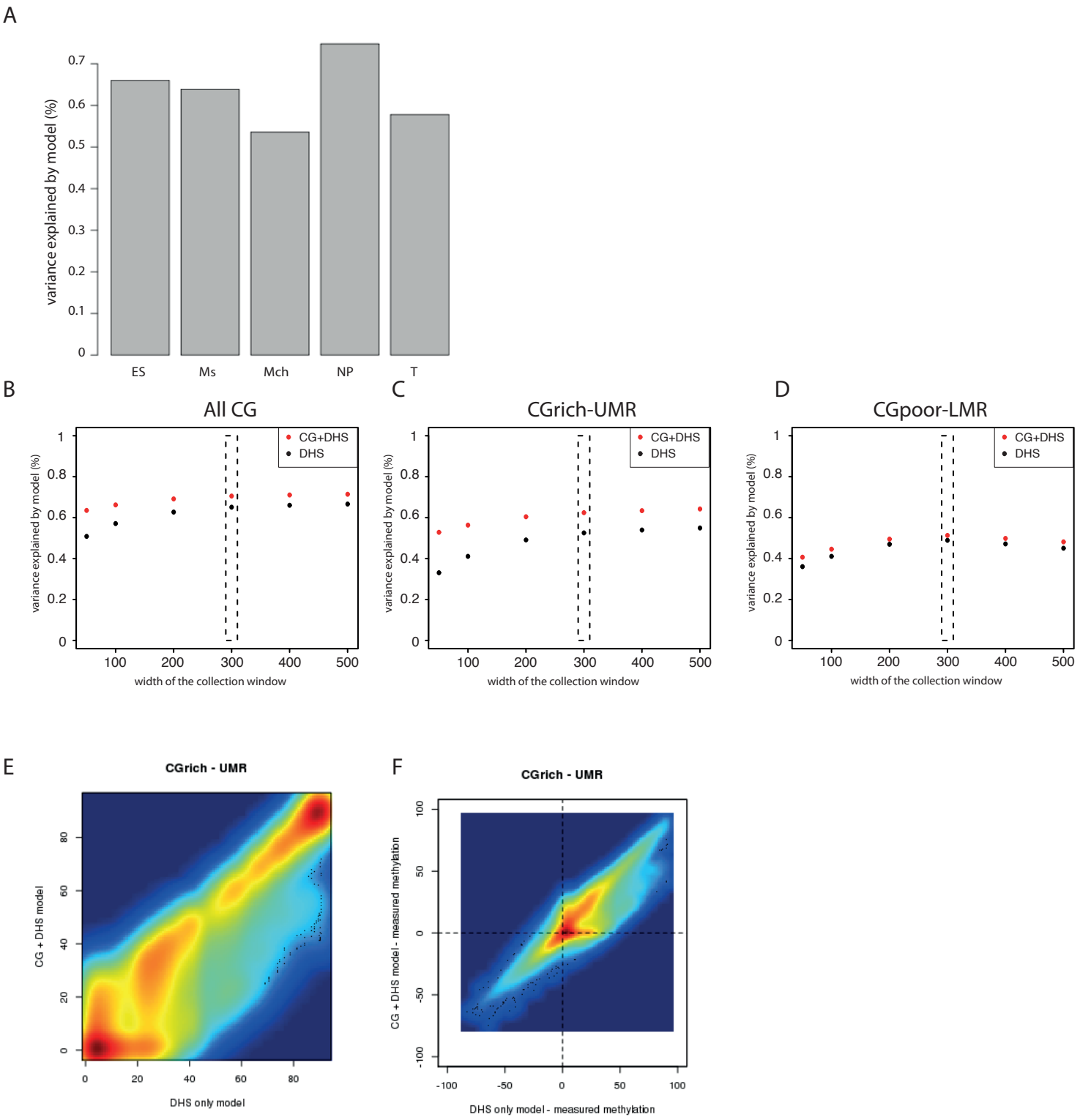

Figure 4-figure supplement 1:
(A) Pearson correlation (R) between the complete model prediction and the measured methylation in human stem cells and four in vitro derived cell types (Mch-Mesenchymal ; Ms-Mesoderm ; NP-Neuronal Progenitor ; T-Trophoblastic). (B) Quantification of the proportion of CGs predicted accurately for each model depending on their genomic context. The prediction of each model was compared to methylation as measured by bisulfite sequencing and prediction accuracy was quantified (with a precision of 20% methylation). The barplot illustrates the improvement gained by each variable used in the modeling. It shows that the combination of CG density and DHS is particularly important to accurately predict methylation at CG rich regions. (C-E) Influence of the size of collecting window for quantifying DHS signal on modeling performance. The coefficient of determination R2 for the DHS only (DHS) or the combined model (CG+DHS) was calculated as a function of the collection window in the DHS dataset. The analysis was performed for all CGs in the genome or for CGs within particular genomic regions. This shows that 300bp is the optimal collection window for DHS regardless of the type of region considered. (F) Comparison of the prediction by the DHS only (DHS) and the combined model (CG + DHS) within CG rich unmethylated regions (UMR). For a significant fraction of the CGs the predicted methylation is lower in for the complete model. (G) Comparison of the prediction accuracy for the two models. The delta between predicted and measured value is calculated for both model and plotted against each other. This reveals that the DHS only model overestimates DNA methylation for a significant part of the CGs within UMRs that are more accurately predicted by the combined model.