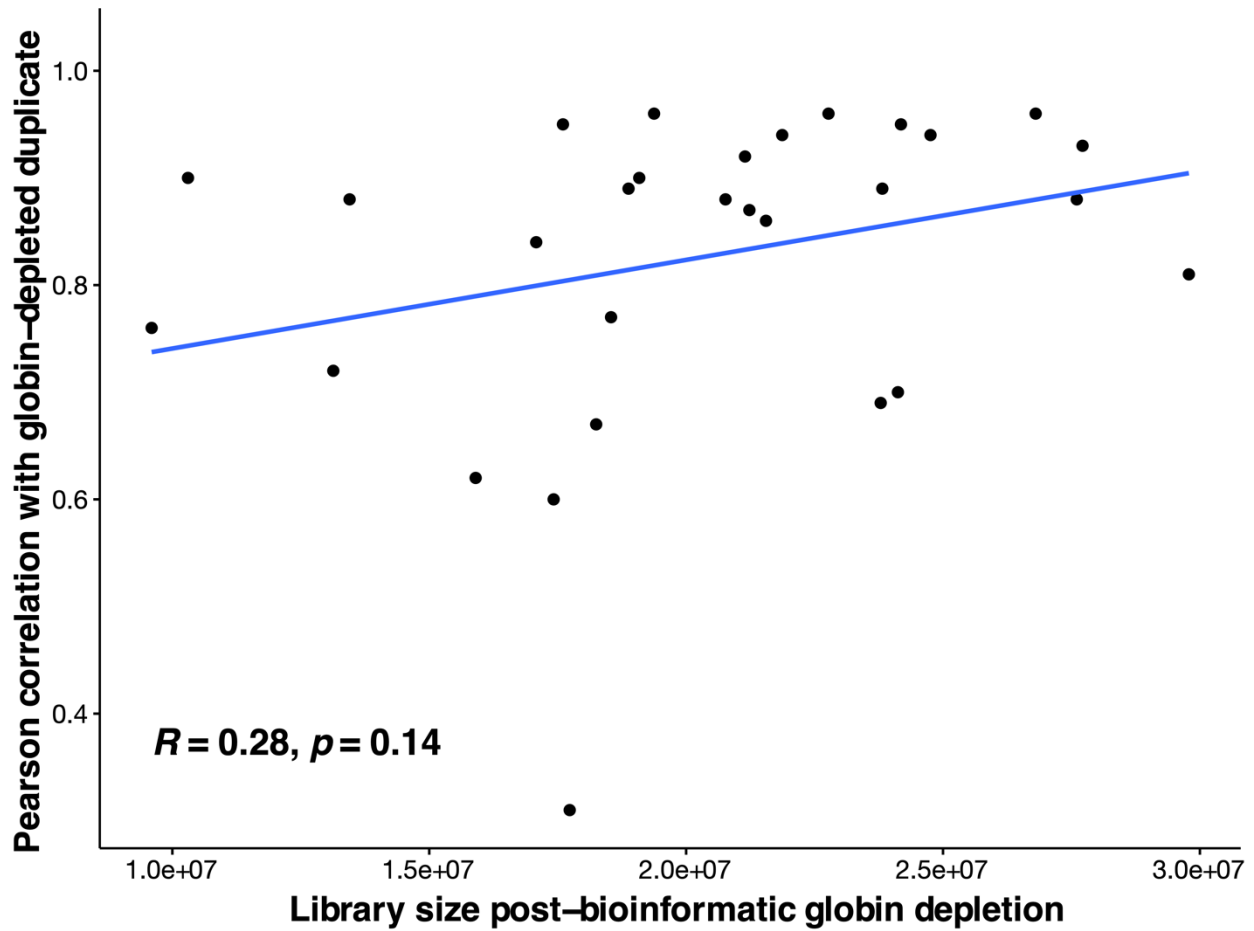




**Supplementary Figure 1. Gene set variation scores associated with each of 70 published tuberculosis (TB) signatures between kit-depleted and bioinformatic-depleted paired samples.**

Gene set variation analysis (GSVA) was performed using the TBSignatureProfiler package<sup>8</sup> in R to generate GSVA scores for 70 TB signatures from the published literature, curated by the package. Each signature is labelled by the first author of the signature and the number of genes in the signature. This analysis was performed on 29 kit-depleted (blue) and 29 matched bioinformatic-depleted (green) samples. Horizontal coloured lines link paired samples. Box plots indicate the median, 1<sup>st</sup> quartile and 3<sup>rd</sup> quartile while the whiskers indicate the lower and upper adjacent values. A paired t-test was performed to compare each group of samples with and without prior globin depletion and the Bonferroni-corrected p-values are displayed; \* $<0.05$ , \*\* $<0.005$ , \*\*\* $<0.00005$ . Signatures with significant differences between the two groups are highlighted in red.





**Supplementary Figure 3. Poor correlation between matched samples cannot be explained entirely by differences in library size.** Pearson correlation between the library size of each sample following bioinformatic globin depletion and the Pearson correlation values derived from a correlation analysis of gene set variation analysis (GSVA) scores generated from 70 tuberculosis signatures between matched pairs of globin-depleted ( $n = 22$ ) and non-depleted samples ( $n = 22$ ). The blue line represents the fitted linear regression.