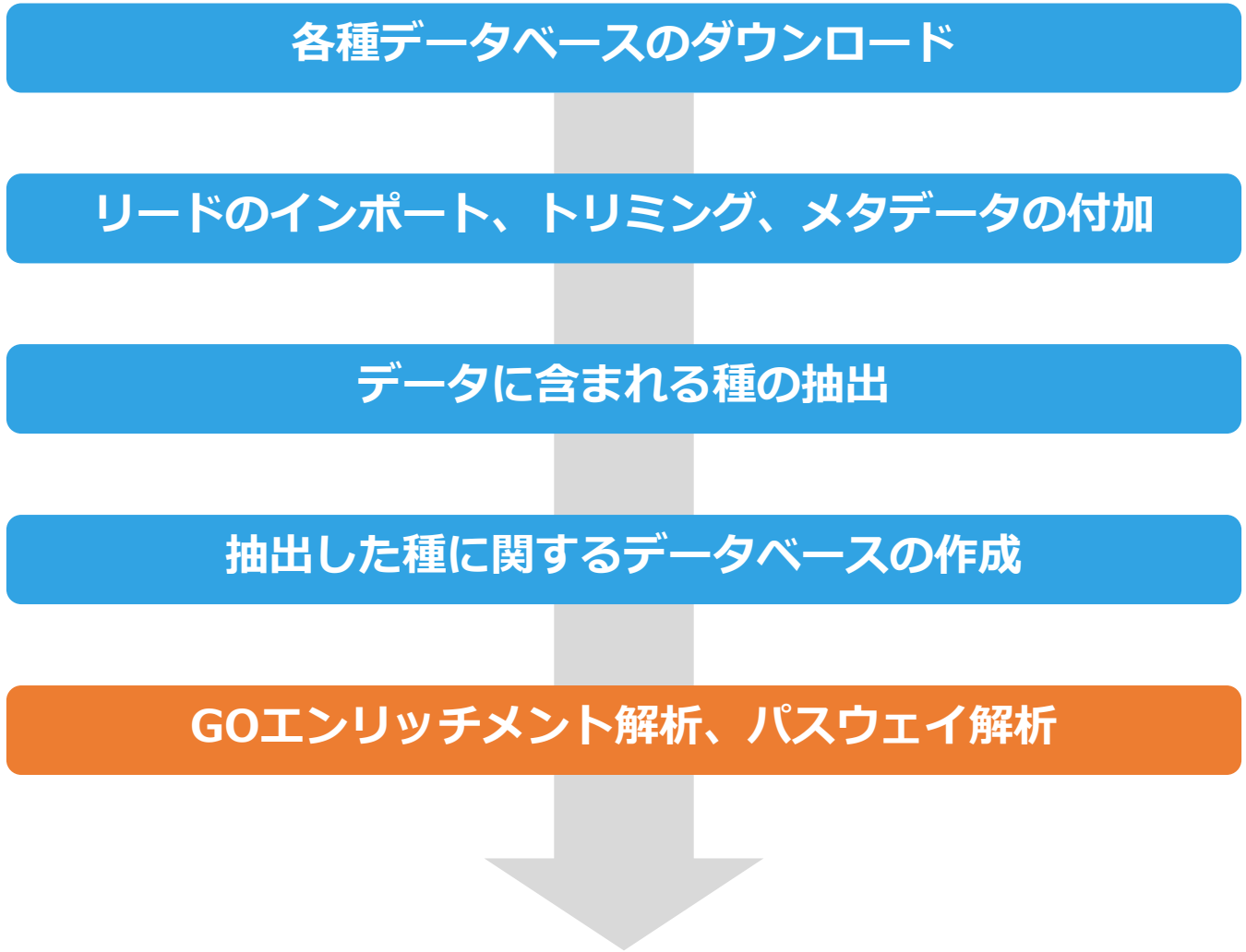
データベースの統合



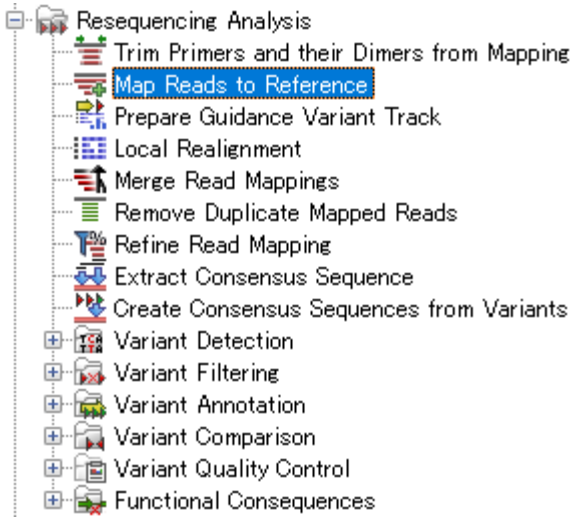


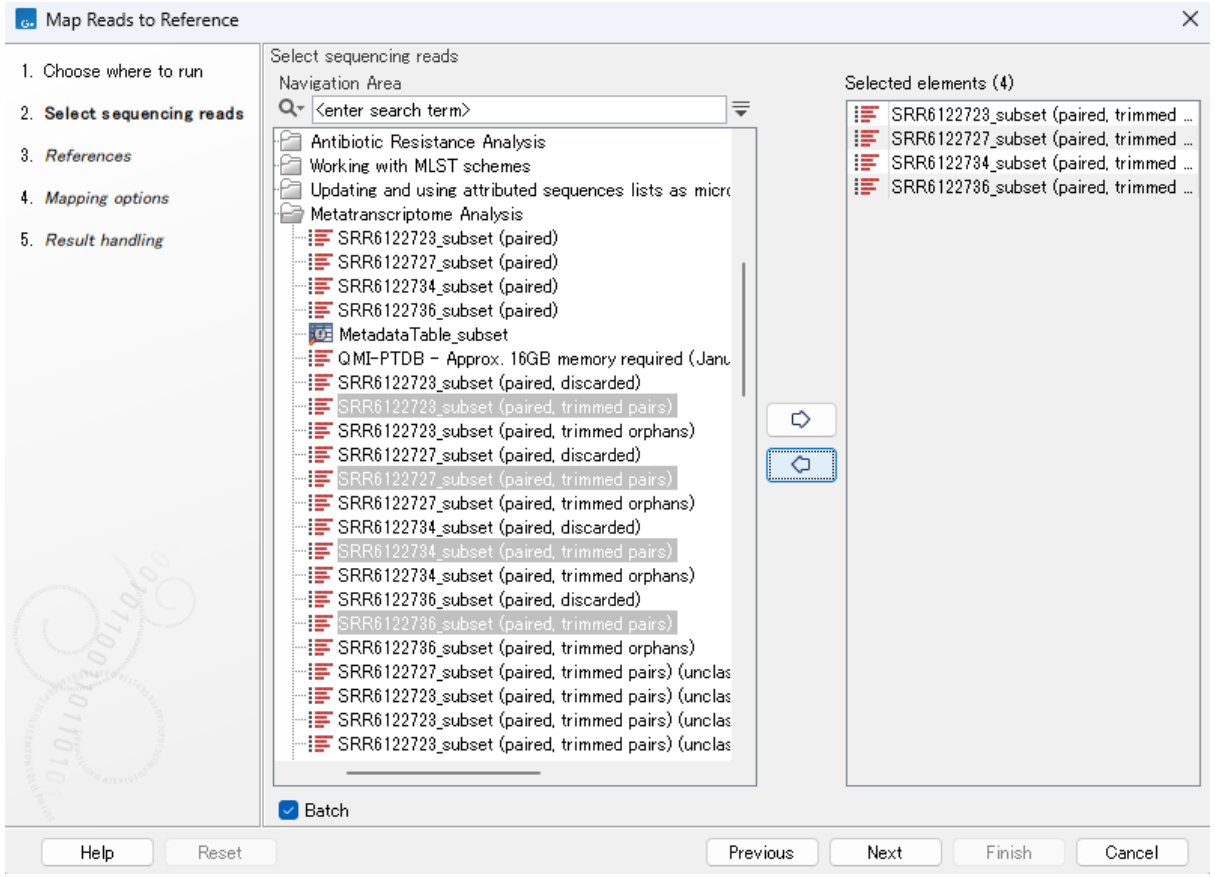
Create Sequence Listツールを起動し、
今抽出した配列リストと、アノテーション付けしたcontigリストを指定します。

解析フロー


各種データベースのダウンロード
リードのインポート、トリミング、メタデータの付加
データに含まれる種の抽出
抽出した種に関するデータベースの作成
GOエンリッチメント解析、パスウェイ解析

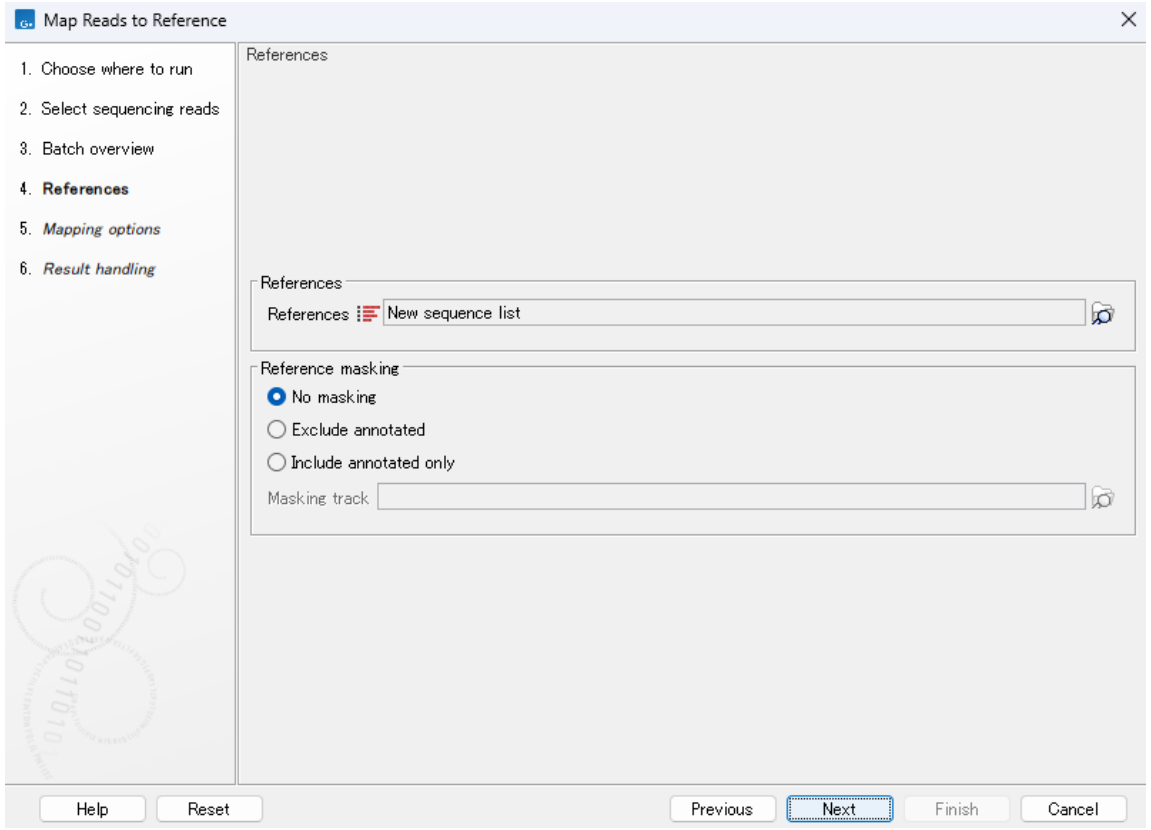
作成したデータベースへリードをマッピング



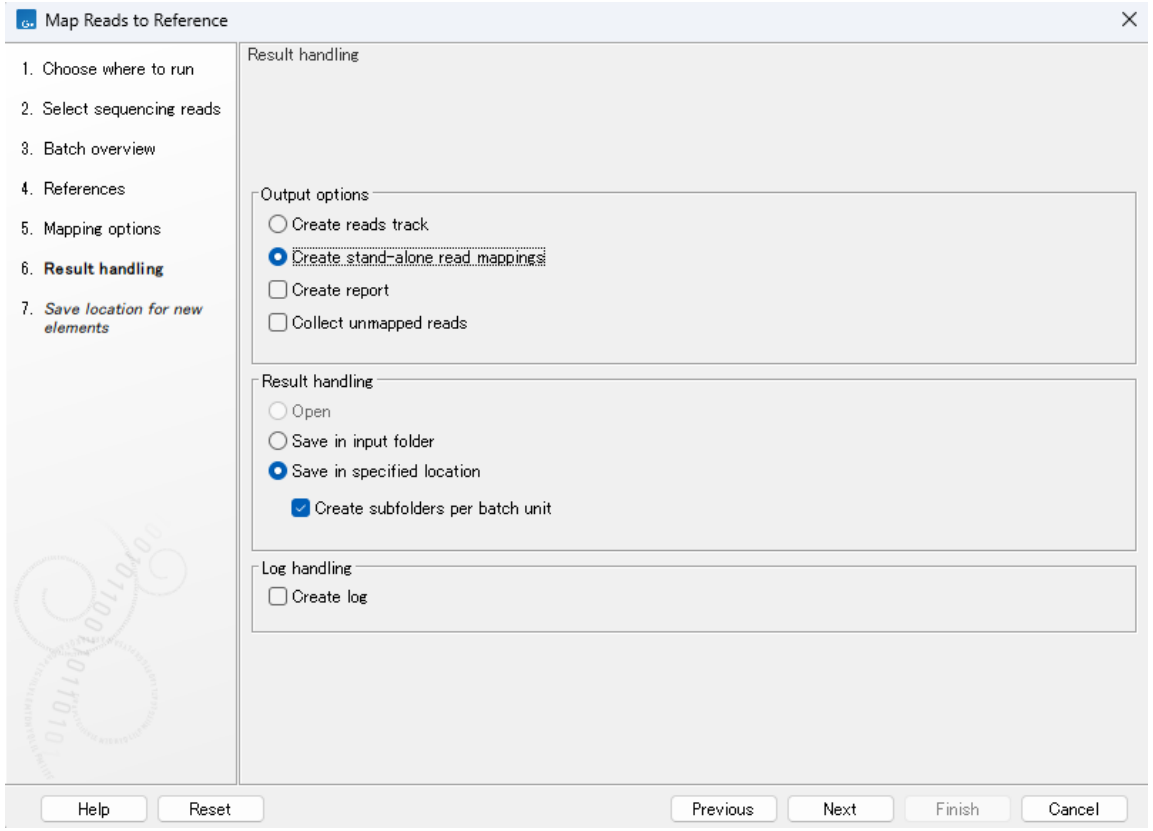


Map Reads to Referenceを起動し、Batchにチェックを入れ、リードデータを選択します。

作成したデータベースへリードをマッピング

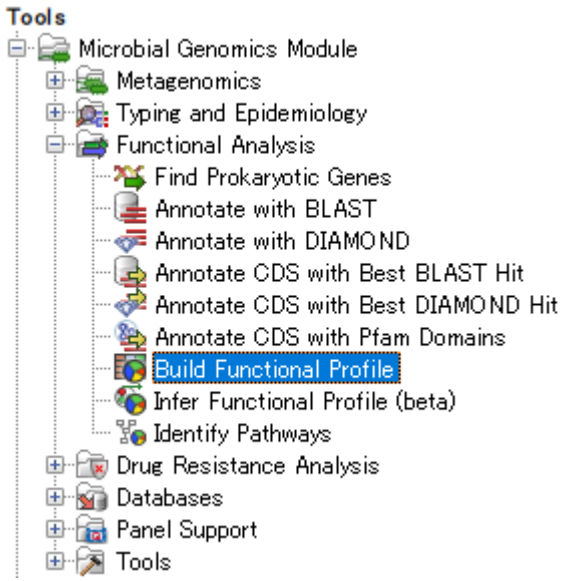


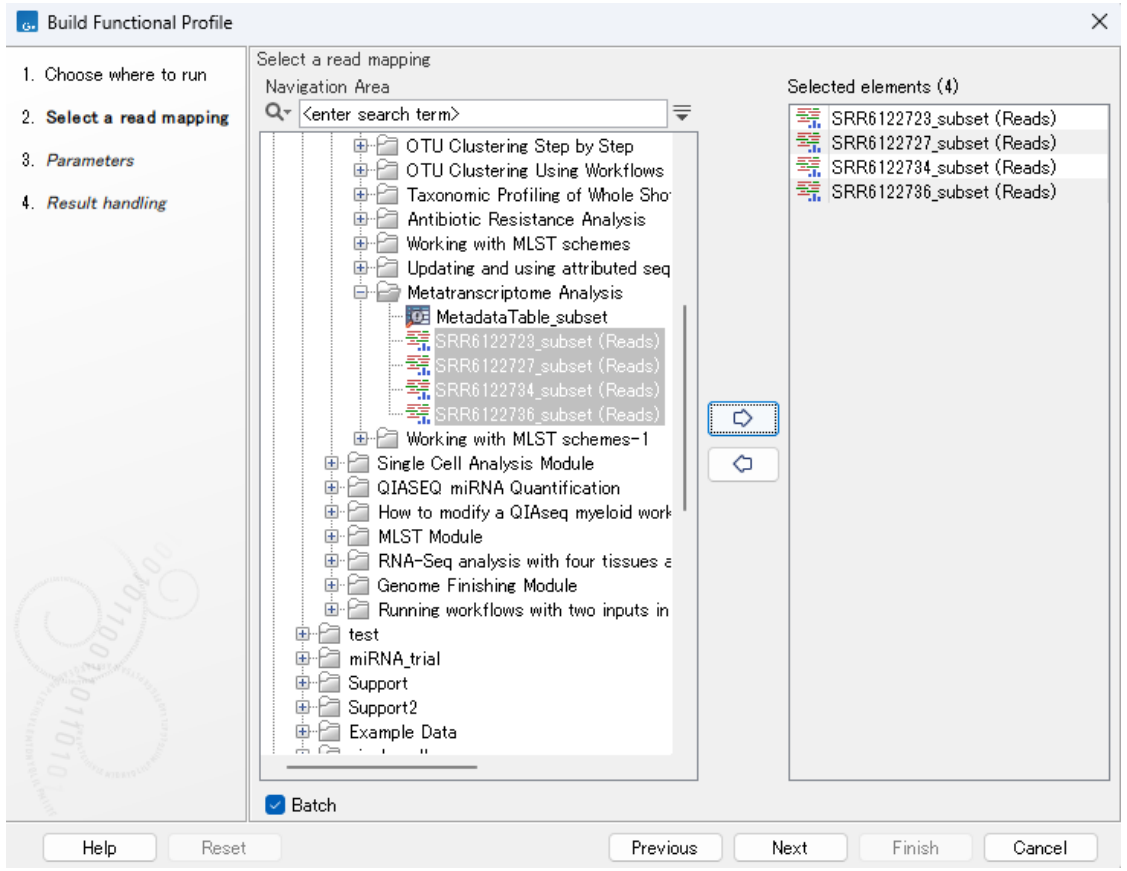
Referenceで先ほど作成したシーケンスリストを指定します。



最後の画面では、Create stand-alone read mapping
にチェックをします。

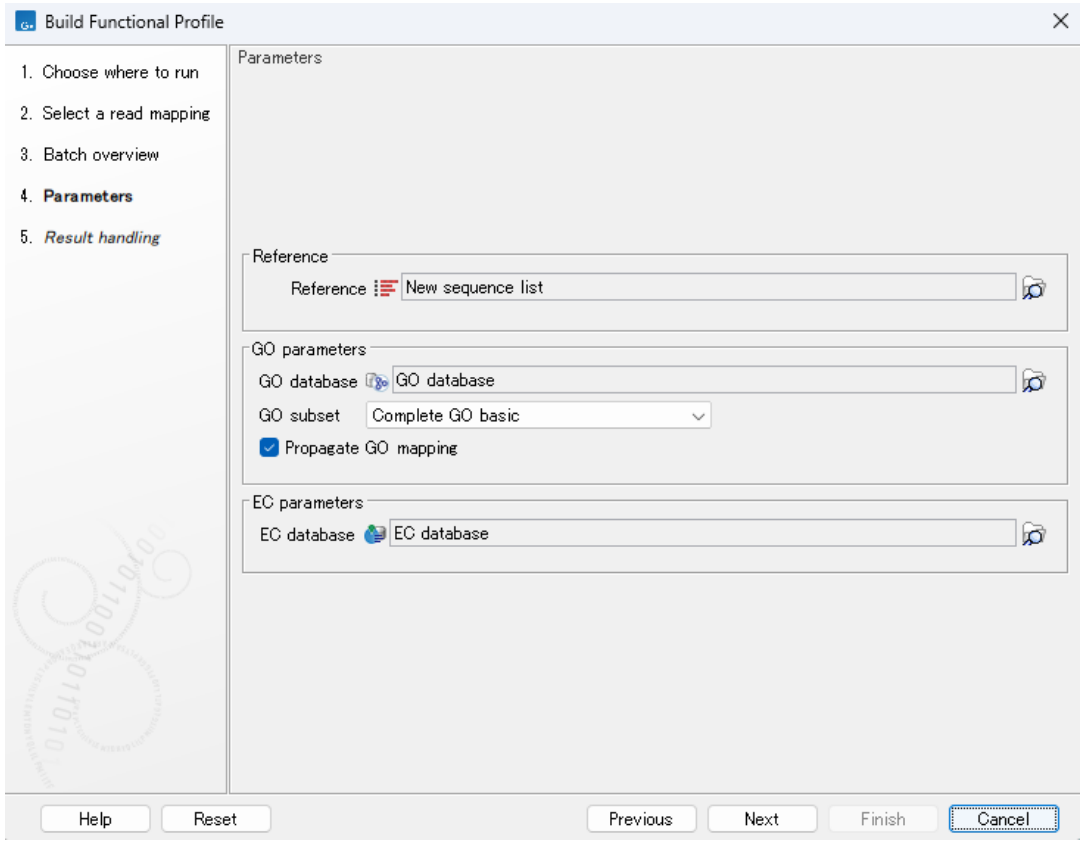
GO Abundance Tableの作成



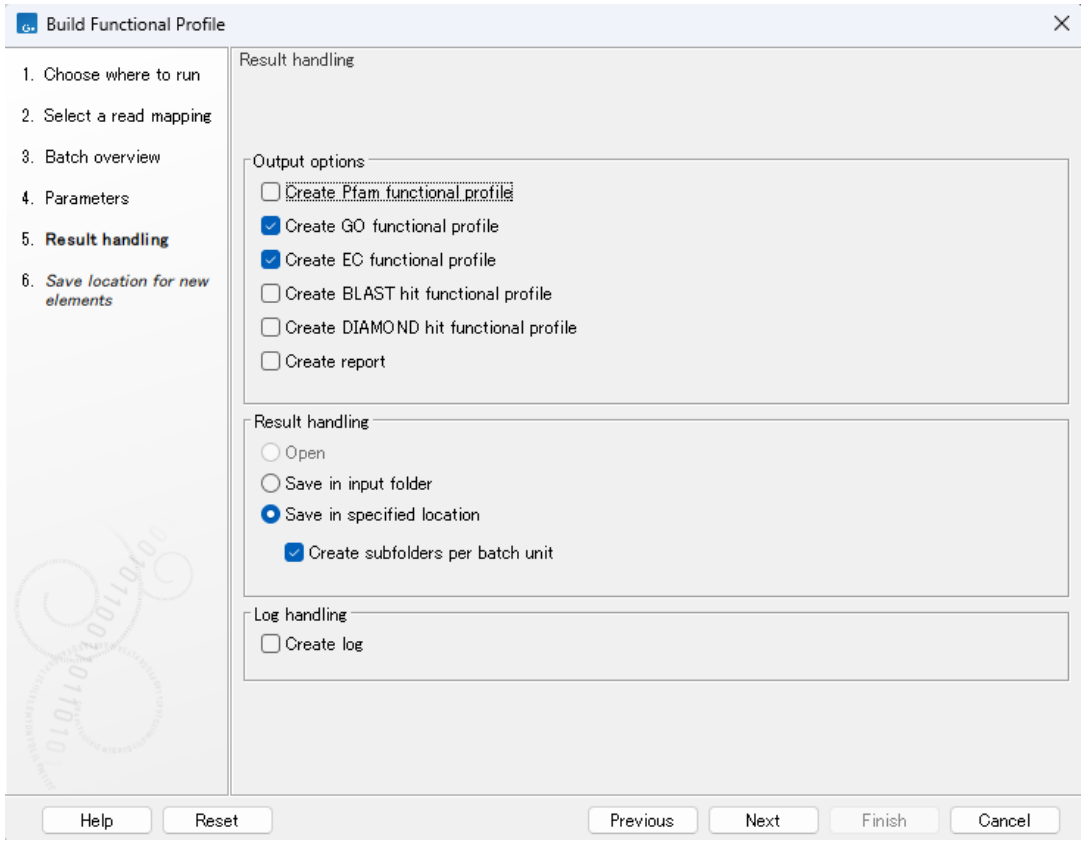


Build Functional Profileを起動し、Batchにチェックを入れ、マッピングデータを指定します。

GO Abundance Tableの作成

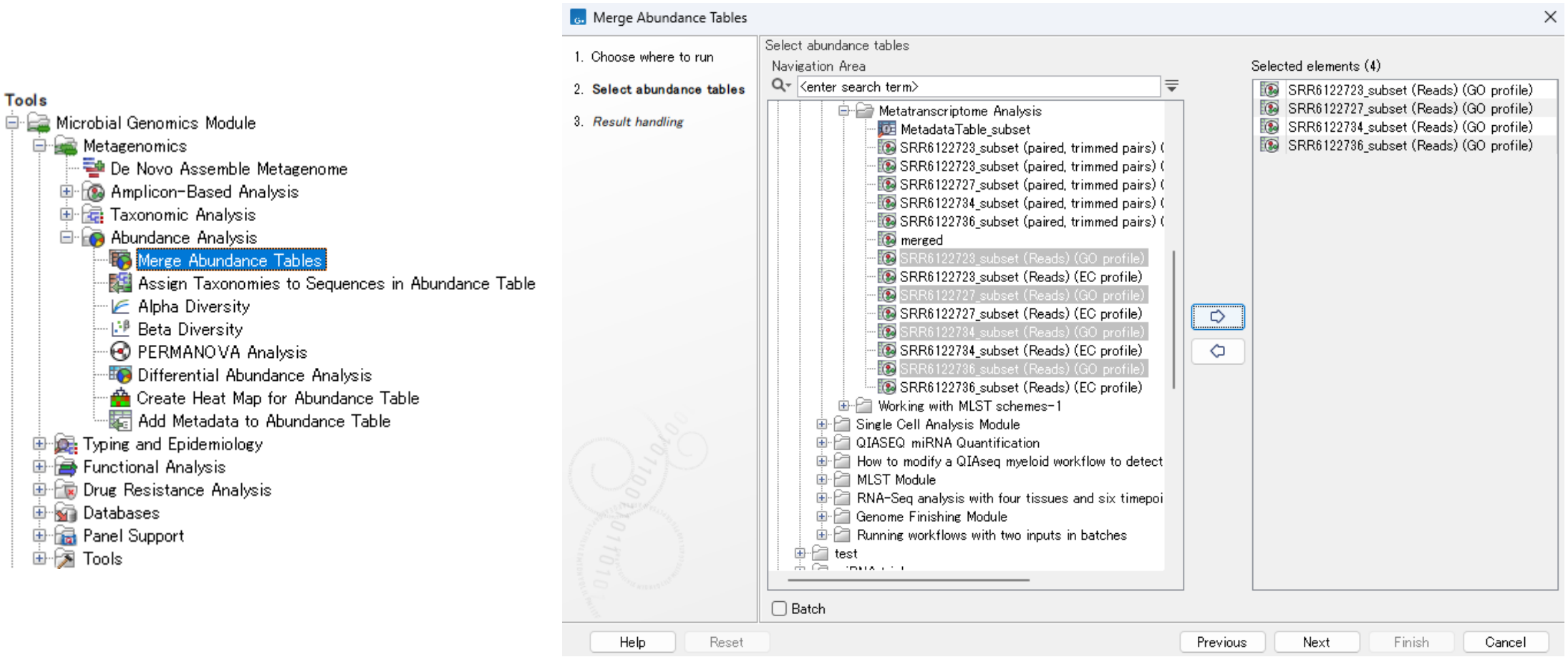


先ほどと同じReferenceデータ、GOデータベース、
ECデータベース（オプション）を指定します。



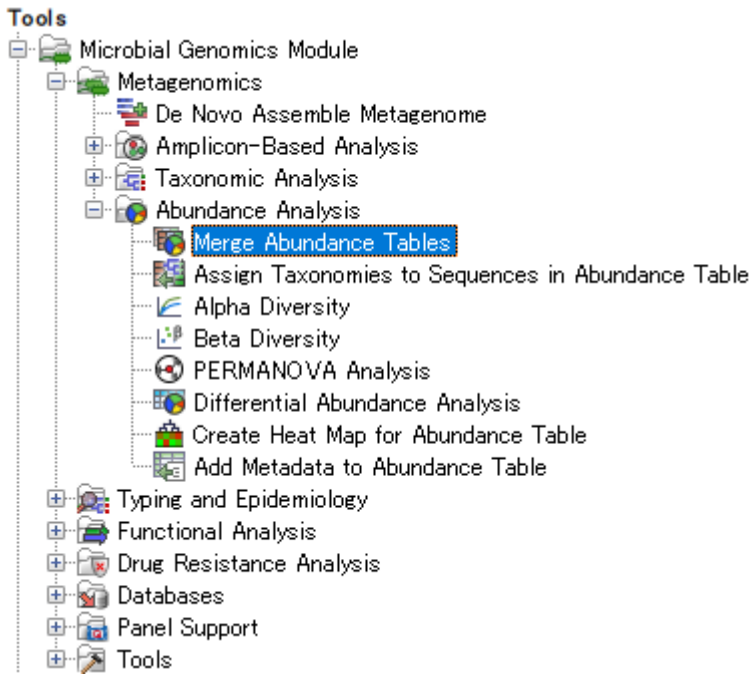
Create GO functional profileと
Create EC functional profileにチェックを入れます。

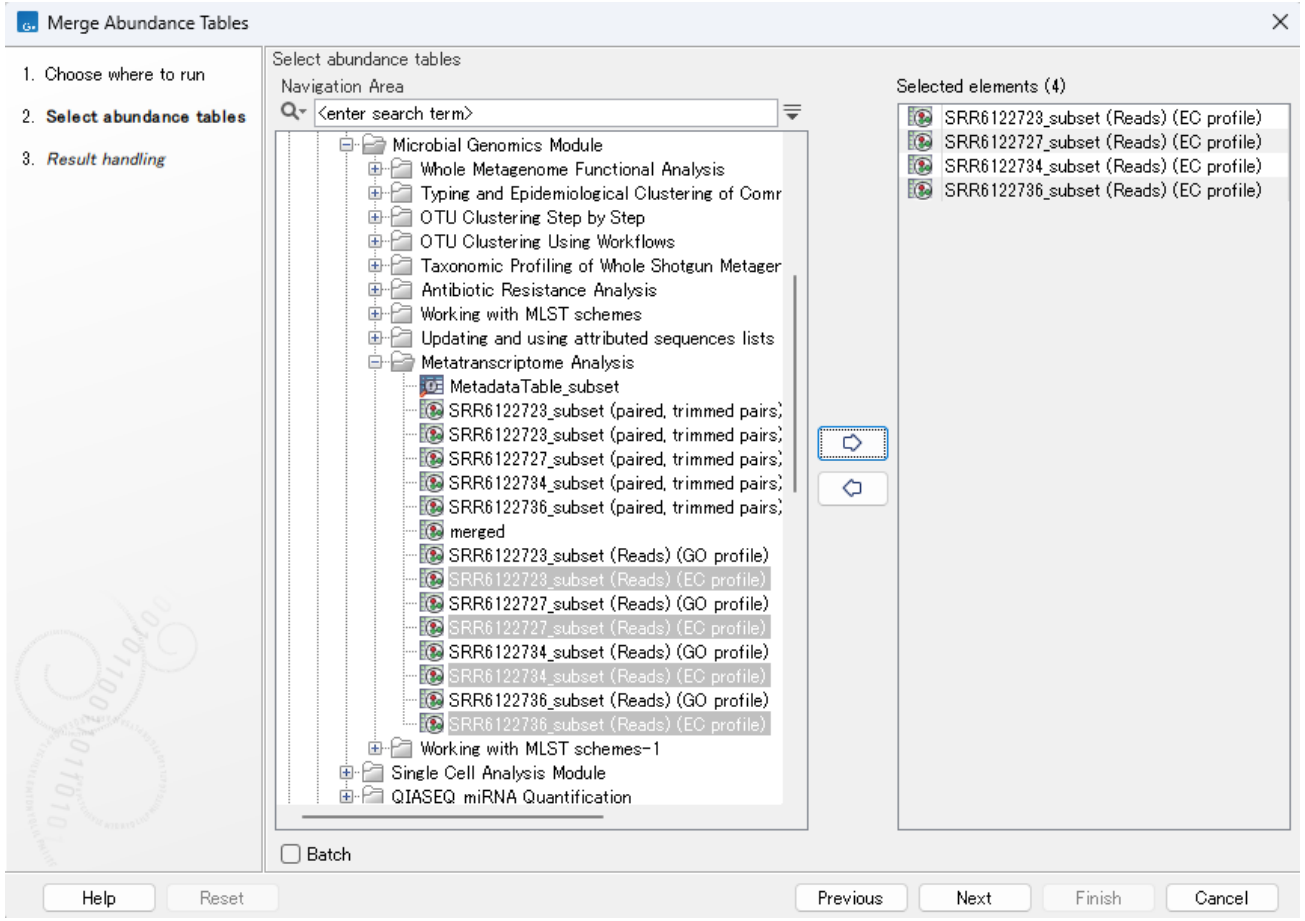
GO Abundance Tableの作成



Merge Abundance Tableを起動し、GOに関するAbundance Tableをすべて選択します。

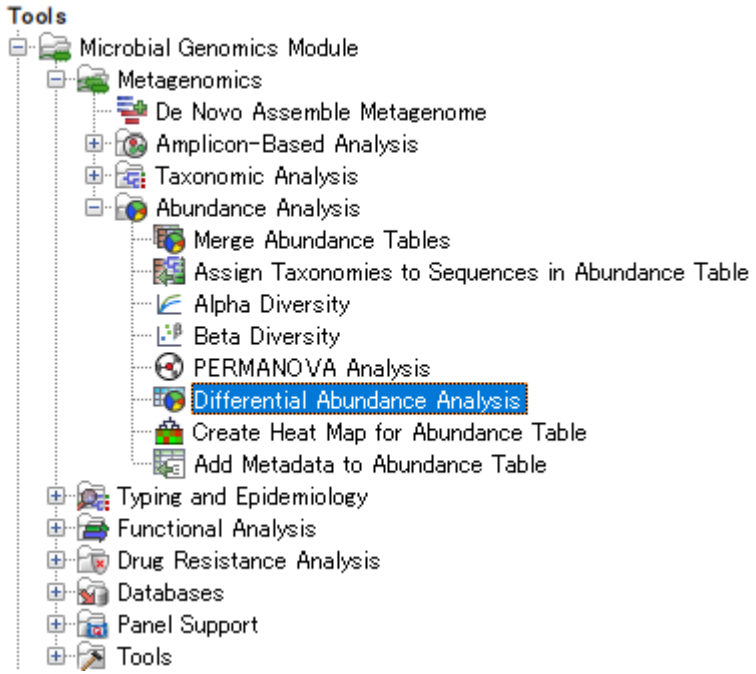
EC Abundance Tableの作成

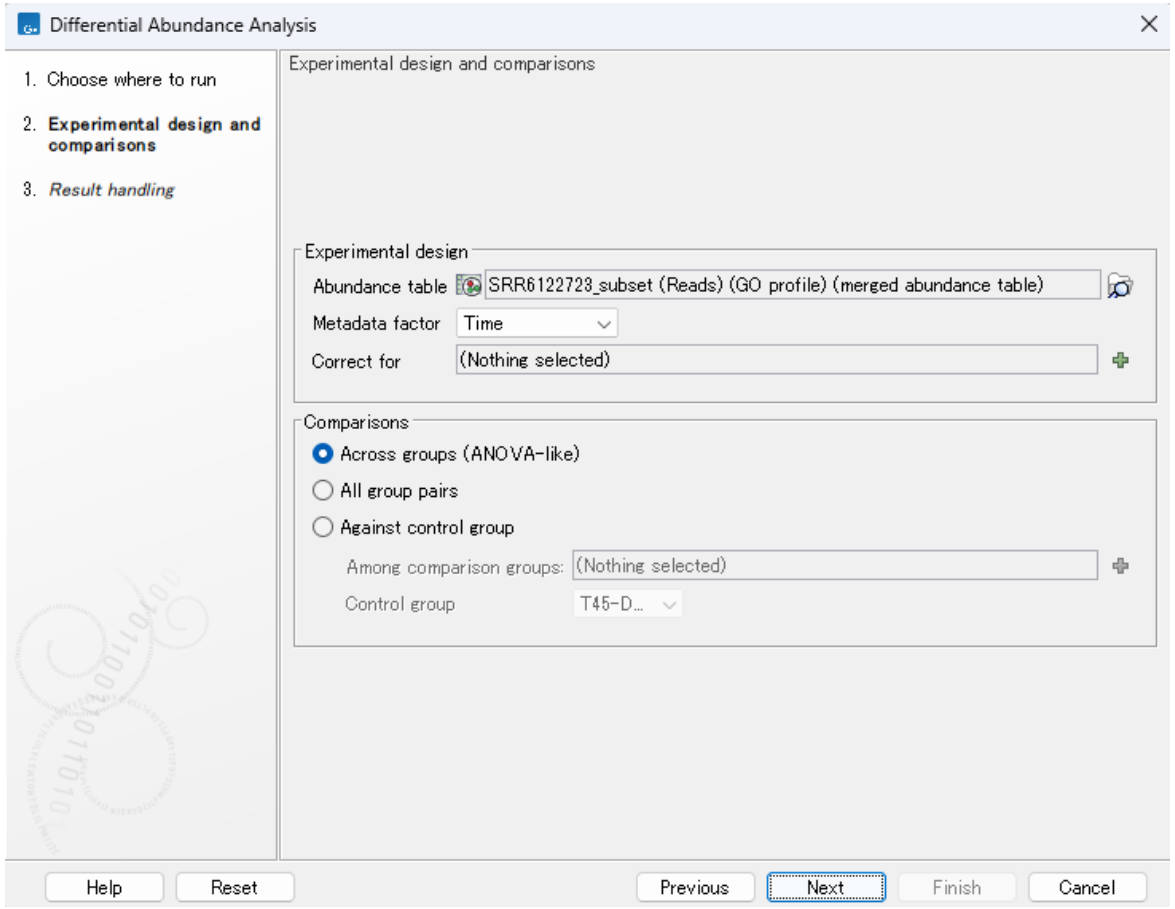




再びMerge Abundance Tableを利用し、ECに関するAbundance Tableをすべて選択します。

GOエンリッチメント解析





Differential Abundance Analysisを起動し、mergeしたGOに関するAbundance Tableを指定します。
Metadata factorで何の要因について比較を行うかを指定します。

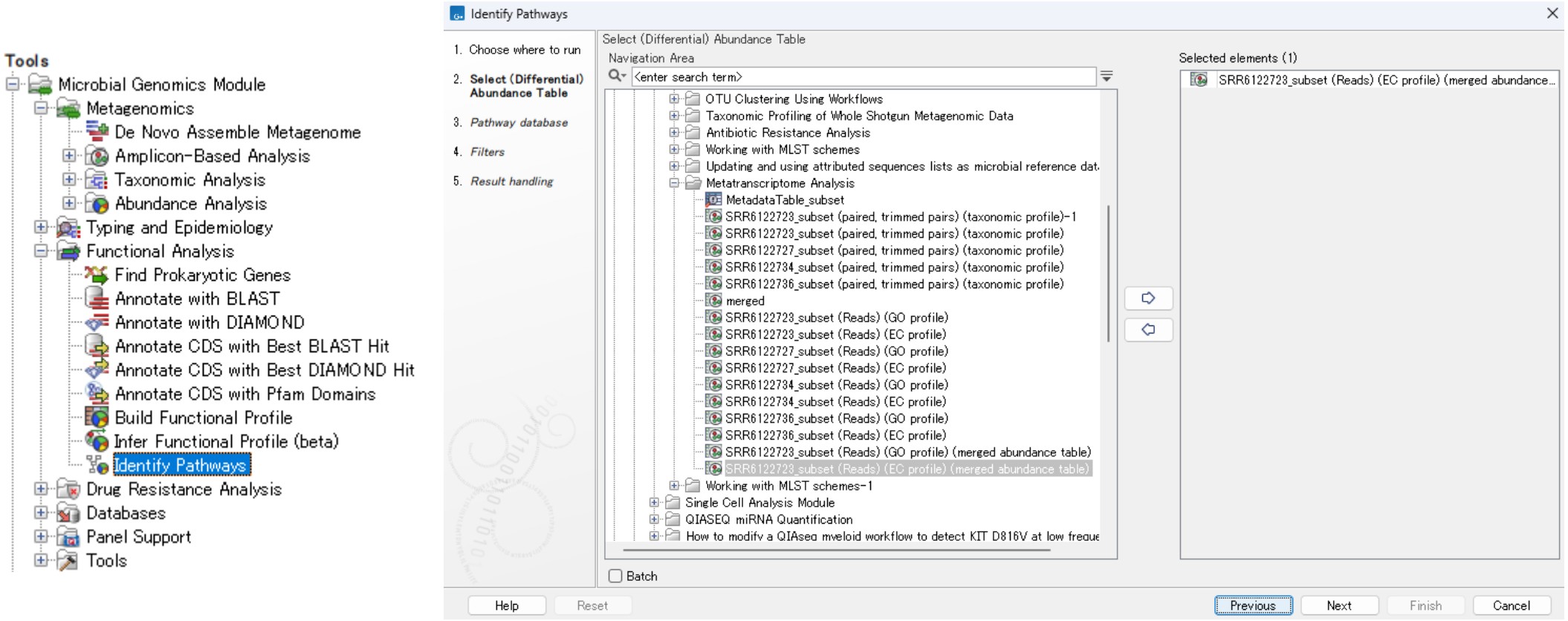
GOエンリッチメント解析

Rows: 168　Venn diagram table　Filter to Selection…　Filter

Name	Due to Time					
	Max group m…	Log_2 fold cha…	Fold change	P-value	FDR p-value	Bonferroni
1902555 // endoribonuclease complex	0.50	-2.65	-6.28	0.45	0.60	1.00
0072521 // purine-containing compound metabolic process	2.50	-4.60	-24.30	0.06	0.16	1.00
0003824 // catalytic activity	40.00	-0.76	-1.69	0.04	0.16	1.00
0045229 // external encapsulating structure organization	10.50	-2.30	-4.93	5.82E-3	0.03	0.98
0003735 // structural constituent of ribosome	326.50	-1.13	-2.19	3.95E-7	1.75E-5	6.63E-5
0003674 // molecular_function	366.50	-1.08	-2.12	1.47E-7	1.75E-5	2.48E-5
0005975 // carbohydrate metabolic process	2.50	-4.60	-24.30	0.07	0.16	1.00
0043038 // amino acid activation	17.00	0.27	1.20	0.52	0.64	1.00
0043039 // tRNA aminoacylation	17.00	0.27	1.20	0.52	0.64	1.00
0055086 // nucleobase-containing small molecule metabolic process	2.50	-4.60	-24.30	0.07	0.16	1.00
0005840 // ribosome	326.50	-1.13	-2.19	9.37E-7	1.75E-5	1.57E-4
0043167 // ion binding	28.50	-0.33	-1.26	0.36	0.60	1.00
0043168 // anion binding	28.50	-0.33	-1.26	0.36	0.60	1.00
0043169 // cation binding	1.00	-3.43	-10.78	0.26	0.48	1.00
0043170 // macromolecule metabolic process	17.50	0.23	1.17	0.59	0.68	1.00
0043228 // non-membrane-bounded organelle	326.50	-1.13	-2.19	9.37E-7	1.75E-5	1.57E-4
0043229 // intracellular organelle	326.50	-1.13	-2.19	9.37E-7	1.75E-5	1.57E-4
0043226 // organelle	326.50	-1.13	-2.19	9.37E-7	1.75E-5	1.57E-4
0043232 // intracellular non-membrane-bounded organelle	326.50	-1.13	-2.19	7.20E-7	1.75E-5	1.21E-4
0005737 // cytoplasm	11.50	-2.43	-5.39	2.85E-3	0.03	0.48
0030680 // dimeric ribonuclease P complex	0.50	-2.65	-6.28	0.45	0.60	1.00
0030677 // ribonuclease P complex	0.50	-2.65	-6.28	0.45	0.60	1.00
0030554 // adenyl nucleotide binding	18.00	0.19	1.14	0.65	0.72	1.00
0005575 // cellular_component	355.50	-1.05	-2.07	6.46E-7	1.75E-5	1.08E-4
0043436 // oxoacid metabolic process	19.50	0.07	1.05	0.86	0.91	1.00
0005524 // ATP binding	18.00	0.19	1.14	0.65	0.72	1.00
0005515 // protein binding	1.00	-3.43	-10.78	0.26	0.48	1.00

サンプル間で変動のあるGOのリストが作成されます。
ECに関しても同様のリストを作成することができます。

パスウェイ解析



Identify Pathwaysを起動し、ECのAbundance Tableを選択します。

パスウェイ解析

Rows: 19　Filter to Selection…　Filter

Name	MetaCyc ID	SRR6122723_subs		
		Min. Solution	Confidence	Coverage
tRNA charging	TRNA-CHARGING-PWY	☑	1.00	0.05
tRNA processing	PWY0-1479	☑	1.00	0.20
UDP-*N*-acetylmuramoyl-pentapeptide biosynthesis Ⅲ (*meso*-diaminopimelate containing)	PWY-7953	☑	0.26	0.12
UDP-*N*-acetylmuramoyl-pentapeptide biosynthesis Ⅱ (lysine-containing)	PWY-6386	☑	0.25	0.12
anhydromuropeptides recycling I	PWY0-1261	☑	0.25	0.08
UDP-*N*-acetylmuramoyl-pentapeptide biosynthesis I (*meso*-diaminopimelate containing)	PWY-6387	☑	0.24	0.12
gluconeogenesis I	GLUCONEO-PWY	☑	0.09	0.08
Calvin-Benson-Bassham cycle	CALVIN-PWY	☑	0.09	0.15
glycolysis I (from glucose 6-phosphate)	GLYCOLYSIS	☑	0.09	0.08
glycolysis V (Pyrococcus)	P341-PWY	☑	0.09	0.10
sucrose degradation V (sucrose α-glucosidase)	PWY66-373	☑	0.08	0.20
sedoheptulose bisphosphate bypass	PWY0-1517	☑	0.08	0.50
formaldehyde assimilation Ⅱ (assimilatory RuMP Cycle)	PWY-1861	☑	0.08	0.11
gluconeogenesis Ⅲ	PWY66-399	☑	0.07	0.08
formaldehyde assimilation Ⅲ (dihydroxyacetone cycle)	P185-PWY	☑	0.07	0.08
glycolysis Ⅱ (from fructose 6-phosphate)	PWY-5484	☑	0.07	0.09
glycolysis Ⅲ (from glucose)	ANAGLYCOLYSIS-PWY	☑	0.07	0.09
glycolysis Ⅳ (plant cytosol)	PWY-1042	☑	0.06	0.10
1,3-propanediol biosynthesis (engineered)	PWY-7385	☑	0.06	0.12

関連するパスウェイのリストが表示されます。

お問い合わせ先：フィルジェン株式会社

TEL: 052-624-4388 (9:00～17：00)

FAX: 052-624-4389

E-mail: biosupport@filgen.jp