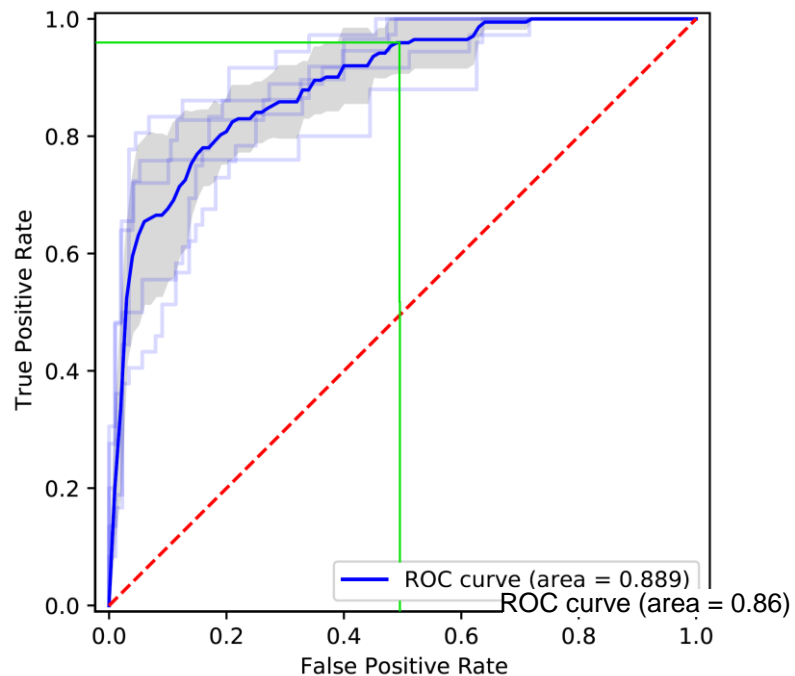
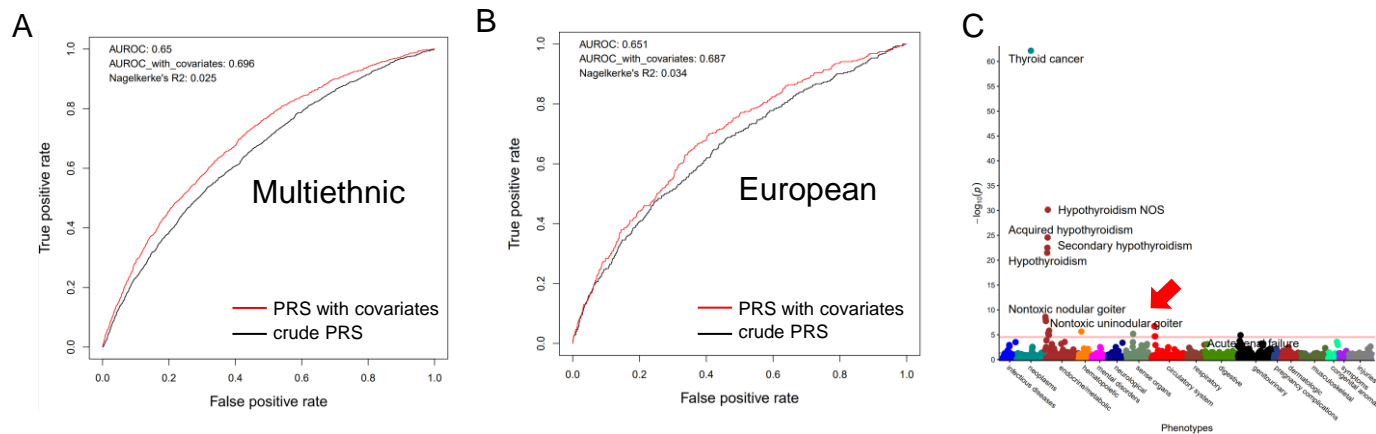


Supplementary Figure 1. Example training curves for the deep-learning thyroid nodule classifier. Changes in AUROC (A) and validation image set loss (B) with training epochs are shown. In this example, training was continued for 600 epochs, and the model was overfitted on validation images after epoch 290. The arrows show a training epoch when the validation loss reached its minimum.



Supplementary Figure 2. The receiver-operating characteristic (ROC) curve for the deep learning thyroid nodule classifier trained using scaling by nodule size and 6 images per video clip per nodule. ROC for 5 cross-validation folds and average are shown. The threshold for binary classification was set to match the sensitivity of thyroid nodule FNA biopsy (~ 0.95).

Supplementary Figure 3. Thyroid nodules of the test image set. For each nodule, transverse and longitudinal grayscale ultrasound images and benign and malignant GradCAM heatmaps are shown. ID: arbitrary thyroid nodule identifier. Class: 0 – benign; 1 – malignant. Pmalign: the probability of malignancy estimated by deep learning classifier. PRS: standardized polygenic risk score.



Supplementary Figure 4. Thyroid cancer polygenic risk score receiver-operating characteristic curve for the entire CCPM Biobank (A) and a subset of individuals of European ancestry (B). C. 26 SNP PRS PheWAS (red arrow highlights the association of 26 SNP PRS with benign nodular goiter phenotypes).