

# Dataset Construction Workflow

## Process:

1. Extract chimeric reads from MDA & Bulk data
2. Compare MDA reads with 3 Bulk references
3. Classify based on support count (0-3)
4. Binary classification: Artificial vs Biological
5. Training data for sequence classification

## Classification:



Artificial (Sup 0)  
Biological (Sup 1,2,3)  
Bulk References  
MDA Data

