Supplemental Figure 3. MCC and ROC AUC statistics across all machine learning algorithms and all prediction models in relation to short- and long-term diabetes incidence (N=2,067). Abbreviations: ANN – artificial neural network; GLM – generalized linear model (refers to logistic regression here); MCC – Matthews correlation coefficient; RF – random forest; ROC AUC – receiver operating characteristic under the curve; SGB – stochastic gradient boosting; SVM-L – support vector machine with linear kernel; SVM-P – support vector machine with polynomial kernel; SVM-R – support vector machine with radial kernel. MCC averages are represented by circles and ROC AUC averages are represented by squares. The averages are calculated from the five obtained MCC and ROC AUC values from the five separate test sets in the nested cross-validation framework. The error bars represent standard deviations of the five obtained MCC and ROC AUC values. The left panel shows discriminative utilities for short-term, while the right panel shows discriminative utilities for long-term diabetes incidence. This figure demonstrates model results for all models (N=2,067).