DBCLS
Database Center for Life Science

DBCLS Galaxy の利用法

ライフサイエンス統合データベースセンター
山口 敦子

Galaxy とは

Galaxy はゲノムなどの生物学データを対象とした，解析ツール組み合わせインタフェイスです．

ペンシルバニア州立大学を中心としたGalaxy team が開発

ユーザは生物学データを，

・用意されたツールを組み合わせて**解析**，

・ツールの組み合わせや計算履歴の**保存と共有**，

・計算履歴，ワークフローの**公開**ができます．

DBCLS Galaxy とは

DBCLS Galaxy は Galaxy をベースとし，主に日本の生物系研究者を対象として，独自のツールや機能を組み込んだGalaxy です．

http://galaxy.dbcls.jp/

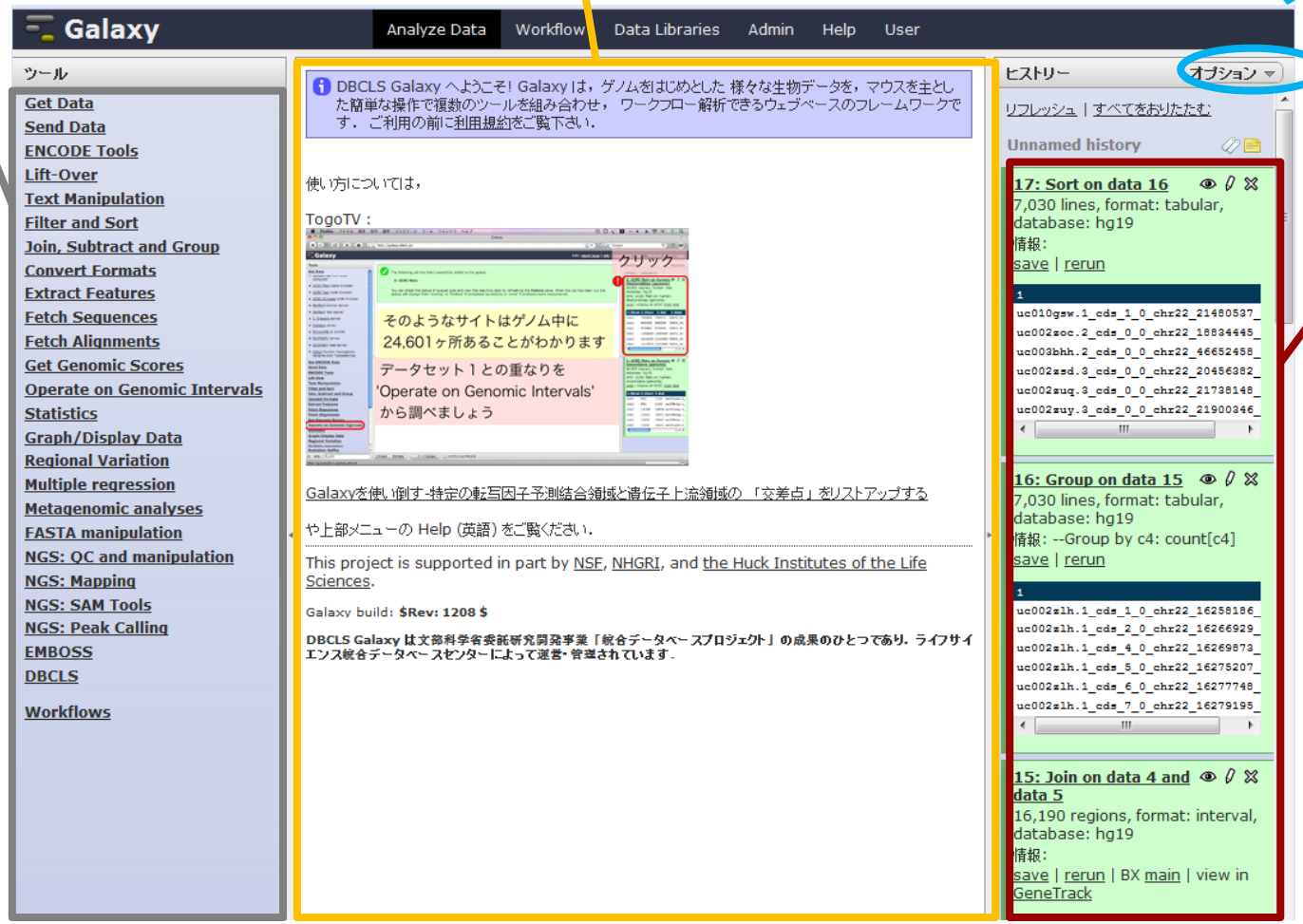
注: DBCLS Galaxy の機能をフルに利用するためには，DBCLS OpenID のアカウントを取得する必要があります．
(データ解析だけならば，アカウントなしでもご利用できます)

DBCLS Galaxy の画面

利用可能な
ツールのリスト

ツール操作・結果表示画面

ヒストリーの操作(共有，新規作成など)
ワークフローの構築
など



ヒストリー
（解析履歴）

DBCLS Galaxy の特徴

DBCLS Galaxy は日本の生物系研究者を主ターゲットとするため,

- 国内の生物系研究者がよく使う国内外 DB をサポート
  (TogoWS経由)
  NCBI, EBI, PDB, KEGG などからデータを取得可能

- 日本人向けのカスタマイズ
  日本語のメニュー
  日本語の文献を検索
  など　　の機能・ツールが組み込まれている

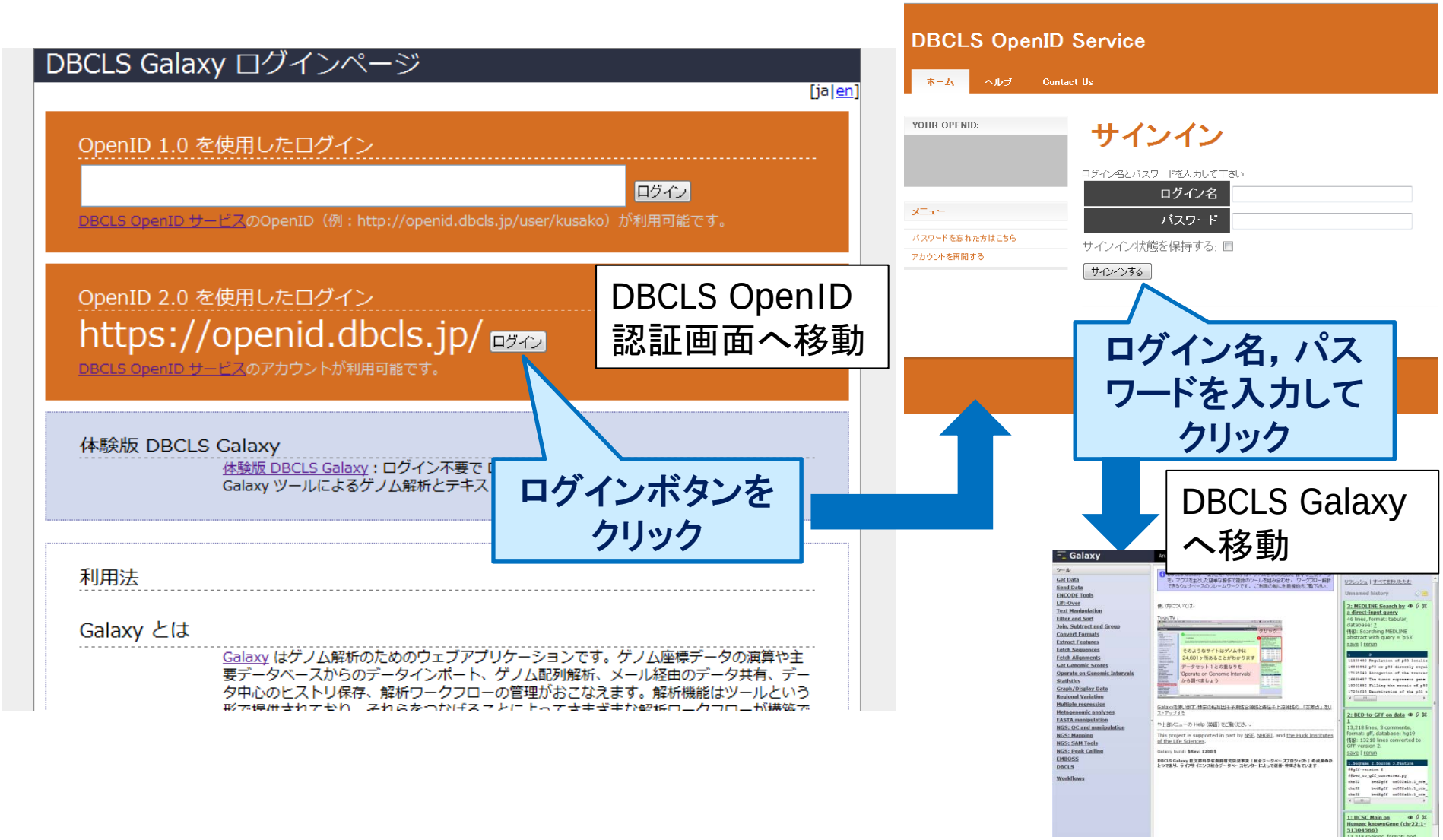
また, DBCLS で開発された
- テキストマイニング系のツールが組み込まれ,
- DBCLS OpenID で認証する

DBCLS Galaxy の利用例

実際に使ってみましょう

Step0. DBCLS OpenID でログインする
Step1. exon のデータと SNPs のデータを用意する
Step2. 「SNPs の数が多い exon を探す」という解析を
DBCLS Galaxyに用意されたツールを組み合わせて行う
Step3. ツールの組み合わせ手順を
ワークフローとして保存する
Step4. 他のユーザと解析履歴，ワークフローを共有する
Step5. 解析履歴，ワークフローを公開する

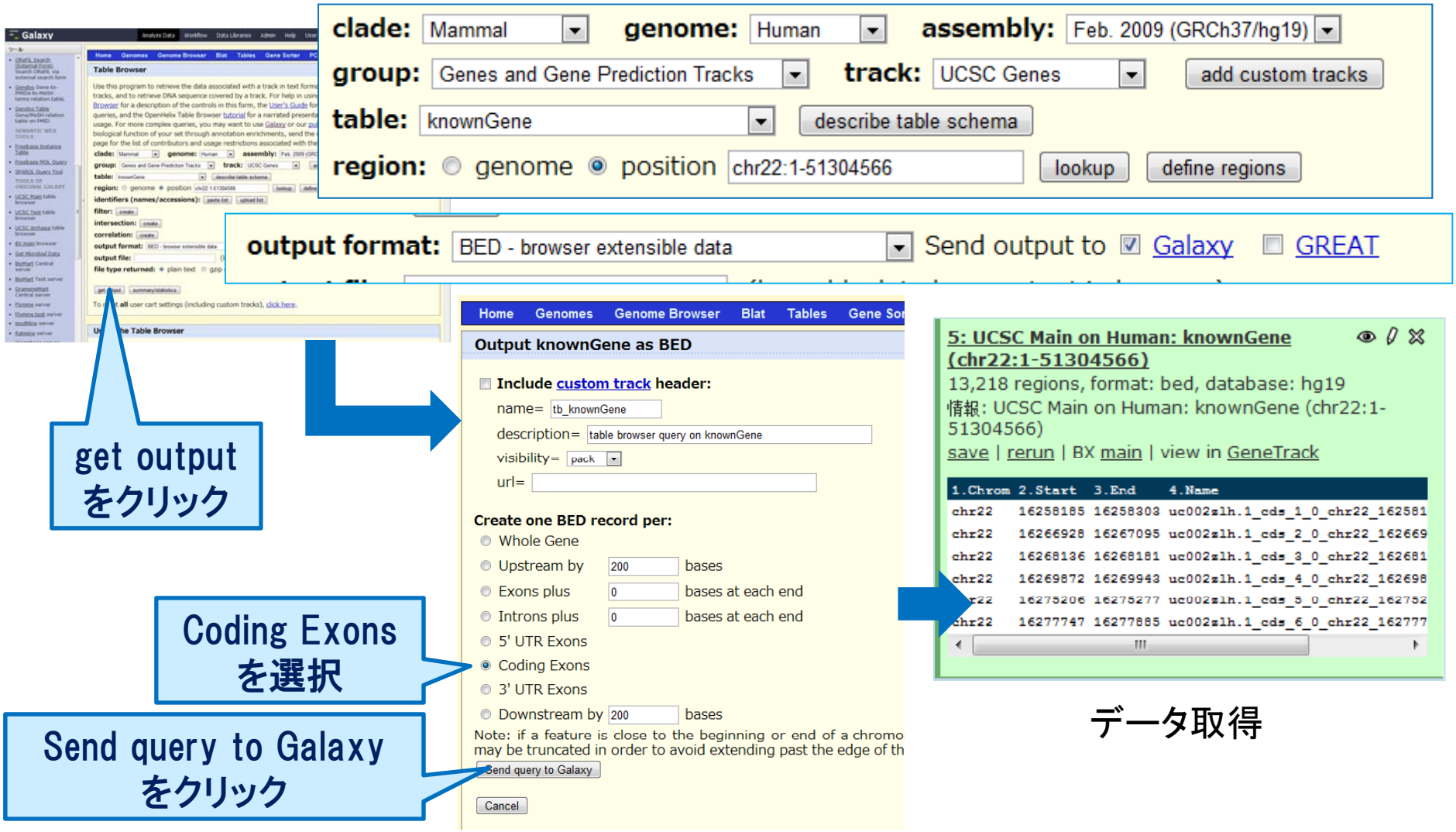
Step0. DBCLS OpenID でログインする



Step1. データの用意

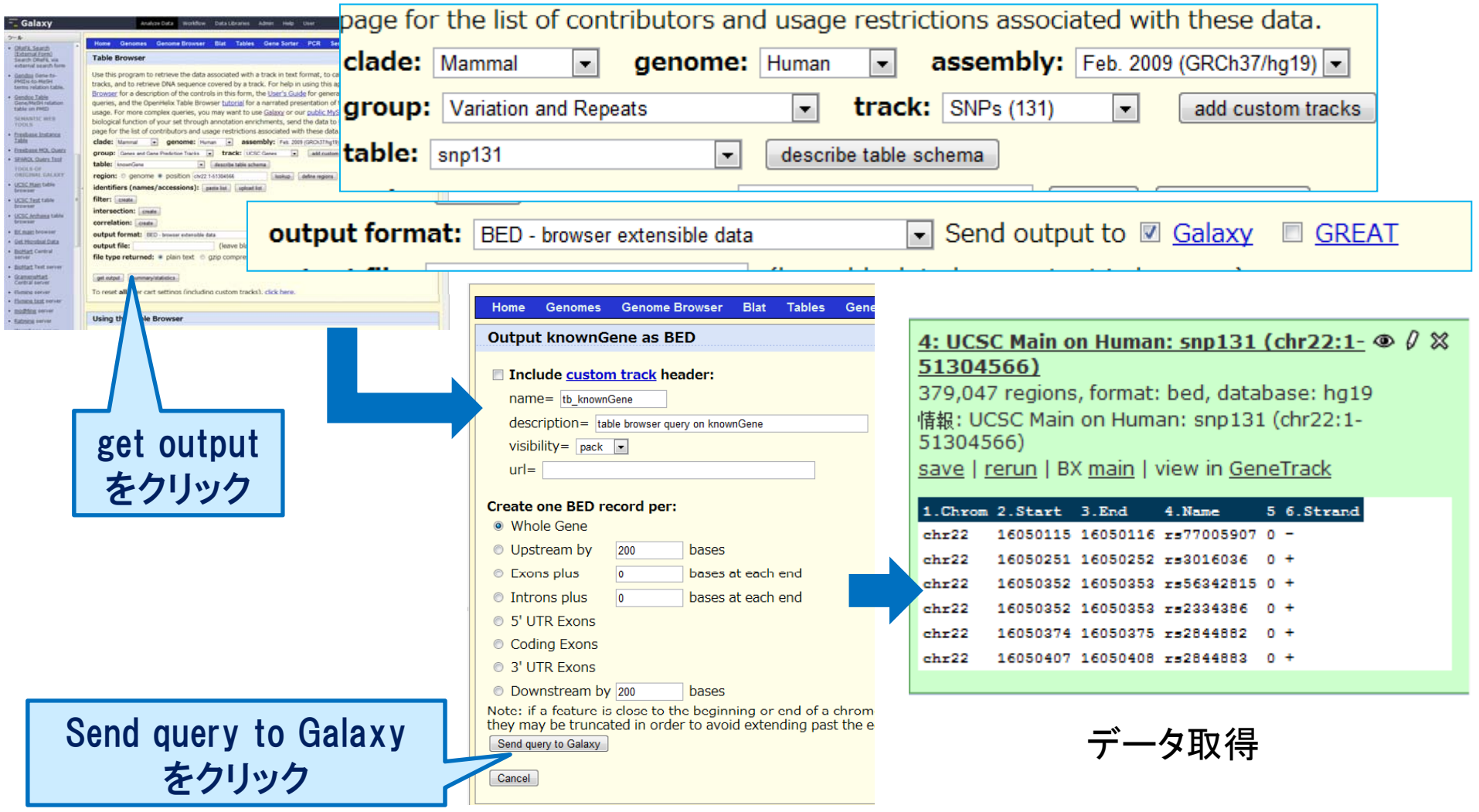
exon のデータと SNPs のデータを用意する

exonデータの取得



データ取得

SNPsデータの取得

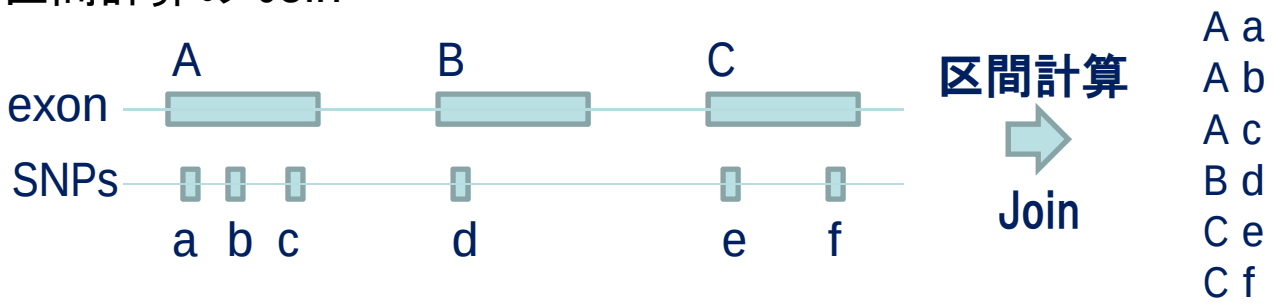

page for the list of contributors and usage restrictions associated with these data.
clade: Mammal
genome: Human
assembly: Feb. 2009 (GRCh37/hg19)
group: Variation and Repeats
track: SNPs (131)
add custom tracks
table: snp131
describe table schema
output format: BED - browser extensible data
Send output to
Galaxy
GREAT
get output をクリック
Home Genomes Genome Browser Blat Tables
Output knownGene as BED
Include custom track header:
name= tb_knownGene
description= table browser query on knownGene
visibility= pack
url=
Create one BED record per:
Whole Gene
Upstream by 200 bases
Exons plus 0 bases at each end
Introns plus 0 bases at each end
5' UTR Exons
Coding Exons
3' UTR Exons
Downstream by 200 bases
Send query to Galaxy
Cancel
Send query to Galaxy をクリック
4: UCSC Main on Human: snp131 (chr22:1-51304566)
379,047 regions, format: bed, database: hg19
情報: UCSC Main on Human: snp131 (chr22:1-51304566)
save | rerun | BX main | view in GeneTrack
1.Chrom 2.Start 3.End 4.Name 5 6.Strand
chr22 16050115 16050116 rs77005907 0 -
chr22 16050251 16050252 rs3016036 0 +
chr22 16050352 16050353 rs56342815 0 +
chr22 16050352 16050353 rs2334386 0 +
chr22 16050374 16050375 rs2844882 0 +
chr22 16050407 16050408 rs2844883 0 +
データ取得

Step2. ツールを組み合わせてデータ解析

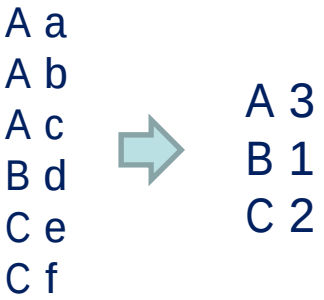
「SNPs の数が多い exon を探す」という解析を
DBCLS Galaxyに用意されたツールを組み合わせて行う

使うツール

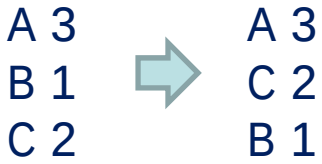
1.区間計算の Join



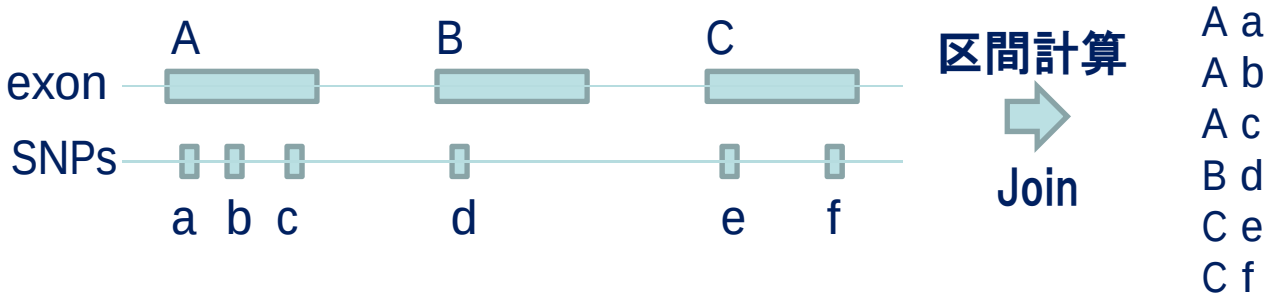
2. グループ化+出現数カウント

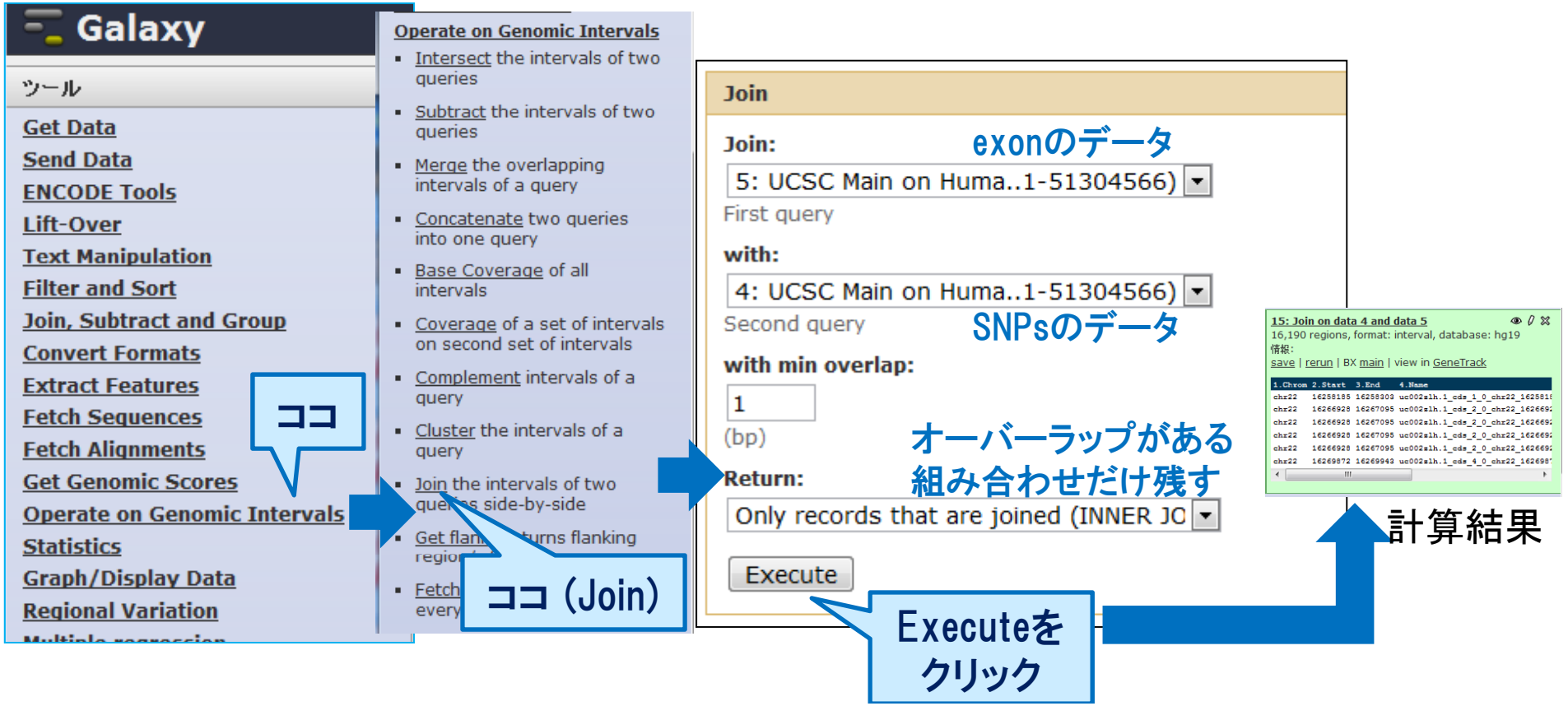


3. ソート

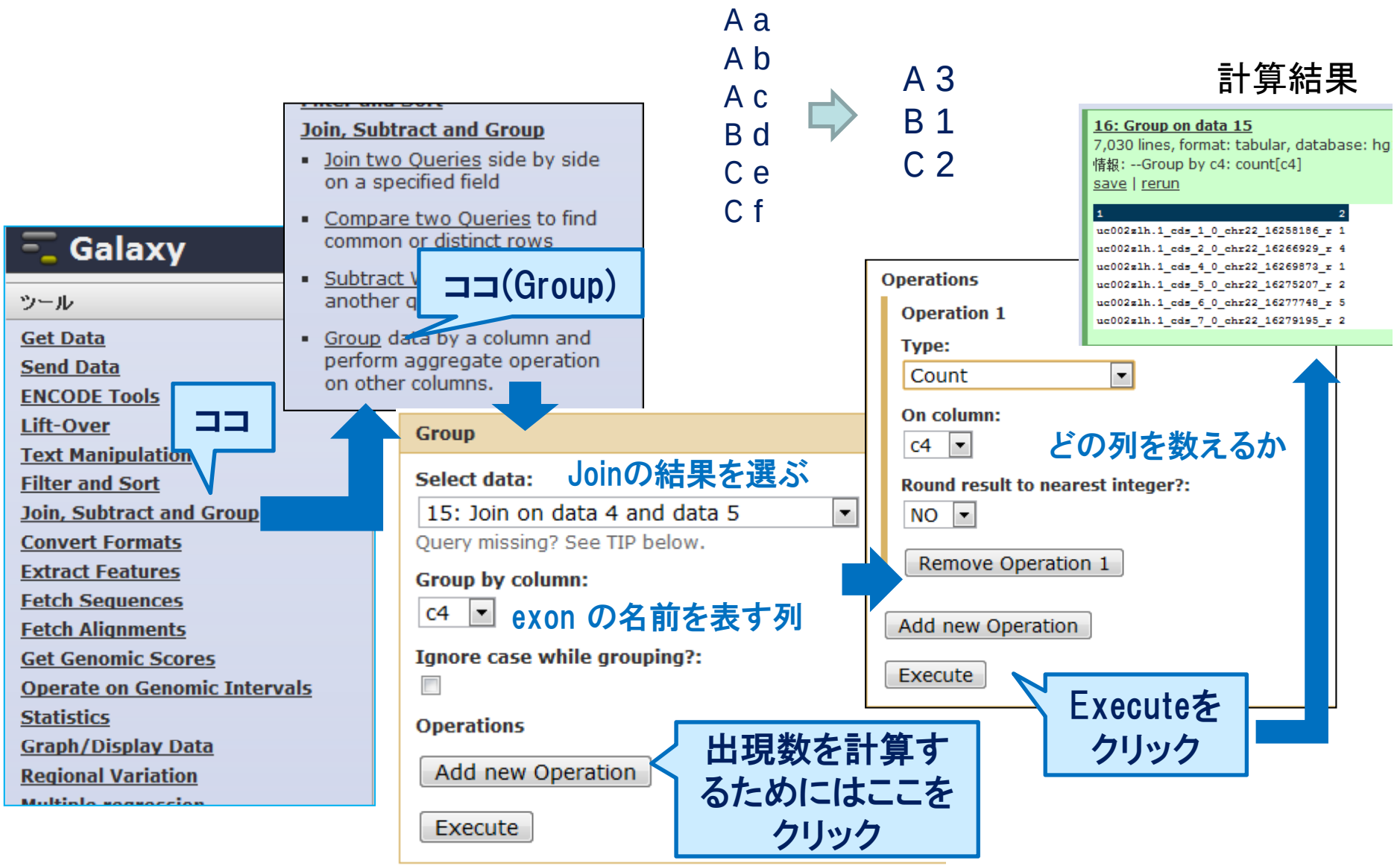


Step2-1区間計算の Join





Step2-2 グループ化＋出現数カウント

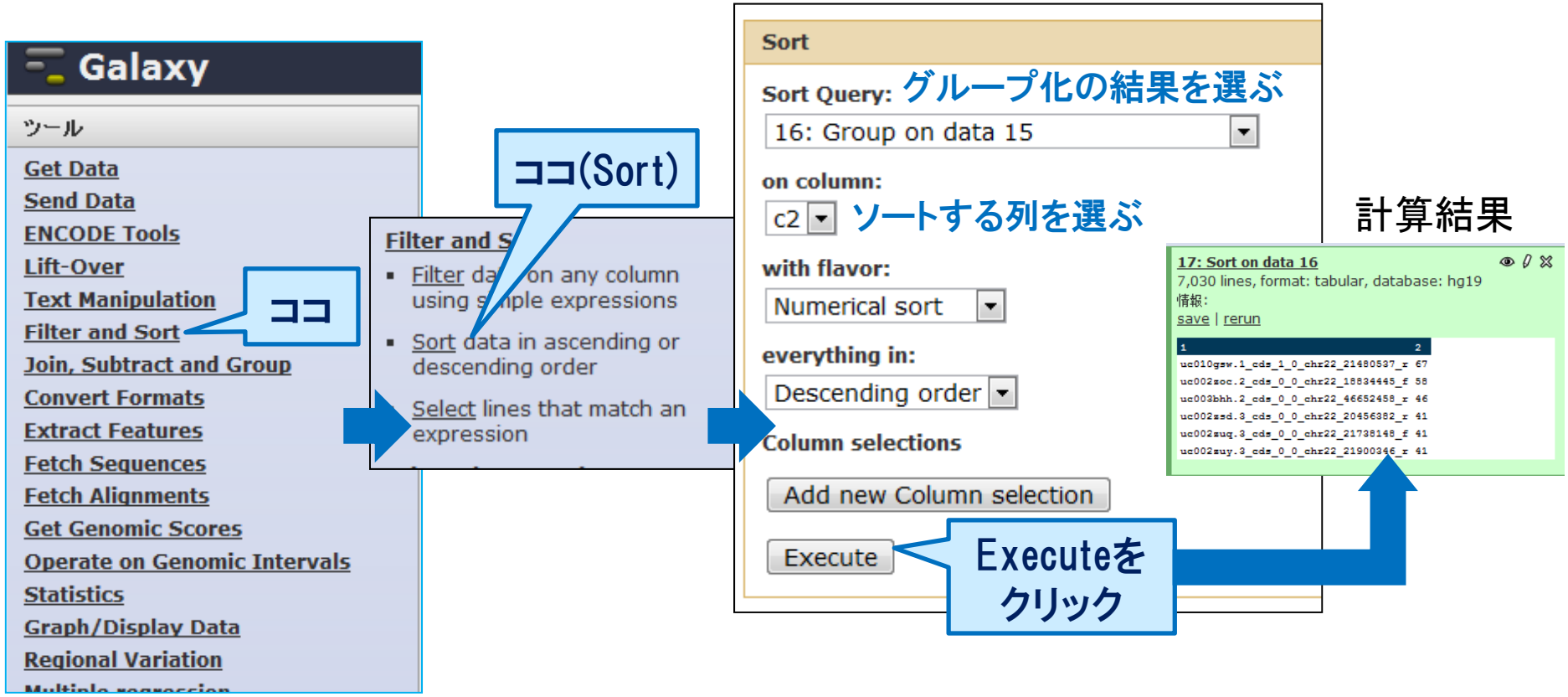


Step2-3 ソート

A 3
B 1
C 2

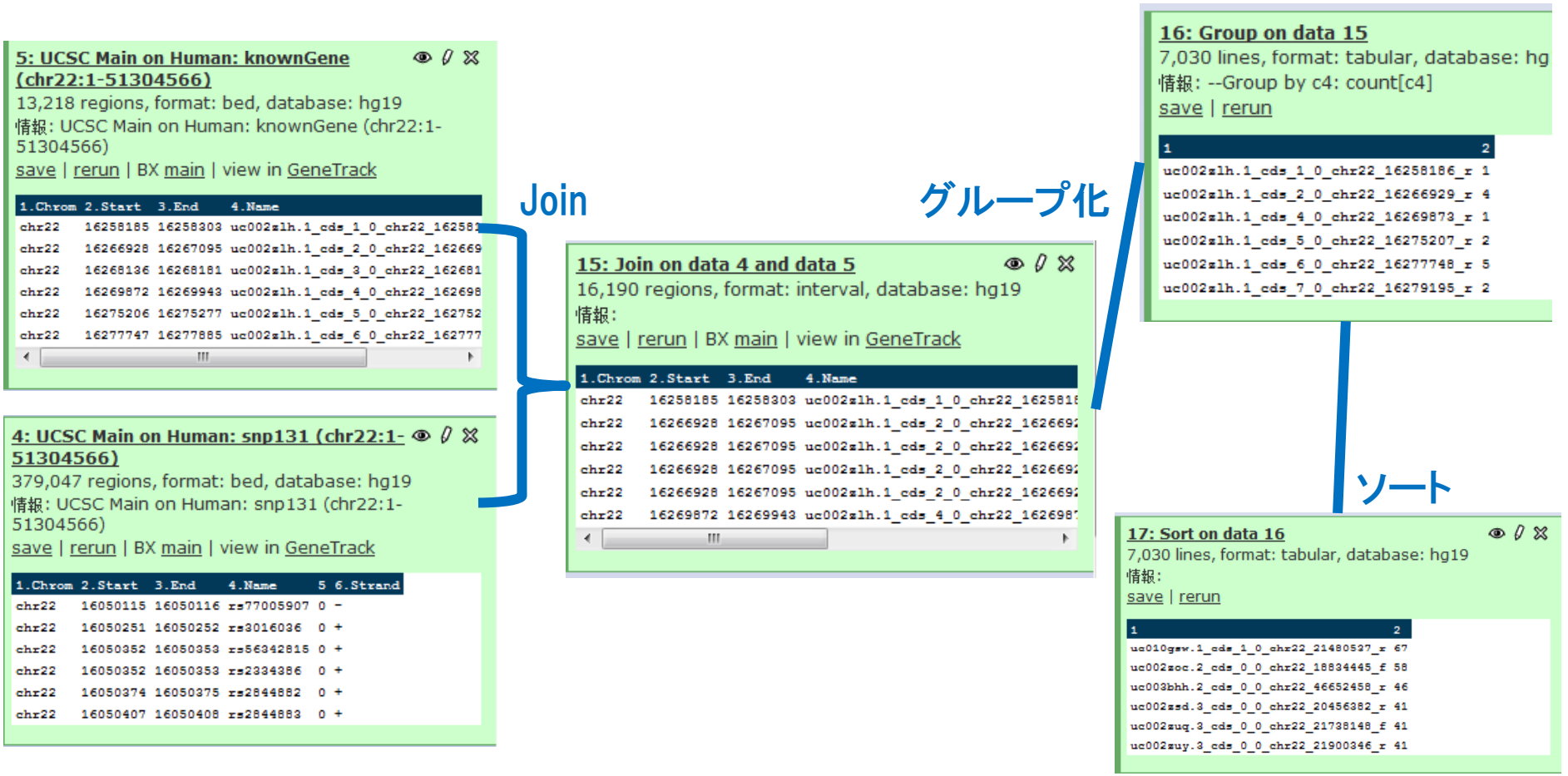
⇒

A 3
C 2
B 1



Step2. ツールを組み合わせてデータ解析

「SNPs の数が多い exon を探す」という解析を
DBCLS Galaxyに用意されたツールを組み合わせて行う



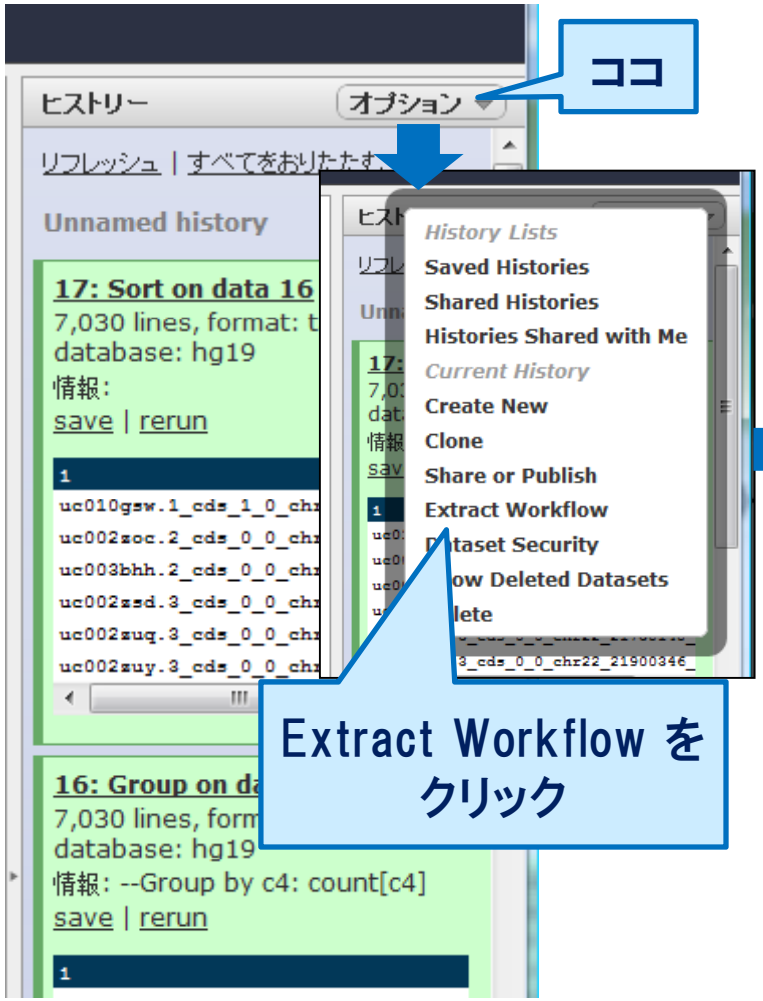
Step2. ツールを組み合わせてデータ解析

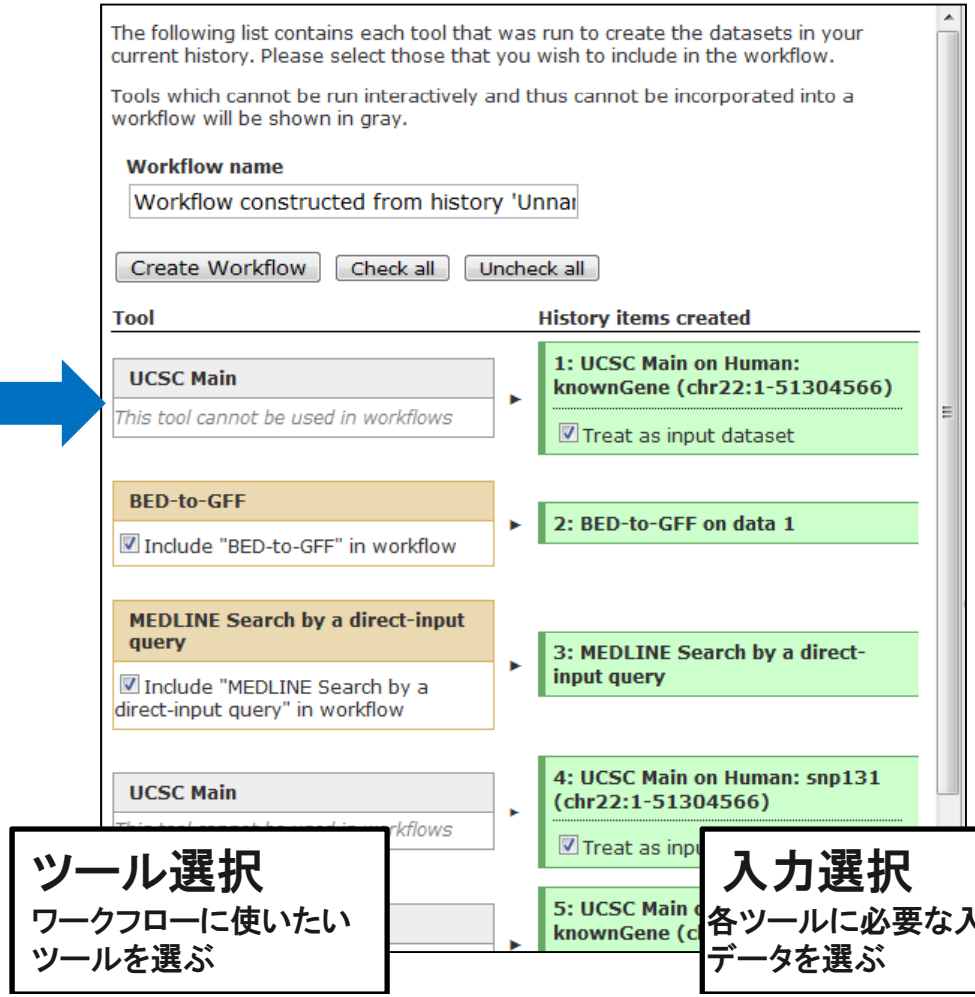
今回は触れませんが…

用意されているゲノム解析のツール群に加え，
キーワードから関連遺伝子→タンパク質構造を見たり，
キーワードから PubMed の検索をし，
それらの文献に関連する日本語の文献を検索したり
できます．

Step3. ワークフローの保存

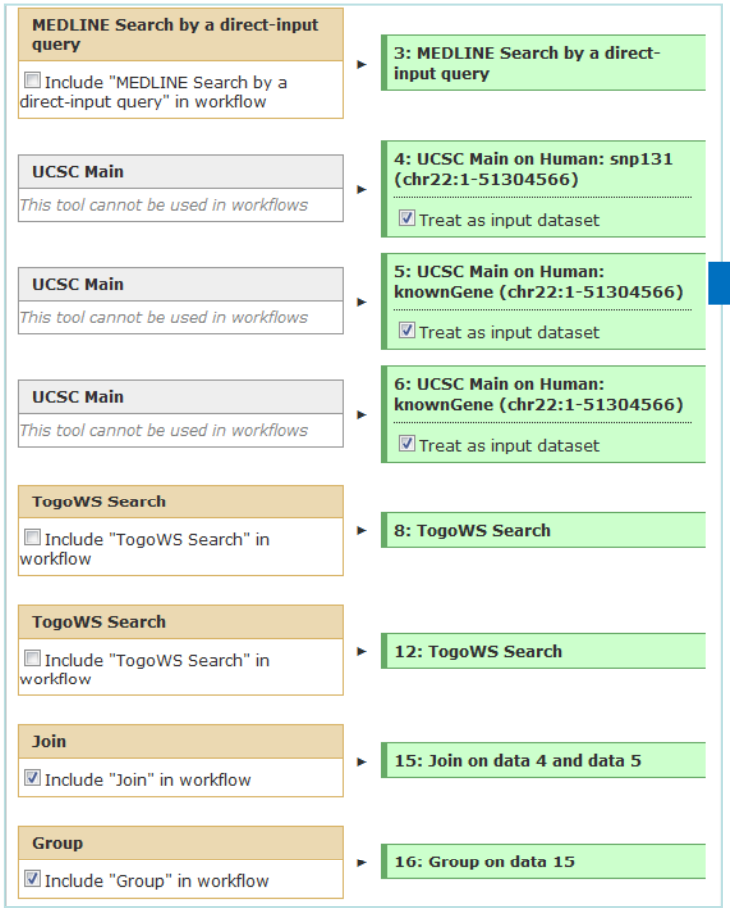
ツールの組み合わせ手順をワークフローとして保存する



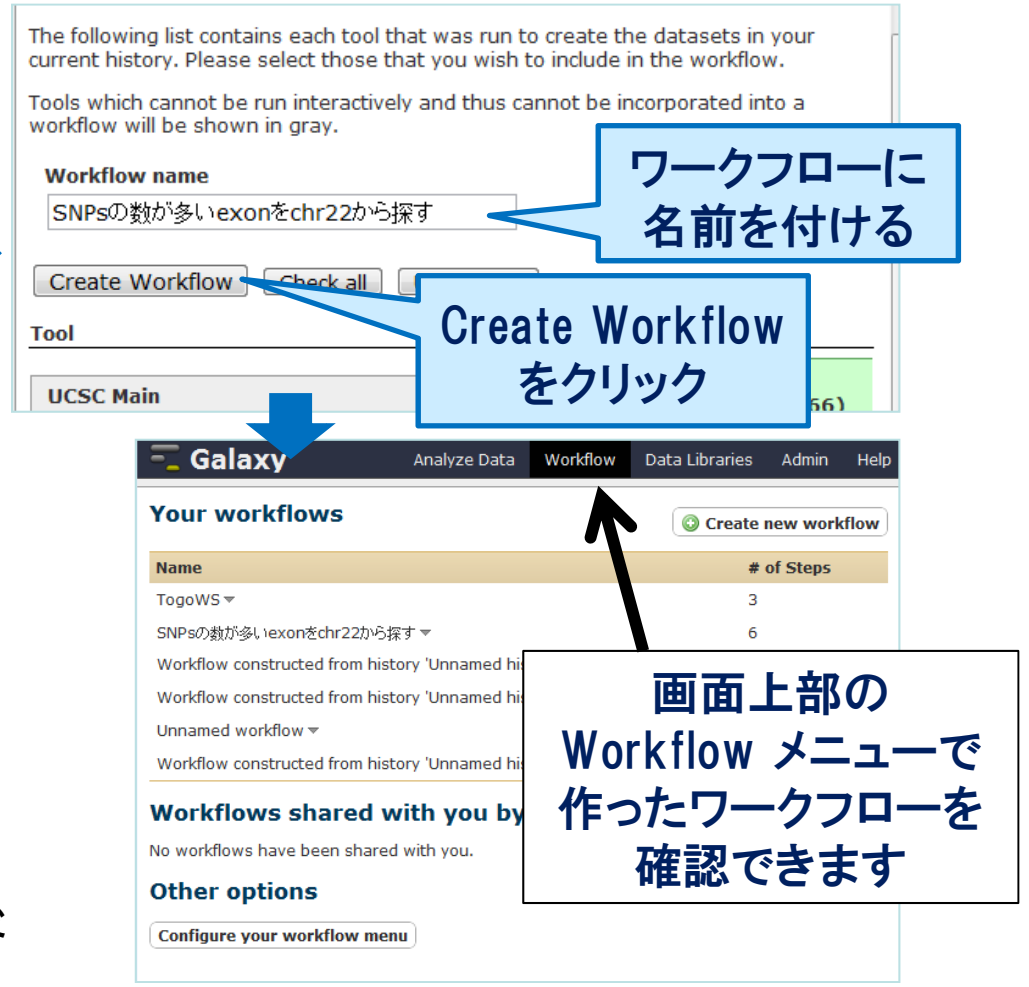


Step3. ワークフローの保存

ツールの組み合わせ手順をワークフローとして保存する

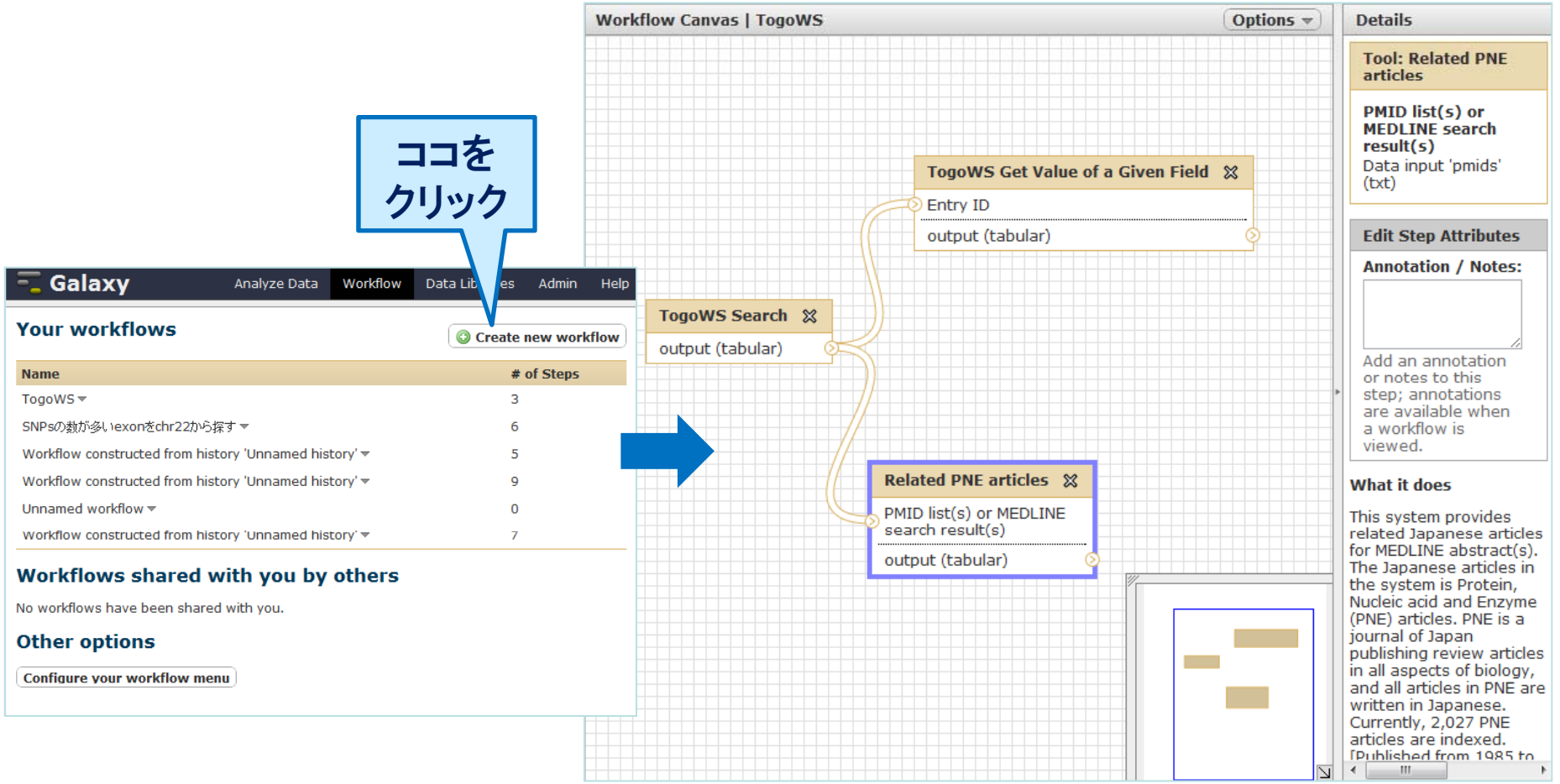


チェックボックスでツールとデータを選択後



ワークフローエディタ

今回は触れませんが…
ヒストリーを使わずワークフローを作るエディタも付いてます



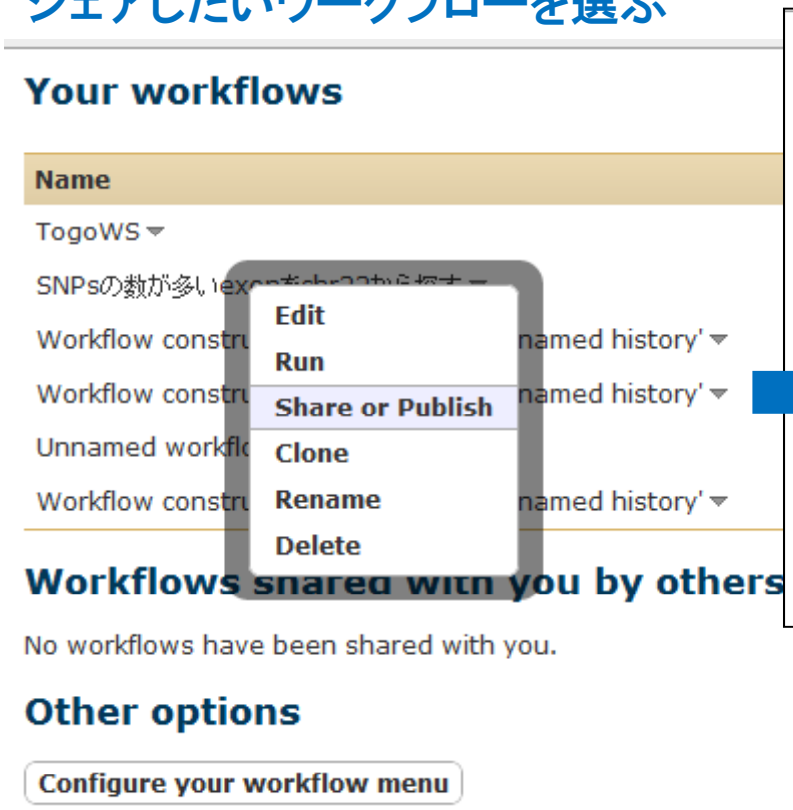
Step4. 他のユーザとの共有

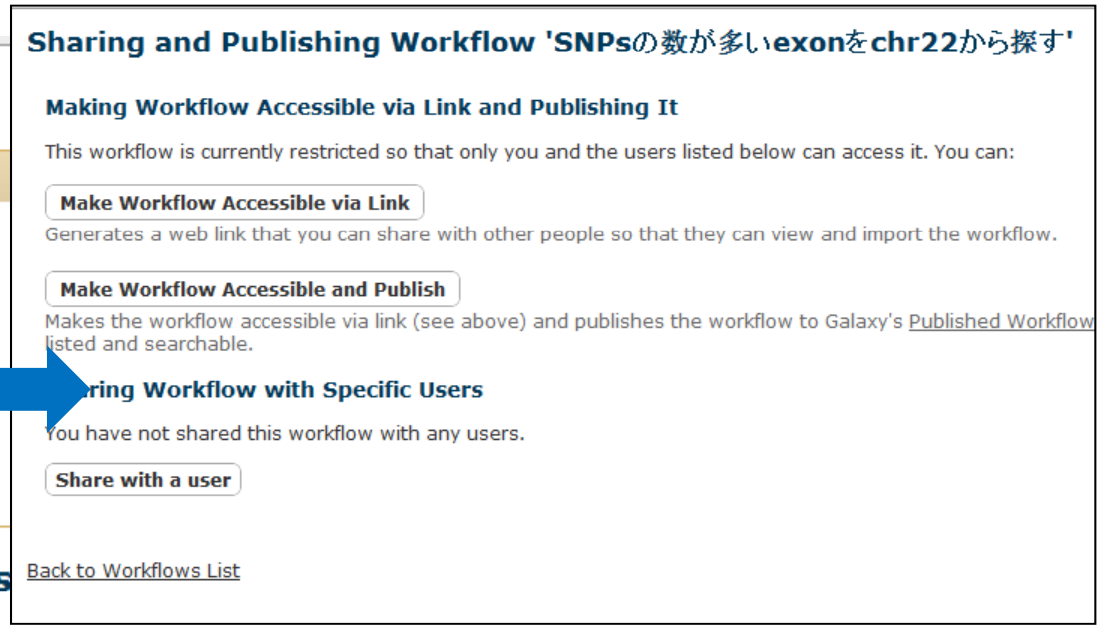
特定のユーザと解析履歴を共有する

Step4. 他のユーザとの共有

特定のユーザとワークフローを共有する

ワークフロー画面から,
シェアしたいワークフローを選ぶ

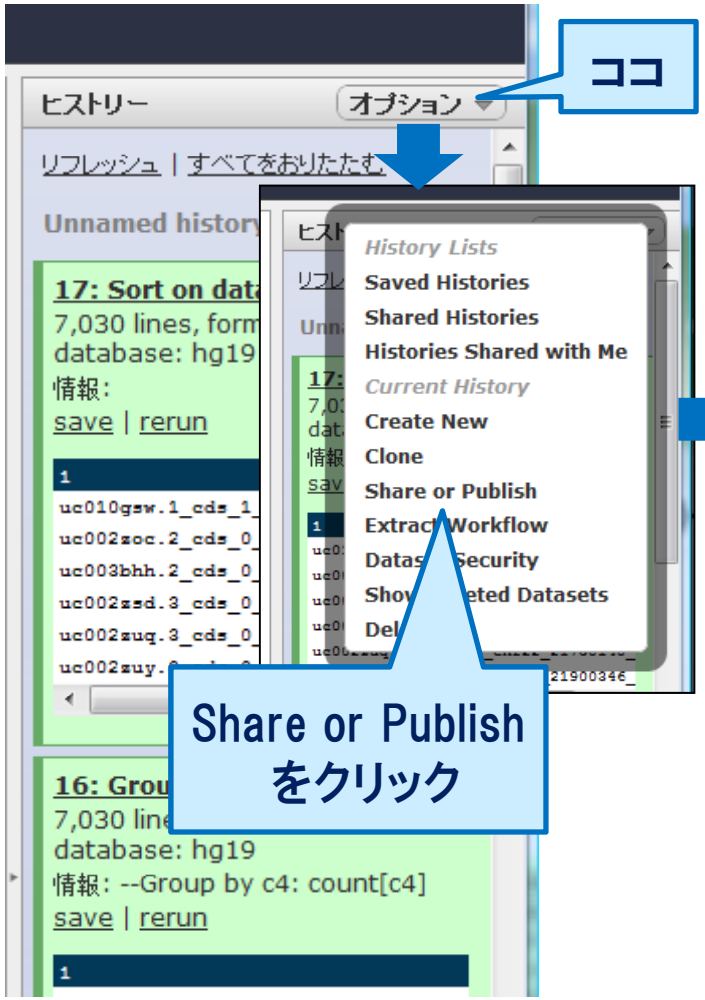


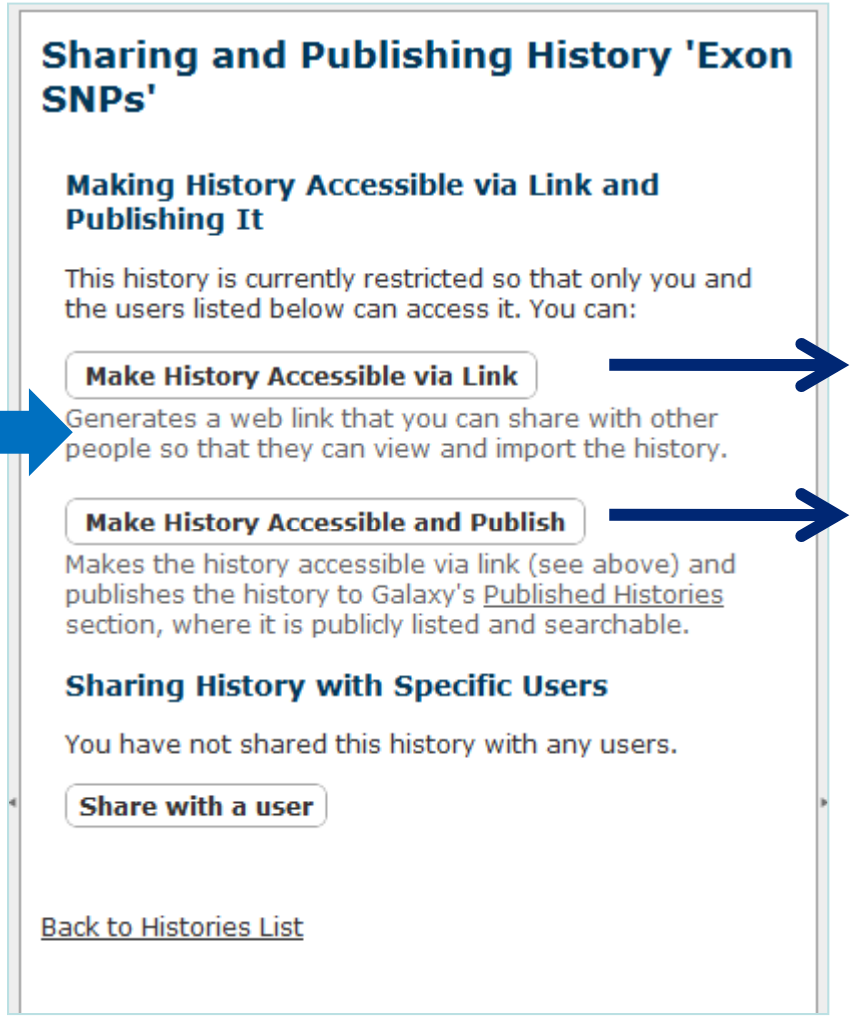


あとはヒストリー(計算履歴)と同様

Step5. 解析履歴，ワークフローの公開

解析履歴を公開する





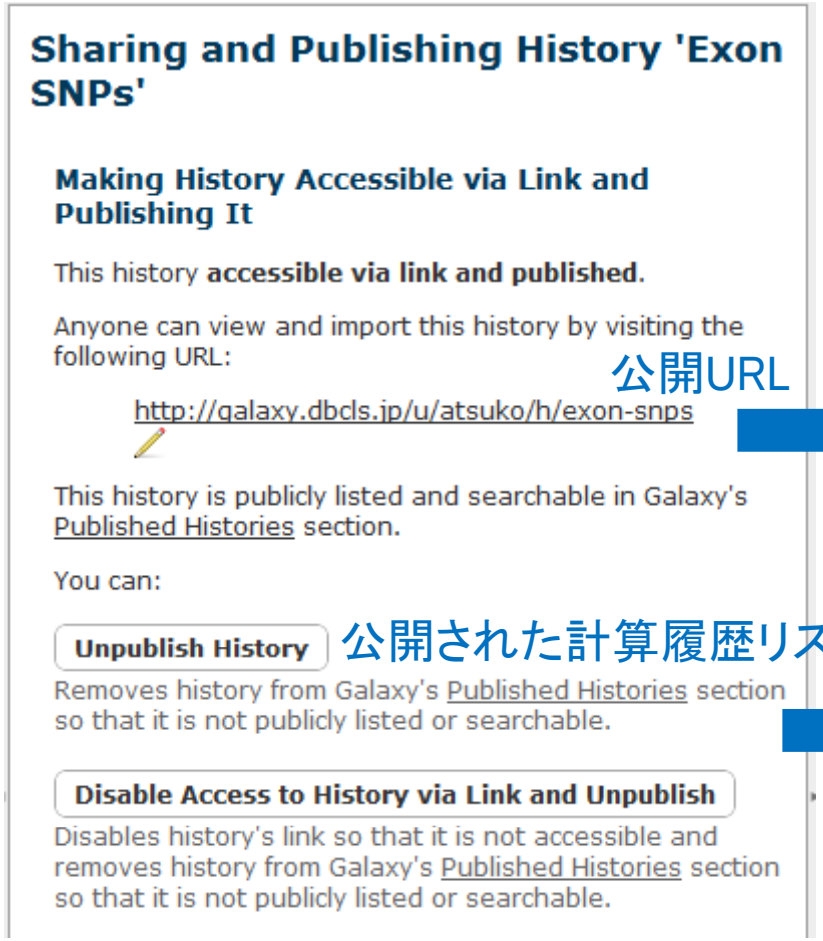
公開のための
URL作成

公開のための
URL作成と
公開リストへ
掲載

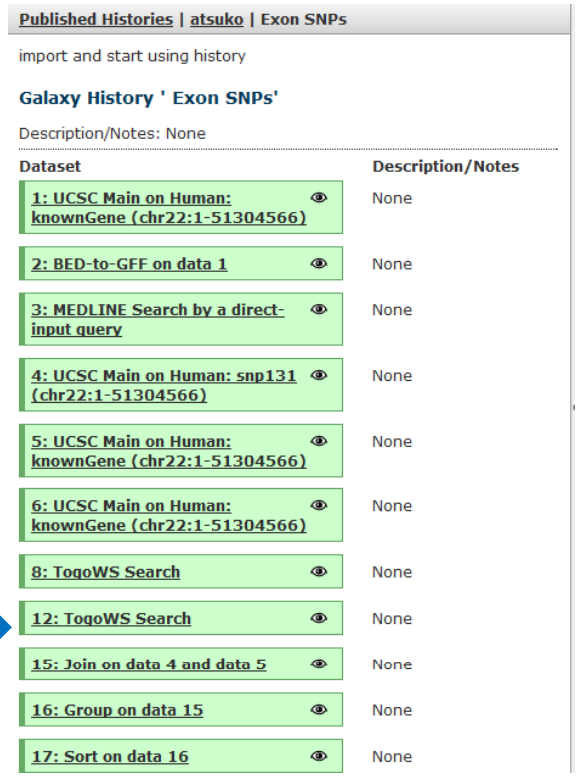
Step5. 解析履歴，ワークフローの公開

解析履歴を公開する

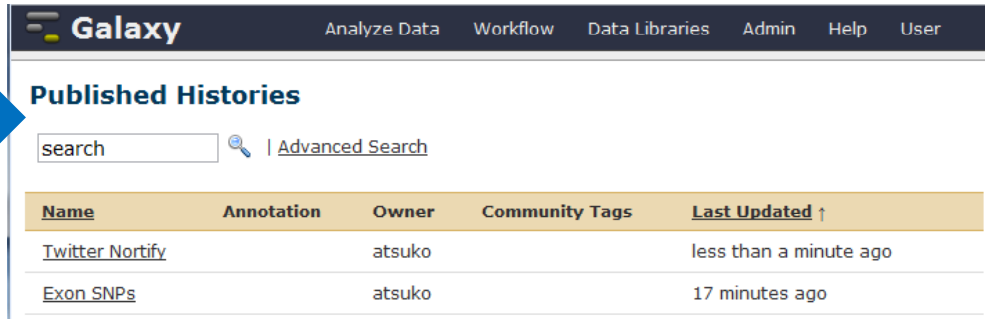
公開のためのURL作成と公開リストへ掲載



公開URL



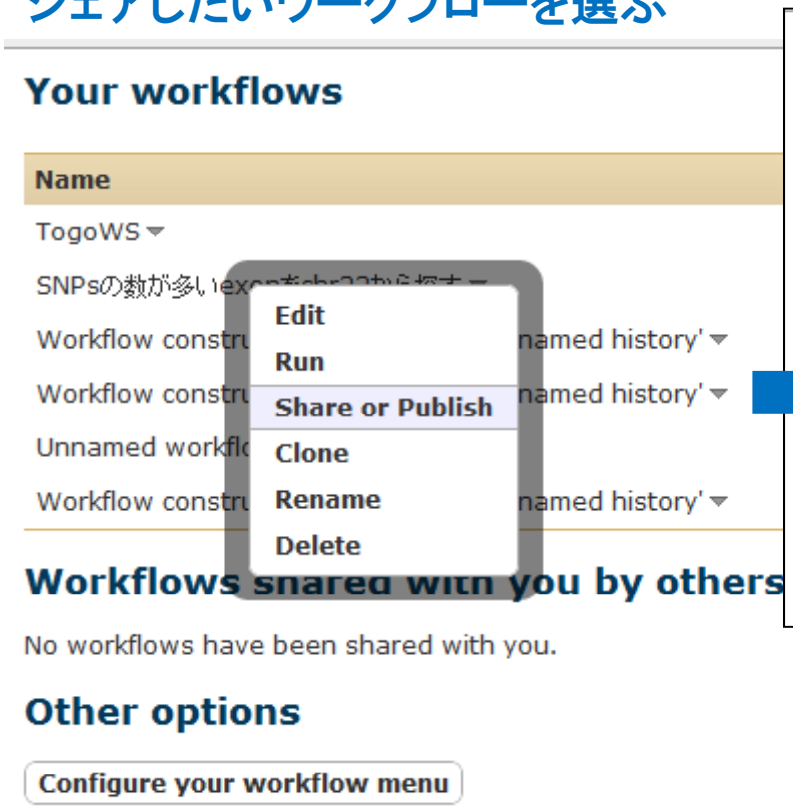
公開された計算履歴リスト

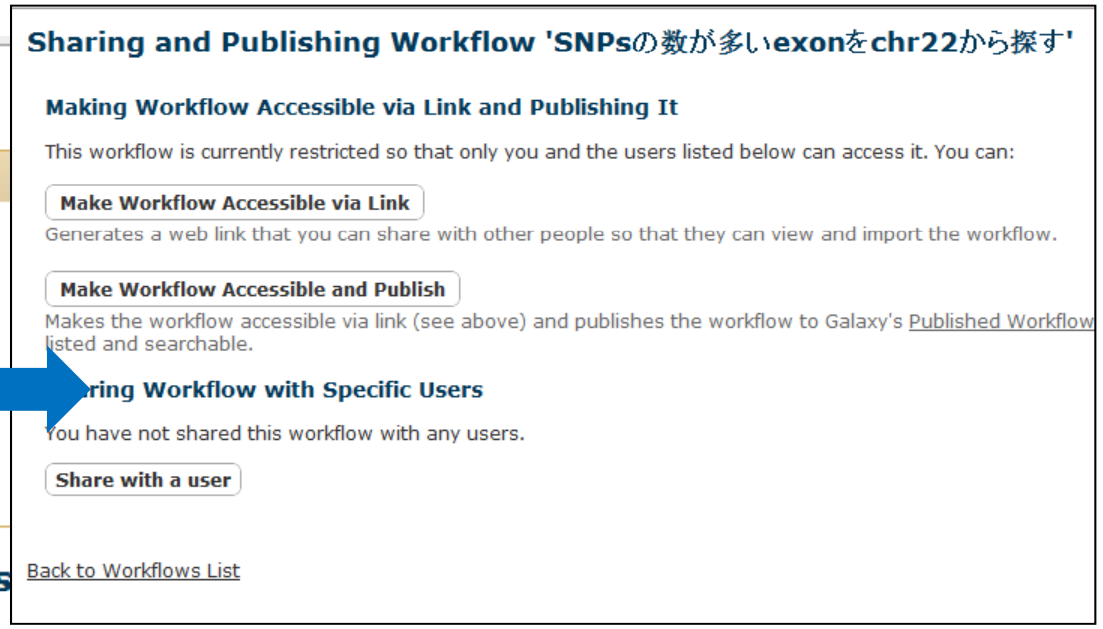


Step5. 解析履歴，ワークフローの公開

ワークフローを公開する

ワークフロー画面から，
シェアしたいワークフローを選ぶ





あとはヒストリー(計算履歴)と同様

まとめ

DBCLS Galaxy はゲノムなどの生物学データを対象とした，解析ツール組み合わせインタフェイスで，

これを使うと，

・用意されたツールを組み合わせて
　生物学データを**解析，**

・ツールの組み合わせや計算履歴の**保存**

・ユーザ間の計算履歴・ワークフローの**共有，**

・計算履歴・ワークフローの**公開**ができます．

今後の予定

データライブラリを整備

LSDBの他のツールとの連携＆組み込み

日本語ドキュメントの充実

組み込みツール，ドキュメント化などについて，
ぜひご意見・ご要望をお寄せ下さい．
support@dbcls.rois.ac.jp