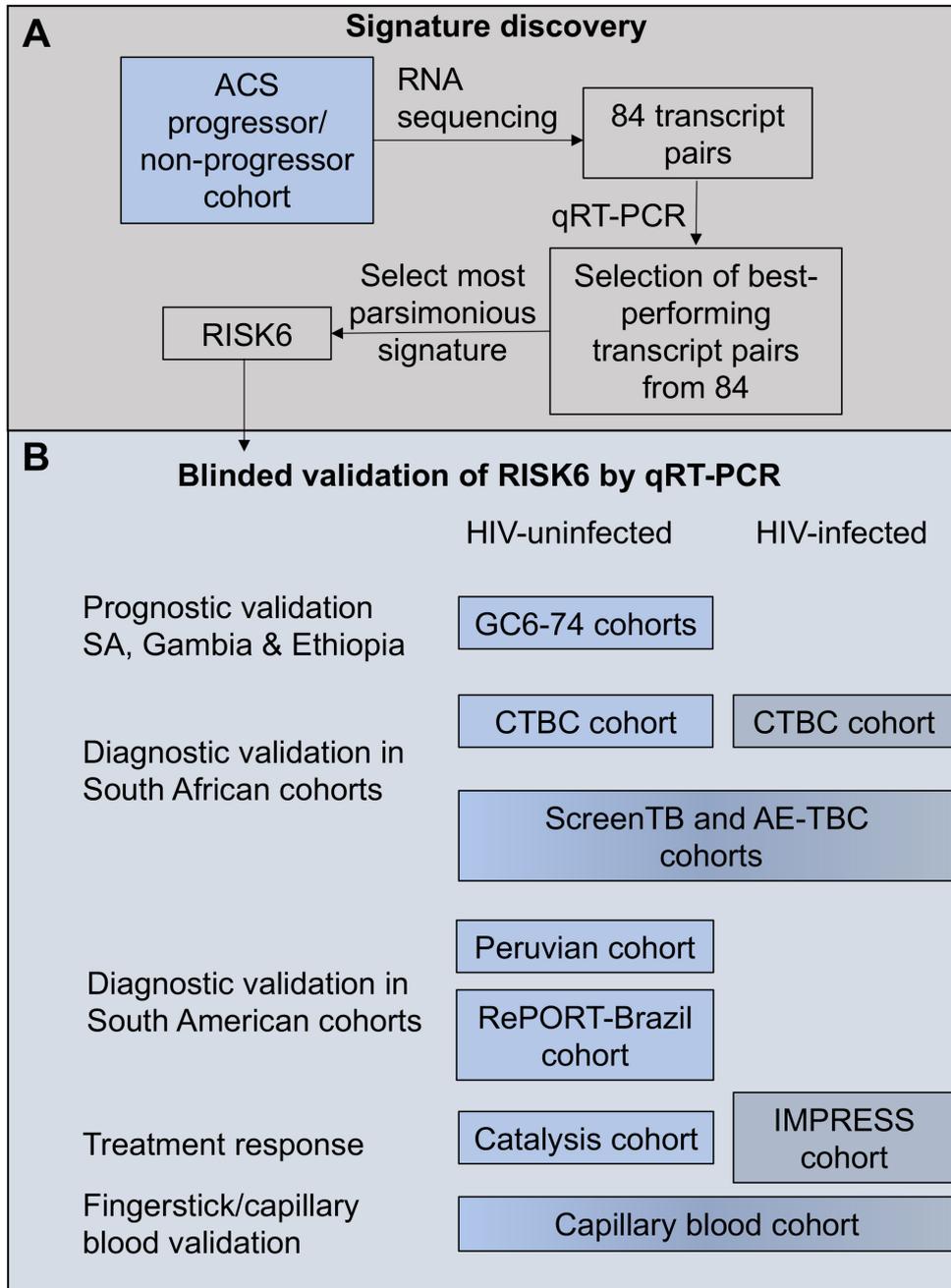


Supplement:

Penn-Nicholson, Mbandi and Thompson et al., RISK6, a universal 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response

Supplementary Figures



Supplementary Figure 1: Discovery and validation of RISK6. (A) Schematic depiction of the process to discover the RISK6 signature in TB progressors and non-progressors from the Adolescent Cohort Study, using samples collected within 360 days of TB disease diagnosis⁶. Differentially expressed exon junctions were identified using RNA-sequencing data and

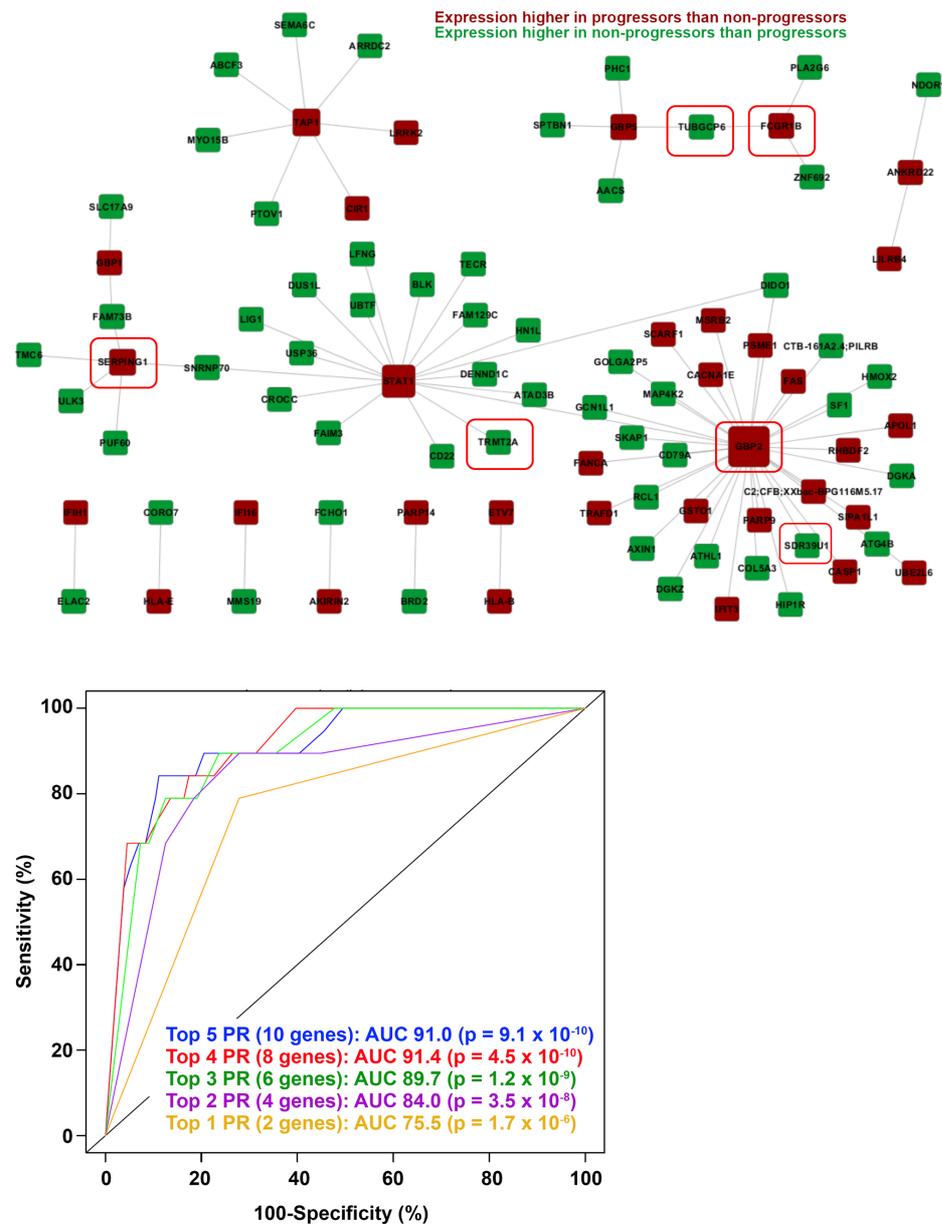
Supplement:

Penn-Nicholson, Mbandi and Thompson et al., RISK6, a universal 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response

support vector machines of all possible pairs of junctions were trained using the Pair-Ratio approach (see methods). 84 junction pairs that differentiated progressors and non-progressors with the highest sensitivity and specificity were selected and quantified by microfluidic qRT-PCR. Using these PCR data, a small set of junction-pairs was selected to train the most parsimonious signature with excellent performance in the ACS cohort, RISK6. (B) RISK6 was validated by blind application using microfluidic qRT-PCR to samples from 5 different cohorts, arranged according to prognostic, diagnostic or treatment response validation and HIV status.

Supplement:

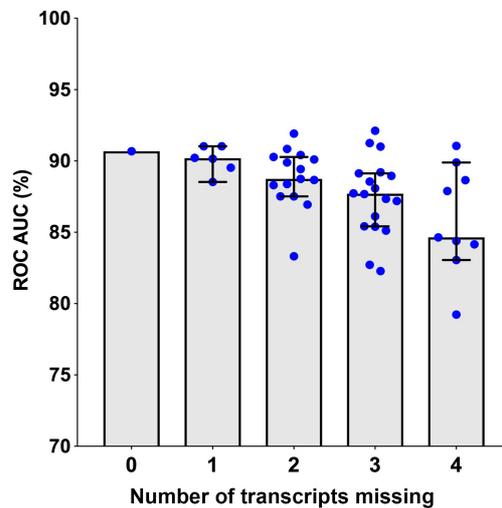
Penn-Nicholson, Mbandi and Thompson et al., RISK6, a universal 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response



Supplementary Figure 2. Network representation of the pairs (depicted by the edges) formed between 84 unique exon junctions (depicted by the nodes) identified from RNA-seq data in the adolescent progressor and non-progressor cohort, by the random subsets approach. This method randomly selects a partition of half the samples with a quarter of the features, to train support vector machines of all possible pairs of junctions using the Pair-Ratio approach. The Pair-Ratio approach pairs transcripts that are regulated in opposite directions in progressors and non-progressors, depicted as red (higher in progressors) and green (higher in non-progressors) nodes. The ROC plot at the bottom depicts the AUCs (and p-values) of 5 parsimonious pair-ratio ensemble signatures comprising between 2 to 10 genes (comprising the top Pair-Ratios, PR), measured by qRT-PCR in samples collected within 6 months of TB diagnosis in ACS progressors versus non-progressors.

Supplement:

Penn-Nicholson, Mbandi and Thompson et al., RISK6, a universal 6-gene transcriptomic signature of TB disease risk, diagnosis and treatment response



Supplementary Figure 3. (A) ROC AUC values for discrimination between TB cases and asymptomatic controls in the Brazilian cohort by the full 6-gene RISK6 signature (nine pairs formed between six transcripts, far left), or after removing 1, 2, 3 or 4, of the transcripts such that every combination of the pairs (represented by individual blue dots) was tested. The grey bar graph represents the mean and the error bar the 95% CI.