



Figure S3 Heuristic for filtering high-confidence data set.

(A) The distribution of the log transformed slopes of the nearly 30,000 barcodes (Figure S1) associated with wild-type BRCA1(2-304) sequences (input read cut-offs represented by color). The poor scoring wild-type variants are thought to be due to loss of individual barcodes that follows a Poisson distribution due to experimental bottlenecks. (B) The 800 input read count cut-off maximizes the number of variants contributing to the analysis (black line) while maintaining the maximum Spearman's rank correlation between the six experimental replicates and minimizing barcode dropout due to bottlenecks (A). Estimates of variance and 95% confidence intervals can be found for each measurement in Table S2.