

DATASET “CONFIDENT” VS “PROMISCUOUS”

- FL \geq 2
- FL \geq 1
- pred. acc \geq 99%
- pred. acc \geq 99%

	Confident	Promiscuous
# of Quiver consensus predicted accuracy \geq 99%	101,284	407,366
# of unaligned (to hg38)	0	15
# of low coverage or identity	2,071	18,531
# after filtering above	99,213	388,820
After collapse step:		
# of final isoforms	21,742	75,490
# of final loci (genes)	9,313	18,107

The “confident” dataset uses the standard cutoffs and stringency for PacBio Iso-Seq data. Users of this dataset are advised to stay with using “confident” unless there are strong reasons to use the “promiscuous” dataset.

DATASET “CONFIDENT” VS “PROMISCUOUS”

- FL \geq 2
- FL \geq 1
- pred. acc \geq 99%
- pred. acc \geq 99%

	Confident	Promiscuous
# of final isoforms	21,742	75,490
Isoform Lengths:		
Min – Max (bp)	352 – 9,457 bp	302 – 9753 bp
Average (bp)	3400 bp	2795 bp
Concordance with hg38	99.8%	99.7%

The “confident” dataset uses the standard cutoffs and stringency for PacBio Iso-Seq data. Users of this dataset are advised to stay with using “confident” unless there are strong reasons to use the “promiscuous” dataset.

DATASET “CONFIDENT” VS “PROMISCUOUS”

- FL \geq 2
- pred. acc \geq 99%
- FL \geq 1
- pred. acc \geq 99%

matchAnnot score	Confident	Promiscuous
score = 4 or 5 (perfect match to reference)	7744 (36%)	11931 (16%)
score = 2 or 3 (potential novel isoforms)	12886 (59%)	49814 (66%)
score = 0 or 1 (novel genes)	1112 (5%)	13745 (18%)

- Comparison was done using matchAnnot against Gencode v25

The “confident” dataset uses the standard cutoffs and stringency for PacBio Iso-Seq data. Users of this dataset are advised to stay with using “confident” unless there are strong reasons to use the “promiscuous” dataset.