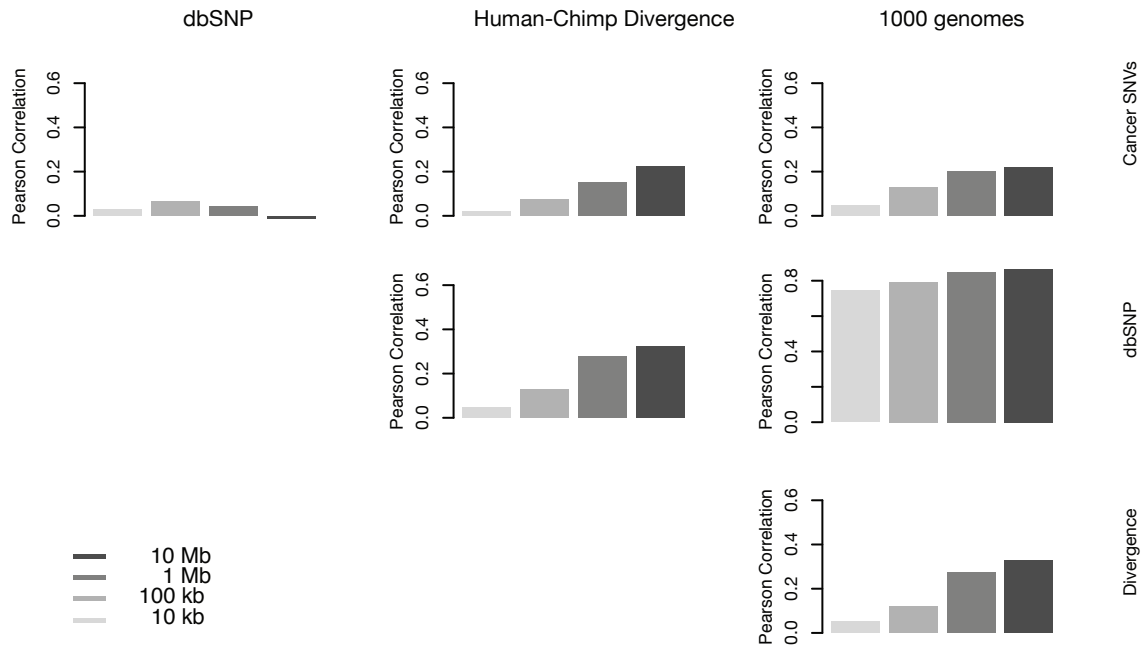
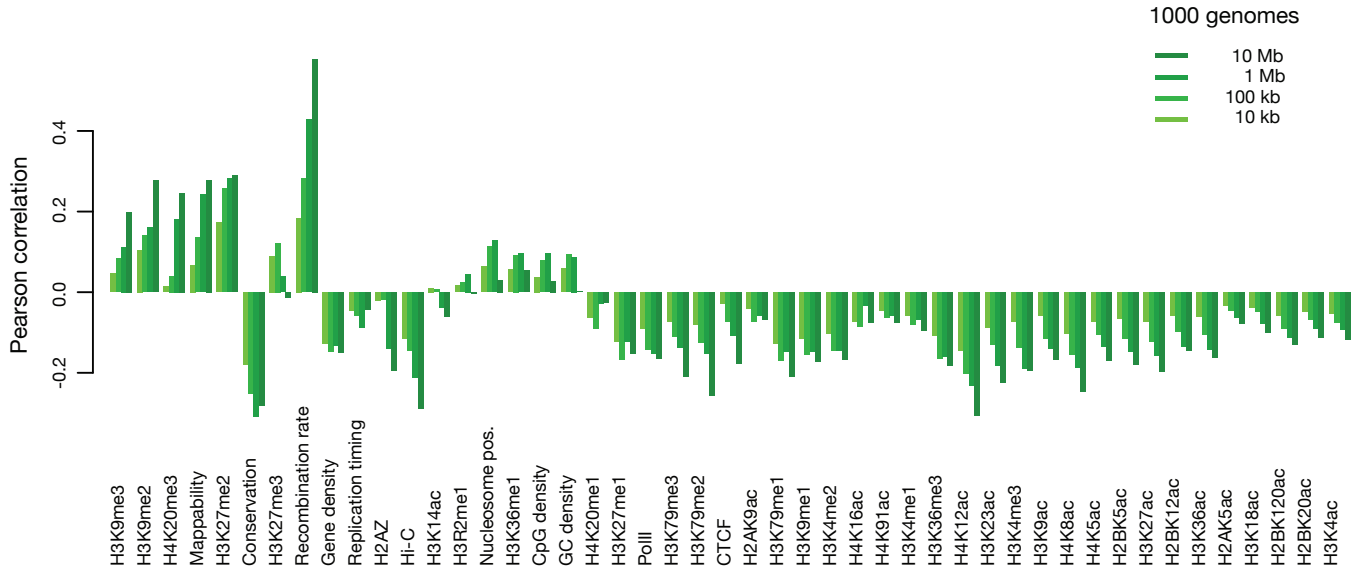
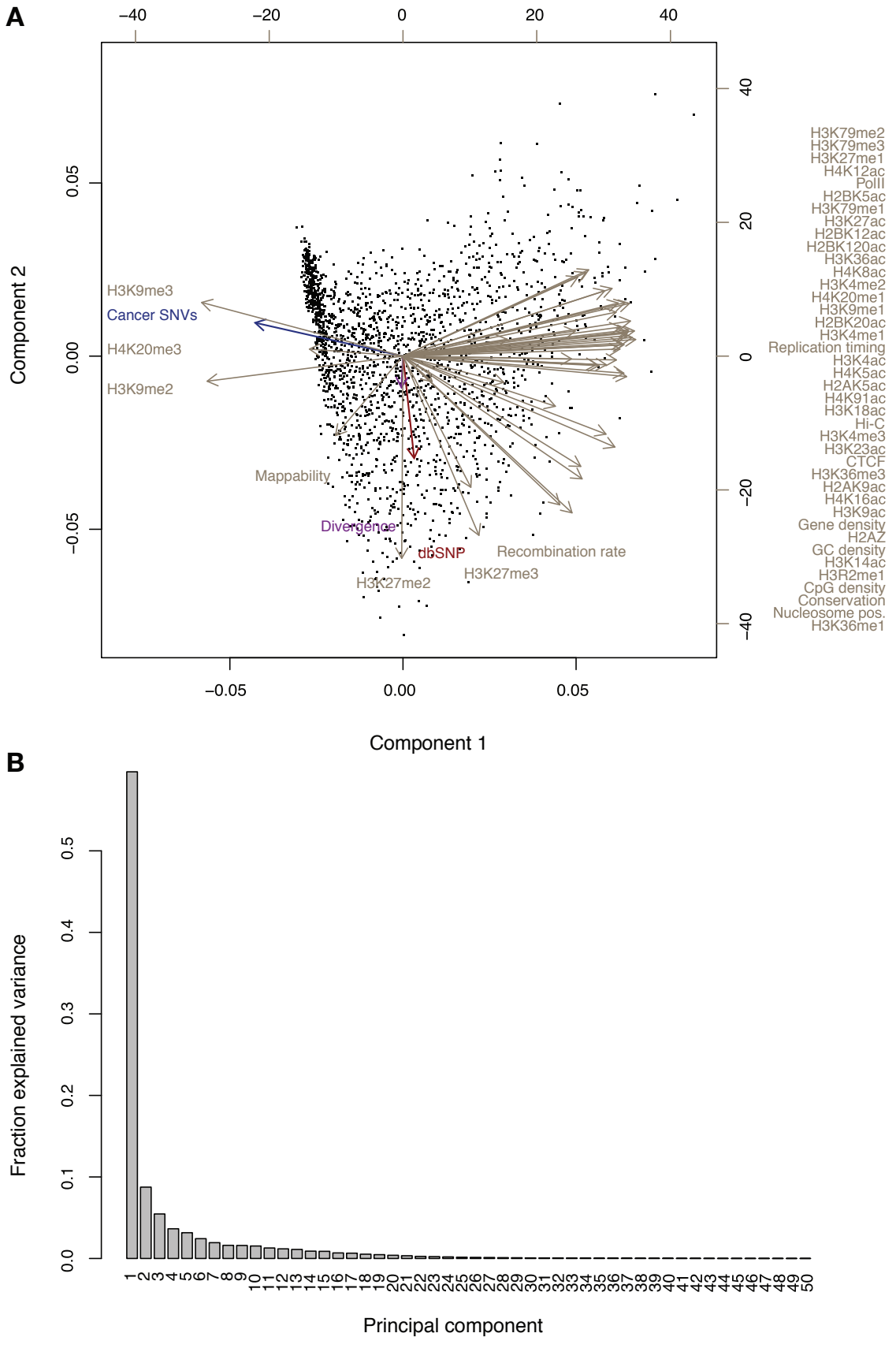
doi:10.1038/nature11273



Supplementary figure 1 Correlations across the human genome between SNV density in cancer genomes, SNP density from dbSNP, and human-chimp divergence quantified in different non-overlapping window sizes.

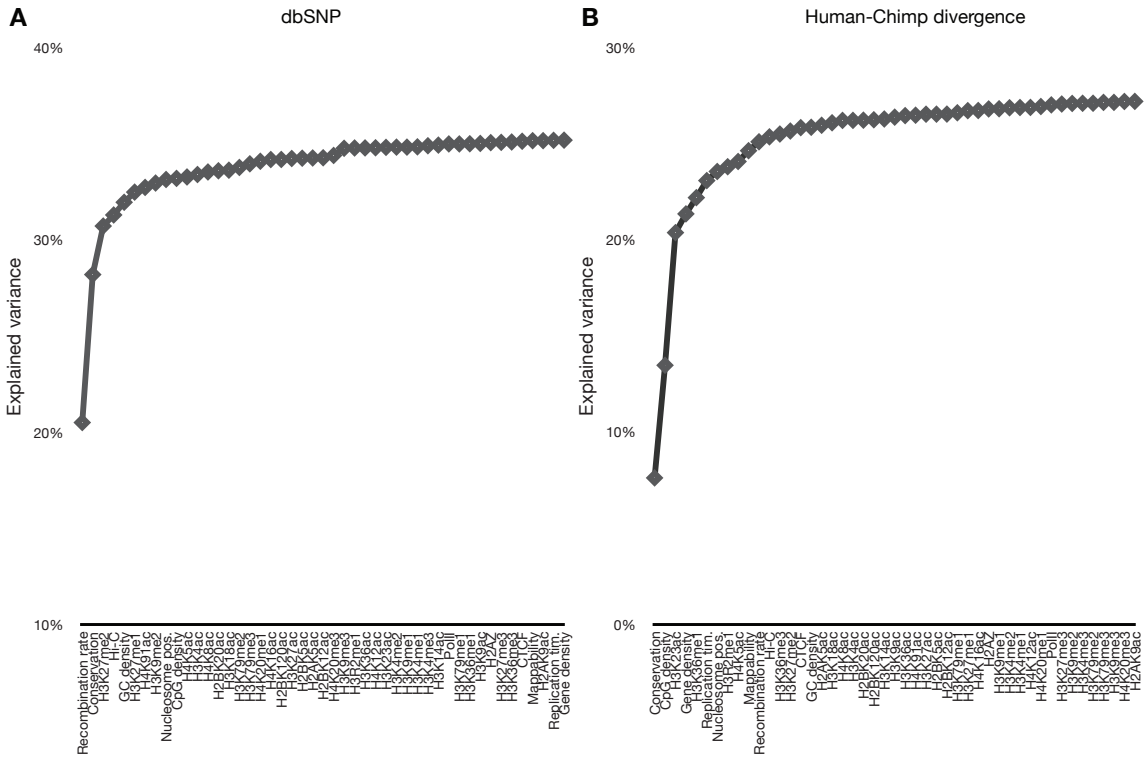


Supplementary figure 2 Correlation coefficients of germline SNP density from the 1000 genomes projects with all genomic features at different window sizes.

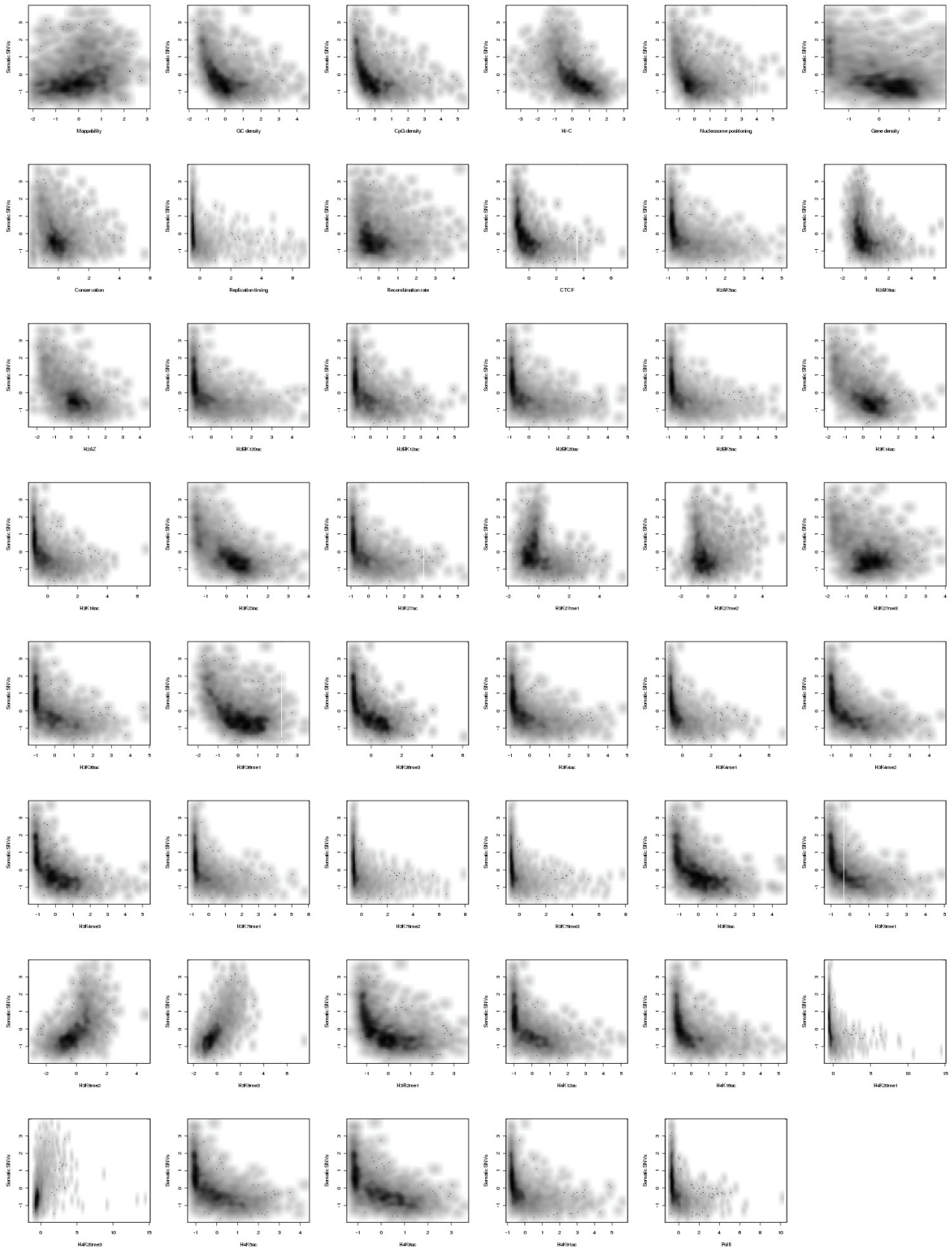


Supplementary figure 3 Principal component analysis of genomic features. **a** Bi-plot

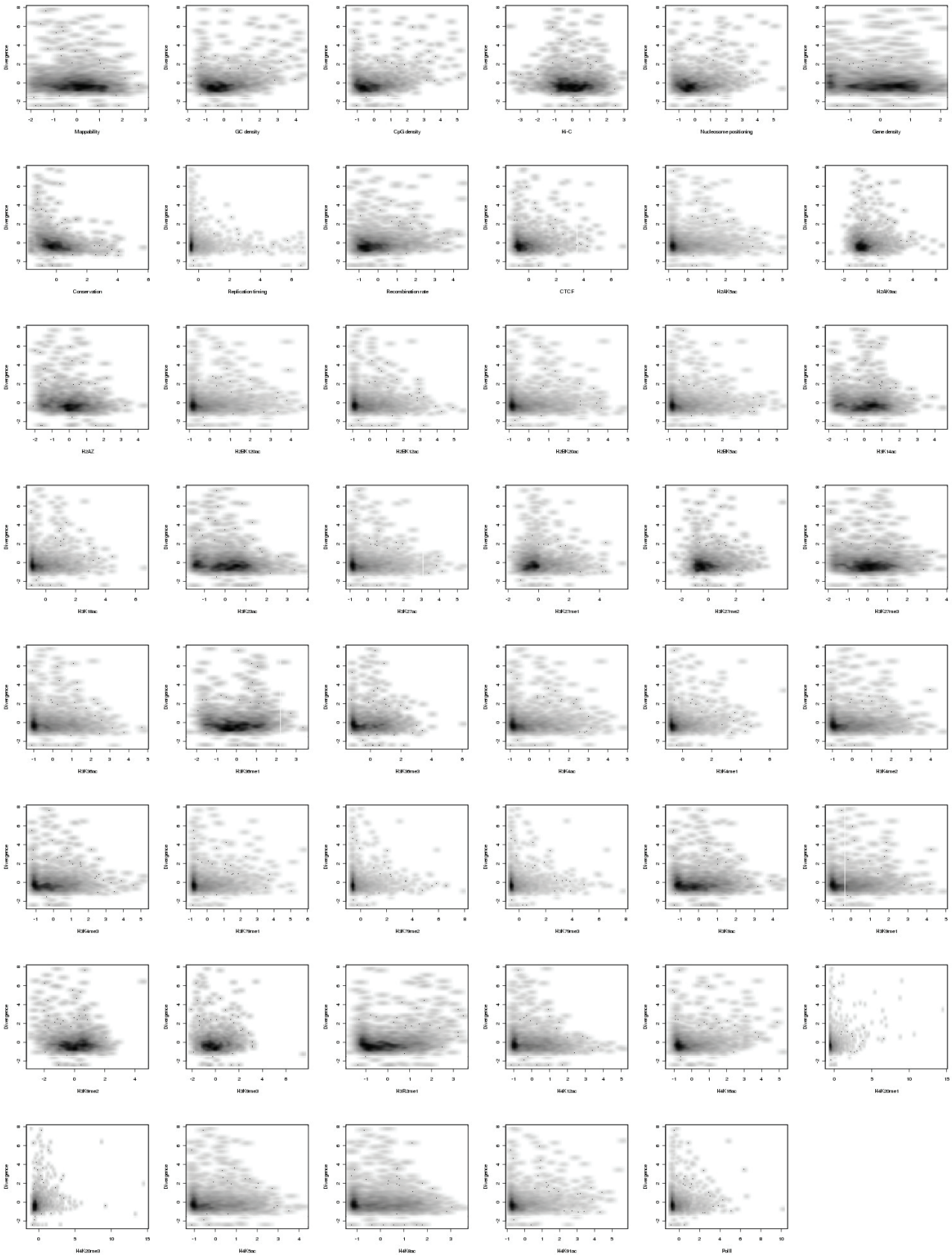
of the first two principal components. Black circles denote transformed values of individual non-overlapping, non-repetitive 1Mb windows, arrows indicate loadings of the genomic features for the respective principal component. The vectors for cancer SNVs, germline SNPs, and human-chimp divergence are highlighted in blue, red, and purple, respectively. **b** Percentage of total variance explained by each principal component.



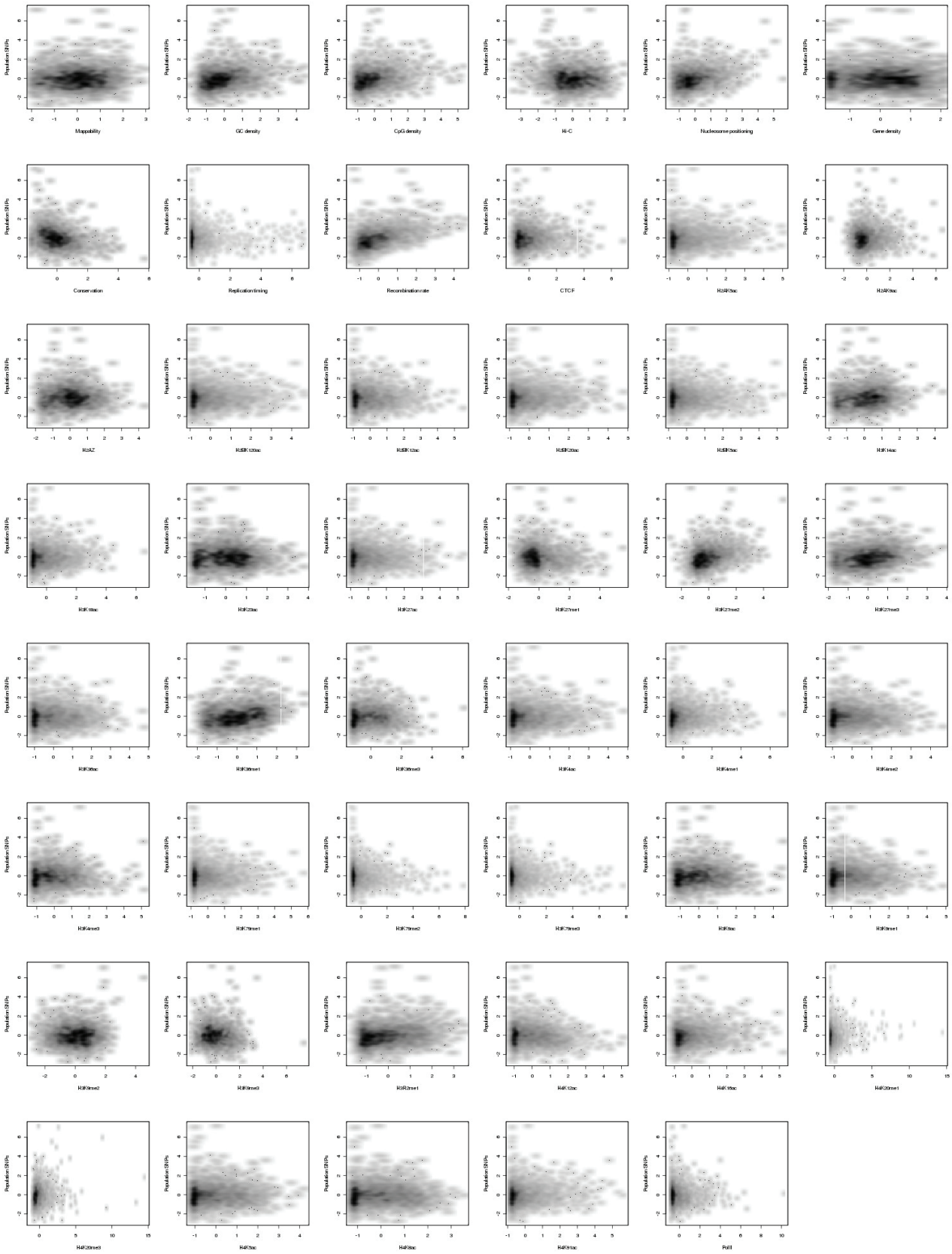
Supplementary figure 4 Prediction of **a** dbSNP SNP density and **b** human-chimp divergence using integrated models. Cumulative R^2 of linear models, adding the feature on the x-axis as a predictor at each step.



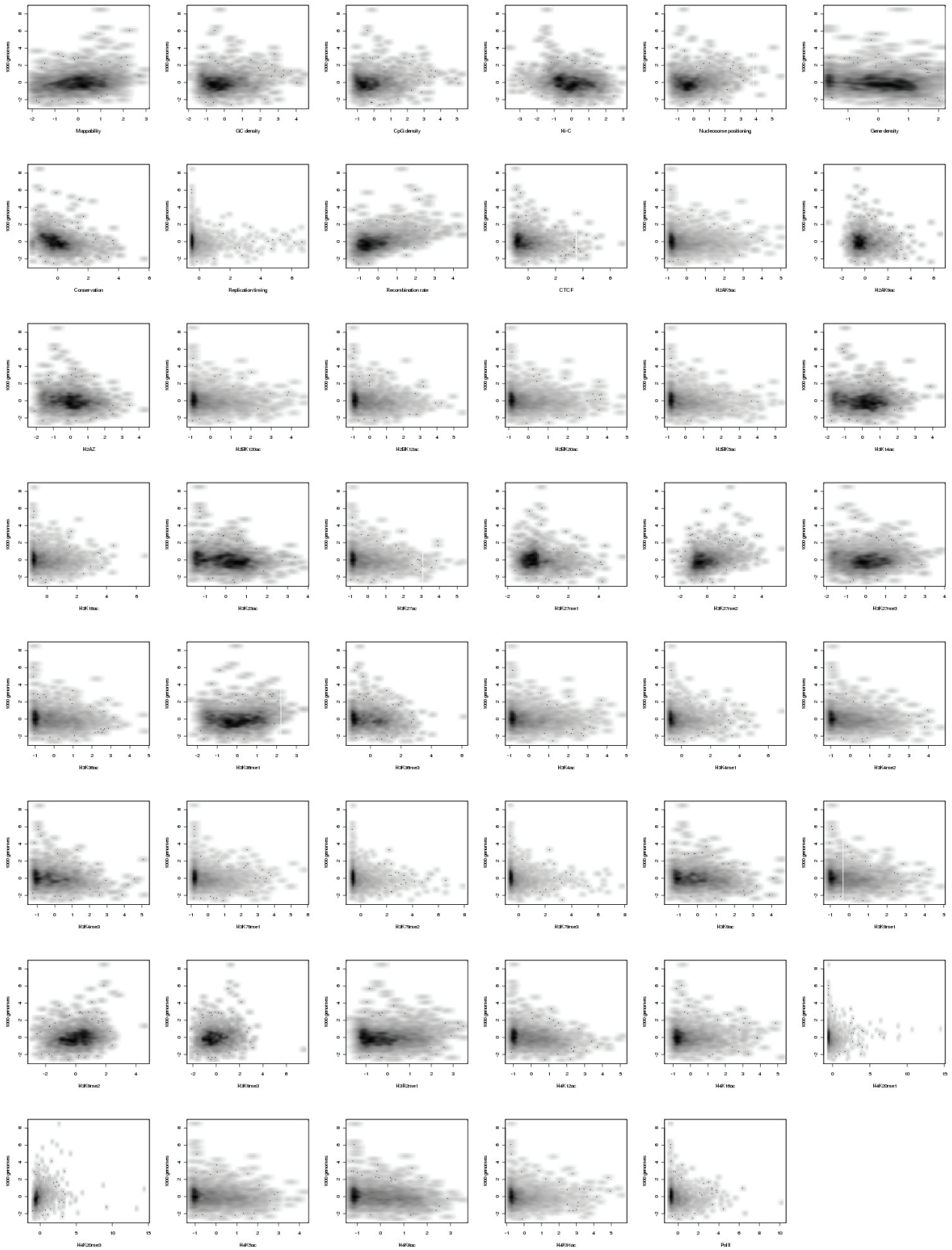
Supplementary figure 5 Correlations between all genomic features and SNV density in cancer genomes in non-overlapping 1Mb windows across the human genome (normalized to standard scores).



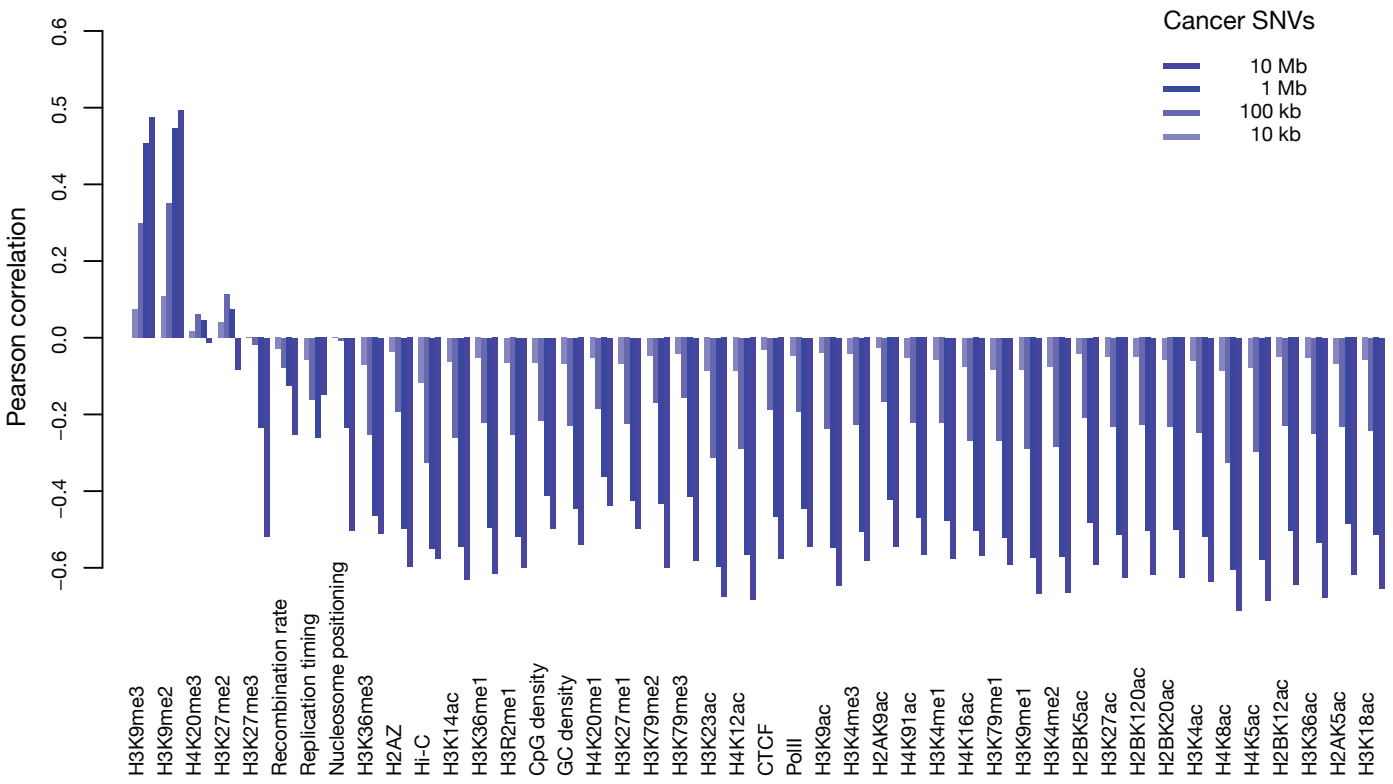
Supplementary figure 6 Correlations between all genomic features and human-chimp divergence in non-overlapping 1Mb windows across the human genome (normalized to standard scores).



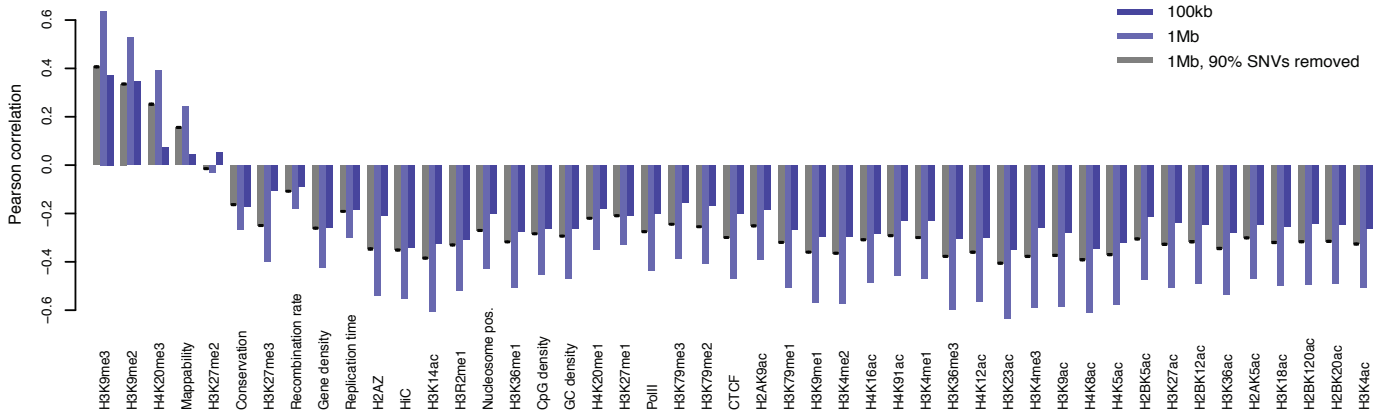
Supplementary figure 7 Correlations between all genomic features and germline SNP density from dbSNP in non-overlapping 1Mb windows across the human genome (normalized to standard scores).



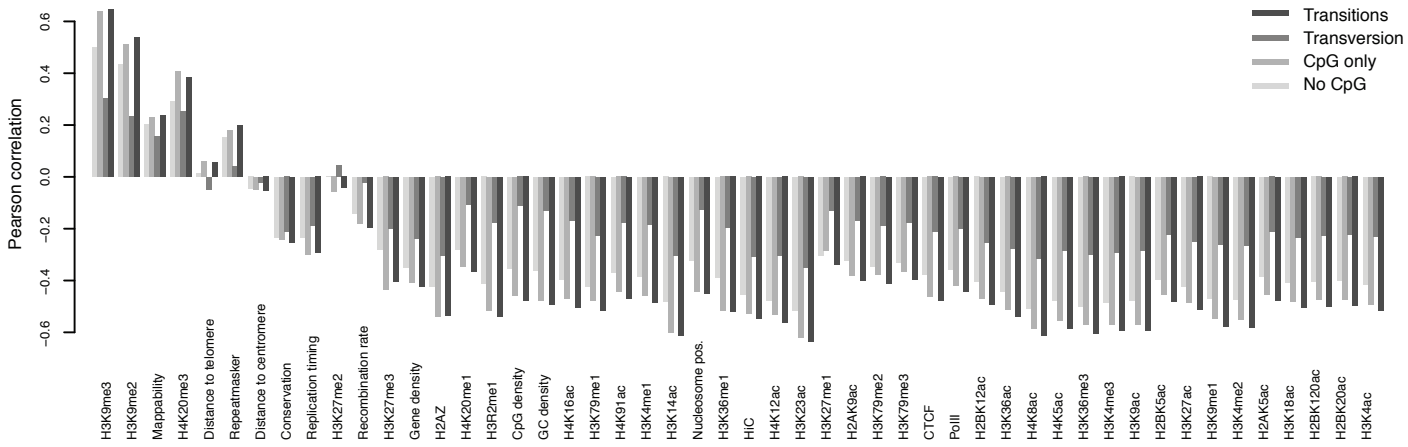
Supplementary figure 8 Correlations between all genomic features and germline SNP density from the 1000 genomes project in non-overlapping 1Mb windows across the human genome (normalized to standard scores).



Supplementary figure 9 Correlations between all genomic features and cancer SNV density, masking out all variants without 20bp of unique flanking sequence in either direction.



Supplementary figure 10 Correlations between all genomic features and cancer SNV density at 1Mb resolution, but randomly removing 90% of all SNVs on each chromosome. Error bars indicate 95% confidence intervals from 100 randomisations. When sub-sampling the SNVs, the correlations reduce to values similar to those detected with the full dataset at 100kb resolution, suggesting that the weaker correlations in smaller window sizes could be due to the reduced number of mutations per window.



Supplementary figure 11 Correlations between all genomic features and cancer SNV density at 1Mb resolution, considering only transitions, transversions, CpG mutations, or non-CpG mutations. Correlations with distance to telomere, distance to centromere, and repeat content as determined using Repeatmasker are also shown.