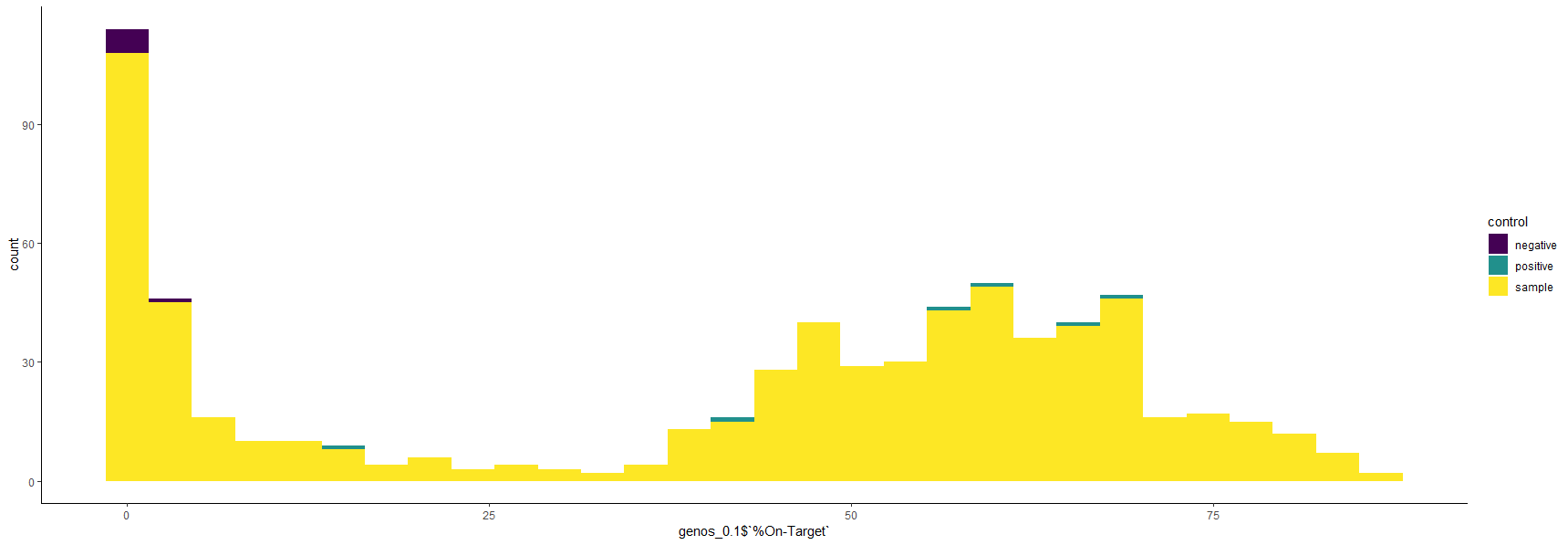
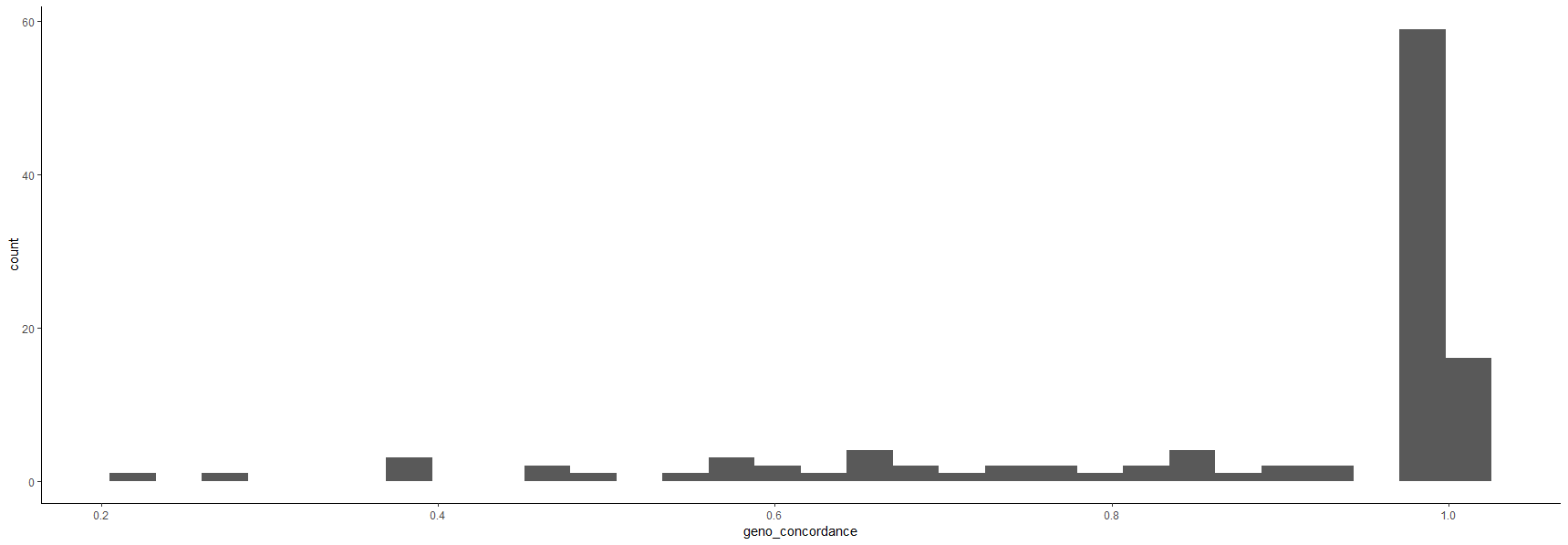
Initial dataset has 673 fastq files genotyped at 352 markers and a sex marker. 7 negative controls and 6 positive controls. 227 duplicate fastq files with repeat sample\_simple IDs are also positive controls. Of the duplicate positive controls 112 are duplicate pairs, one triplicate (OtsAC23UMPR\_0064).

% of reads on target is low in negative controls, and higher in samples and positive controls. Looks good.



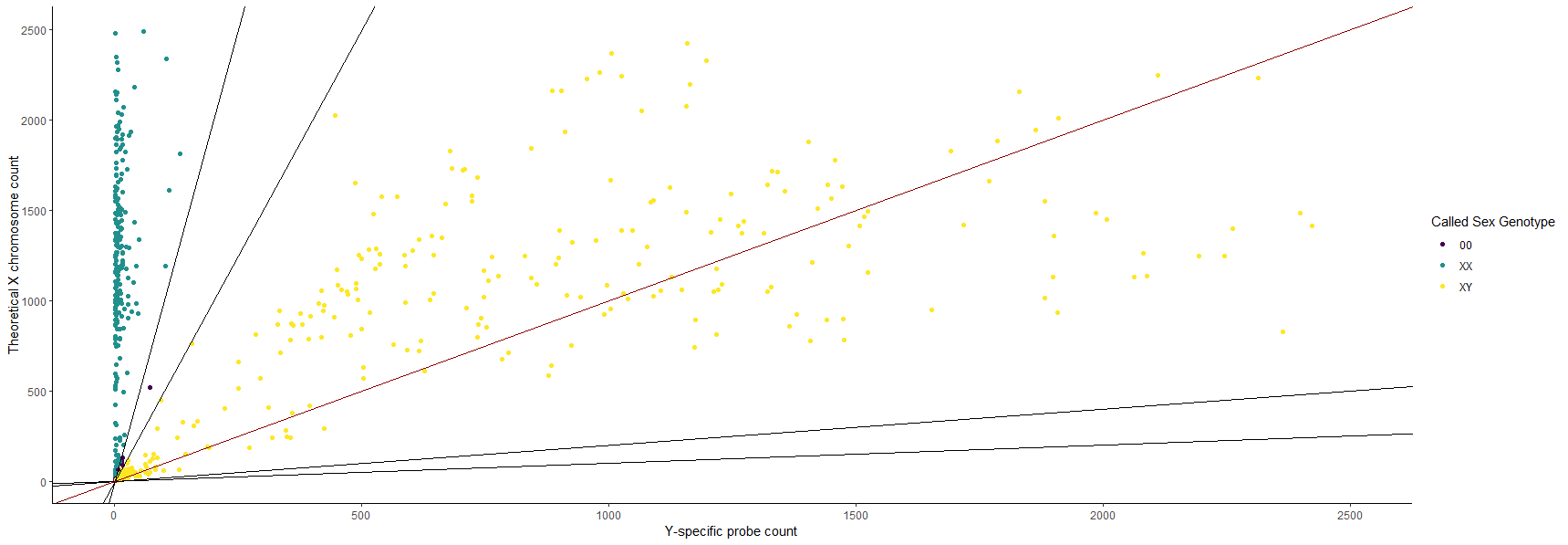
##most duplicate samples have high concordance but some pairs do not



##27 pairs with concordance below 80%

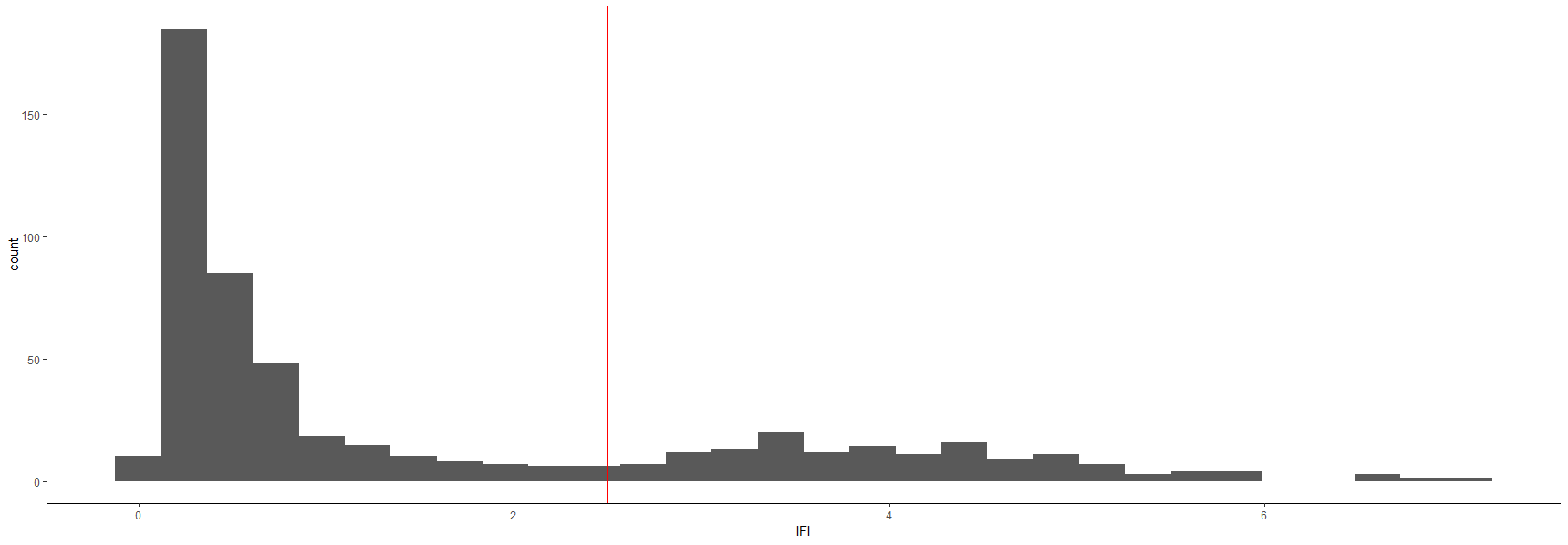
##of those 27, 25 are carcass samples so low concordance is likely due to degradation

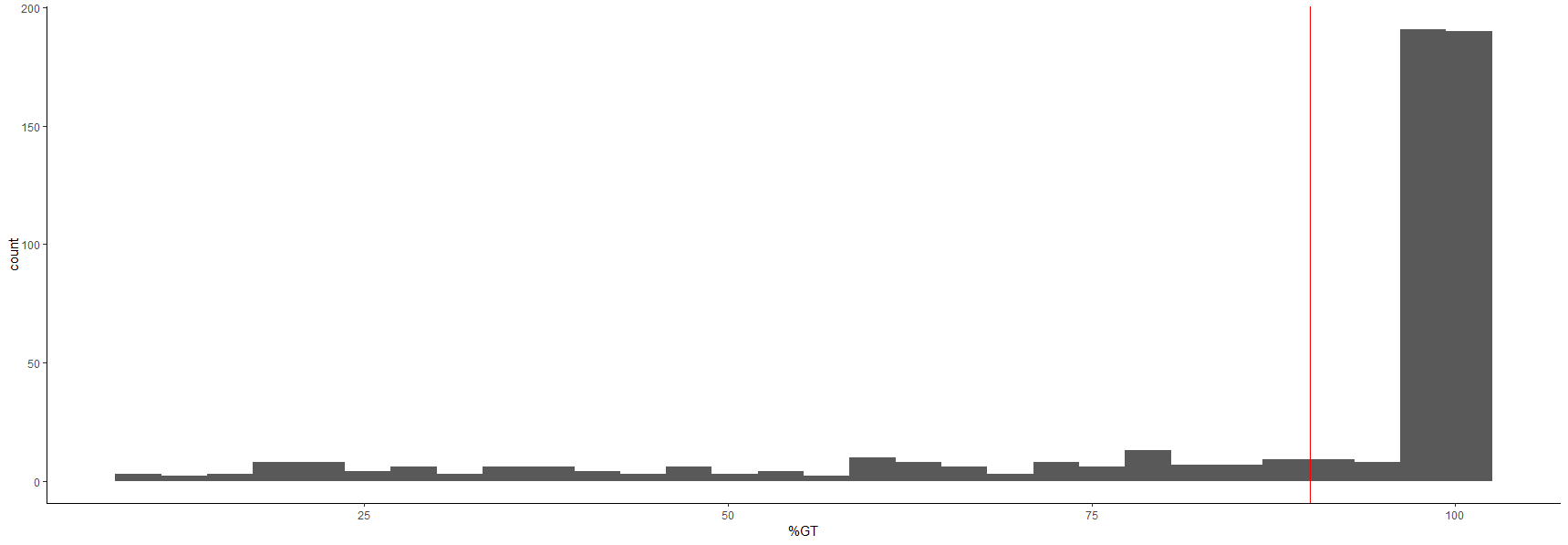
###Sex marker looks good. Leave as is without "correction" script##



Filter individuals/markers:

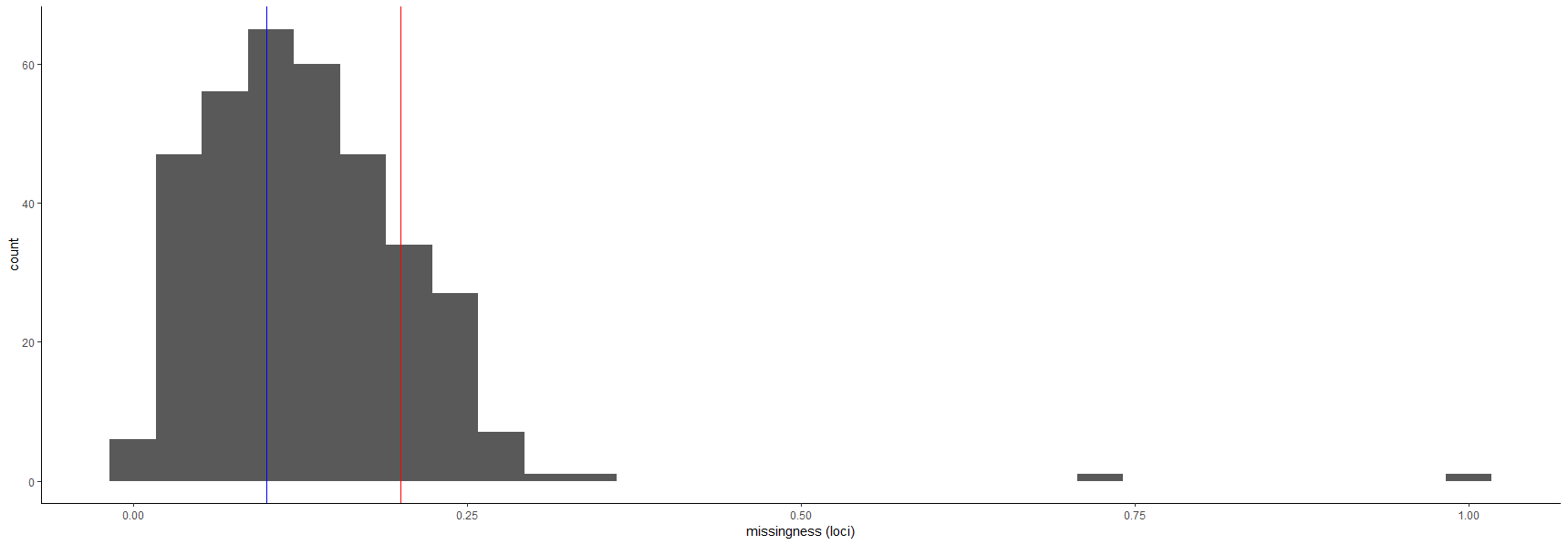
Quite a few samples with high IFI scores and low genotyping call rates





##remove 98 samples with call rates below 70%

Two markers missing in more than 50% of samples (Ots17\_1486479\_C6, Ots\_wenYhap\_33126)



Remove them and recalculate IFI for each sample.

0 samples have IFI scores greater than 10.

Recalculate sample genotyping call rate.

#49 individuals have call rates below 90%##. Remove them.

Recalculate marker call rates.

#all markers called in greater than 80% of samples

Recalculate IFI scores.

##12 samples have IFI greater than 2.5. Remove them.

##Of these 12, 9 are carcass samples again pointing to degradation

##At this point the dataset has 387 Individuals and 350 loci and a sex marker##

##Individual call rates range 91 - 100%; Marker call rates range 81 - 100%

##IFI ranges 0.10 - 2.44##

Evaluate for potential PSVs:

##4 markers removed for being potential paralogs.

"Ots17\_1066109\_C6", "Ots\_CHI06105101\_16717", "Ots\_110495-380", "Ots\_GPDH-338"

12 monomorphic markers:

"Ots\_GH2" "Ots\_IsoT" "Ots\_crRAD23631-48" "Ots\_crRAD26081-28" "Ots\_crRAD46751-42" "Ots\_txnip-321" "Ots\_wenYhap\_106664\_9" "Ots\_wenYhap\_25067\_92" "Ots\_wenYhap\_71572" "Ots\_zn593-346" "Ots\_IGF-I.1-76" "Ots\_u07-64.221"

Normally would remove these but they may be polymorphic when integrating samples from other years so will leave for now and reassess monomorphic markers after combing data across years.

Looked for duplicate genotypes with estimates of relatedness in the program coancestry.

##37 instances of high pairwise relatedness indicative of duplicate genotypes

##Remove 36 samples to resolve duplicate genotypes:

"OtsAC23CHER\_1035", "OtsAC23CHER\_0004", "OtsAC23CHER\_0007", "OtsAC23CHER\_1044", "OtsAC23CHER\_0013", "OtsAC23CHER\_1001", "OtsAC23CHER\_1049", "OtsAC23CHER\_1038", "OtsAC23CHER\_1009", "OtsAC23CHER\_1045", "OtsAC23CHER\_1004", "OtsAC23CHER\_1032", "OtsAC23CHER\_1050", "OtsAC23CHER\_0027", "OtsAC23CHER\_0028", "OtsAC23CHER\_1018", "OtsAC23CHER\_0034", "OtsAC23CHER\_1017", "OtsAC23CHER\_0036", "OtsAC23CHER\_1023", "OtsAC23CHER\_0039", "OtsAC23CHER\_0041", "OtsAC23CHER\_0044", "OtsAC23CHER\_0045", "OtsAC23CHER\_1007", "OtsAC23CHER\_1015", "OtsAC23CHER\_1011", "OtsAC23CHER\_0053", "OtsAC23CHER\_1027", "OtsAC23CHER\_1039", "OtsAC23CHER\_0063", "OtsAC23CHER\_1031", "OtsAC23CHER\_1041", "OtsAC23CHER\_1034", "OtsAC24ELKR\_0016", "OtsCC23SUMP\_0012"

Final filtered dataset has 351 individuals genotyped at 346 markers and a sex marker.