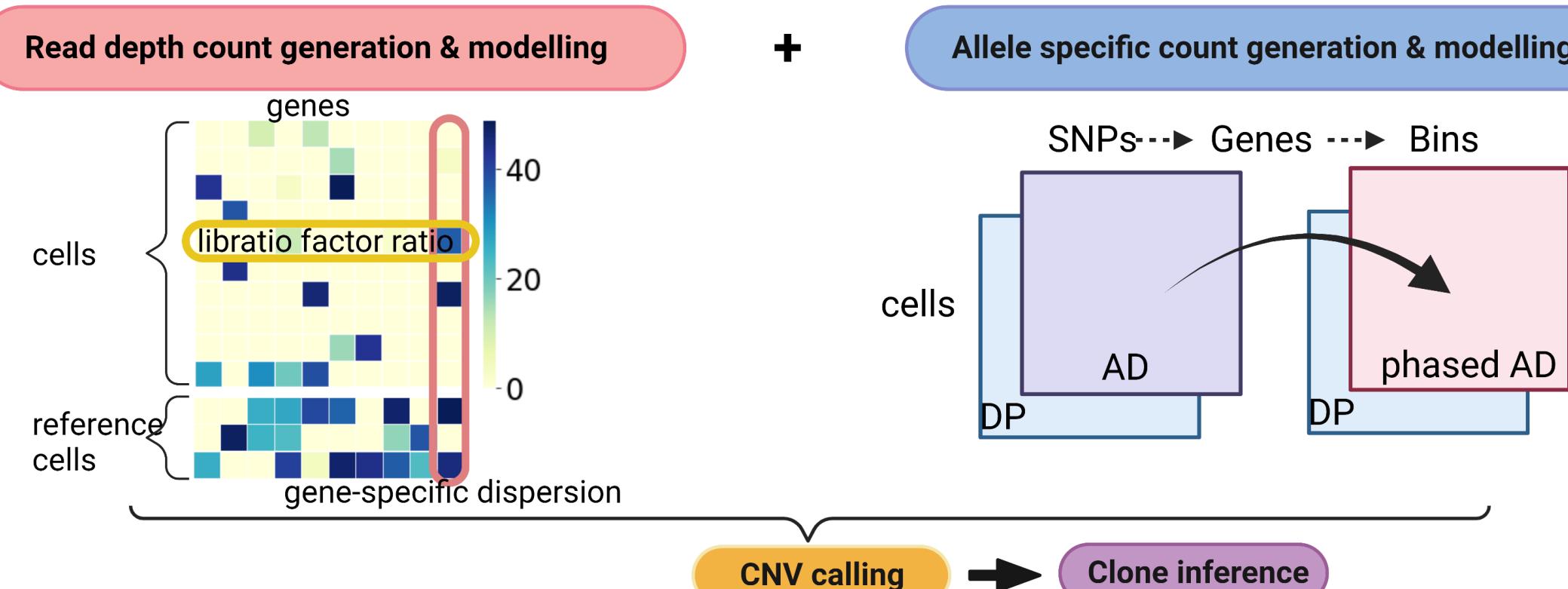




Rongting Huang, Xianjie Huang, Oliver Stegle and Yuanhua Huang

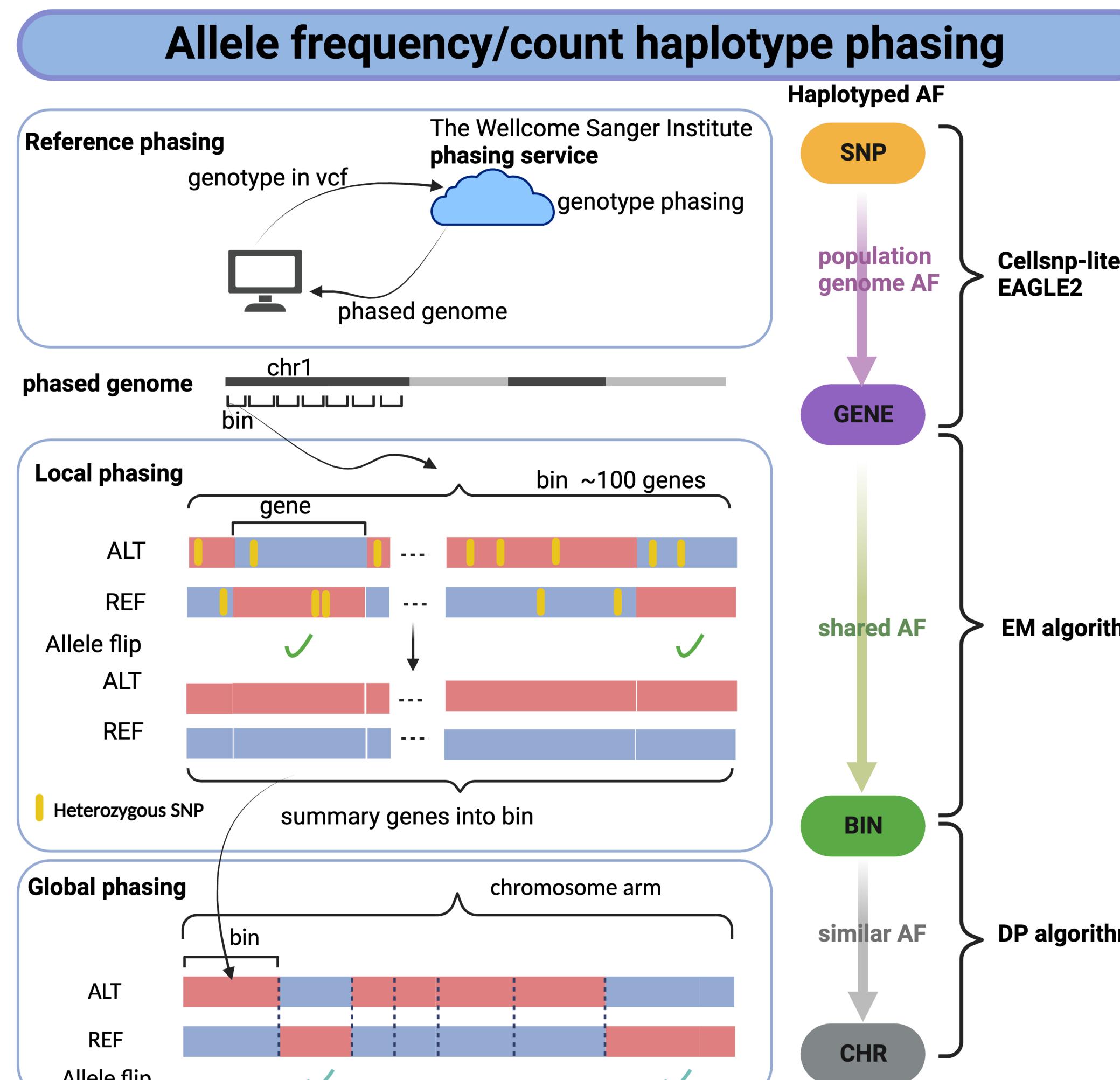
## MOTIVATIONS

- The copy number variations (CNVs), including copy gain, loss and copy-neutral loss-of-heterozygosity (CNLOH/LOH), have been widely shown associations in various cancers.
- Accurate detection of CNVs is crucial for studying the genetic factors that contribute to the development and progression of cancers.
- Single-cell RNA-seq (scRNA-seq) technologies provide a way to get a dual readout of transcriptome and genetic makeups.
- High technical sparsity makes it challenging to call CNVs at a single-cell resolution accurately.



**Figure 1:** We introduce XClone, a statistical model to detect allele-specific CNVs on individual cells by modelling the raw read/UMI counts from scRNA-seq data. This method accounts for two modules: the sequencing read depth ratio(RDR) of individual genes and the B-allele frequency (BAF) of heterozygous variants .

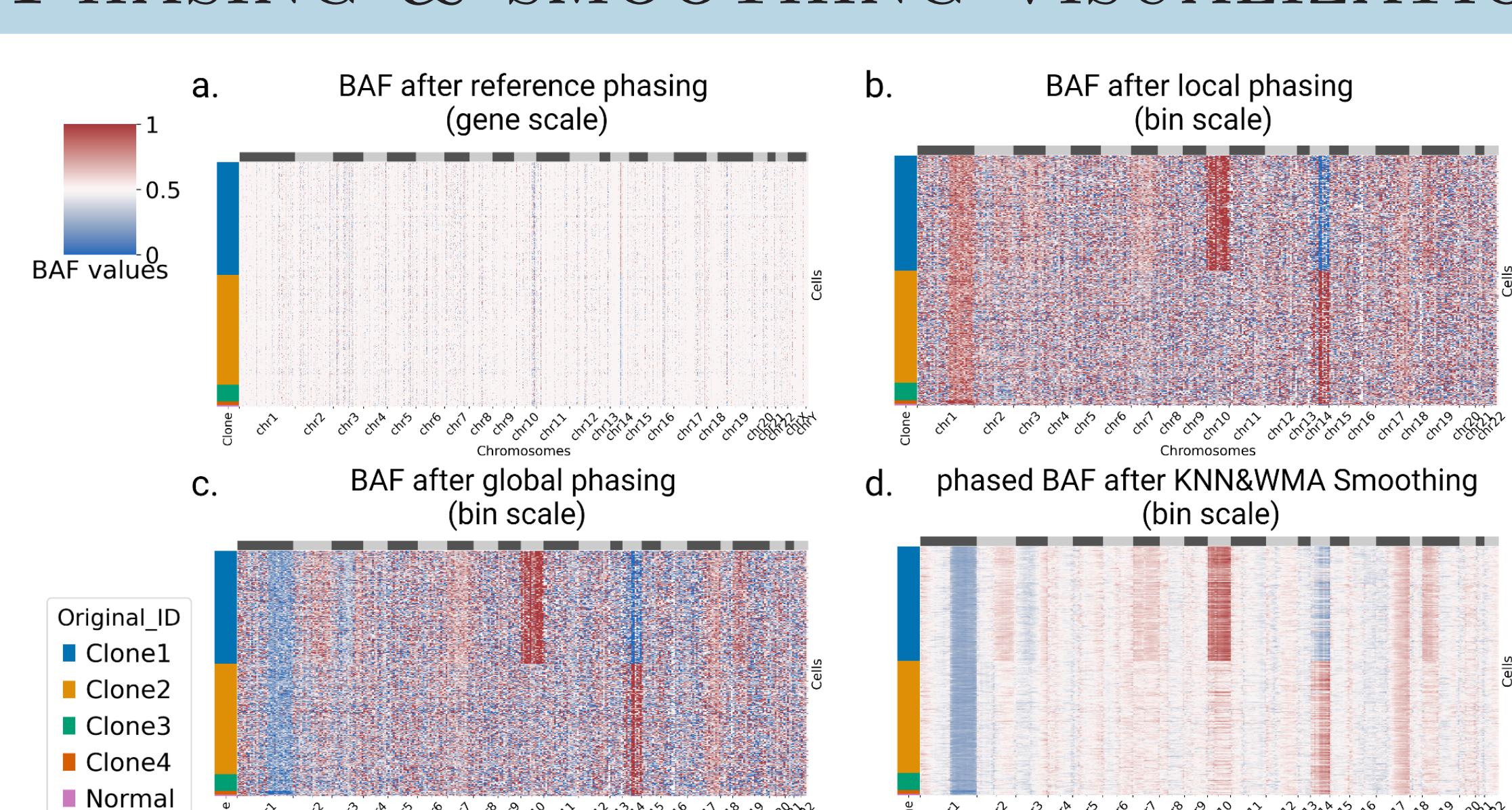
## METHODS HIGHLIGHT



**Figure 3:** We utilized a three-level allele phasing approach to enhance the signal-to-noise ratio when analysing the allele information.

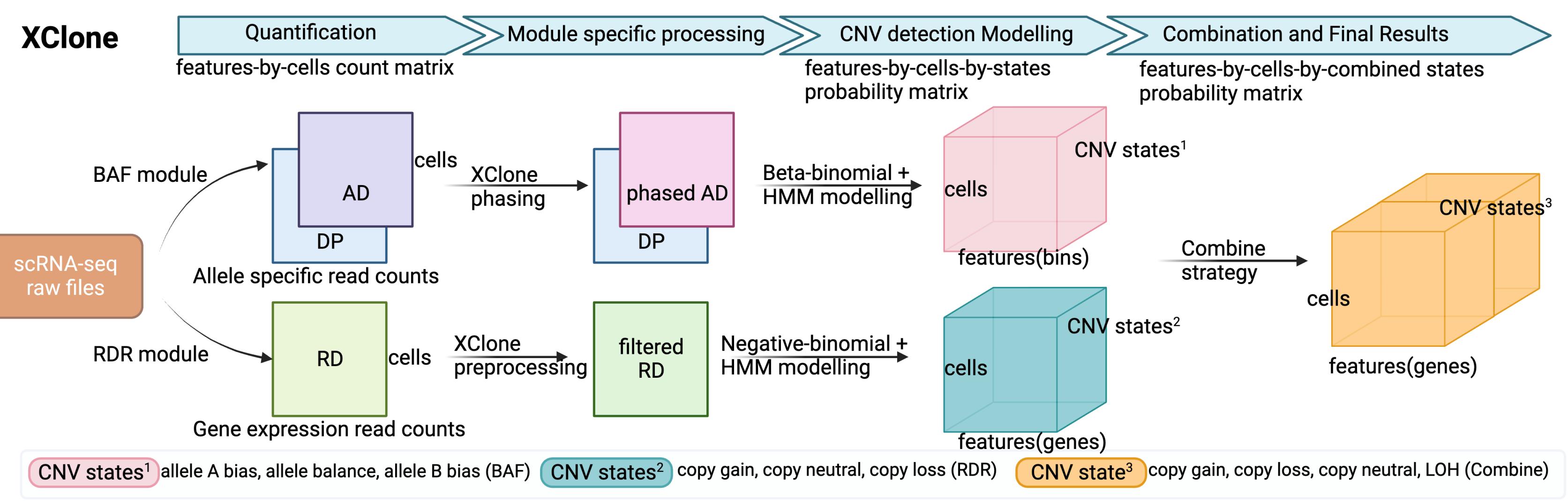
- Reference-based phasing by utilizing the human population haplotype reference to phase individual SNPs.
- Gene-bin phasing (local phasing) to link a group of proximate and consecutive genes.
- Chromosome-range phasing (global phasing) to further link all gene bins on each chromosome arm for synchronising the BAF for visualisation.

## PHASING & SMOOTHING VISUALIZATION



**Figure 4:** A glioma sample BCH869 scRNA-seq B Allele Frequency (BAF) visualization before and after phasing and smoothing generated by XClone.(a) After reference phasing. (b) After local phasing. (c) After global phasing. (d) Horizontal (weighted moving average, WMA) and vertical (k-nearest neighbour, KNN) smoothing on phased BAF.

## METHODOLOGY OF XCLONE



**Figure 2:** Overview of XClone workflow.

### Preprocessing and Modelling, Combine RDR and BAF to infer CNV states

- Preprocessing: extract count matrices with `xcltk`
- Pre-analysis: three-level **allele phasing** for haplotype-specific BAF
- Pre-analysis: horizontal and vertical **smoothing** for visualization
- Modelling: negative binomial mixture model for expression data (RDR module)

$$p(X_{c,g}|z_{c,g,k} = 1) = NB(X_{c,g}|\mu_{c,g,k}, \phi_g) \quad (1)$$

$$\mu_{c,g,k} = l_c \times X_{ref,g} \times C_k,$$

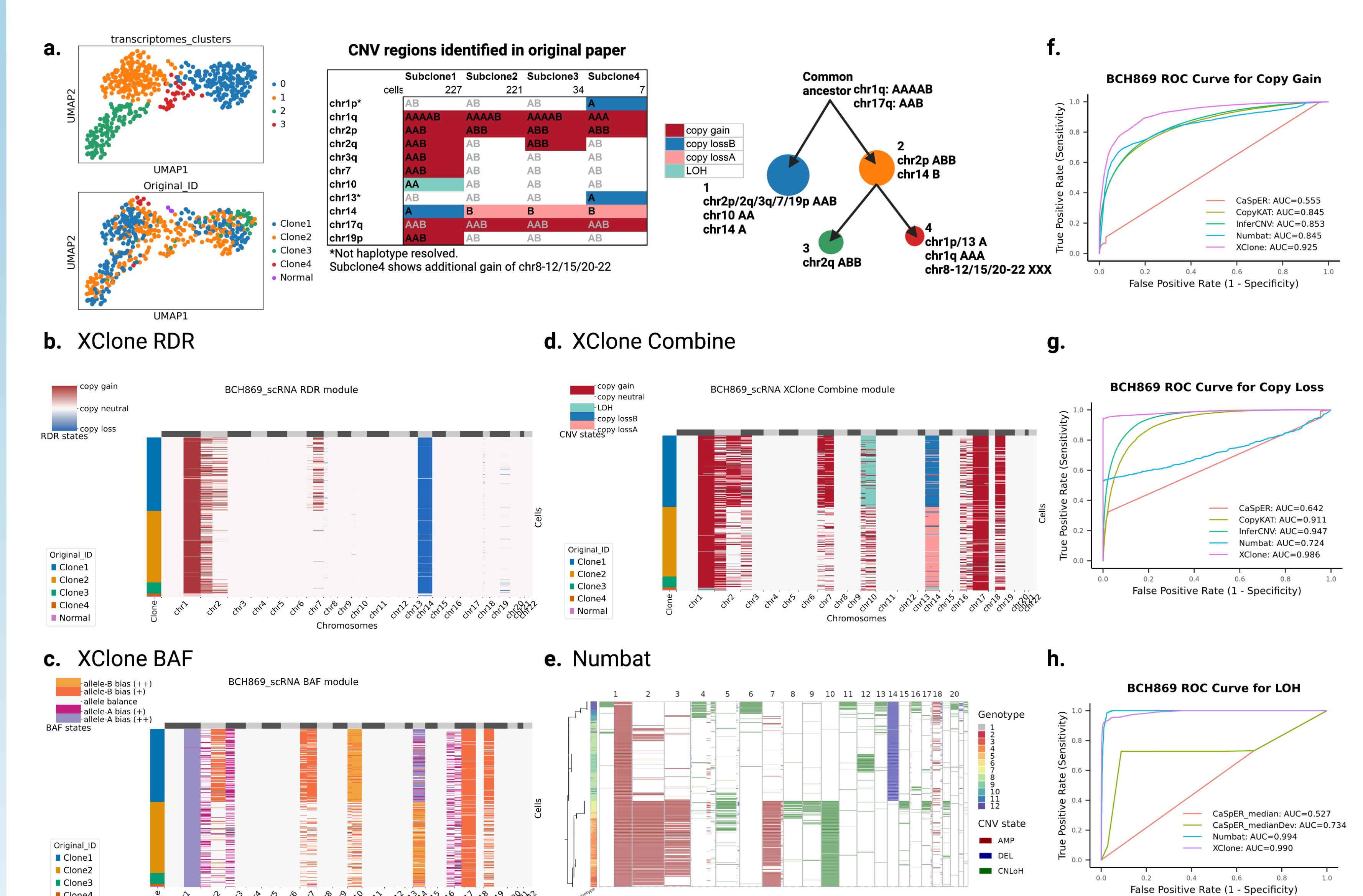
- Modelling: beta-binomial mixture model for phased allele-specific data (BAF module)

$$p(a_{c,g}|d_{c,g}, z_{c,g,k} = 1) = BB(a_{c,g}|d_{c,g}, \rho_{g,k}, \tau_{g,k}), \quad (2)$$

- Post-analysis: CNV states smoothing, combination and denoise

**Notation:**  $X_{c,g}$ : raw read/UMI count at gene  $g$  cell  $c$ ;  $a_{c,g}$  and  $d_{c,g}$ : phased B allele count and total count at gene-bin  $g$  in cell  $c$

## ASSESSMENT OF CNV DETECTION ON A GLIOMA



**Figure 5:** CNV identification of a glioma sample BCH869 with clone-specific allele loss by XClone and performance comparison between methods. (a) UMAP of BCH869 with different types of cell annotation: cell clustering by transcriptomes and CNV tumor clone ID identified by the original paper, CNV region identified for each clone and the inference clonal tree in the original paper. (b-d) Heatmap visualisation of CNVs by XClone RDR module, BAF module and the combination of the 2 modules, respectively. (e) Heatmap visualisation of CNVs generated by Numbat. (f-h) Assessment of performance in identification of copy number gain, copy number loss and loss of heterozygosity on BCH869 glioma sample. With ground truth of BCH869 sample, we benchmarked XClone with four other CNVs detection tools specifically designed for scRNA-seq data( XClone v.0.3.4, InferCNV v.1.8.0, CopyKAT v.1.0.4, CaSpER v.0.2.0 and Numbat v.1.2.1) using their default parameters.

## REFERENCES

- Huang, Rongting and Huang, Xianjie and Tong, Yin and Yan, Helen YN and Leung, Suet Yi and Stegle, Oliver and Huang, Yuanhua. XClone: detection of allele-specific subclonal copy number variations from single-cell transcriptomic data. *bioRxiv*, 2023-04.
- XClone Python package: <https://pypi.python.org/pypi/xclone>; Github: <https://github.com/single-cell-genetics/XClone>
- Xcltk Python package: <https://pypi.org/project/xcltk/>