Supplemental Material to:

Vladimir A Naumov, Edward V Generozov, Natalya B Zaharjevskaya, Darya S Matushkina, Andrey K Larin, Stanislav V Chernyshov, Mikhail V Alekseev, Yuri A Shelygin, and Vadim M Govorun

Genome-scale analysis of DNA methylation in colorectal cancer using In nium HumanMethylation450 BeadChips
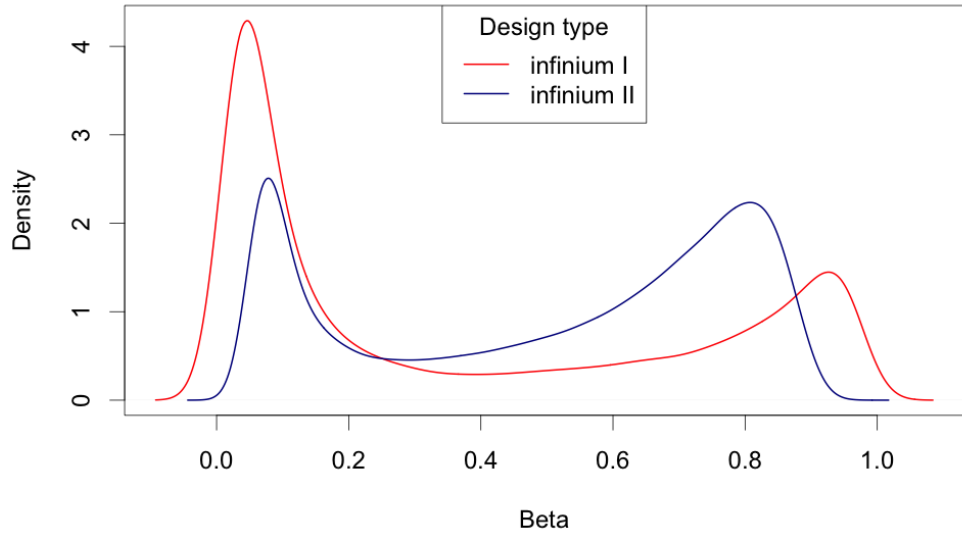
Figure S1.

A

B
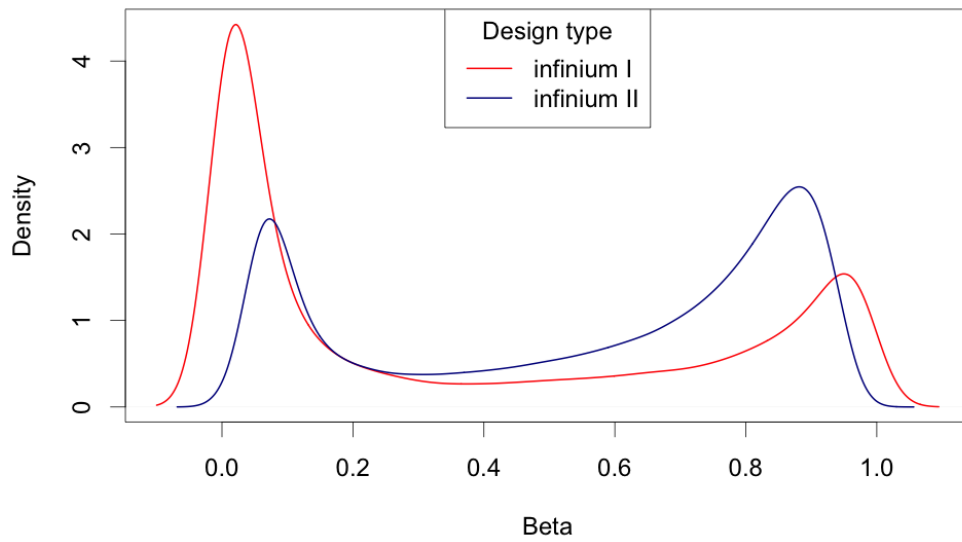
C
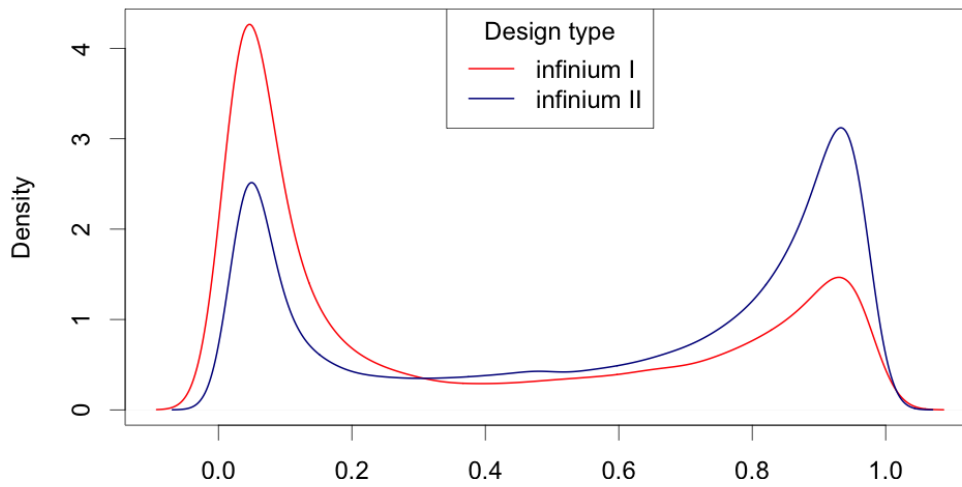
Figure S1. Density plots of the β-values distribution for the InfI and InfII probes.

A. No data correction. B. Illumina GenomeStudio data correction. C. Illumina GenomeStudio data correction and IMA-R peak correction by S. Dedeurwaerder et al. [10]
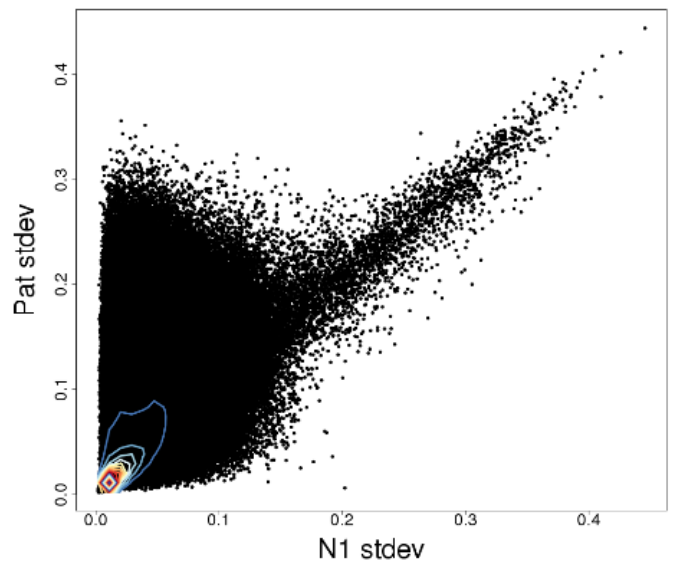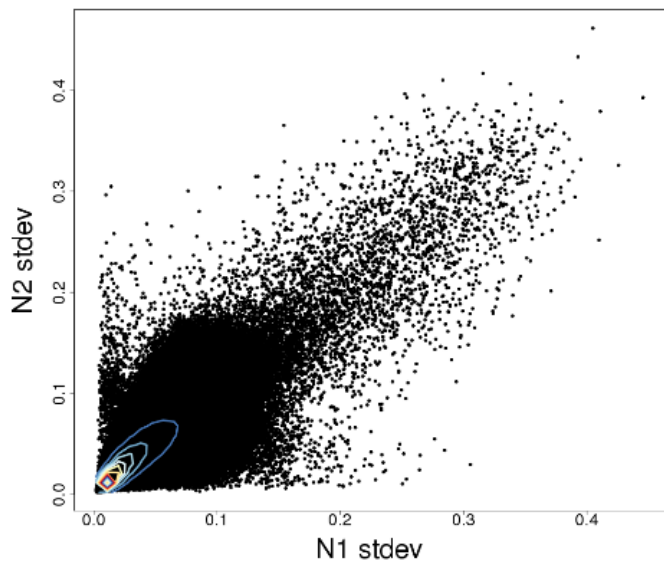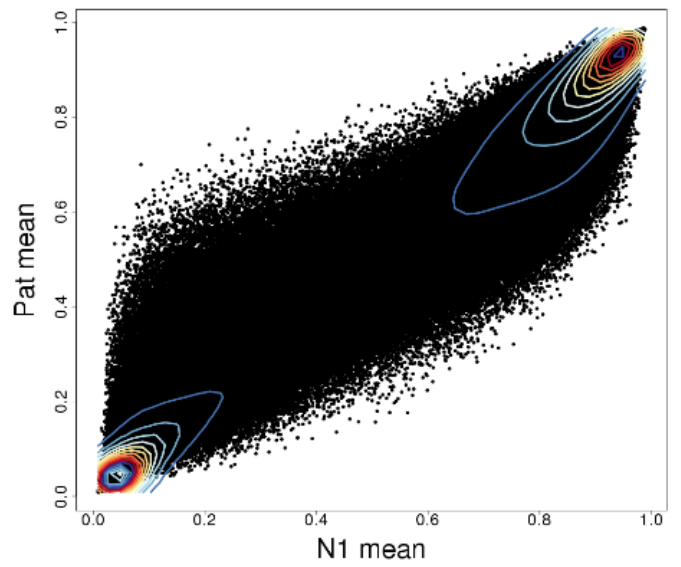
Figure S2.
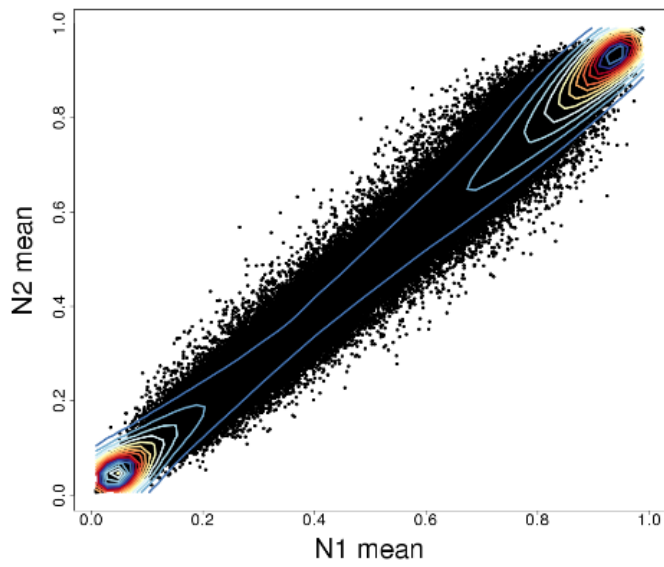
Figure S2. Scatter plots of mean β-values (upper plots) and standard deviations of β-values (lower plots) between compared groups (C – tumor, N1 – healthy tissue from CRC patients, N2 – healthy tissue from cancer-free donors).

Figure S3.
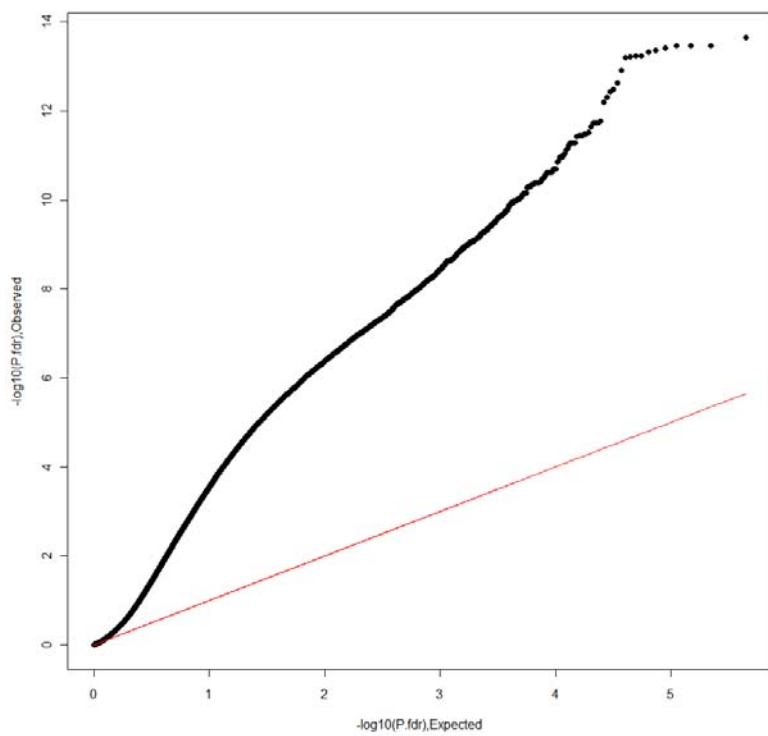
Figure S3. Q-Q quantile plot of –log10 (p-values) for the 444,888 CpGs sites analyzed.

Figure S4.

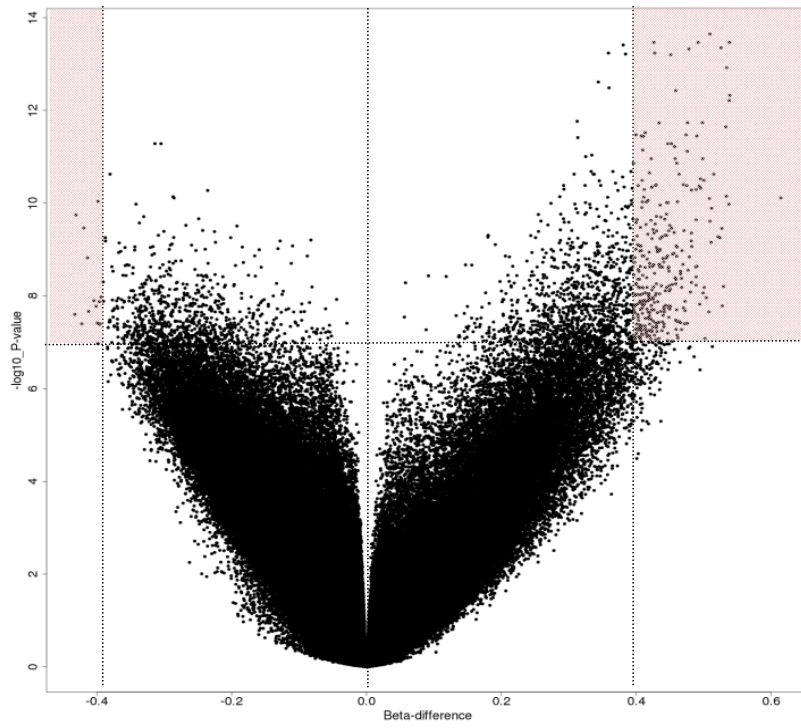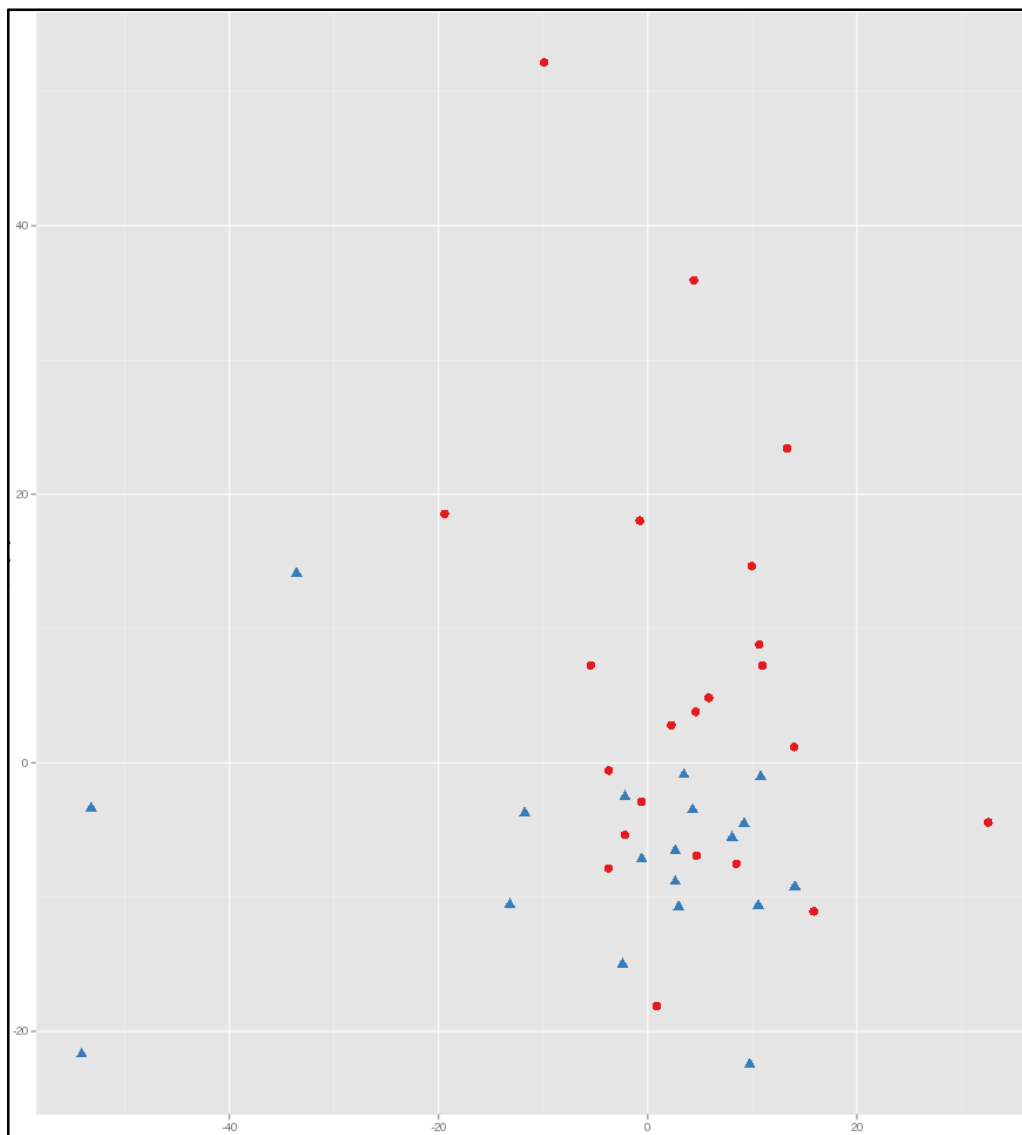**Figure S4.** Volcano plot distribution of significant CpG sites identified in the site-level test.

The delta β-value is shown on the x-axis, and the p-value is shown on the y-axis. Red markers indicate high significance in the combined groups that showed large methylation differences between normal and CRC tissues.

Figure S5.

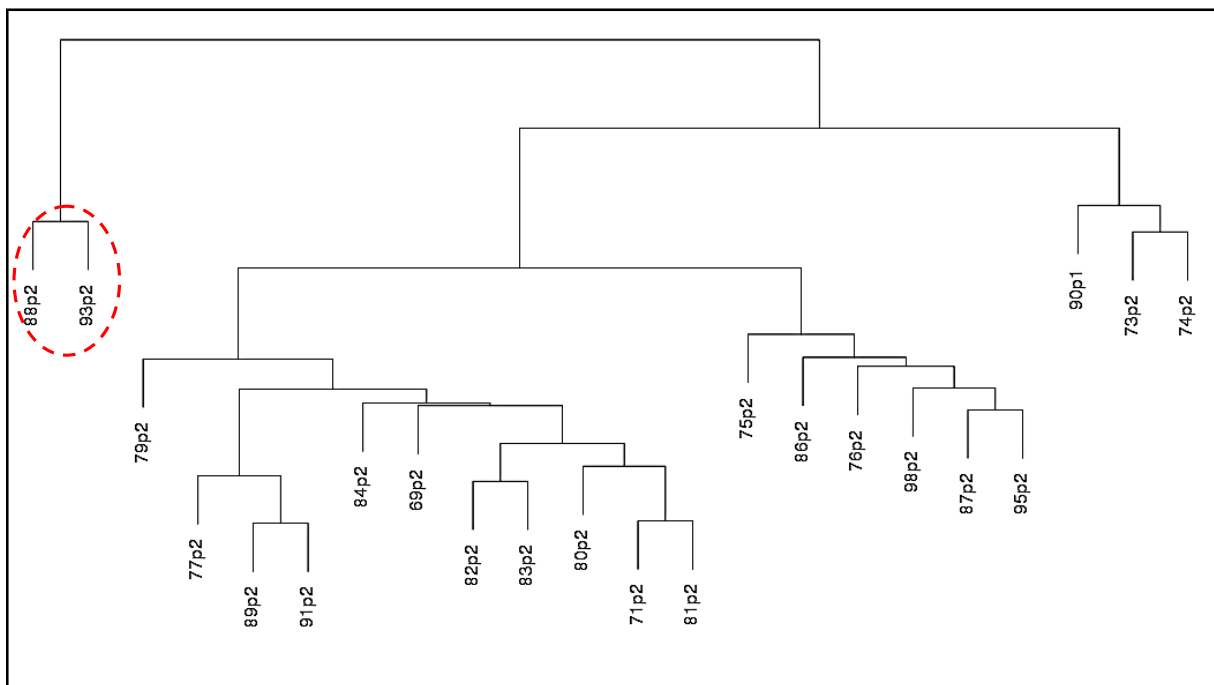**Figure S5.** PCA scores plot for the methylation status of samples from different variants of normal colon tissue: CRC patients (red circles) and healthy donors (blue triangles).

Figure S6.
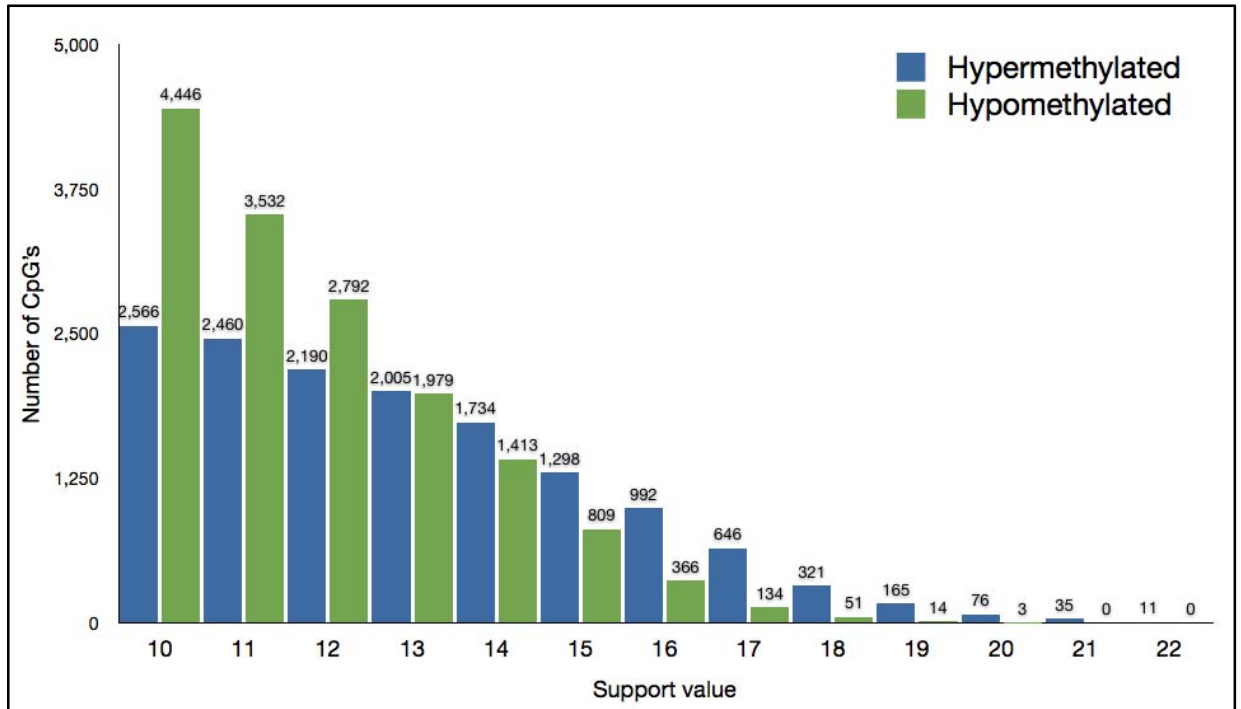
**Figure S6.** Distribution of differentially (hyper- and hypo-) methylated CpG sites common in sample pairs. The number of sample pairs required for support (10 to 22) is indicated on the x-axis.

Figure S7.

Figure S7. Cluster dendrogram of tumor samples based on Euclidean distances calculated using β-values of all CpG sites. CIMP samples forming a separate cluster are rounded.
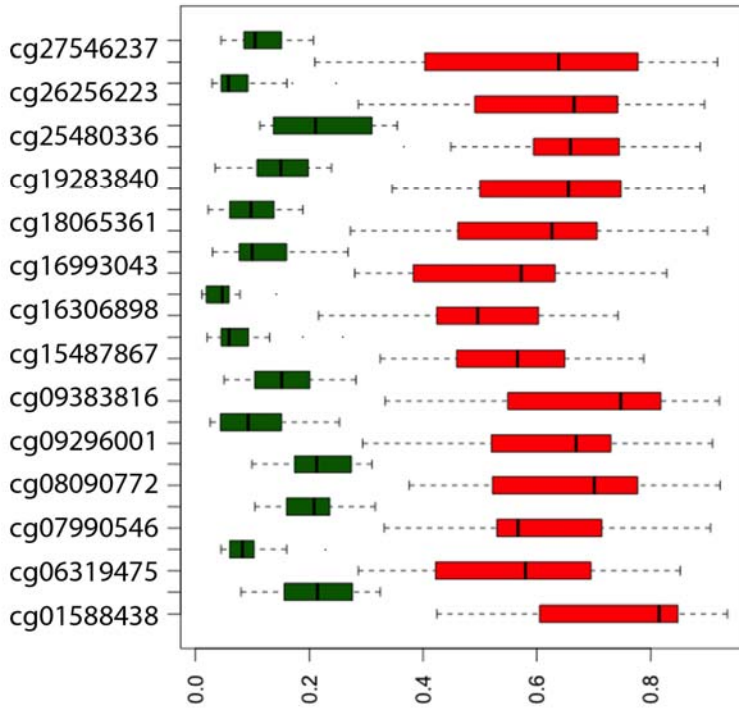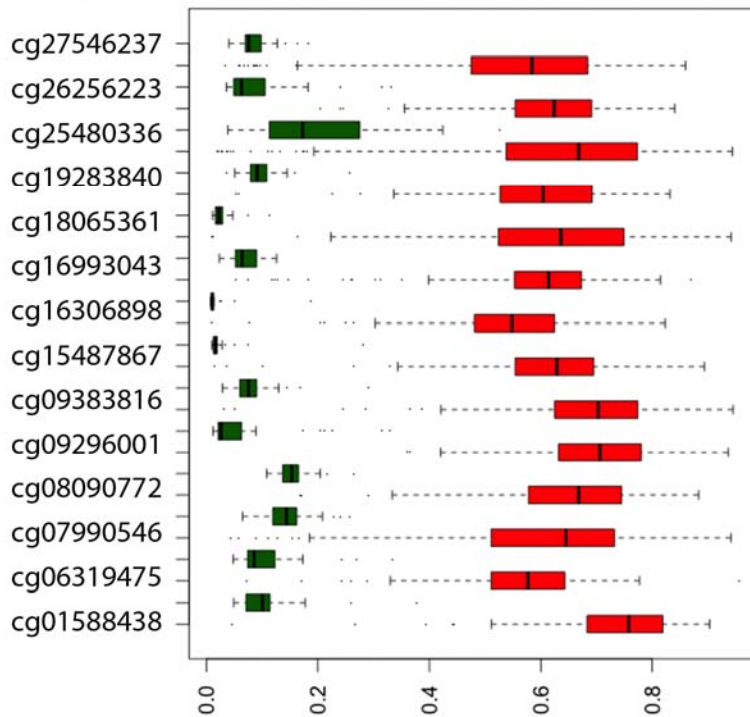
Figure S8.
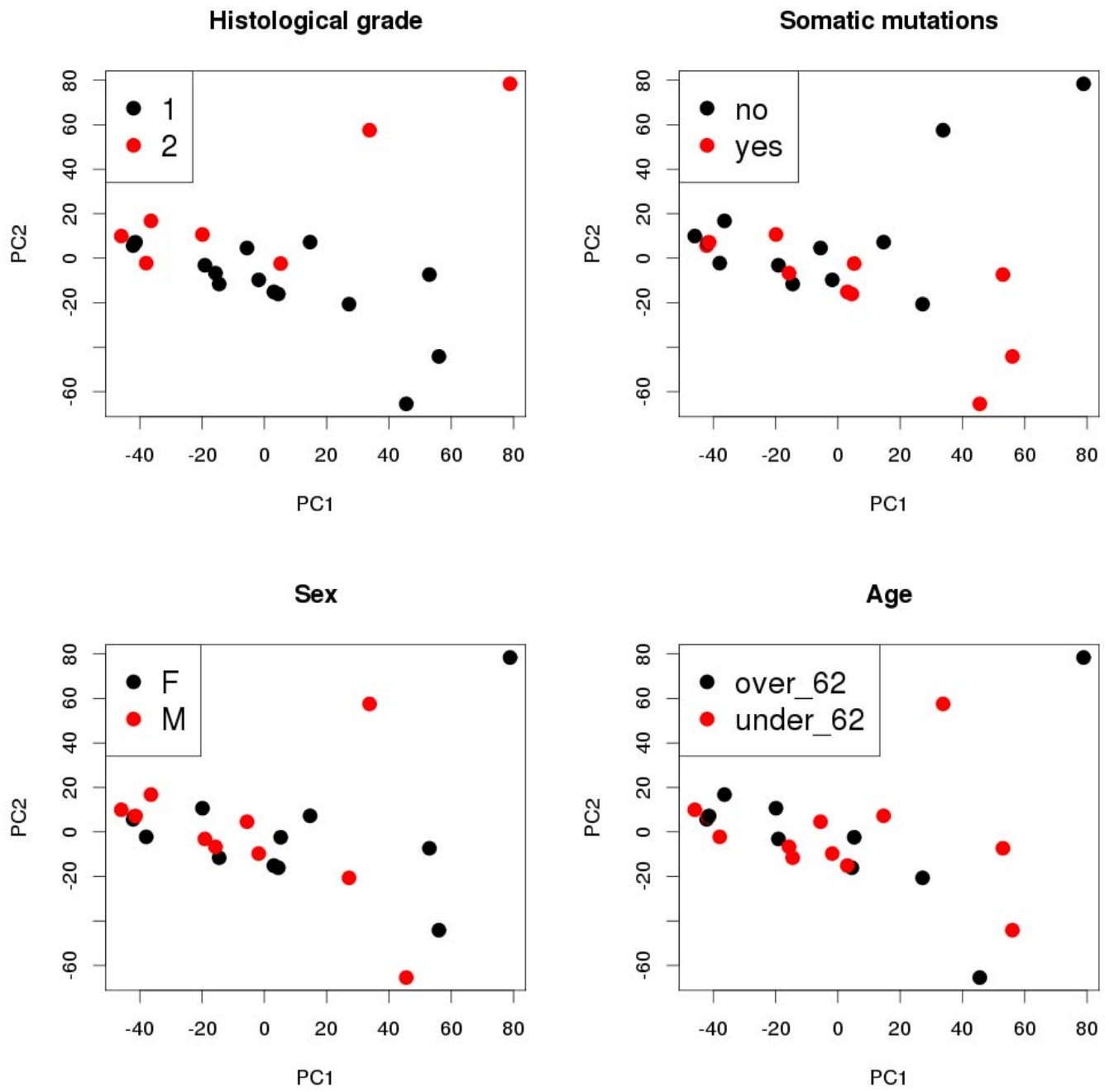
Figure S8. Plot of PCA scores for the methylation status of CRC samples based on different molecular and clinical features: A - histological grade, B - somatic mutations, C - sex, D - age.
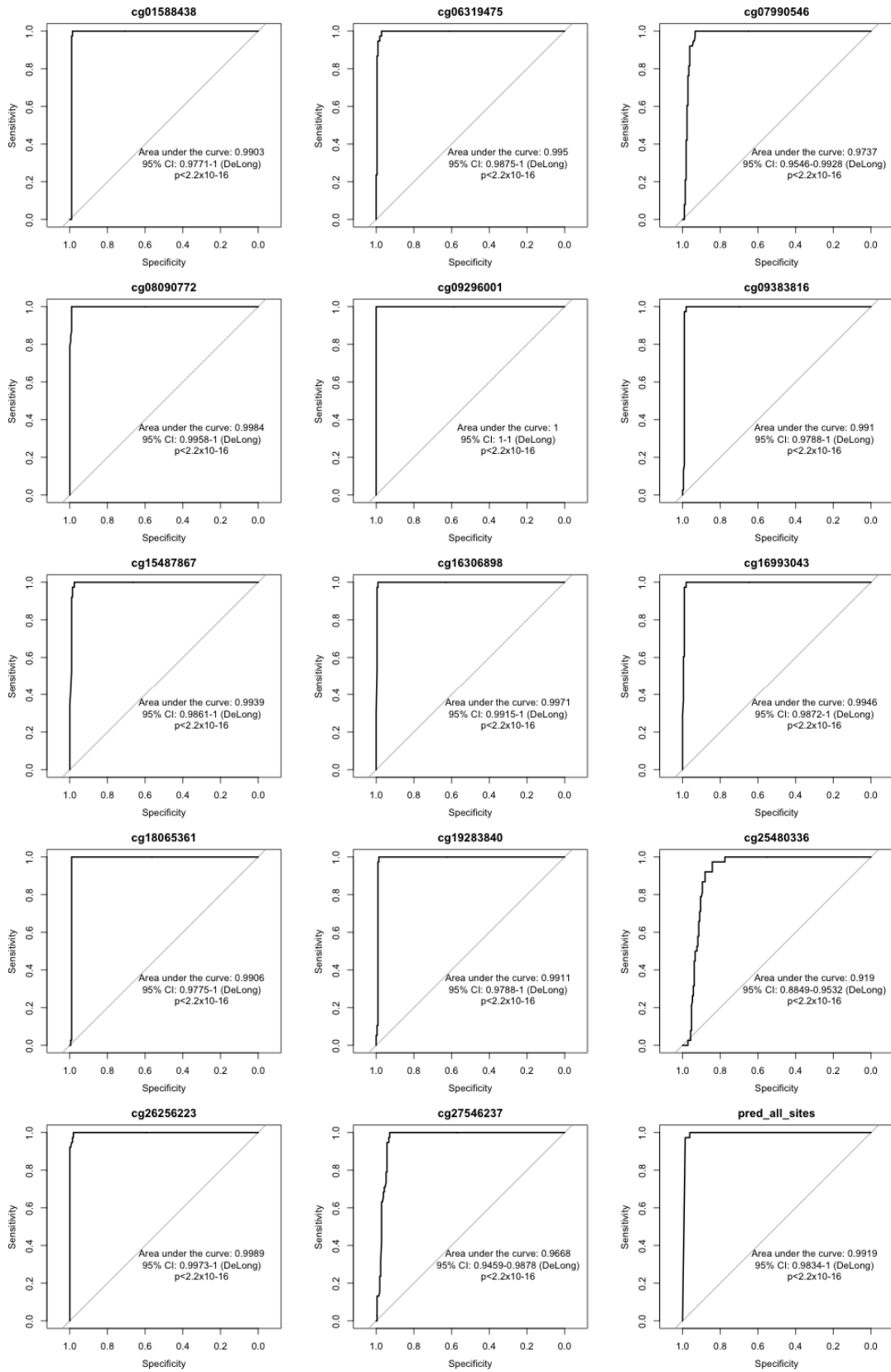
Figure S9.

Figure S9. Distribution of methylation β-values for CpG sites selected as diagnostic markers. A – experimental data (22 CRC vs. 22 normal colon tissue samples); B – TCGA dataset (209 CRC vs. 38 normal colon tissue samples). Red box plots – CRC samples, green – normal tissue. The β-value is plotted on the x-axis. The median (bar), 25–75 percentile range (box), and 10–90 percentile range (whisker) values of all CpG sites are displayed for each tissue category.

Figure S10.

Figure S10. ROC curve-based evaluation of the diagnostic accuracy of methylation markers. Mean AUROC and 95% confidence interval (CI) are shown, as are the p-values corresponding to the comparison to the non-discriminative curve.
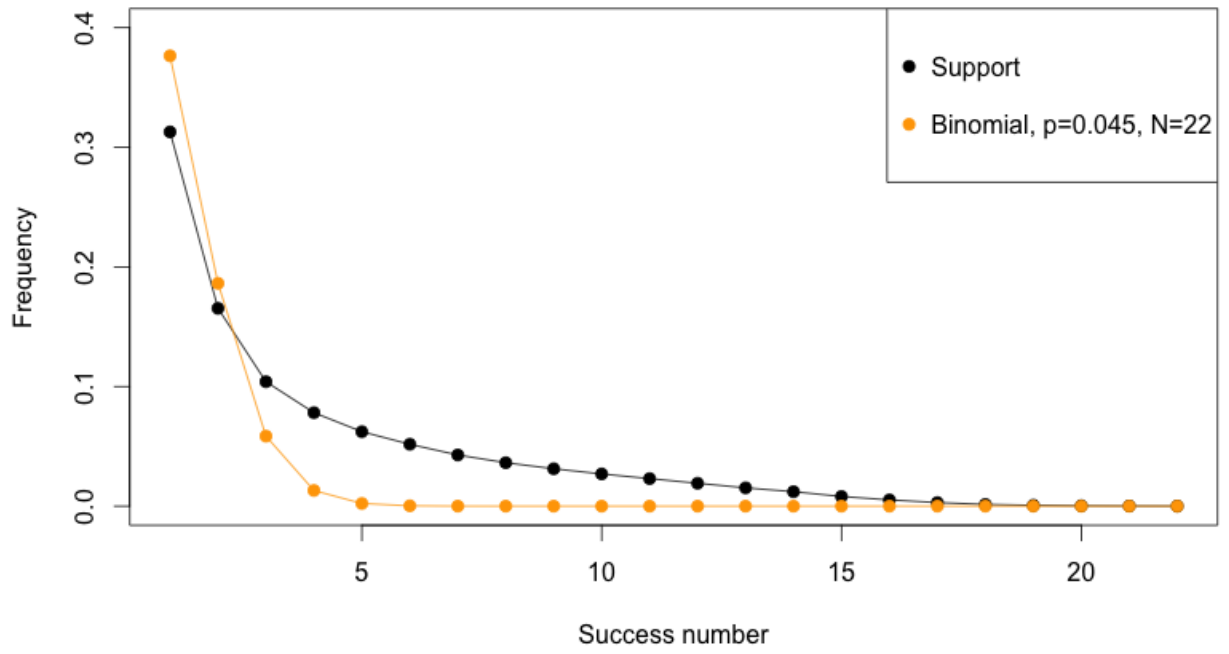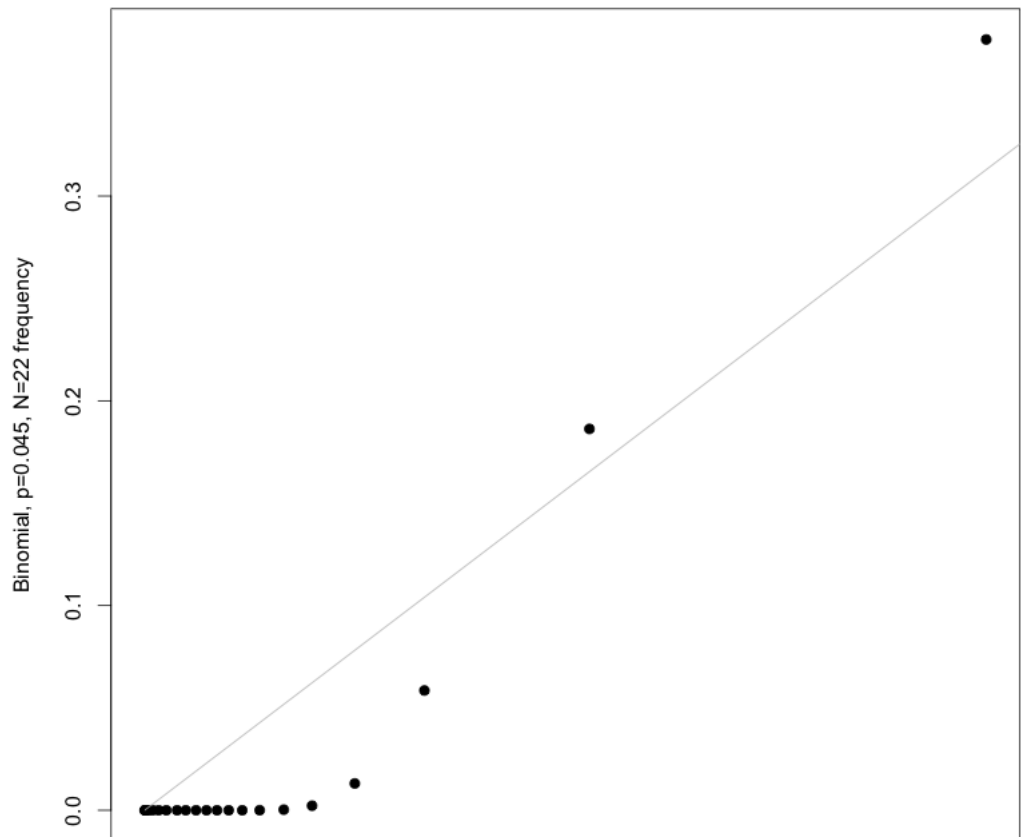
Figure S11.
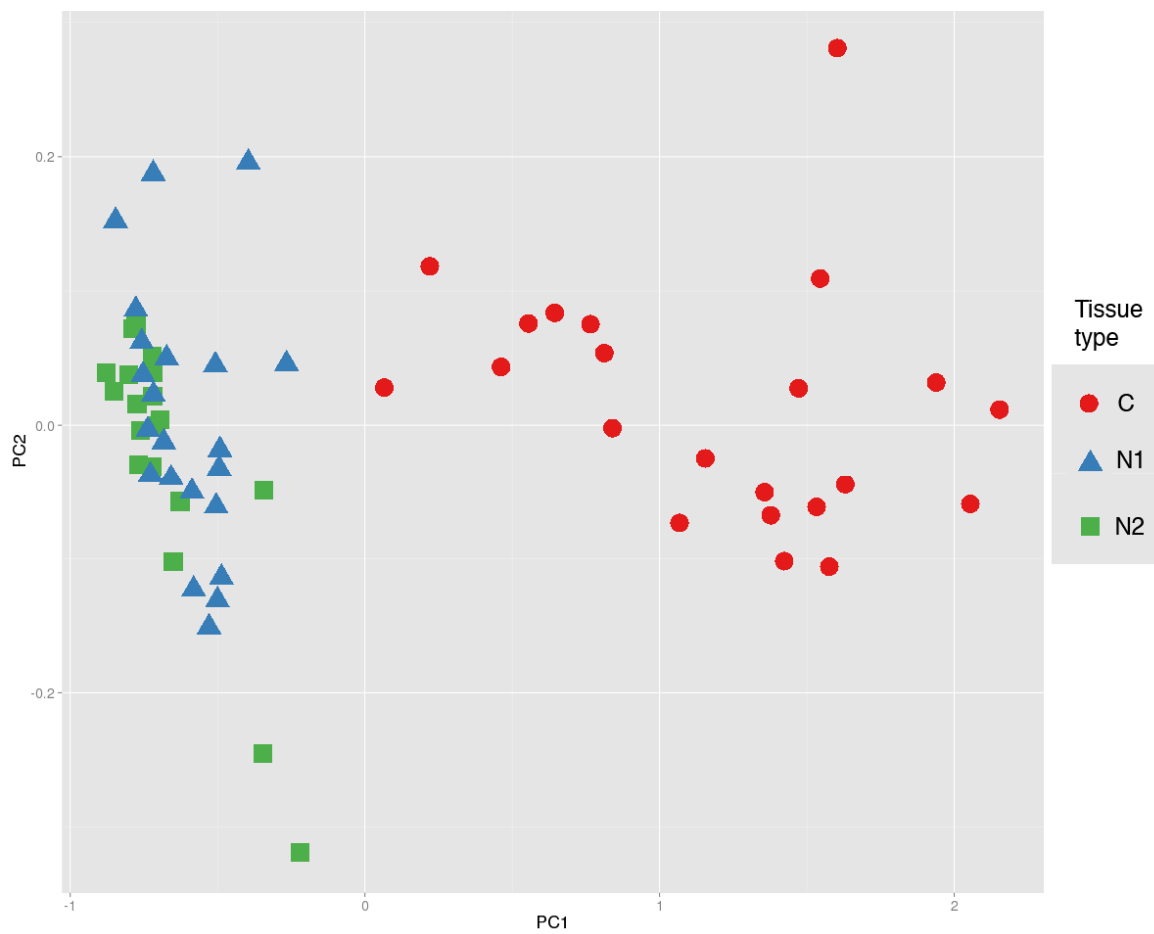
Figure S11. PCA performed on 14 selected CpG sites distinguished tumor samples from healthy colon samples. Red dots indicate tumors (C); blue triangles indicate normal colon samples from patients with CRC (N1); green squares indicate normal colon samples from healthy donors (N2).

Figure S12.

A

B

Figure S12. Approximation of support distribution with binomial distribution. A. Comparison of support experimental distribution (black) with binomial distribution (p=0.045, N=22). Success number means number of successes in Bernouli traits for binomial distribution or number of pairs in that CpG sites differs. Frequency for binomial distribution was calculated using parameters p=0.045, N=22, and frequency of support is from experiment. B. QQ-plot of support distribution vs binomial distribution (p=0.045, N=22).