Dataset Construction Workflow

MDA Data
Contains Chimeric Artifacts
Nanopore

Bulk Data 1
No Chimeric Artifacts
Reference

Bulk Data 2
No Chimeric Artifacts
Reference

Bulk Data 3
No Chimeric Artifacts
Reference

MDA Chimeric Reads

Bulk 1 Chimeric Reads

Bulk 2 Chimeric Reads

Bulk 3 Chimeric Reads

Comparison Engine
Match Counting Algorithm
For each MDA read:
Count bulk matches

Match 0
Artificial Chimeric
0/3 Bulk Match

Match 1
Biological Chimeric
1/3 Bulk Match

Match 2
Biological Chimeric
2/3 Bulk Match

Match 3
Biological Chimeric
3/3 Bulk Match

Artificial
Class 0

Biological
Class 1