# Genotyping-in-Thousands by sequencing (GT-seq): A low cost, high-throughput, targeted SNP genotyping method

Nathan R. Campbell, Stephanie A. Harmon, Shawn R. Narum
Columbia River Inter-Tribal Fish Commission

**Abstract**
GT-seq is a genotyping method which leverages large read numbers from Illumina sequencers to genotype hundreds of single nucleotide polymorphisms within pools of multiplex PCR amplicons generated from thousands of individual samples (Campbell *et al.* 2015). This method produces genotypes that are 99.9% concordant to those produced using TaqMan™ assays at approximately 1/4$^{th}$ the cost. Since its development, GT-seq panels have been created for several species (Chinook salmon, coho salmon, sockeye salmon, rainbow trout, and pacific lamprey) and has become the preferred SNP genotyping method in our laboratory. New genotyping software allows genotypes and summary figures to be produced from a lane of raw sequencing data in under an hour using a desktop linux computer.

**Figure 1**: Overview of the GT-seq genotyping technique.

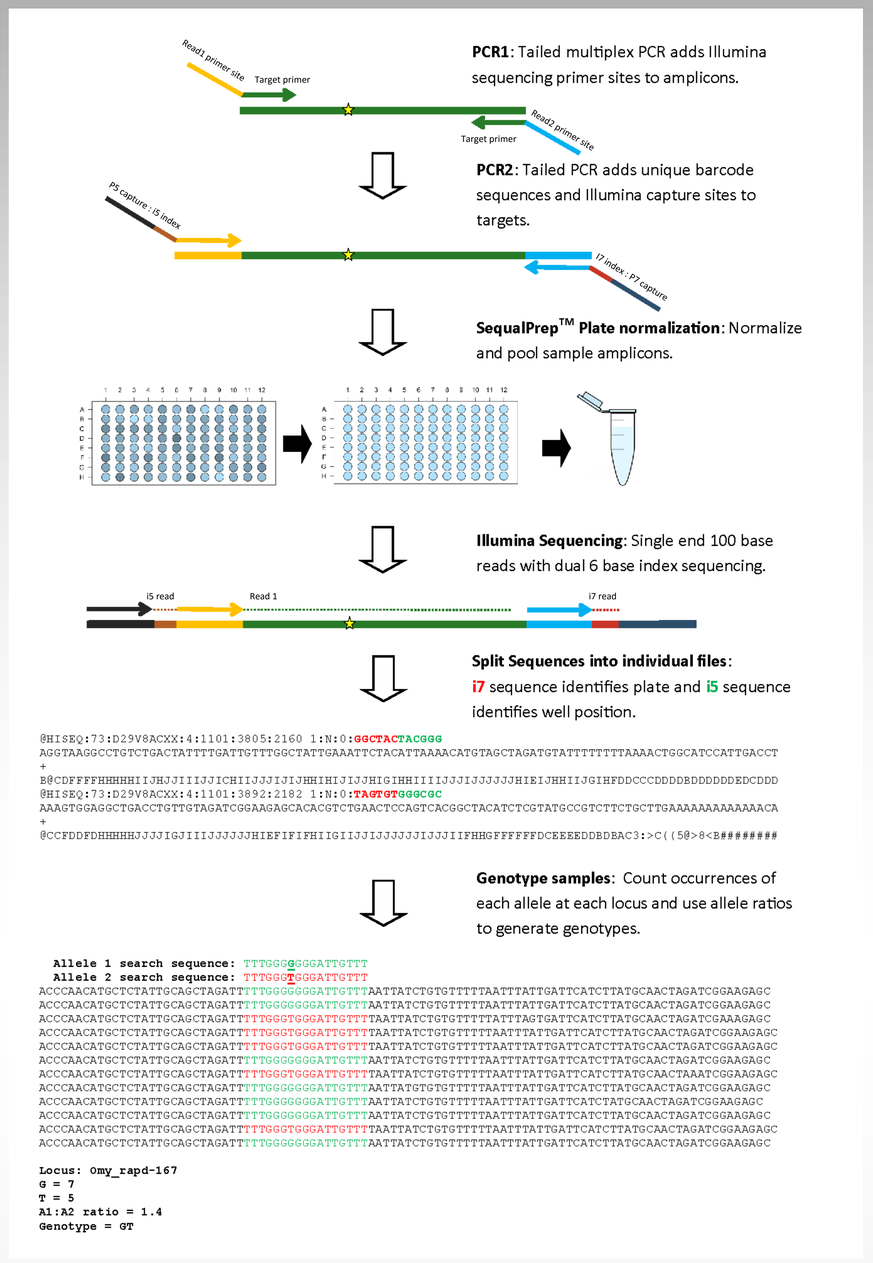



**Figure 2**: Illustrates the relationship between genotyping percentage and sequencing reads.

A.




B.




**Figure 3:** (A) Read depth at each of the 192 target loci was expressed as a percentage of the total "on-target" reads for each of the 1,006 individuals included in the library. The red line indicates the read depth if all loci were evenly represented. (B) This illustrates the percentage of "On-Target" reads for each SNP locus. Calculated as [Number of reads beginning with forward primer AND containing an expected allele sequence] / [All reads containing forward primer].

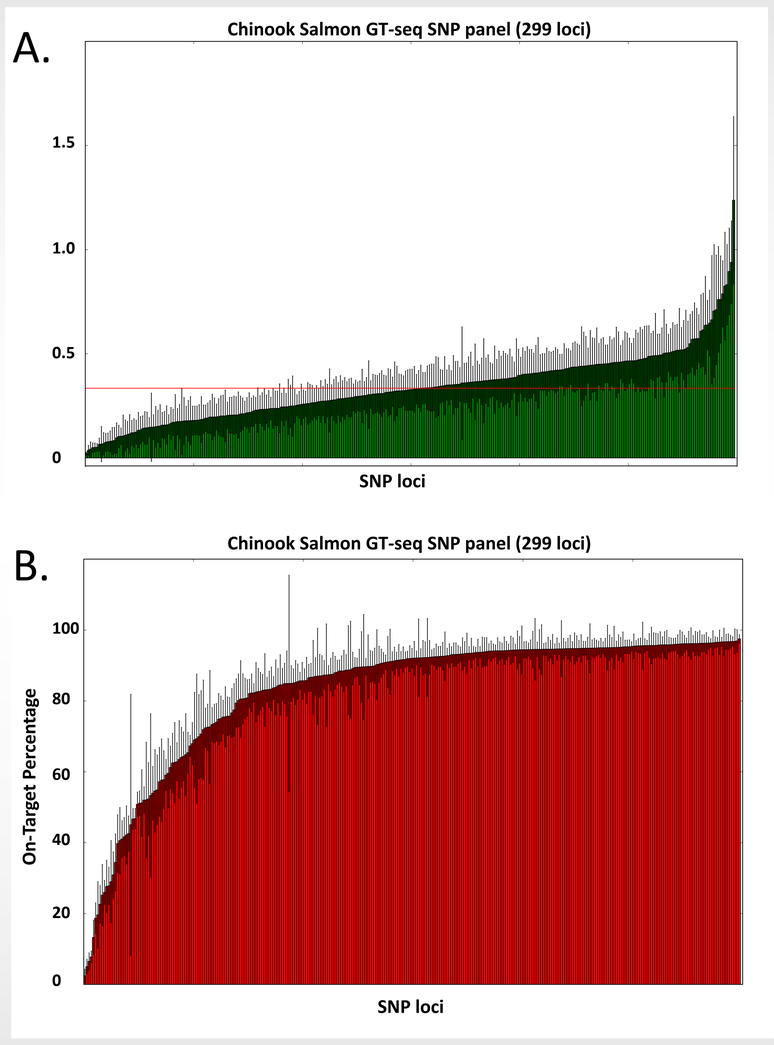



**Figure 4:** This graph plots allele 1 vs. allele 2 read counts for all target loci illustrating very low background signal and predictable allele ratios using GT-seq. The Chinook salmon panel contains 299 loci including a sex determination marker. Solid lines indicate cutoff boundaries for generating genotypes. The purple line represents an even allele ratio that would be expected for heterozygous samples.

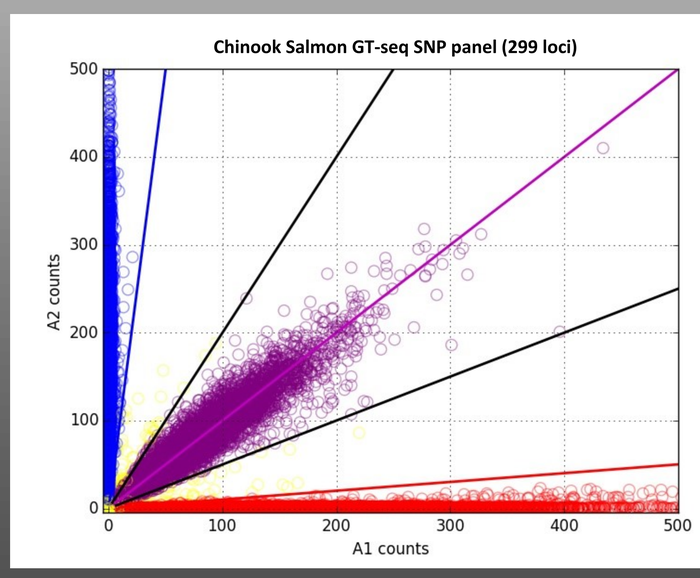



**Figure 5:** Genotyping plots for all included SNP loci can be generated using the GTseq_SummaryFigures python script. Two examples shown.

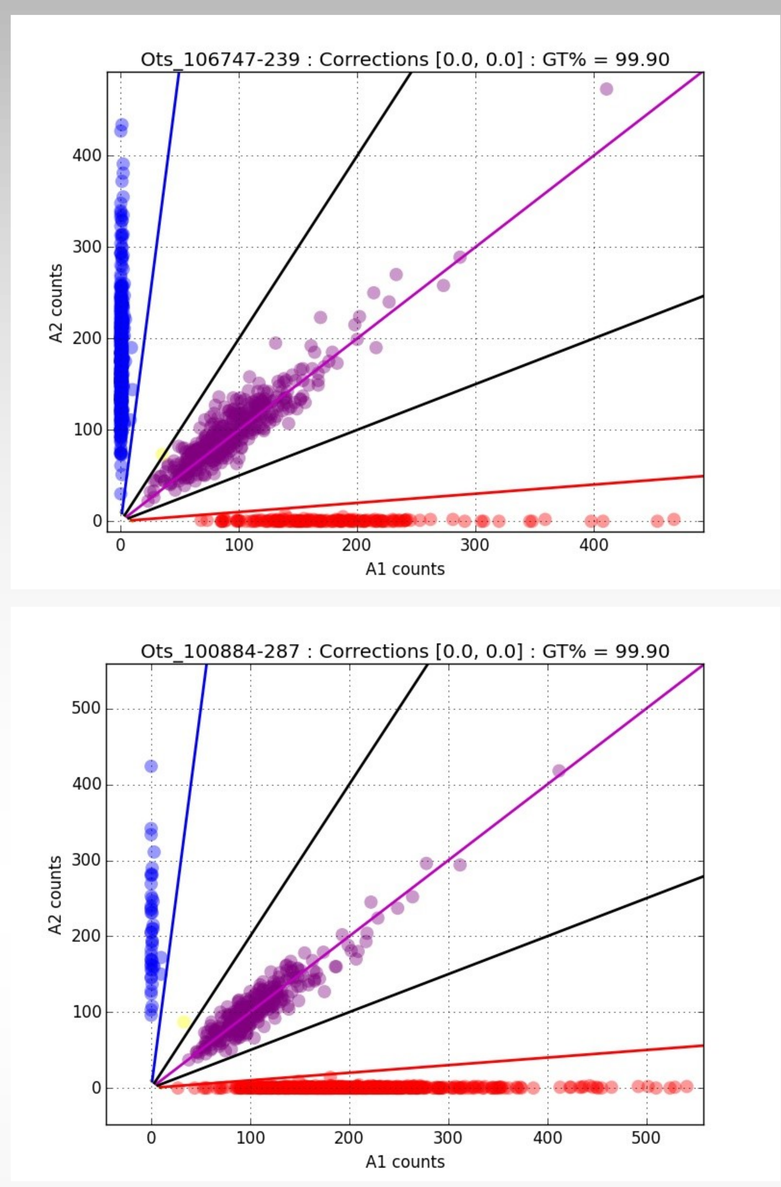
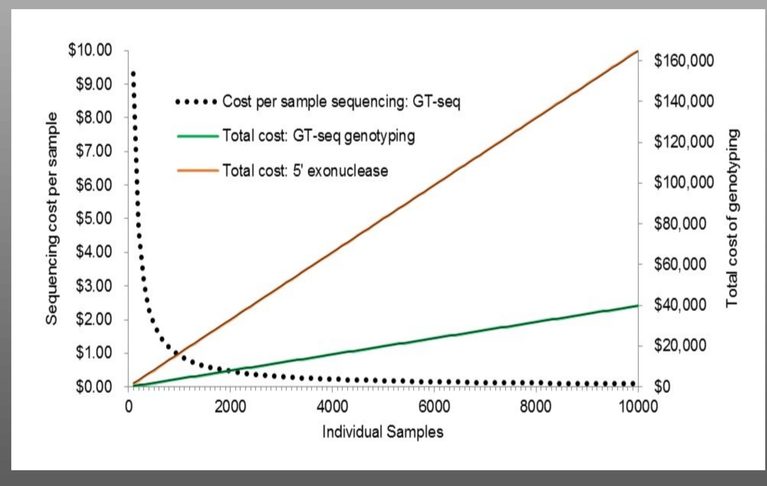



**Figure 6:** Estimated genotyping costs with GT-seq protocol. The dotted line (primary y-axis) represents only the costs of *sequencing* per sample, specific to one lane of single read 100 on an Illumina HiSeq. The dotted line shows that at a minimum of one 96-well plate of samples in a single lane, sequencing costs are $9.30 per sample and this amount decreases exponentially to $0.44 per sample for 2,068 samples in one lane as shown in the current study. The secondary y-axis shows the total cost of genotyping 192 SNPs including DNA extracts, library preparation, and sequencing costs for both the 5' exonuclease method at $16.50 per sample versus the total cost of GT-seq at $3.98 per sample with 2,068 samples per library.




**Campbell, N. R., Harmon, S. A., Narum, S. R. (2015) Genotyping-in-Thousands by Sequencing (GT-seq): A Cost Effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources* (15) 855-867.**