



HAL
open science

Annotation transfer in pangenome graphs

Nina Marthe

► **To cite this version:**

Nina Marthe. Annotation transfer in pangenome graphs. XI JORNADES DE BIOINFORMÀTICA I GENÒMICA, Dec 2023, Barcelona, Spain. hal-04490506

HAL Id: hal-04490506

<https://hal.science/hal-04490506v1>

Submitted on 5 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

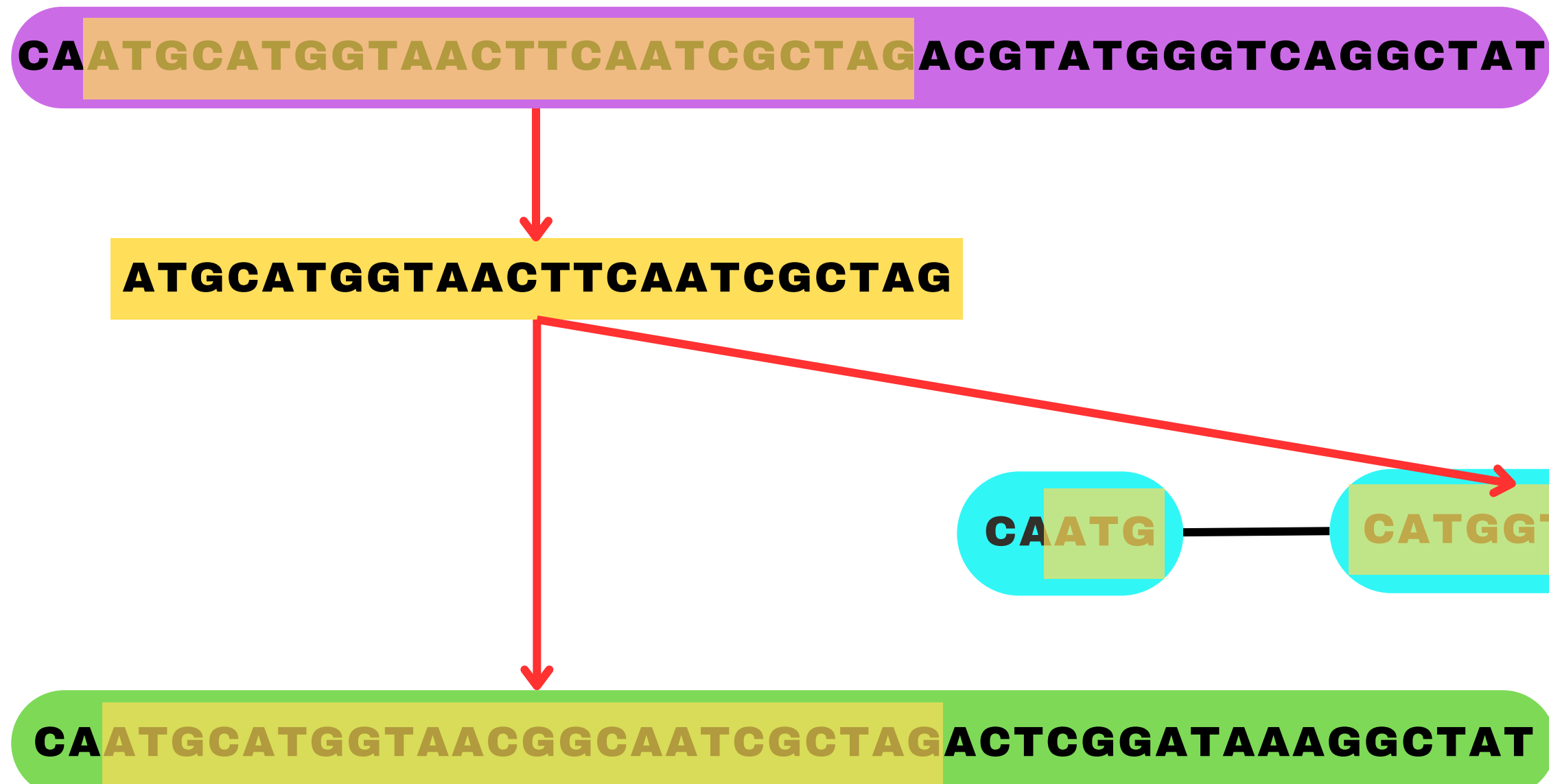


Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Nina Marthe, nina.marthe@ird.fr

Annotation transfer usually consists in aligning the annotated sequences on the genome we want to transfer to. This approach can be applied to pangenome graphs, however it sacrifices precision to reduce runtime. Pangenome graphs allow another approach for annotation transfer that doesn't rely on alignment, since it simply converts the annotated region's coordinates to get their position in the graph.

Traditional annotation transfer

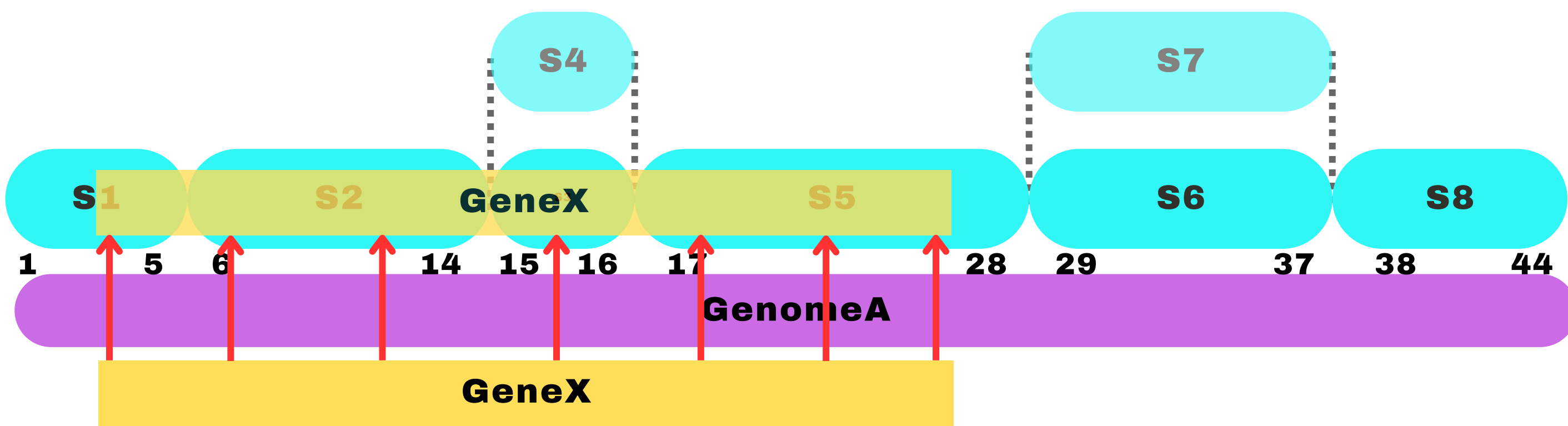


- Extract the sequence of the annotated region
- Align it on the target genome/pangenome
- Report the annotation where the sequence aligned best

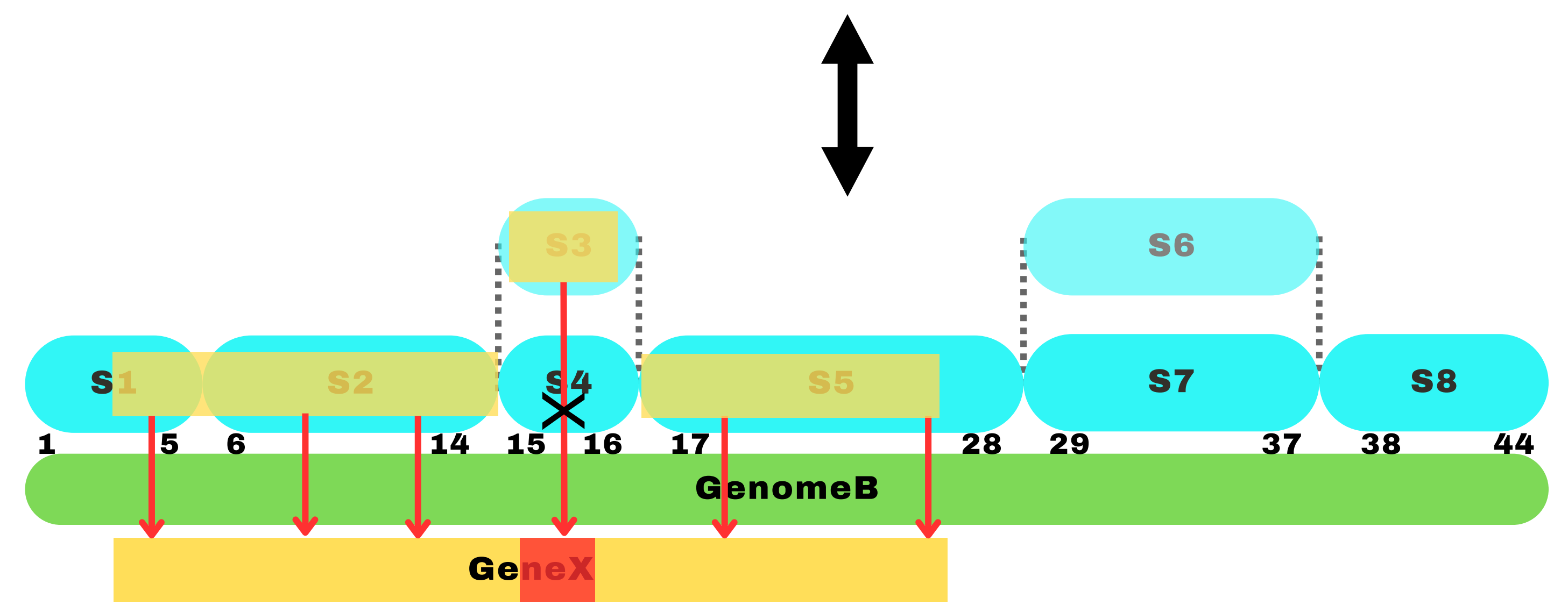
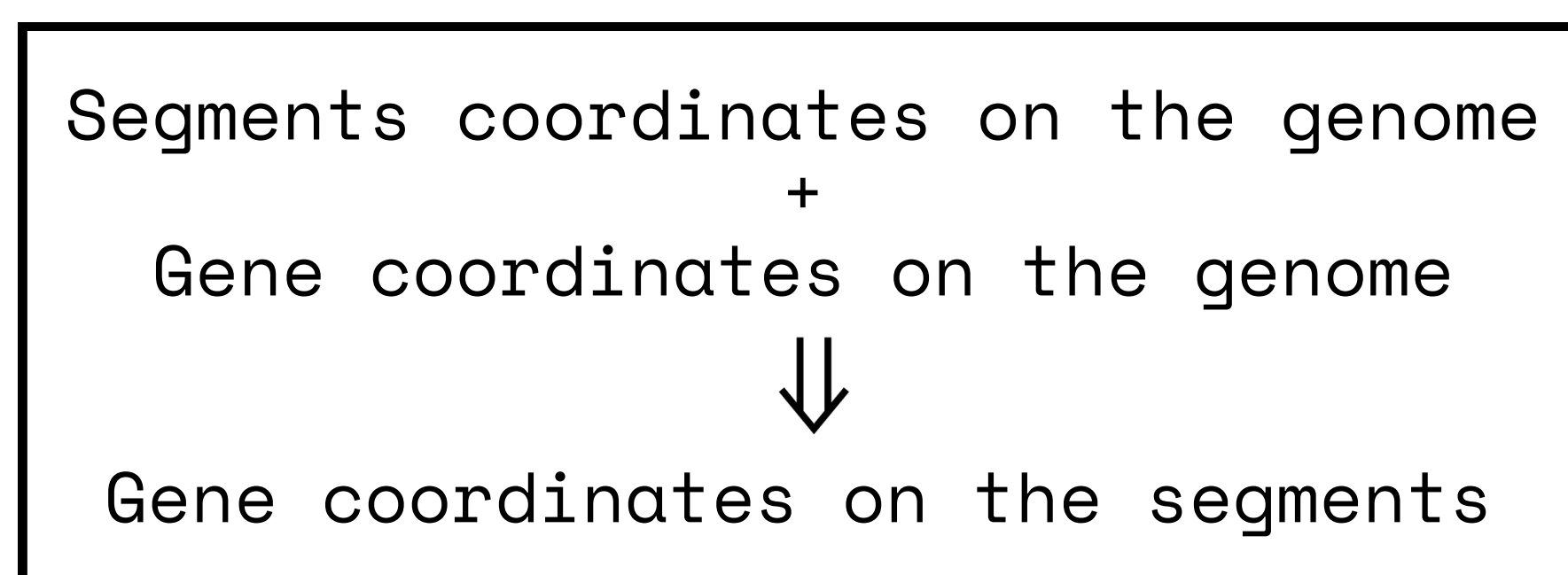
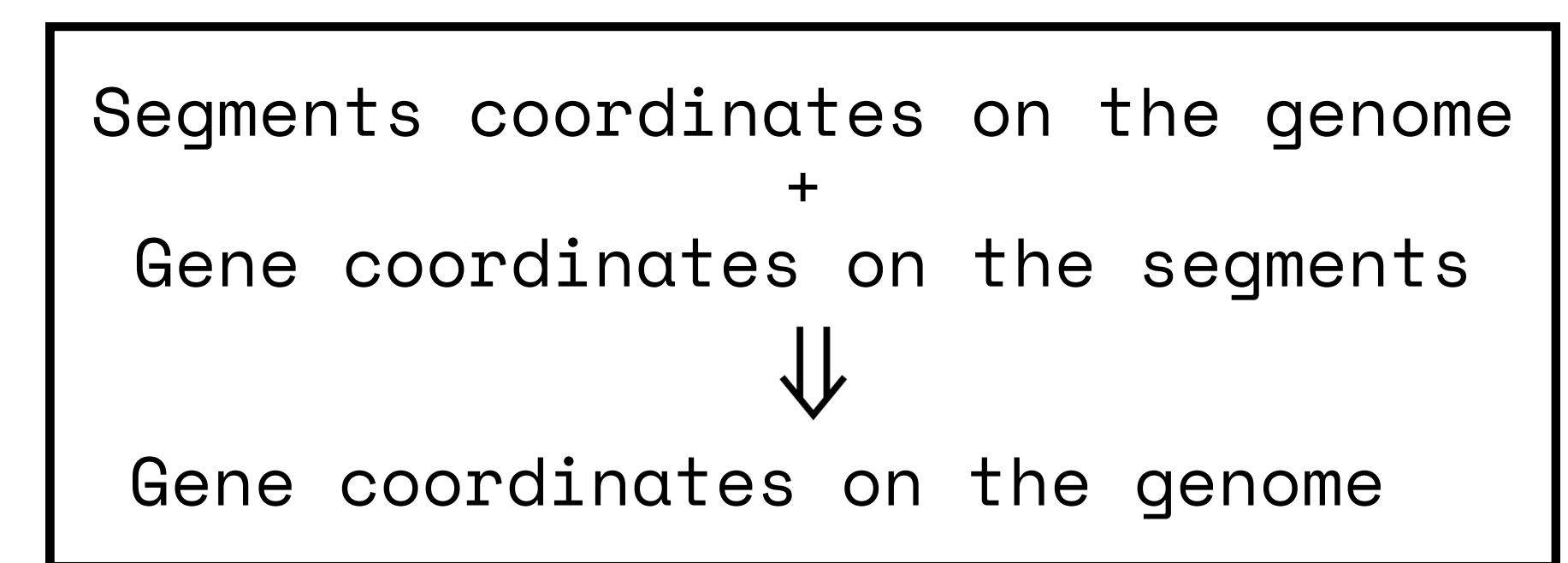
- Approximate solution (heuristic solution)
- Sacrifices precision to reduce runtime

Annotation transfer using coordinate conversion

Annotation transfer from genome to graph



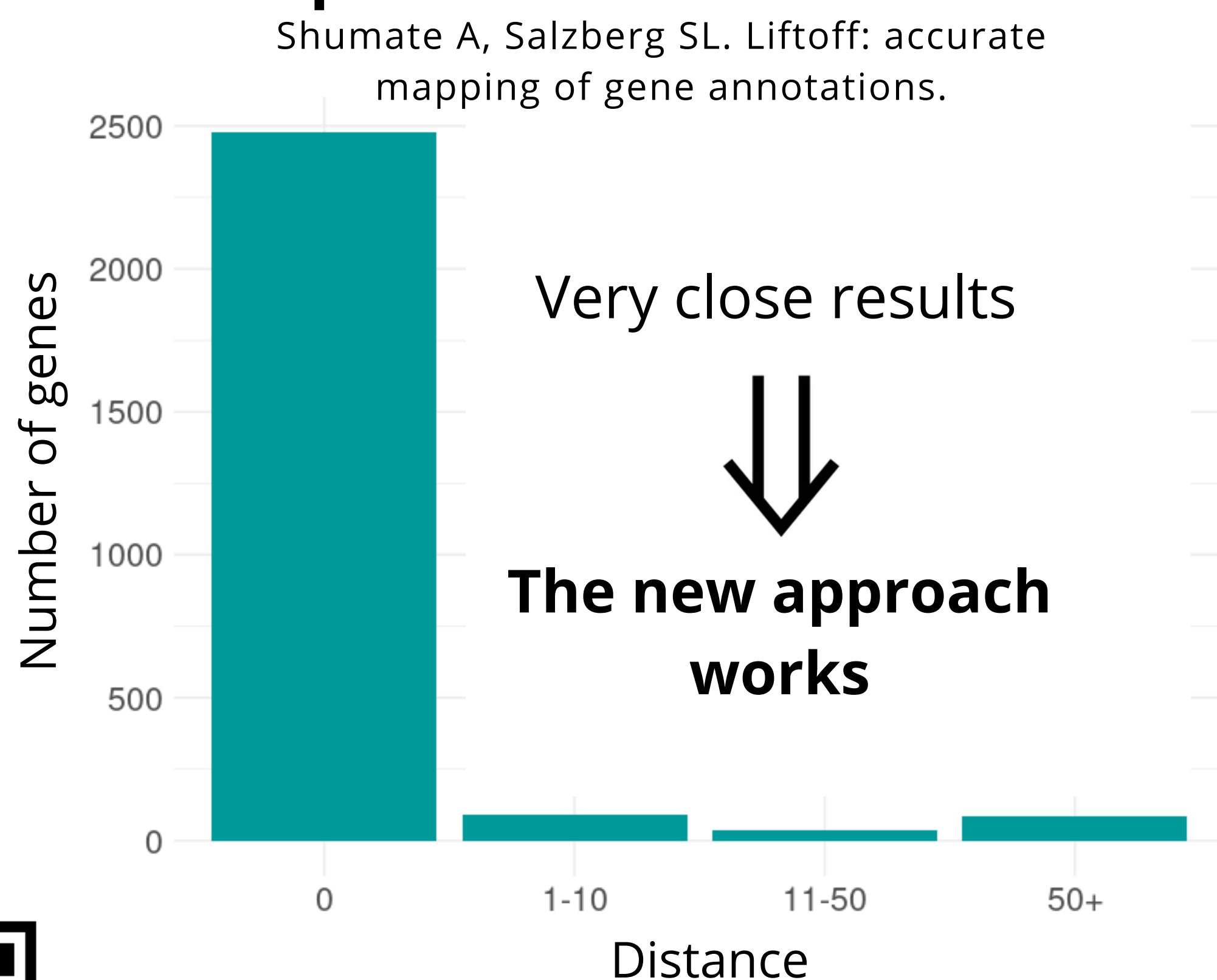
Annotation transfer from graph to genome



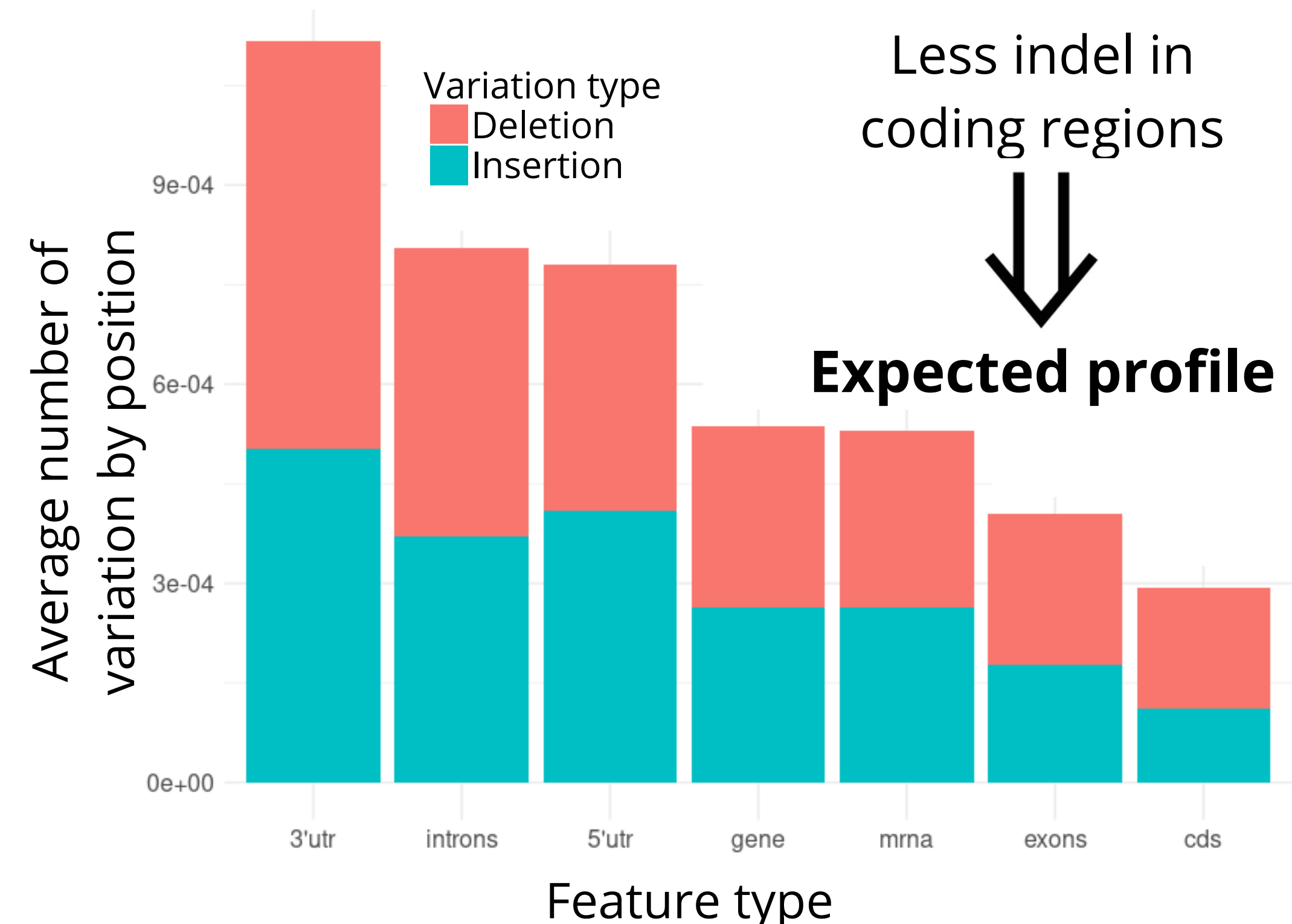
- High precision (exact solution)
- No sequence information needed
- Uses simple operations

- High precision (exact solution)
- Cannot miss a part of the gene
- Can detect modifications of the gene

Comparison with Liftoff transfer



Indel distribution in genes



This new approach provides an exact solution to annotation transfer. It should prove valuable to both genome and pangenome studies, and allow for a better harnessing of the information contained within the pangenomes.

Implemented in python
within GrAnnot