

Additional TR region Sources

- Already have 5 sources, found 3 more: pbsv, trgt, vamos
- Check how much already captured by Adotto TR regions v0.2
- Expand annotations with new regions

Adotto TR region annotation pipeline:

- Merge regions within then between sources
- Add 25bp slop
- Remove regions within 5kb of reference gaps or outside 10bp to 50,000bp
- Remove regions without non-SNPs in 86 sample pVCF
- Annotate region with TandemRepeatFinder

New Sources intersection with v0.2

		v0.2 Intersection		v0.2 Intersection (50% overlap)	
source	count	count	percent	count	percent
Vamos	467,104	247,925	53.08%	247,924	53.08%
TRGT	163,400	163,083	99.81%	163,083	99.81%
PBSV	316,944	211,921	66.86%	150,227	47.40%

Results

Region counts

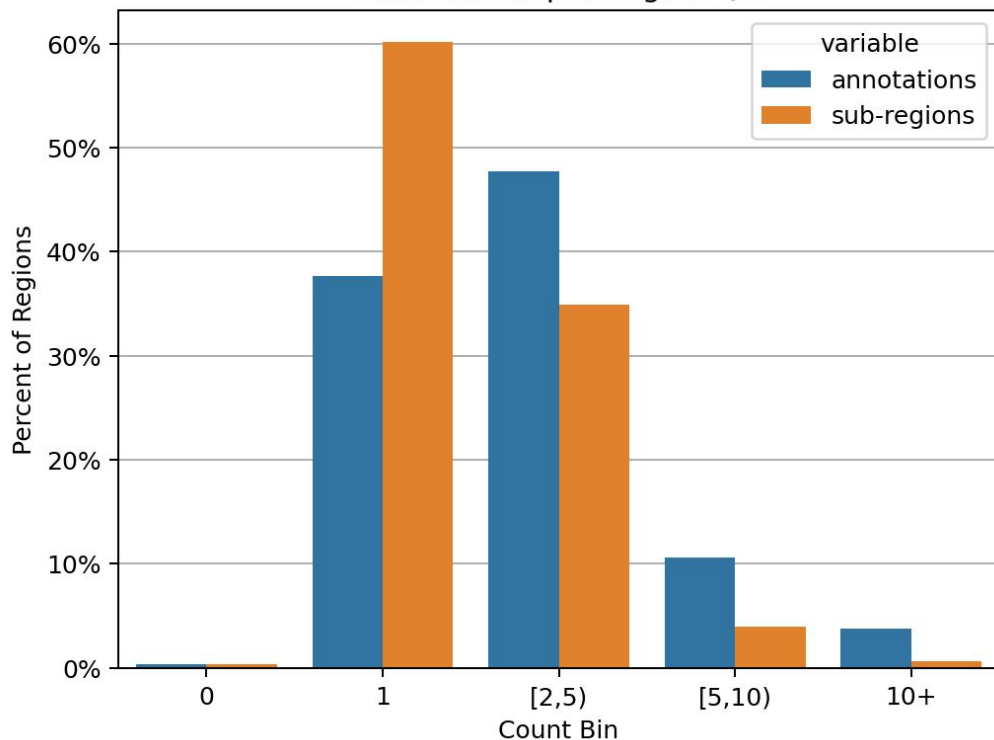
version	count	span	genome percent
v0.1	2,232,565	238,052,458	7.44%
v0.2	938,088	121,788,538	3.81%
v0.3	1,360,416	226,136,695	7.07%

Source intersection

source	count	v0.1	v0.2	v0.3
Vamos	467,104	99.92%	53.08%	68.43%
TRGT	163,400	99.97%	99.81%	99.86%
PBSV	316,944	89.32%	66.86%	85.63%

QC

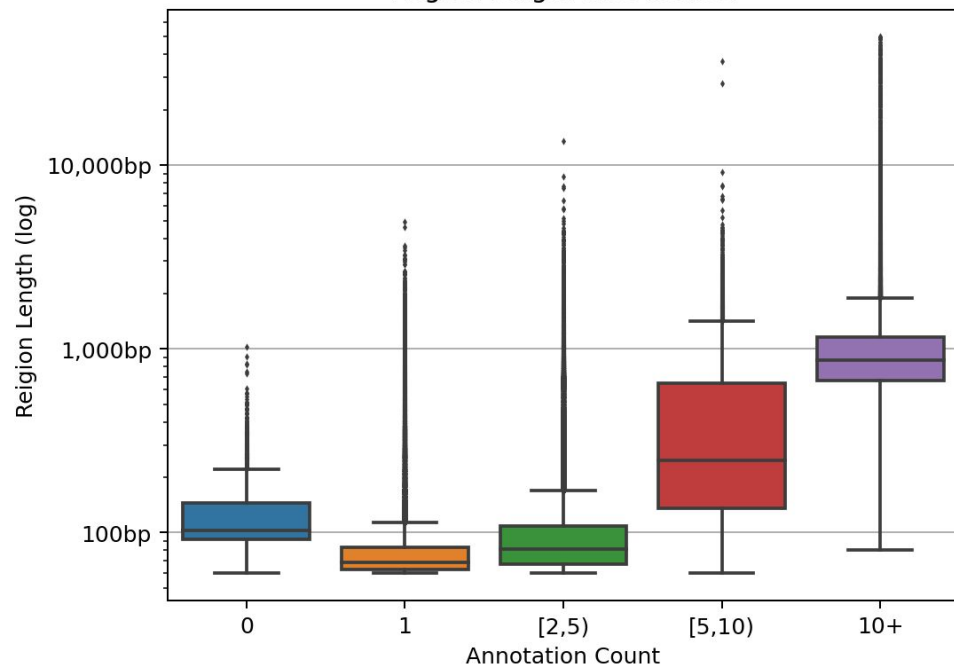
Annotations per-Region QC



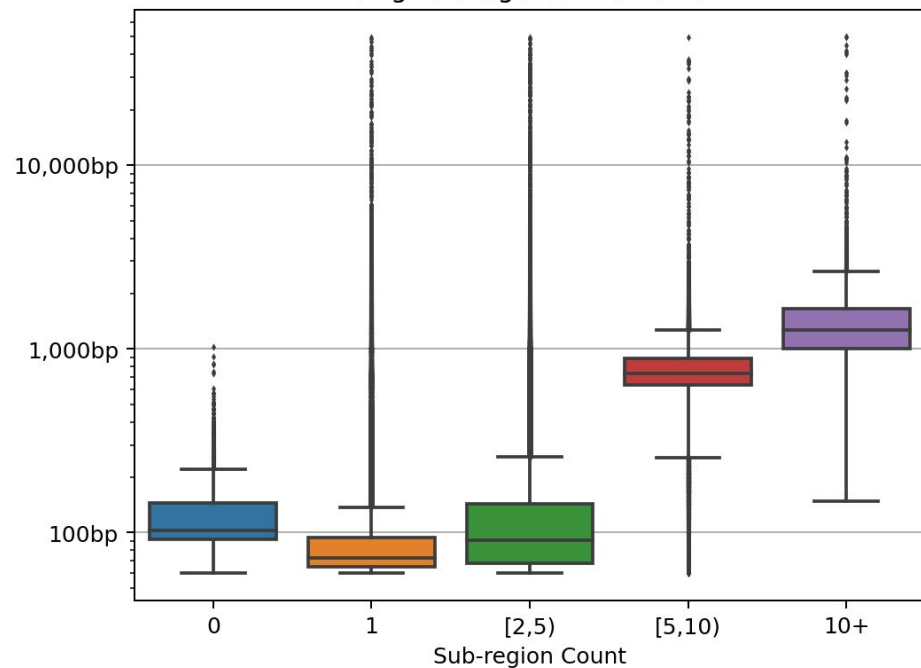
- In a simple world, each region would have a single TR annotation.
- 37.7% of regions have a single annotation.
- 60% of regions have all annotations overlapping (single sub-region)

QC

Region length distribution



Region length distribution

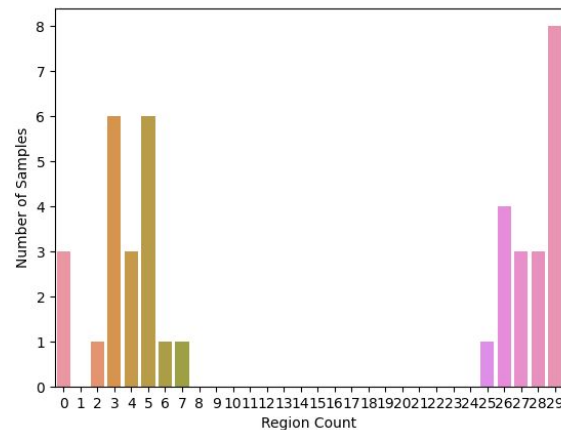


Combined DNA Index System (CODIS)

- 52 CODIS regions on GRCh38
 - 23 autosome regions
 - 29 chrY regions
 - All within adotto tr regions v0.2 and v0.3
- Most autosome regions adequately covered across pVCF
- Males (N=40) chrY have mixed coverage.
 - HPRC assemblies have most regions covered. Eichler/Li do not.
 - HG002 has all but two chrY regions covered
- PentaE, DYS522, and DYS460 regions have no variants
- SE33 has most variants with 109

Autosomal Regions	
Covered Count	Sample Count
23	78
22	7
21	1

ChrY Region Coverage



Pathogenic TRs

- 56 Pathogenic TRs from TRGT catalog
 - All within v0.3
 - 49 within v0.2
- 37 Pathogenic TRs from an ONT study [Stevanosan, 2022 (Garvan)]
 - 3 not in TRGT list
 - Only one is missing from v0.3
 - VMW1 chr1:1435798
 - nearest Adotto region chr1:1435664-1435761 (37bp upstream)
- Total: 58 Patho TRs