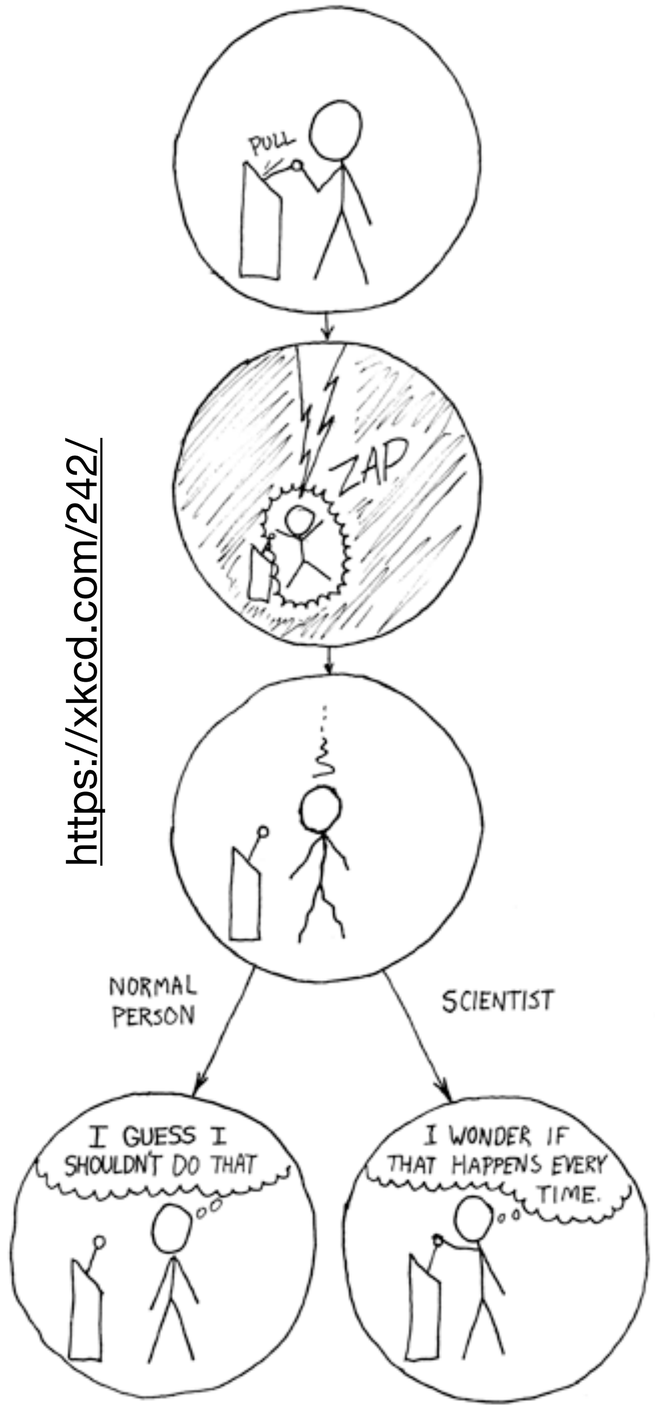
ChIP-Seq: Handling replicates



https://xkcd.com/242/

Reproducibility between replicates

- As with any high-throughput experiment, any single assay is often subject to a substantial amount of variability
- Two replicates measuring the same underlying biology should have high consistency
- Metrics that objectively assess the reproducibility of high-throughput assays are important for producing reliable scientific discoveries

Peak Calls

Nanog IP (Condition1)

Replicate1 Replicate2

Peak Calls

Pou5f1 IP (Condition2)

Replicate1 Replicate2

Peak Calls

Nanog IP (Condition1)

Replicate1 Replicate2

vs.

Peak Calls

Pou5f1 IP (Condition2)

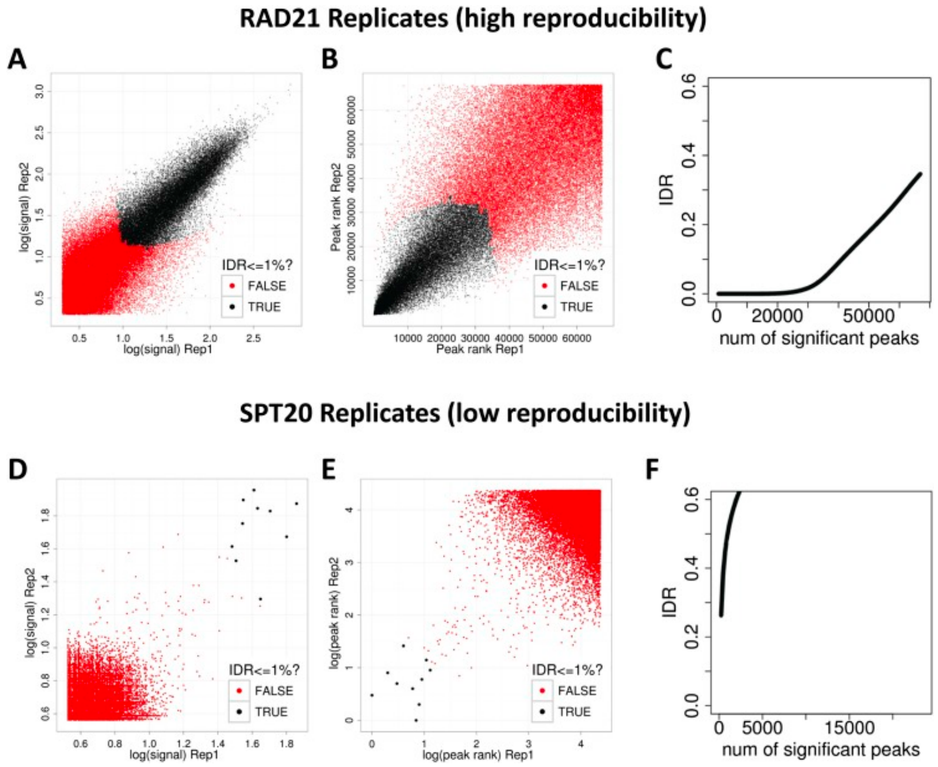
Replicate1 Replicate2

Methods for combining replicates

- Merge peak calls across replicates and then assess differences in binding regions between two groups
- Assess differential binding for pairs of samples and look for consensus set of peaks across pairs of replicate
- Identifying statistically significantly differentially bound sites using statistical routines developed in an RNA-Seq context
- …

Irreproducibility Discovery Rate (IDR)

- A framework developed by Qunhua Li and Peter Bickel's group that compares a pair of ranked lists of regions/peaks and assigns values that reflect its reproducibility



Why IDR?

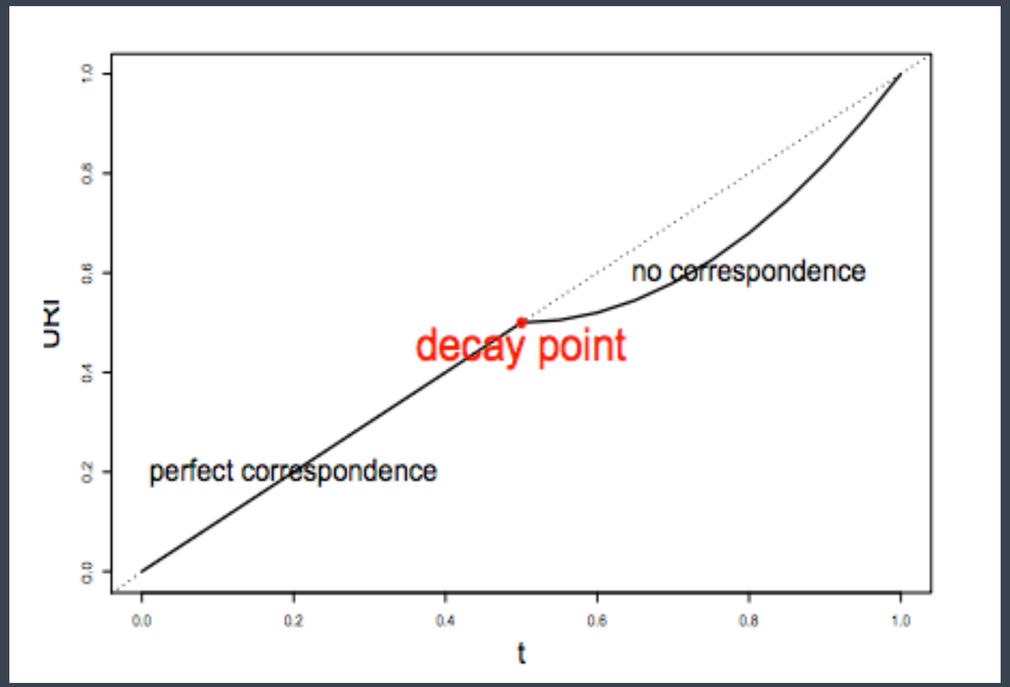
- It is **extensively used by the ENCODE and modENCODE** projects and is part of their ChIP-seq guidelines and standards.
- Avoids choices of initial cutoffs, which are not comparable for different callers. IDR does **not depend on arbitrary thresholds** and so all regions/peaks are considered
- Based on ranks, so does **not require** the input signals to be calibrated or with a **specific fixed scale** (only order matters)

Peak callers tested with IDR

- SPP - Works out of the box
- MACS1.4 - DO NOT use with IDR
- MACS2 - Works well with IDR with occasional problems of too many ties in ranks for low quality ChIP-seq data.
- HOMER - developers have a detailed pipeline and code (in beta) for IDR analysis with HOMER at https://github.com/karmel/homer-idr
- PeakSeq - Run with modified PeakSeq parameters to obtain large number of peaks
- HotSpot, MOSAiCS, GPS/GEM, …

Components of IDR

1. A **correspondence curve**: a graphical representation of matched peaks as you go down the ranked list. Qualitative, *not* adequate for selecting signals.



Components of IDR

1. A **correspondence curve**: a graphical representation of matched peaks as you go down the ranked list. Qualitative, *not* adequate for selecting signals.

2. An inference procedure: summarizes the proportion of reproducible and irreproducible signals. Quantitative, using a copula mixture model.

Components of IDR

1. A **correspondence curve**: a graphical representation of matched peaks as you go down the ranked list. Qualitative, *not* adequate for selecting signals.

2. An inference procedure: summarizes the proportion of reproducible and irreproducible signals. Quantitative, using a copula mixture model.

    - What proportion of identifications have a poor correspondence, i.e. falling into ”noise”?

Components of IDR

1. A **correspondence curve**: a graphical representation of matched peaks as you go down the ranked list. Qualitative, *not* adequate for selecting signals.

2. An inference procedure: summarizes the proportion of reproducible and irreproducible signals. Quantitative, using a copula mixture model.

   - What proportion of identifications have a poor correspondence, i.e. falling into ”noise”?
   - How consistent are the identifications before reaching breakdown?

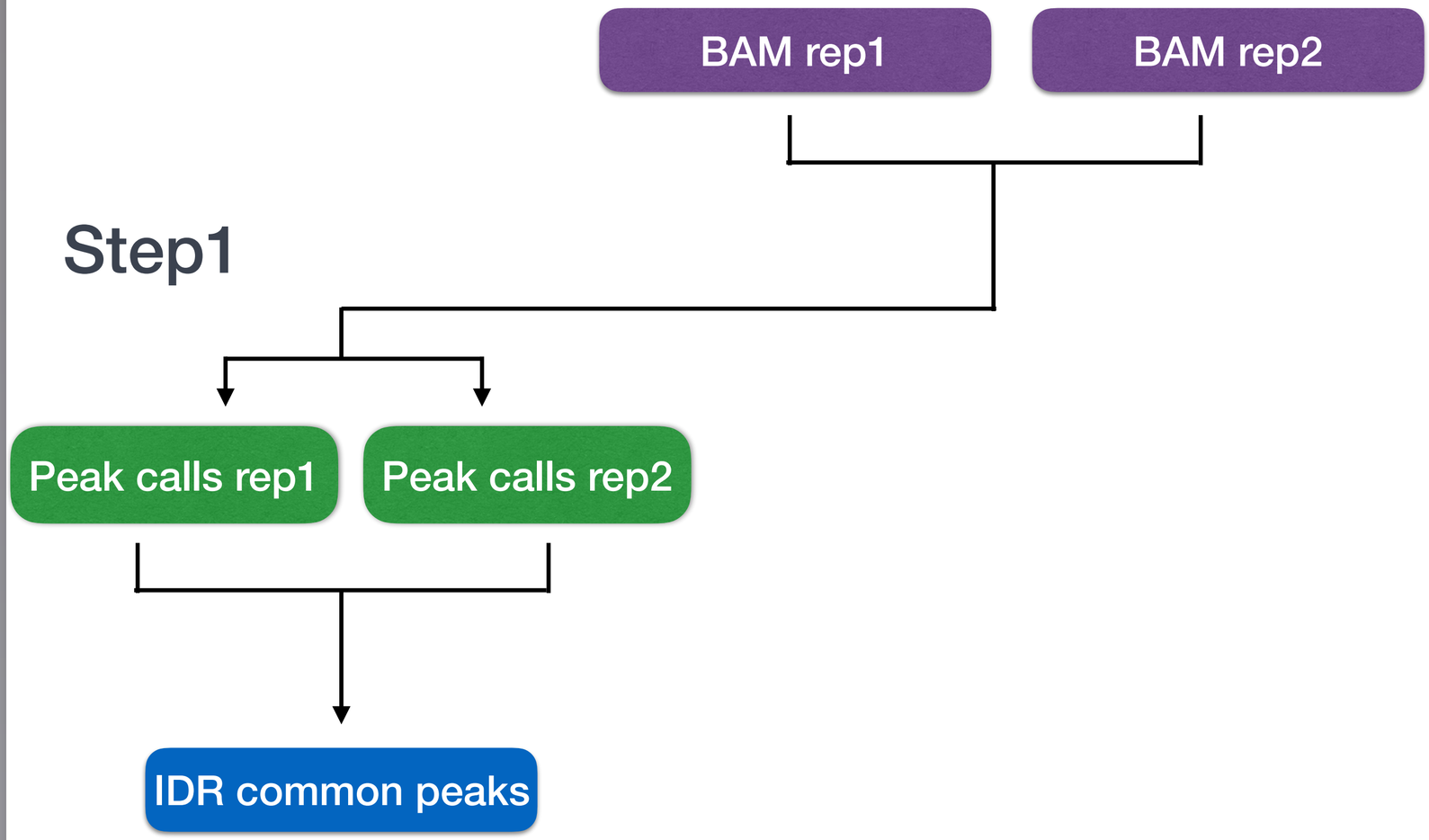
Components of IDR

1. A **correspondence curve**: a graphical representation of matched peaks as you go down the ranked list. Qualitative, *not* adequate for selecting signals.

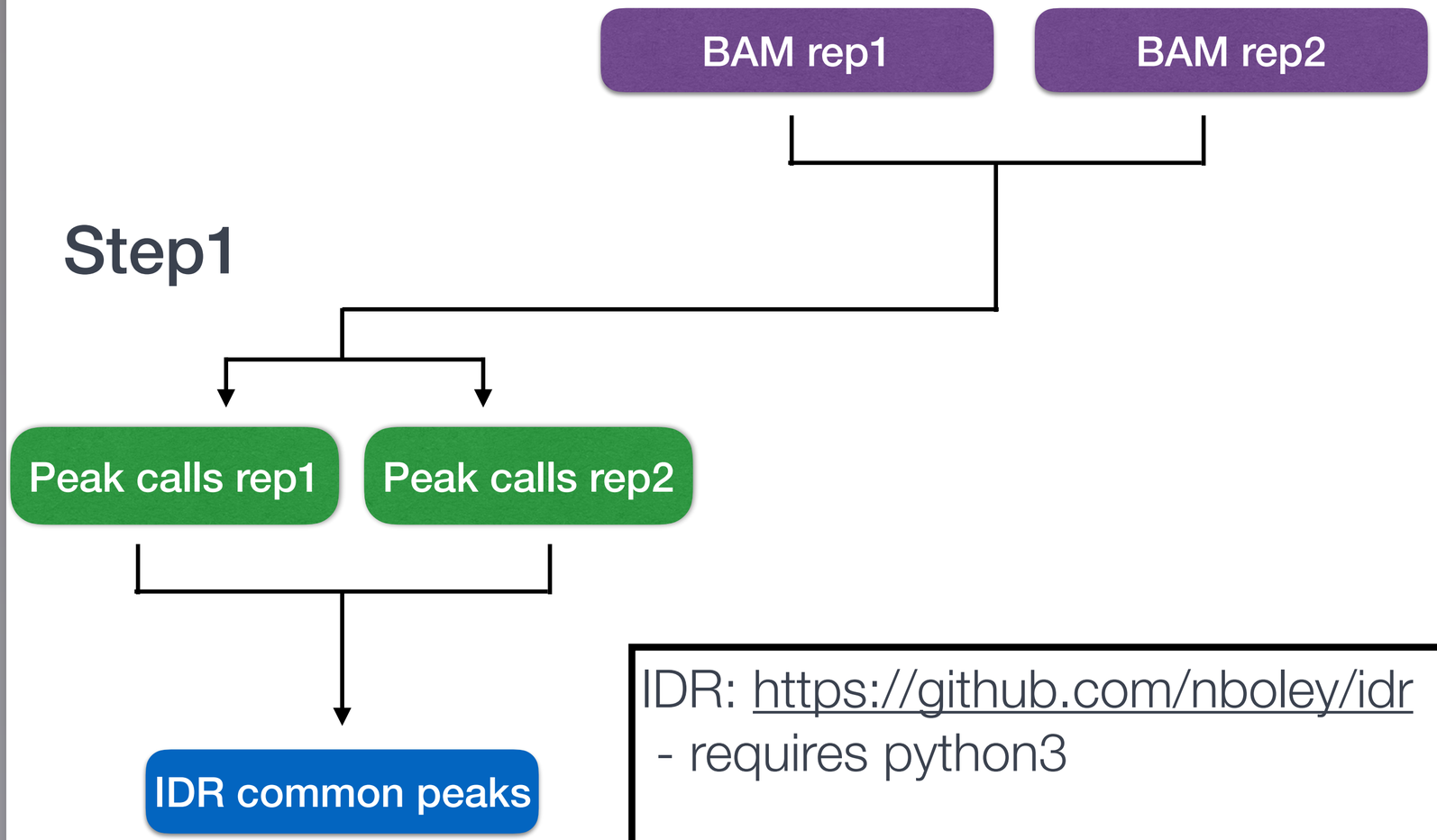
2. An inference procedure: summarizes the proportion of reproducible and irreproducible signals. Quantitative, using a copula mixture model.

3. **Irreproducible Discovery Rate** (IDR): Derive a significance value from the inference procedure (#2) in a fashion similar to FDR, and can be used to control the level of irreproducibility rate when selecting signals.

    - i.e. 0.05 IDR means that peak has a 5% chance of being an irreproducible discovery



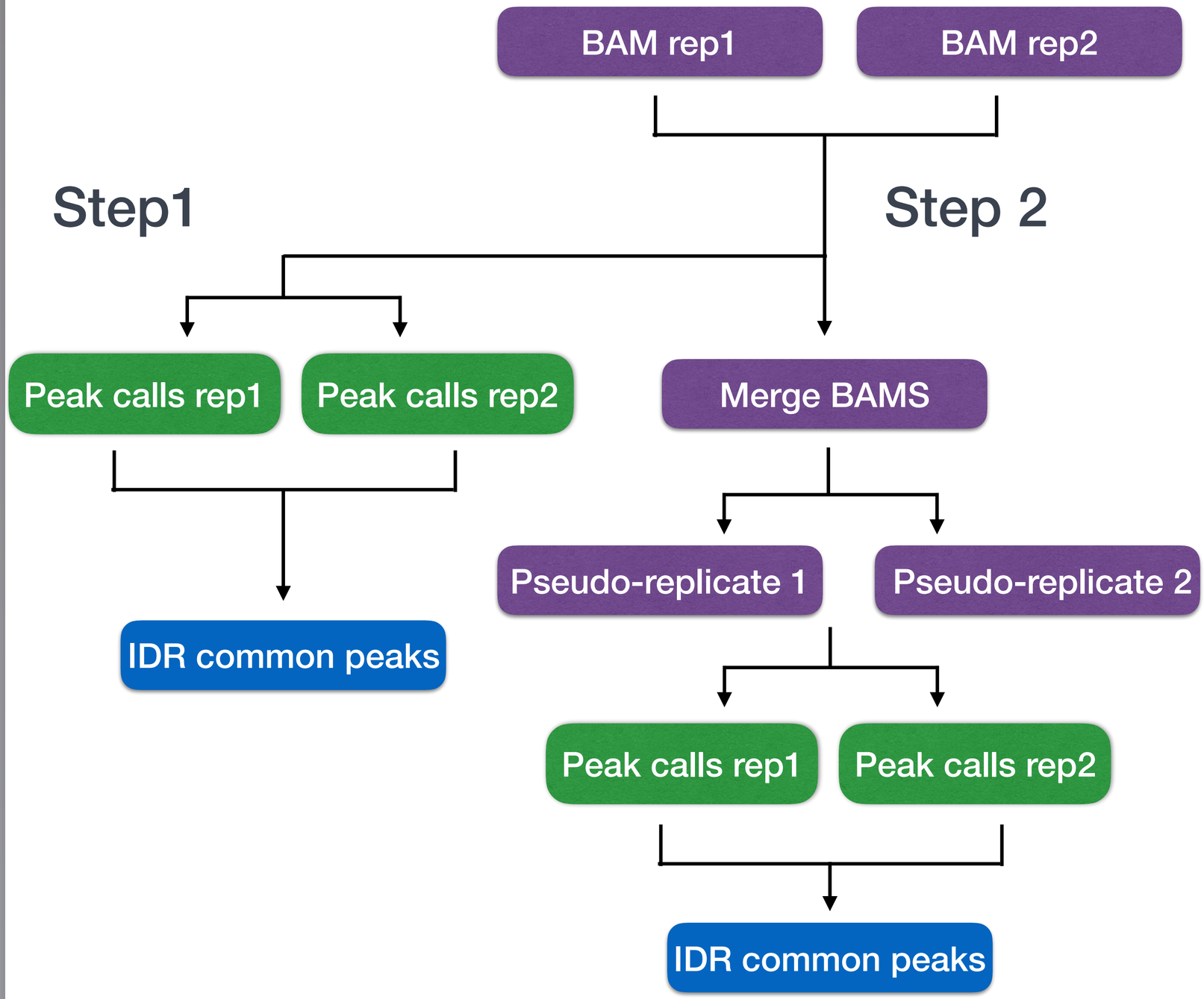
IDR pipeline



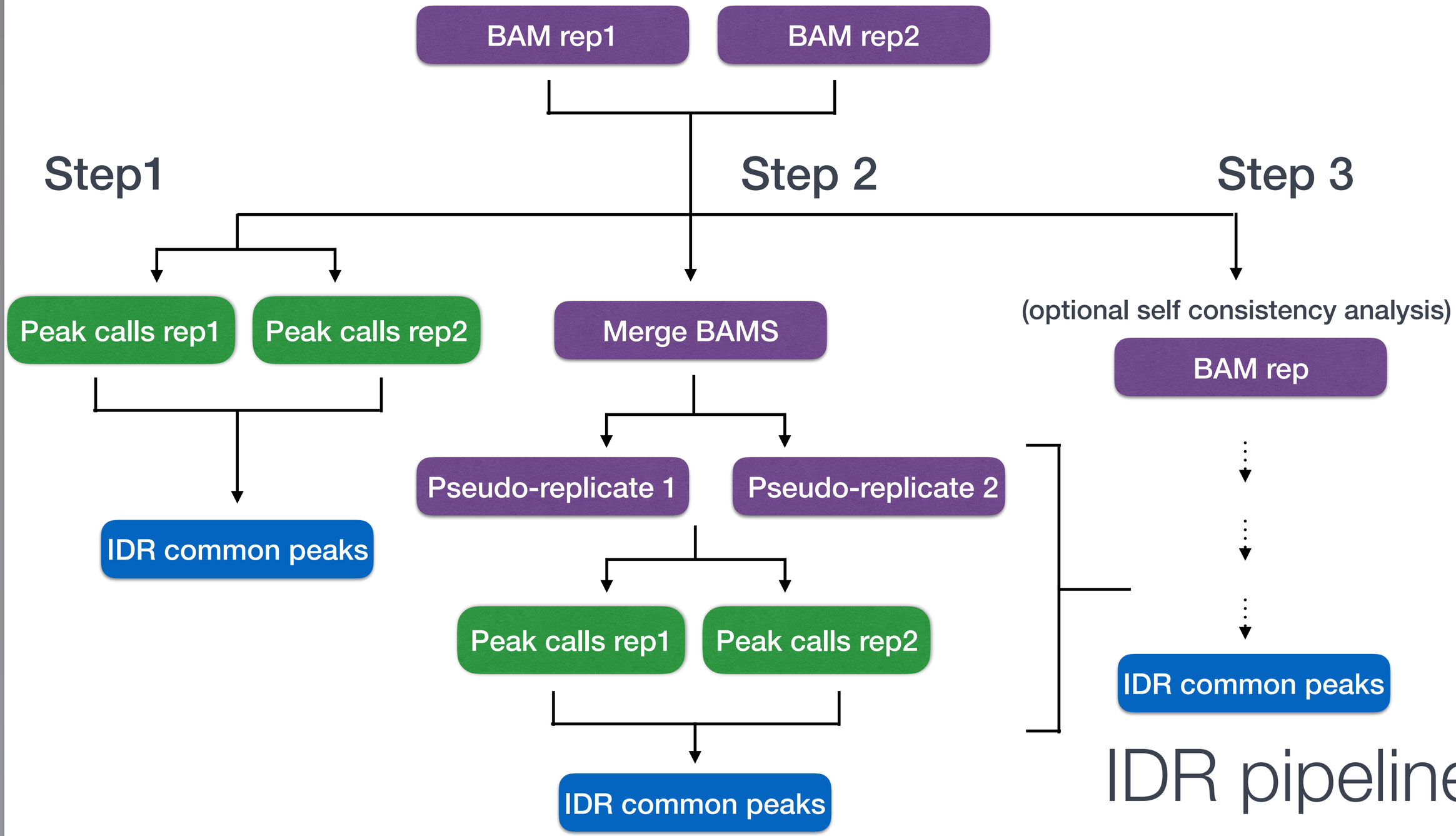
IDR: https://github.com/nboley/idr
- requires python3

```
# Usage
idr -h
idr --samples ../idr/rep1.narrowPeak ../idr/rep1.narrowPeak
```

IDR pipeline



IDR pipeline


BAM rep1
BAM rep2
Step1
Step 2
Step 3
Peak calls rep1
Peak calls rep2
Merge BAMS
(optional self consistency analysis)
BAM rep
IDR common peaks
Pseudo-replicate 1
Pseudo-replicate 2
Peak calls rep1
Peak calls rep2
IDR common peaks
IDR common peaks
IDR pipeline

IDR pipeline

- If starting with **< 100K pre-IDR peaks** for large genomes (human/mouse):

- If starting with **< 100K pre-IDR peaks** for large genomes (human/mouse):
  - For true replicates and self-consistency replicates an **IDR threshold of 0.05** is more appropriate

- If starting with **< 100K pre-IDR peaks** for large genomes (human/mouse):
  - For true replicates and self-consistency replicates an **IDR threshold of 0.05** is more appropriate
  - Use a tighter threshold for pooled-consistency since pooling and subsampling equalizes the pseudo-replicates in terms of data quality. Err on the side of caution and use more stringent **IDR threshold of 0.01**

- If starting with **< 100K pre-IDR peaks** for large genomes (human/mouse):
  - For true replicates and self-consistency replicates an **IDR threshold of 0.05** is more appropriate
  - Use a tighter threshold for pooled-consistency since pooling and subsampling equalizes the pseudo-replicates in terms of data quality. Err on the side of caution and use more stringent **IDR threshold of 0.01**
- If your replicates are truly good replicates then the results from the two runs (replicate IDR vs. pooled pseudo-replicate IDR) are usually similar (**within a factor of 2 at max**)

IDR pipeline

- If starting with **< 100K pre-IDR peaks** for large genomes (human/mouse):
  - For true replicates and self-consistency replicates an **IDR threshold of 0.05** is more appropriate
  - Use a tighter threshold for pooled-consistency since pooling and subsampling equalizes the pseudo-replicates in terms of data quality. Err on the side of caution and use more stringent **IDR threshold of 0.01**
- If your replicates are truly good replicates then the results from the two runs (replicate IDR vs. pooled pseudo-replicate IDR) are usually similar (**within a factor of 2 at max**)
- Anything beyond that indicates the true replicates are substantially different and is almost always a sign that one replicate is messed up or is a sample swap or some other issue.

IDR pipeline