

Transcriptome Annotation by Modular Algorithms TAMA

Richard I. Kuo¹, Yuanyuan Cheng²³, Runxuan Zhang⁴, John W. S. Brown^{5,6}, Jacqueline Smith¹, Alan L. Archibald¹, and David W. Burt^{1,2}

¹The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Midlothian EH25 9RG, UK

²The University of Queensland, St. Lucia, Brisbane, QLD 4072, Australia

³School of Life and Environmental Sciences, University of Sydney, Sydney, New South Wales, Australia

⁴Information and Computational Sciences, The James Hutton Institute, Invergowrie, Dundee, Scotland, UK

⁵Plant Sciences Division, School of Life Sciences, University of Dundee, Invergowrie, Dundee, Scotland, UK

⁶Cell and Molecular Sciences, The James Hutton Institute, Invergowrie, Dundee, Scotland,



THE UNIVERSITY *of* EDINBURGH



TAMA tools



Transcriptome Annotation by Modular Algorithms

TAMA
Collapse

Mapped reads to transcript annotation

TAMA
Merge

Merge annotations

ORF/NMD
Predictor

Identify coding region and associate
with known genes

TAMA-GO

Bunch of helpful little tools

<https://github.com/GenomeRIK/tama>



@GenomeRIK

T.A.M.A.G.O

BBSRC
20 Years of Pioneering Great British Bioscience

BMC Genomics

Home About Articles Submission Guidelines

Research article | **Open Access** | Published: 30 October 2020

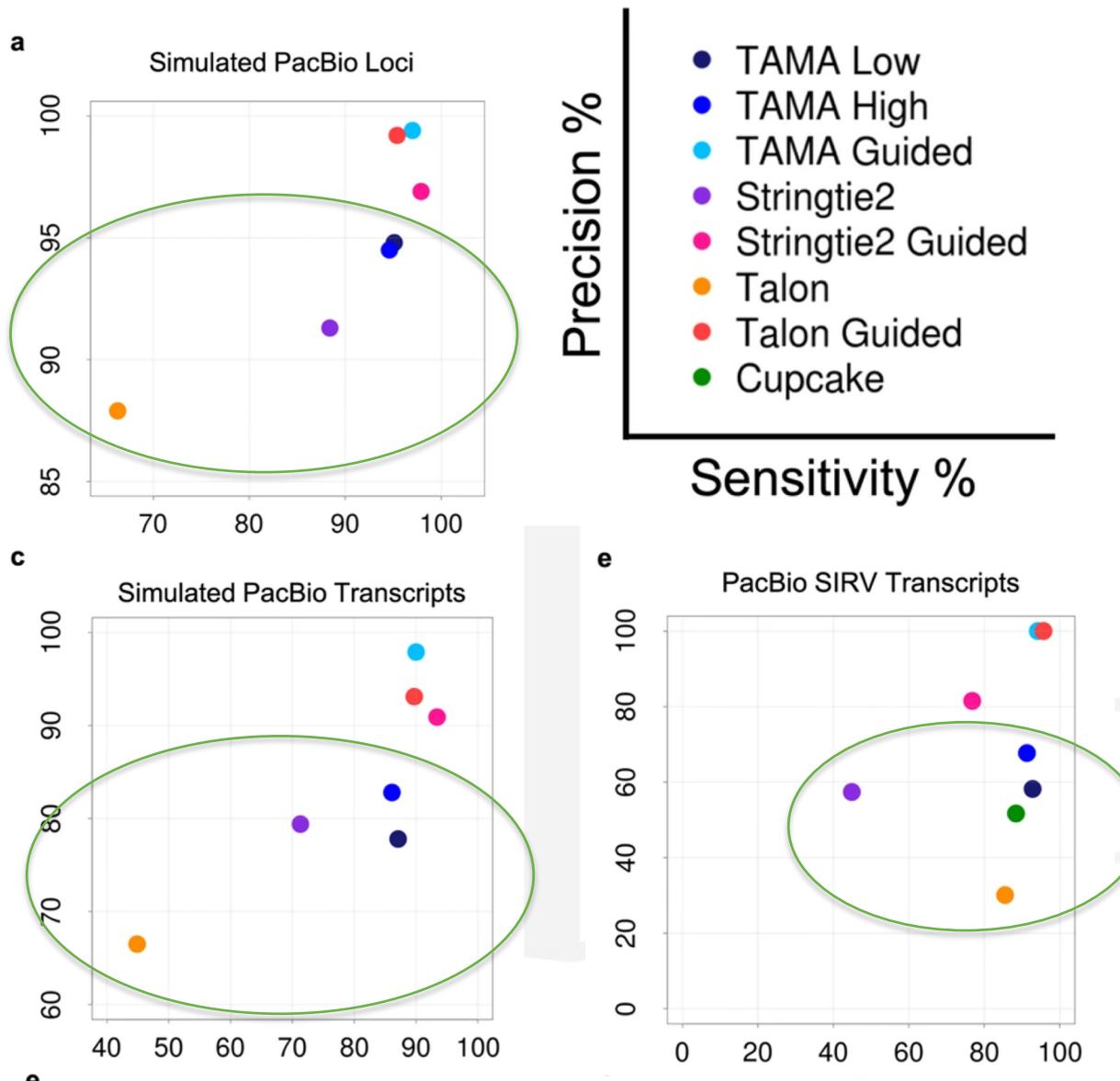
Illuminating the dark side of the human transcriptome with long read transcript sequencing

Richard I. Kuo , Yuanyuan Cheng, Runxuan Zhang, John W. S. Brown, Jacqueline Smith, Alan L. Archibald & David W. Burt

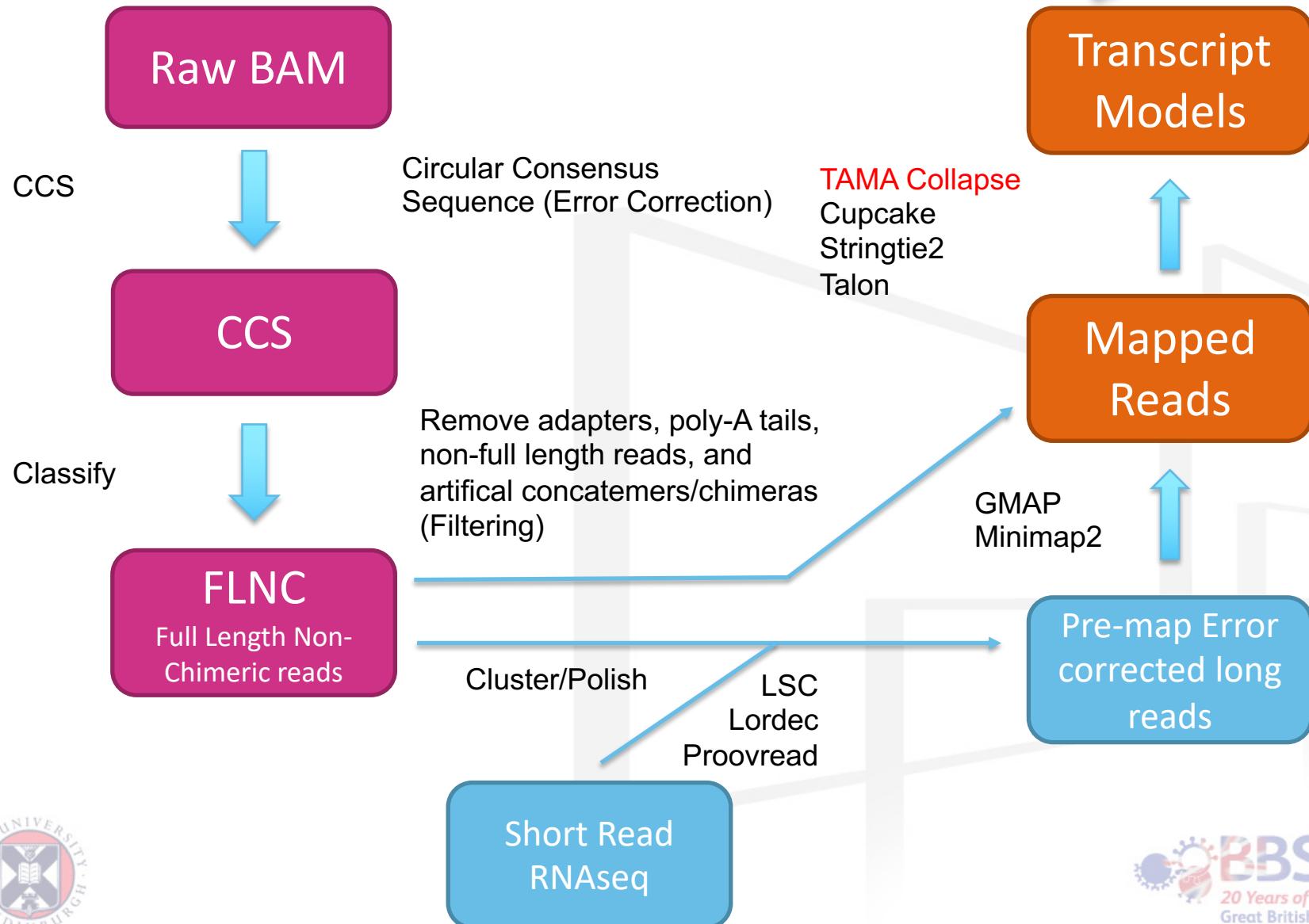
BMC Genomics **21**, Article number: 751 (2020) | Cite this article

3695 Accesses | **29** Altmetric | Metrics

TAMA Benchmarking



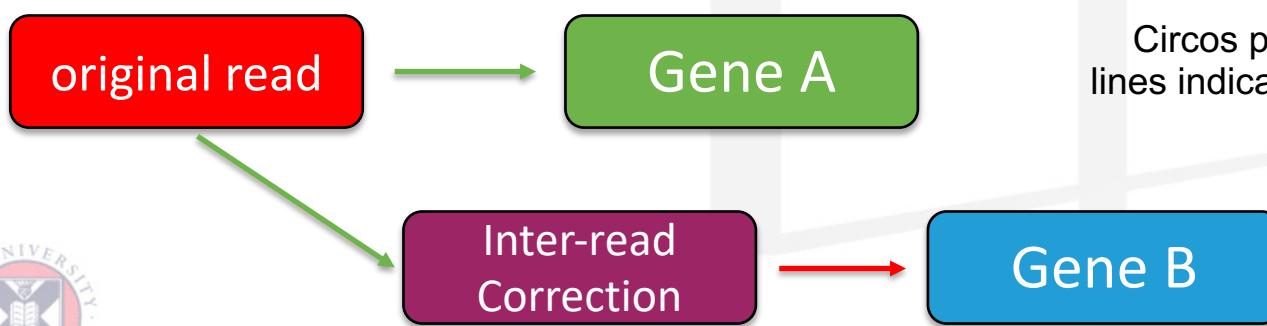
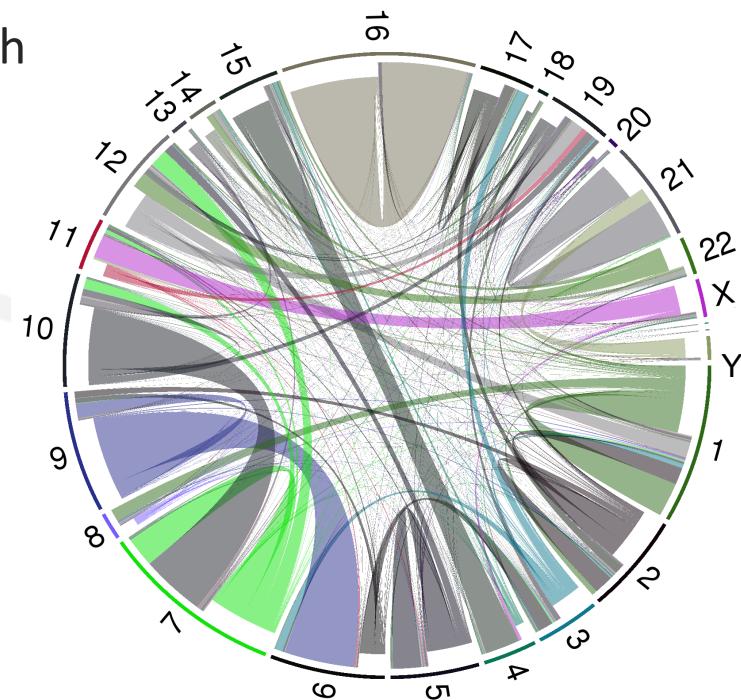
Iso-Seq Analysis



Inter-read error correction



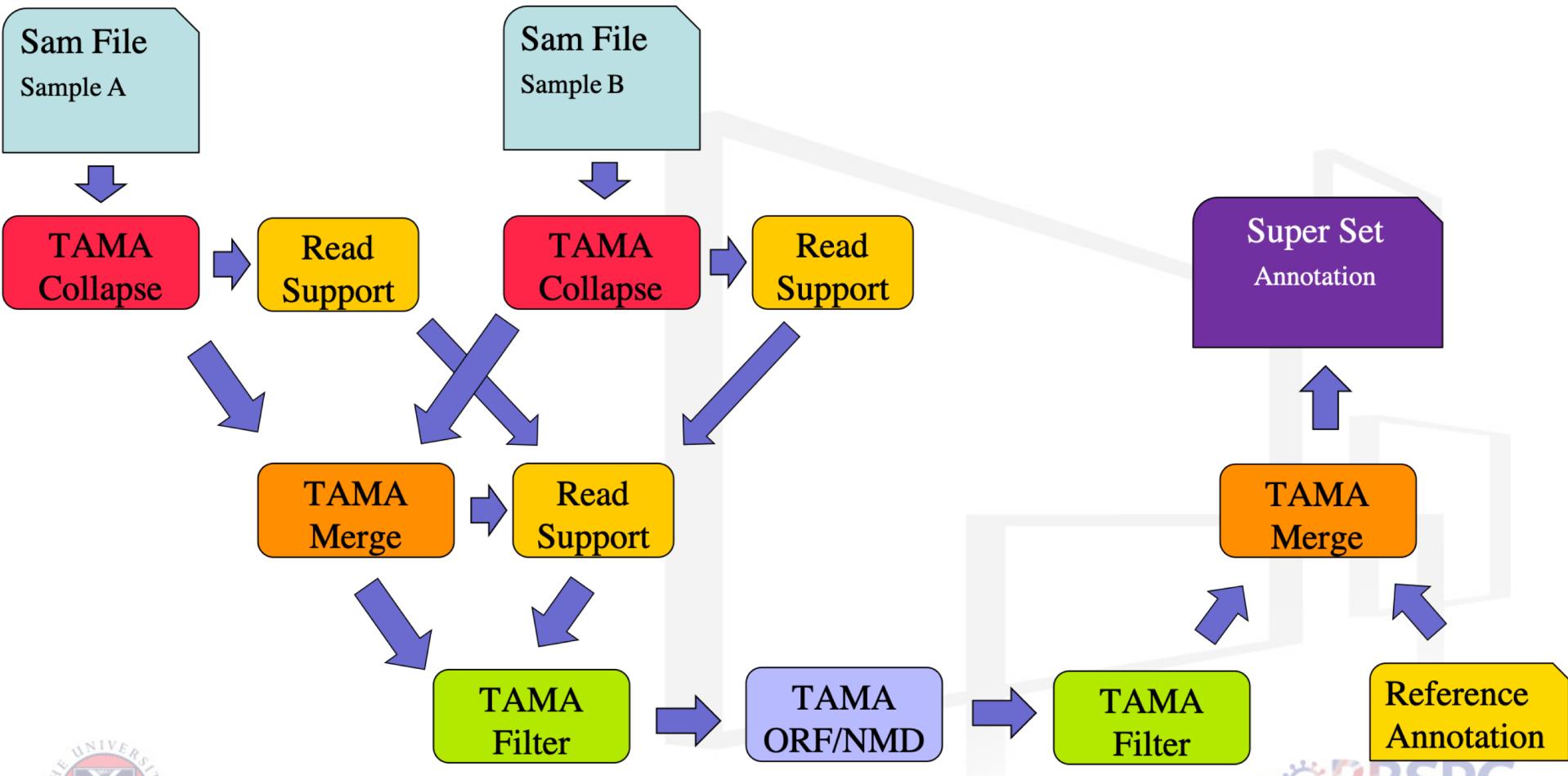
- 6,774 genes affected when using Cluster/Polish
- 3,476 genes affected when using Lordec
- Indicates sequence ambiguity between gene families
- May be used to get better understanding of genome relationship



Circos plot with chromosome axis lines indicating each mis-assigned read
+100K reads



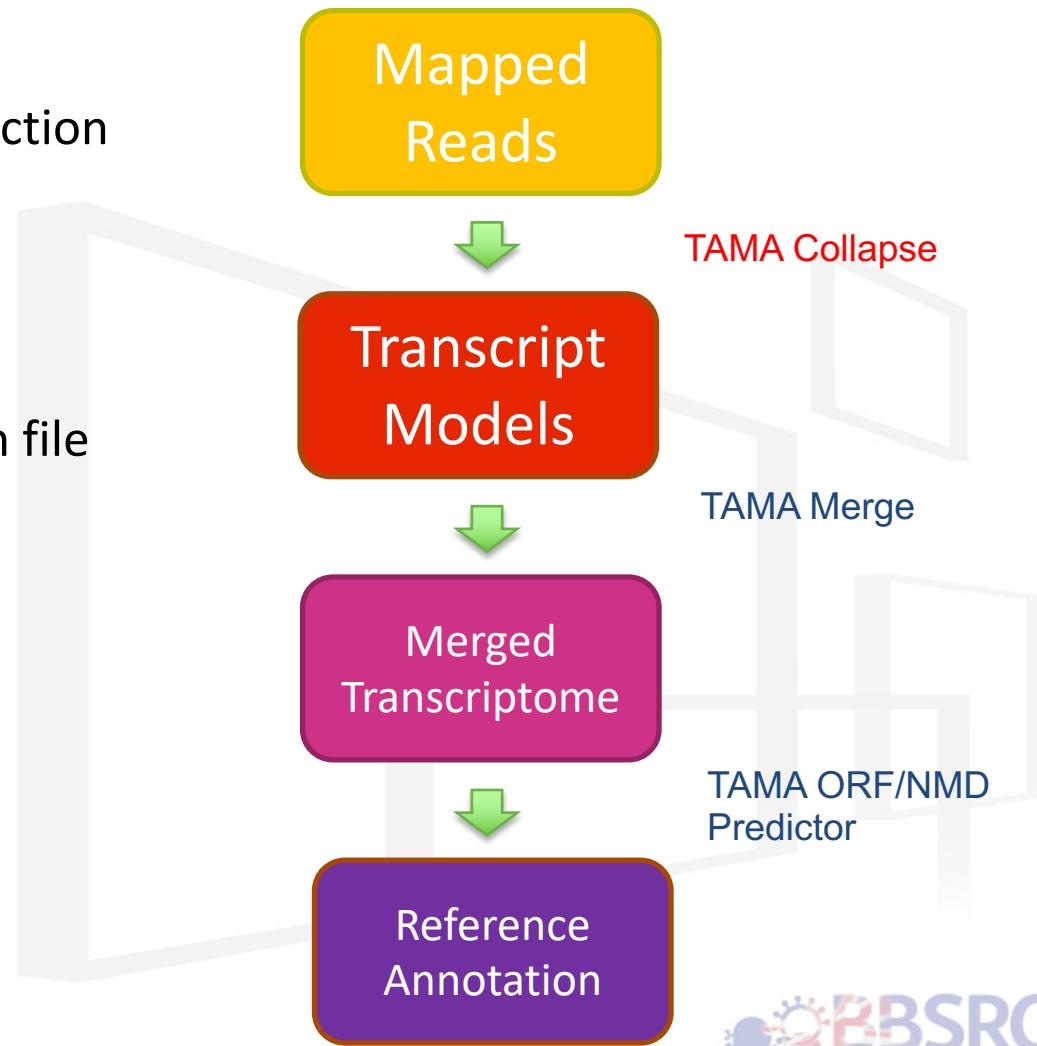
TAMA pipeline



TAMA Collapse



- Removes redundant models
- Reference based error correction
- Highly tunable
- Detailed reports
- Expression ready annotation file



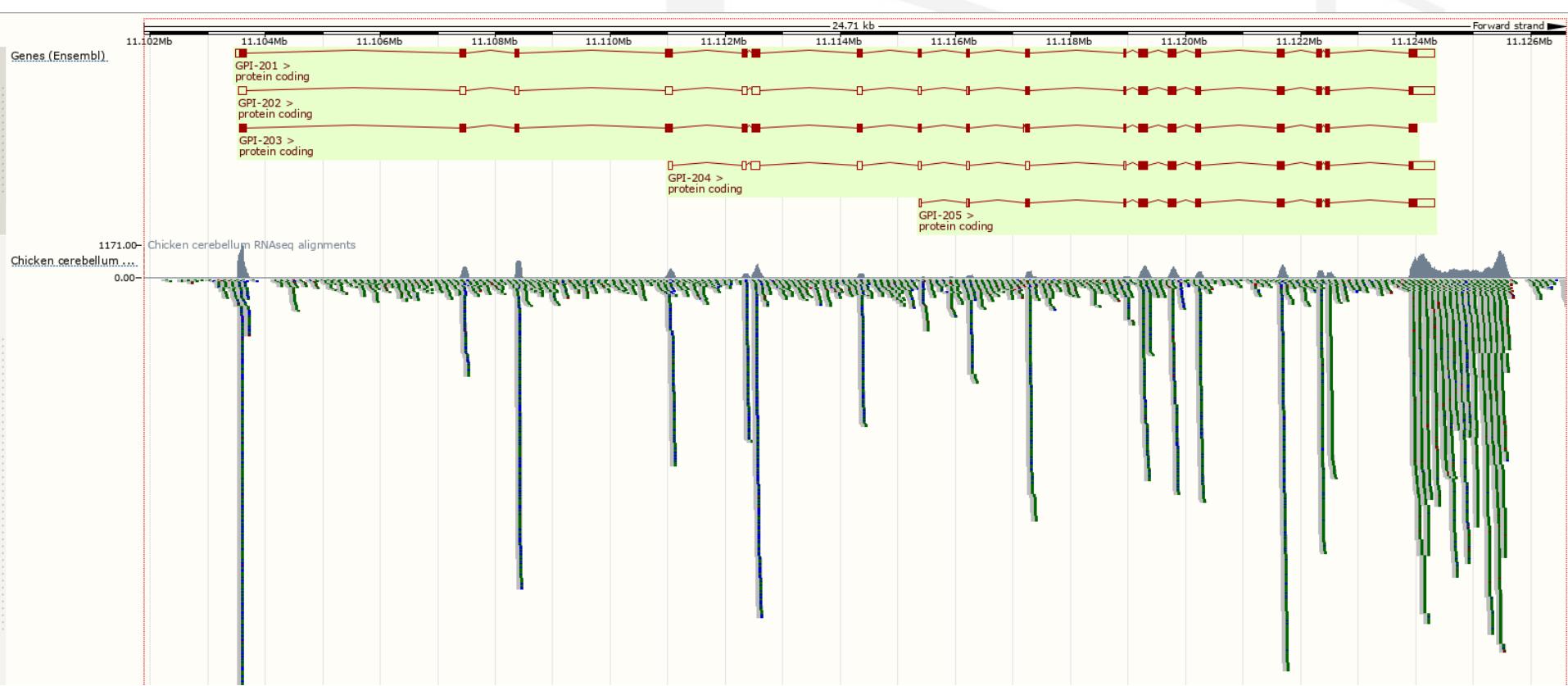
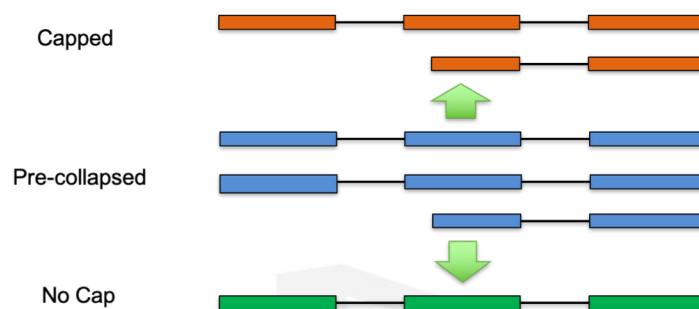
Collapsing



Ensembl
TAMA
Collapse



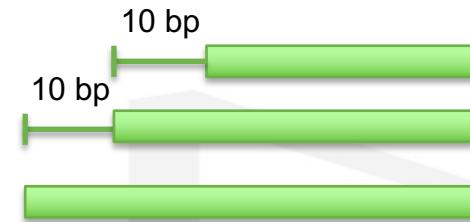
TAMA Collapse: Capped



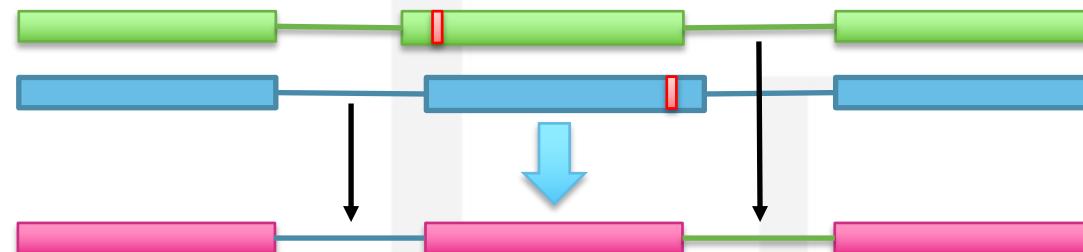
TAMA collapse: Wobble/SJP



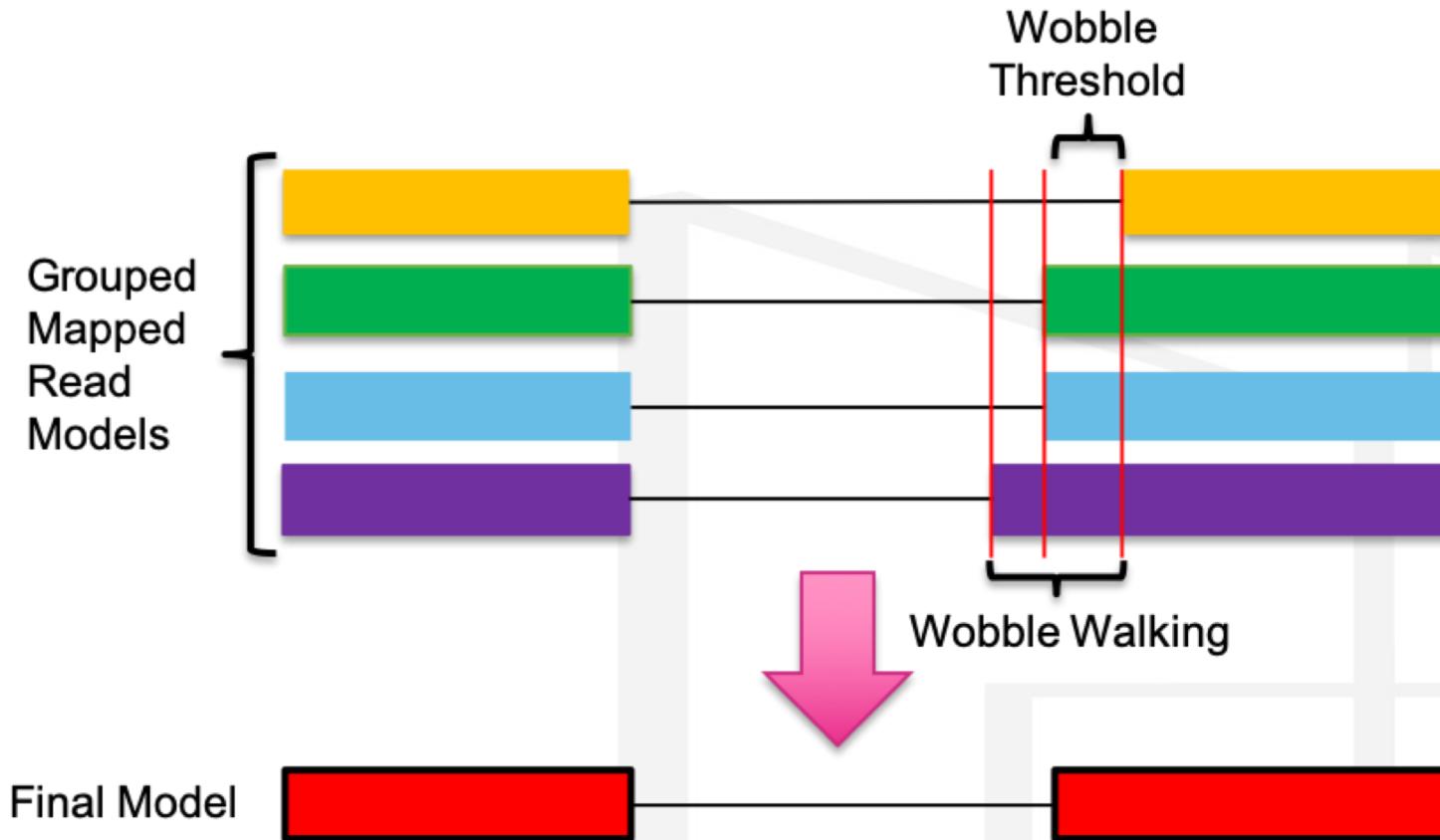
Wobble



Splice junction priority



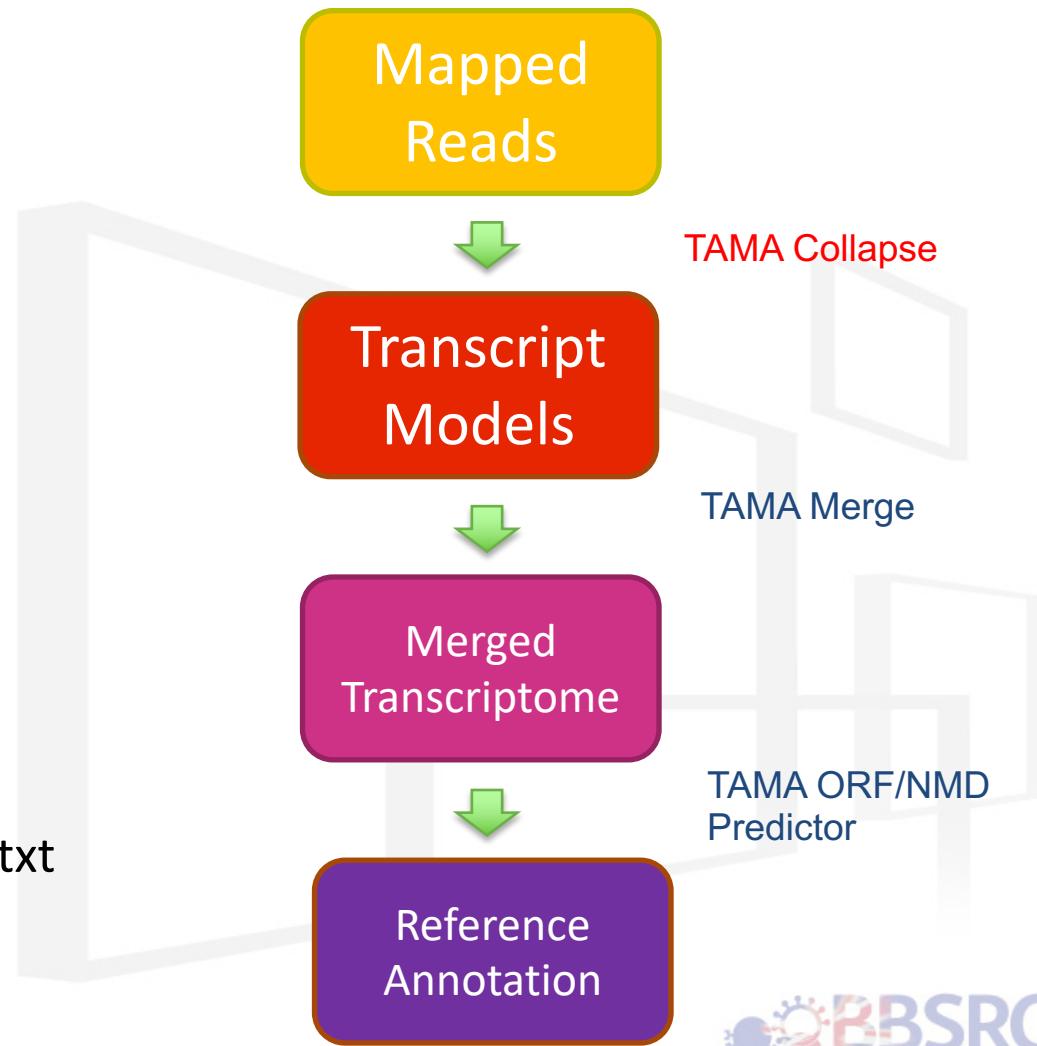
TAMA Collapse: Wobble



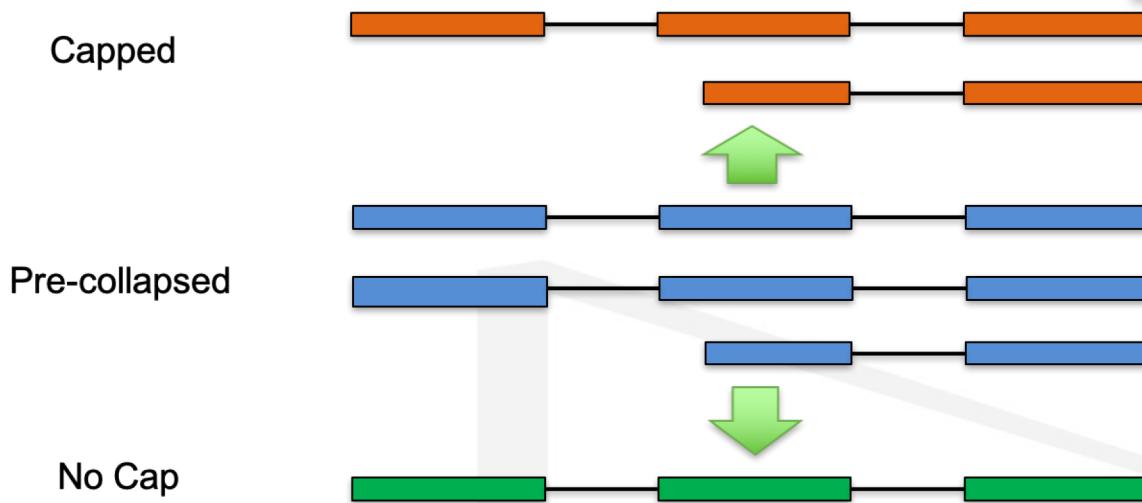
TAMA Collapse: outputs



- prefix.bed
- prefix_read.txt
- prefix_polya.txt
- prefix_strand_check.txt
- prefix_trans_read.bed
- prefix_trans_report.txt
- prefix_varcov.txt
- prefix_variants.txt
- prefix_local_density_error.txt



Degradation Signature



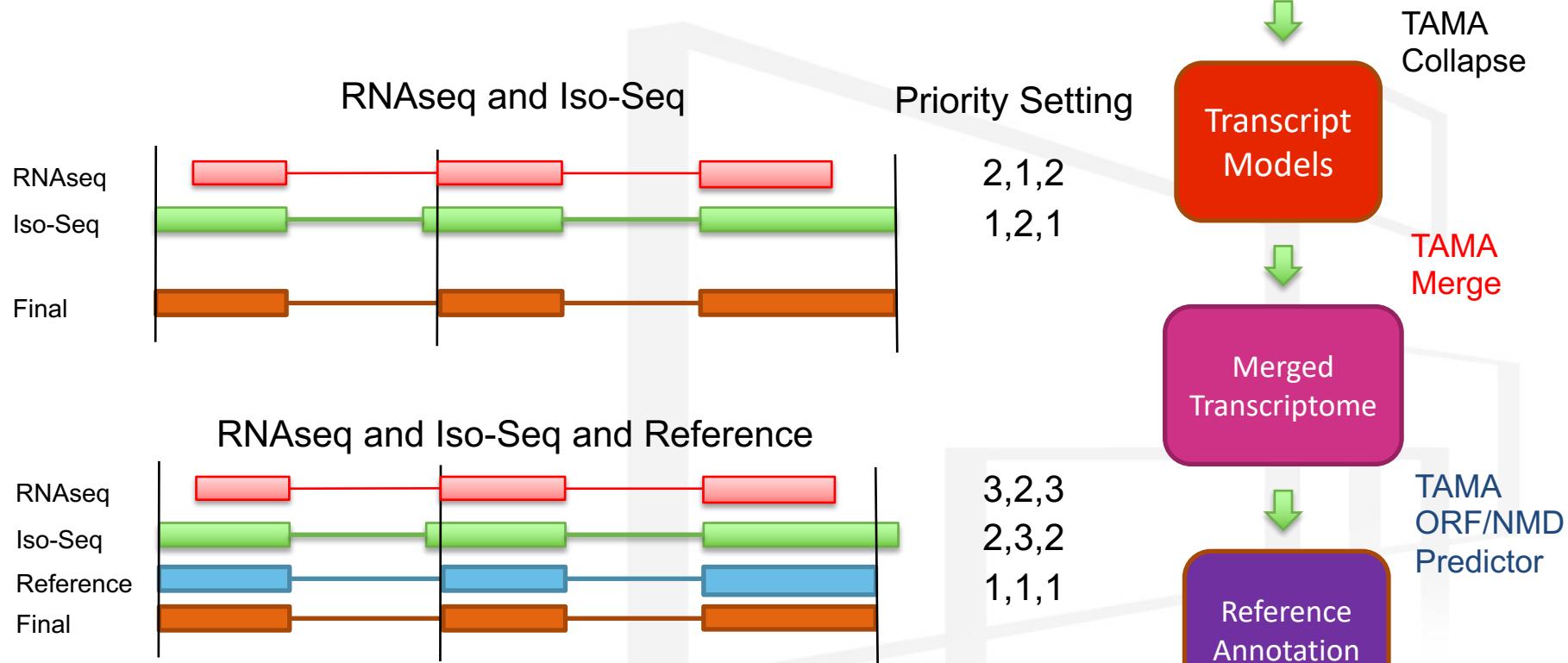
		Multi-read gene transcripts			Degradation Signature
	# FLNC	Capped	No Cap	% decrease	
Clontech	308,283	80,640	43,123	47%	
Teloprime	432,635	10,041	8,696	13%	

- Number of collapsed transcript models
- Only evaluating multi-exon models
- 0% Degradation Signature is impossible

TAMA Merge



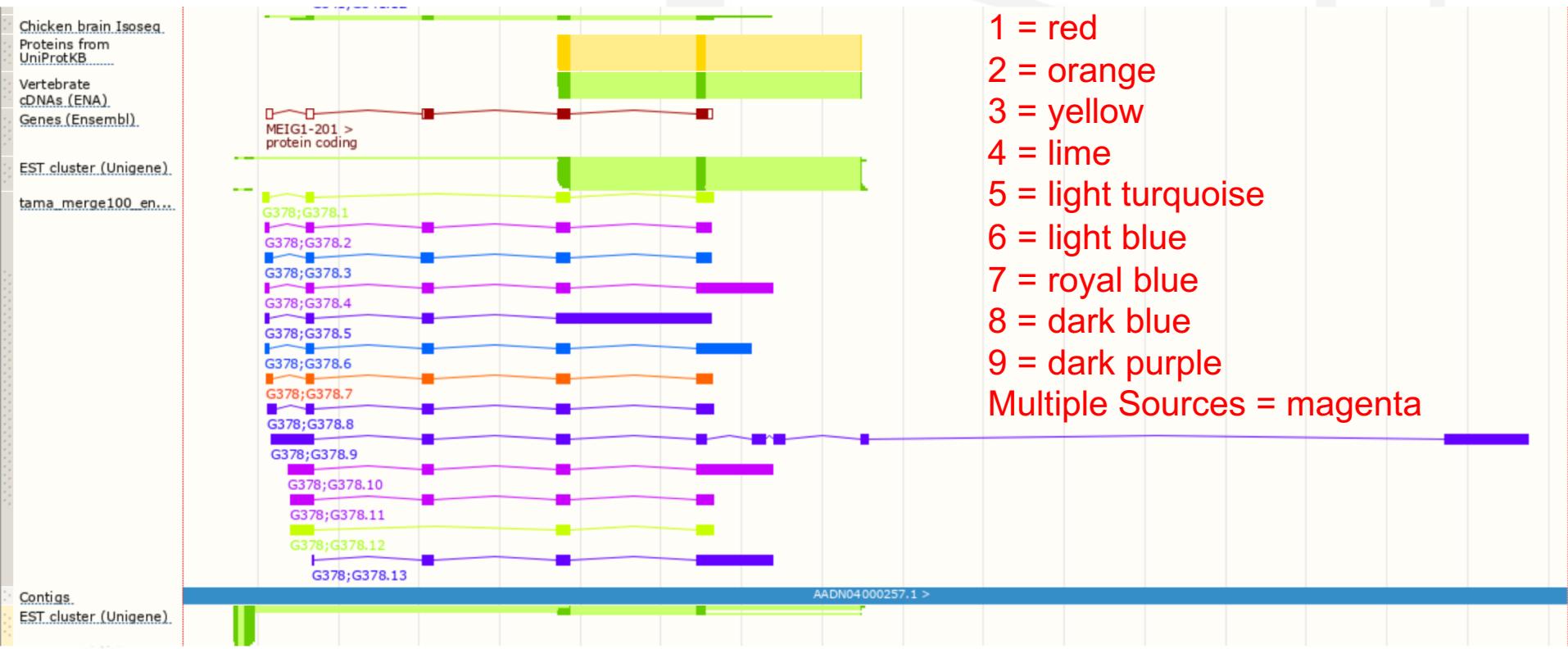
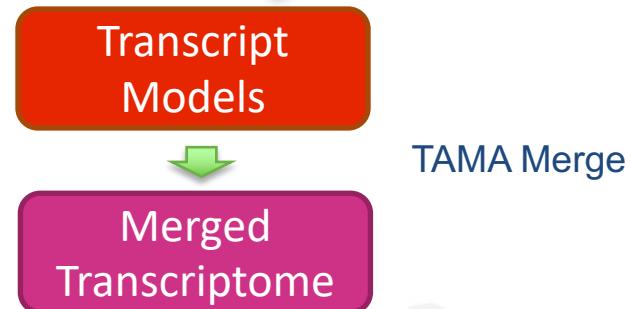
- Merge Iso-Seq projects or multiple sources
- Gives feature priority based on input parameters



TAMA Merge: prefix.bed



- prefix.bed
- GTF convertor available
- Colours = source support



TAMA Merge: Comparison



- Public Annotation Comparison

Ensembl = red

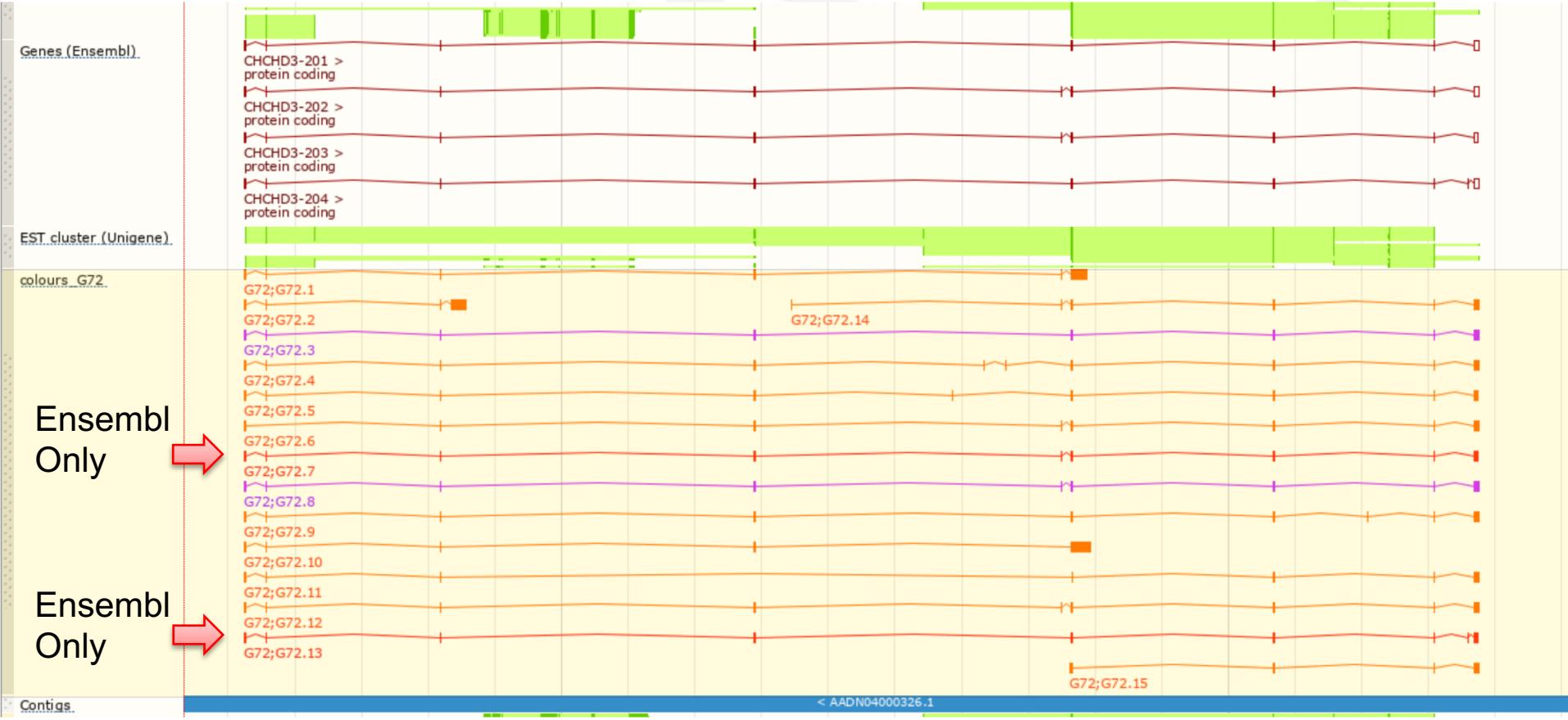
Iso-Seq = orange

Multiple Sources = magenta

Transcript
Models

TAMA Merge

Merged
Transcriptome



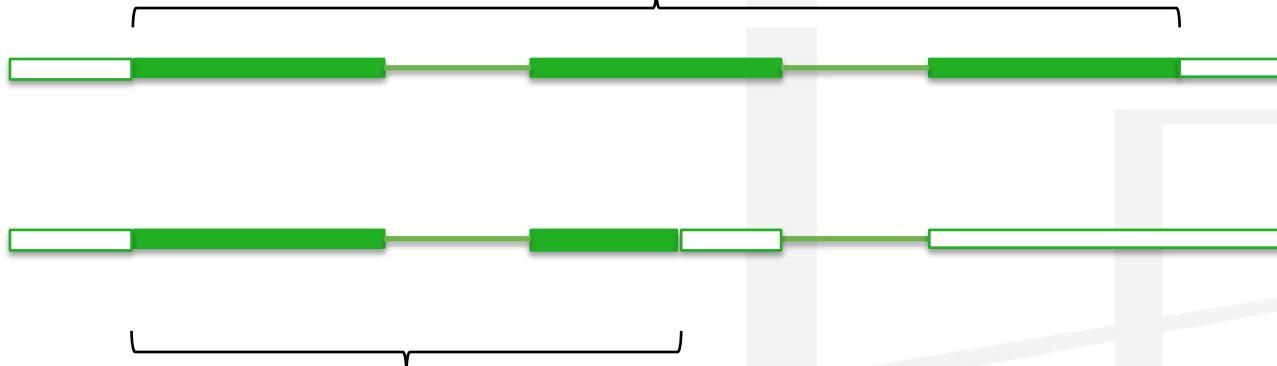
ORF/NMD Prediction



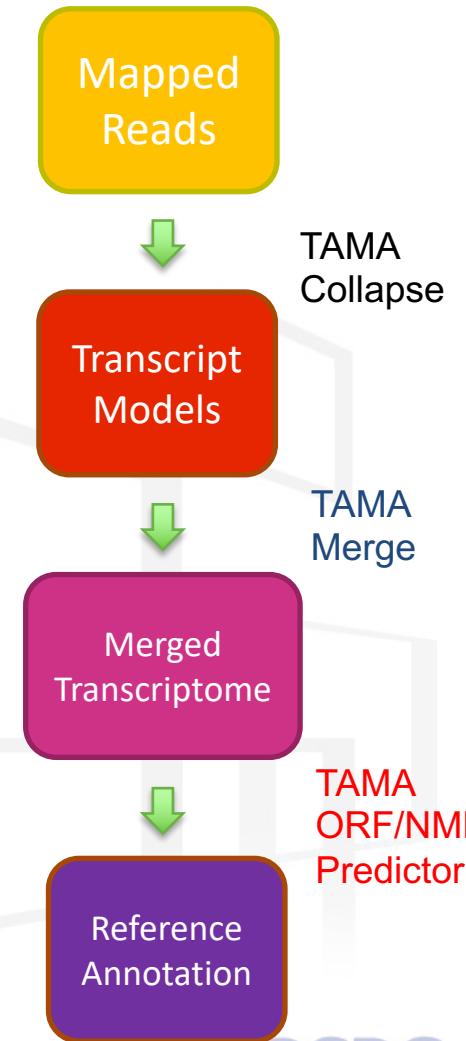
- Protein match
- Open Reading Frame/Coding Region
- Non-Sense Mediated Decay products
- Possible degradation

G1;G1.2;R4GIW7;full_length;50_match;prot_ok;F1

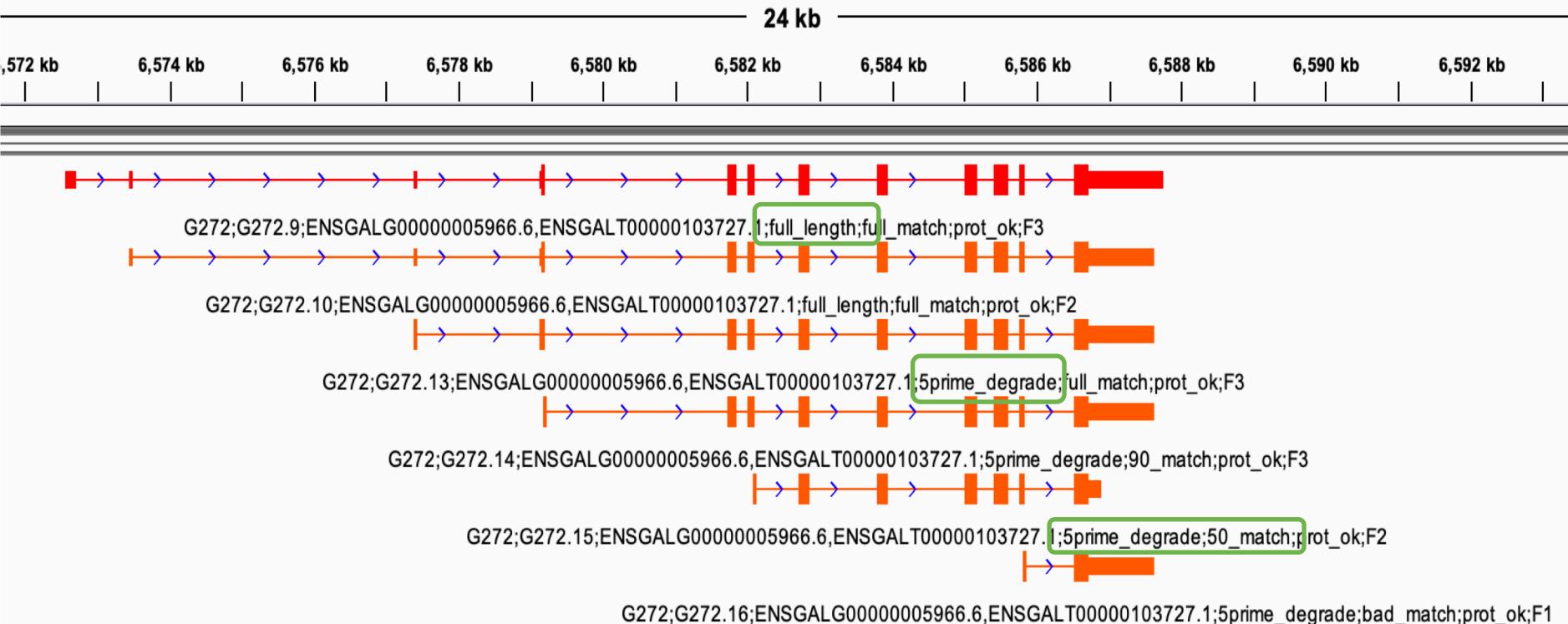
ORF/CDS Region



ORF Region terminates 50bp upstream of last splice junction. Possible NMD.



ORF/NMD Results



Filter by Read
Count

Filter Poly-A
Truncation

Degradation
Signature

Read Jumble
Analysis

Remove
Fragments

Saturation
Curves

More to come!



Acknowledgement



Professor Alan Archibald

Jacqueline Smith

Katarzyna Miedzinska



Elizabeth Tseng



Professor Dave Burt



Runxuan Zhang



Yuanyuan Cheng



Professor John Brown



Robert Kraus Ralf Mueller

