2016年度第3回
バイオインフォマティクス実習

HTSeqでマッピングしたシーケンスタグのカウント

前回

- bowtieを使ってSRR1805875のデータ（サイズを縮小）をマッピング
- sam→bam形式にフォーマット変換
- データのソートとインデックスファイルの作成
- IGVで可視化

今回

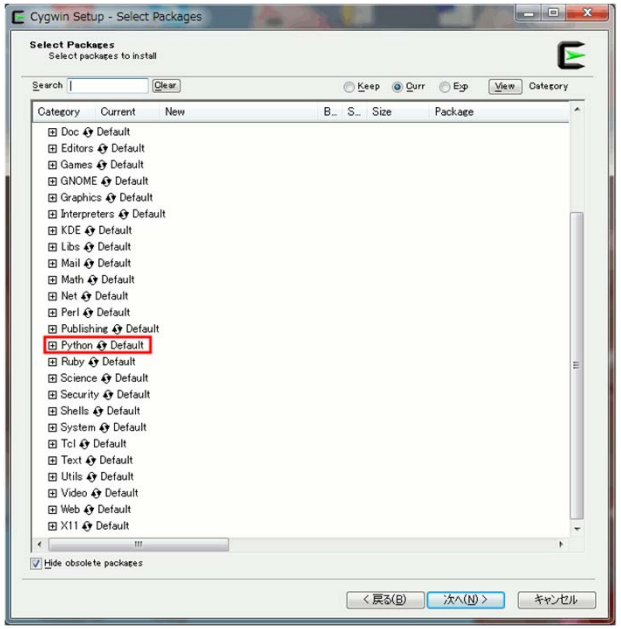
- HTSeqを使ってタグを計測→タグ数の数値データを取得する
- プログラム言語Pythonを実行する

Python

- グイド・ヴァンロッサムが開発した汎用プログラム言語の一つ
- 標準ライブラリや様々な用途に使える専用の解析用ライブラリが充実している

ウィンドウズPCでの実行

- cygwin をインストール時にpythonのパッケージをチェック



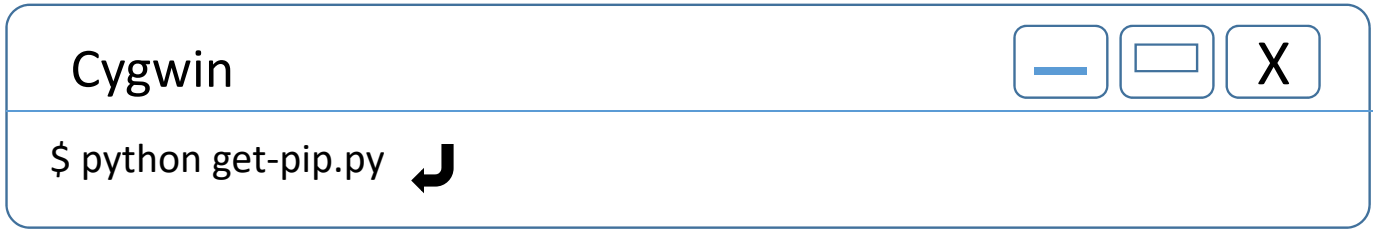
- 実習室用PCにはインストール済み
  cygwinコンソール上で実行できます

Pythonのパッケージ管理

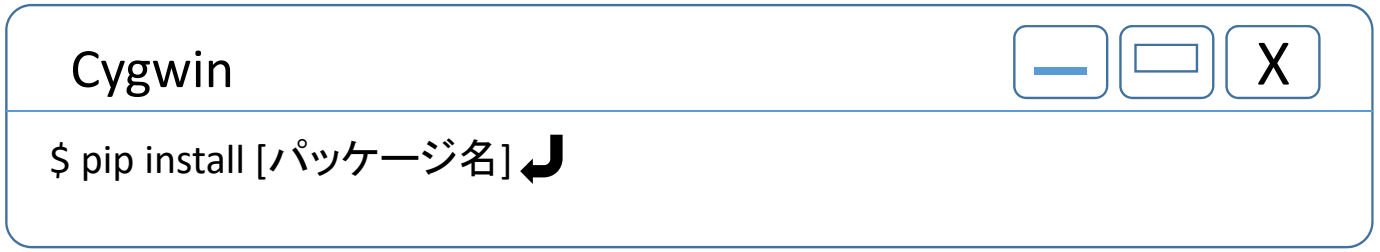
- Pipによるパッケージの管理

get-pip.pyの最新バージョンをダウンロード

get-pip.pyを保存したディレクトリで

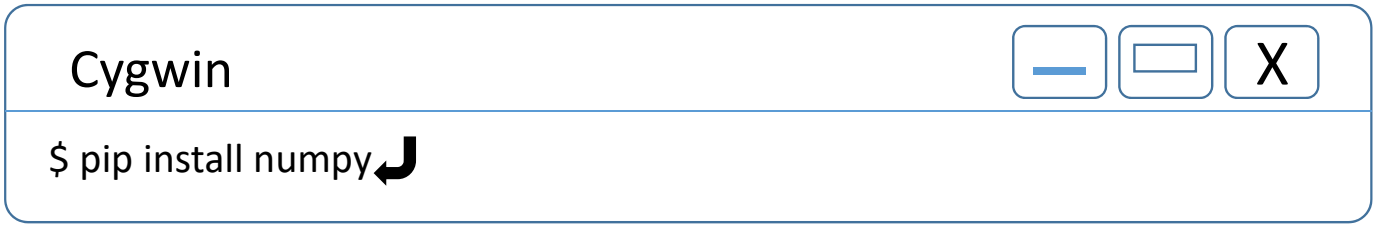


その後はコンソール上でpipコマンドでパッケージをインストールできる

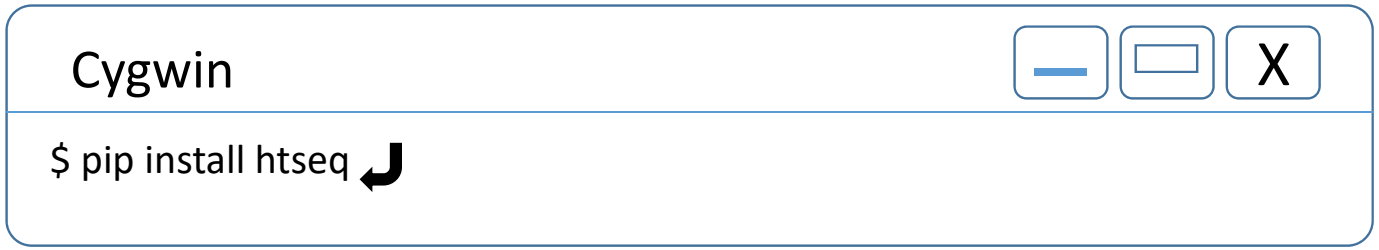


HTSeqのインストール

- numpy ライブラリのインストール



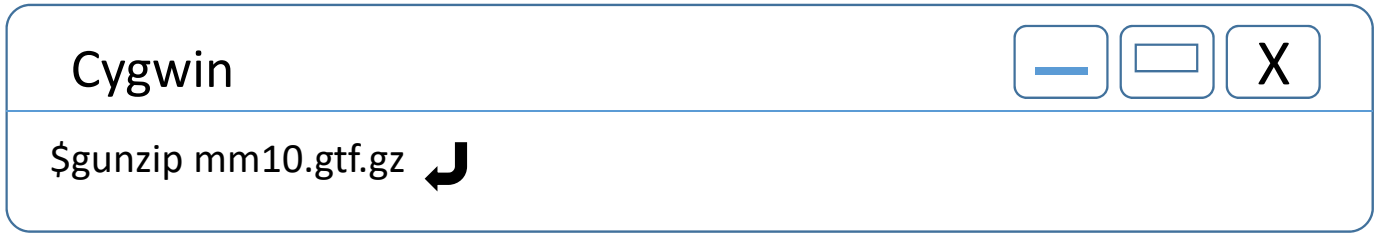
- HTSeqのインストール



annotation file の取得

https://sourceforge.net/projects/ngsonestop/files/annotation/

からmm10.gtf.gzをダウンロード



で解凍

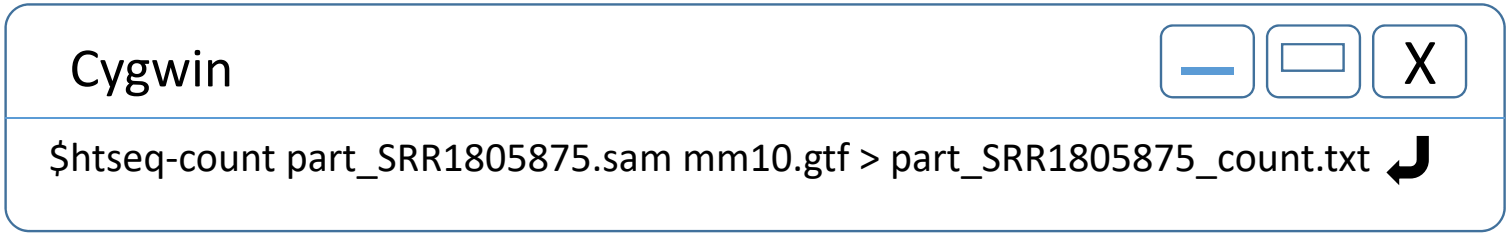
GTF ファイルフォーマット：9項目

染色体	データの種類	feature	開始	終了	score	ストランド	コドン	ID
chr1	unknown	stop_codon	3216022	3216024	.	−	.	gene_id ″Xkr4″
chr1	unknown	CDS	3216025	3216968	.	−	2	gene_id ″Xkr4″
chr1	unknown	CDS	3421702	3421901	.	−	1	gene_id ″Xkr4″
chr1	unknown	exon	3421702	3421901	.	−	.	gene_id ″Xkr4″

HTSeqによるタグ数のカウント

- htseq-countコマンド

htseq-count マッピング後のsamファイル名　アノテーション用gtfファイル名　＞　出力ファイル名



まとめ

- python のインストールと実行
- HTSeqのインストール
- HTSeqでマッピングしたシーケンスタグをカウント