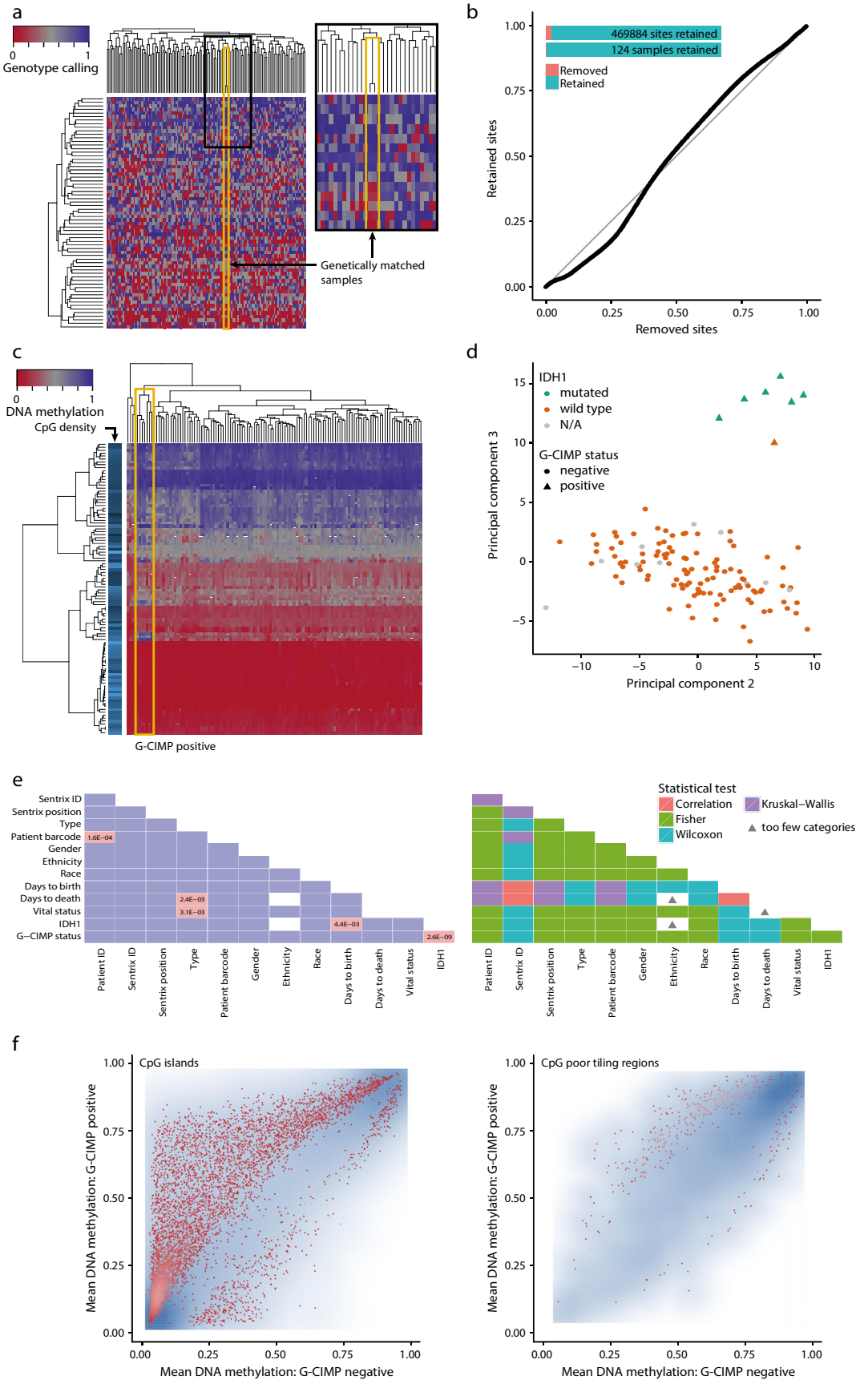

a
Genotype calling
Genetically matched samples
b
469884 sites retained
124 samples retained
Removed
Retained
Retained sites
Removed sites
c
DNA methylation
CpG density
G-CIMP positive
d
IDH1
mutated
wild type
N/A
G-CIMP status
negative
positive
Principal component 3
Principal component 2
e
Sentrix ID
Sentrix position
Type
Patient barcode
Gender
Ethnicity
Race
Days to birth
Days to death
Vital status
IDH1
G–CIMP status
Statistical test
Correlation
Fisher
Wilcoxon
Kruskal–Wallis
too few categories
f
CpG islands
CpG poor tiling regions
Mean DNA methylation: G-CIMP positive
Mean DNA methylation: G-CIMP negative

Supplementary Figure 1

Analysis of DNA methylation in a cancer cohort based on Infinium 450K data.

RnBeads was used to rediscover a clinically distinct subgroup of glioblastoma patients characterized by increased DNA methylation levels (termed G-CIMP+), and to predict the G-CIMP status for a total of 124 patients using Infinium 450k data obtained from the TCGA project (http://cancergenome.nih.gov).

(a) Detection of genetic duplicates among the patient samples (columns) using a clustered heatmap of intensity values for the genotyping probes that are present on the Infinium microarray (rows). The inset shows that two samples exhibit a high level of genetic identity, and they are indeed derived from tumors of the same patient.

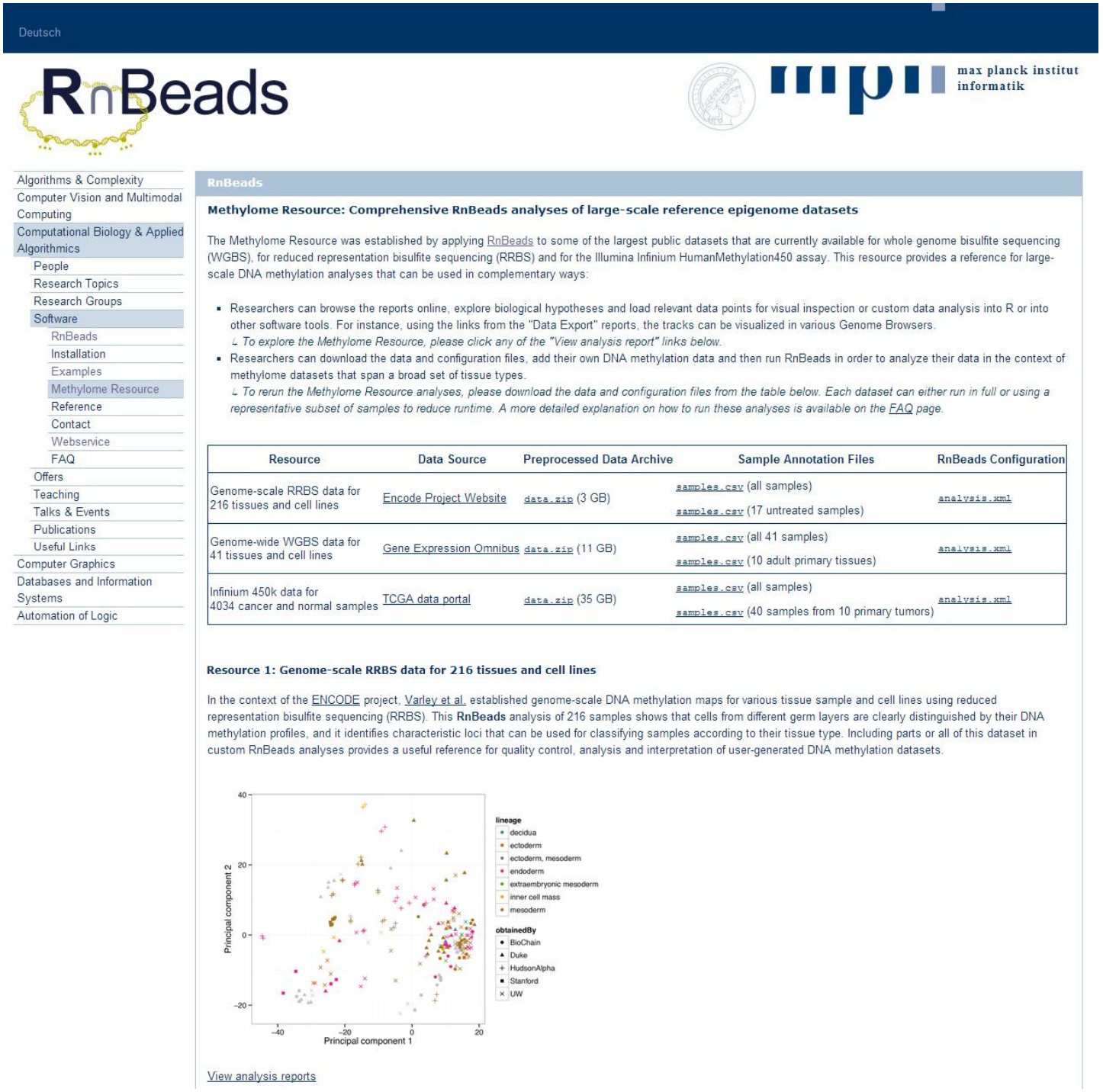
(b) Quality control plot summarizing the outcome of the data filtering. The bar plots on the top left show that the majority of CpG sites (top) and samples (bottom) are of good quality and can be retained. The relatively straight line in the quantile-quantile plot indicates that the probe filtering does not have a major impact on the distribution of DNA methylation in the dataset.

(c) Identification of a small but clearly distinguished cluster of G-CIMP+ glioblastoma samples with elevated DNA methylation levels especially in CpG-rich genomic regions (light blue in the leftmost column). In the heatmap, blue colors denote high levels of DNA methylation, red indicates low levels and grey represents intermediate levels. For visualization purposes, only the 100 gene promoters (rows) with the highest levels of inter-sample variation in DNA methylation are shown (columns), but the hierarchical clustering is based on the full set of promoters.

(d) Global assessment of the similarity between the DNA methylation profiles, plotting all glioblastoma samples according to their second and third principal components. The samples exhibit strong separation according to the G-CIMP status (denoted by point shape) and IDH1 mutation status (denoted by point color).

(e) Analysis of significant associations between all user-provided sample annotations. Significant p-values (<0.05) are highlighted in the left triangle, and the corresponding statistical tests are annotated in the right triangle (orange: Pearson correlation followed by permutation-based estimation of the p-value; green: Fisher's exact test; blue: Wilcoxon rank sum test; violet: Kruskal-Wallis one-way analysis of variance).

(f) Genome-scale comparison between the DNA methylation levels of G-CIMP positive (y-axis) and G-CIMP negative (x-axis) tumor samples, focusing on CpG islands (left scatterplot) and on 5-kilobase tiling regions with a CpG content in the bottom quartile (right scatterplot), respectively. Genomic regions that are differentially methylated with an FDR below 0.05 are presented as red points. All other regions are displayed in blue, and color brightness denotes point density.



Supplementary Figure 2

RnBeads-based Methylome Resource of reference epigenome data sets.

Screenshot of the Methylome Resource (http://rnbeads.mpi-inf.mpg.de/methylomes.php), which makes large DNA methylation datasets more readily available for follow-up research. On the one hand, it provides detailed analysis reports for publicly available methylome datasets that can be explored interactively. On the other hand, the Methylome Resource website lets RnBeads users download all data and configurations that are needed to re-run all or part of the DNA methylation analyses in their local or cloud-based computing environment. These re-runnable analysis configurations make it straightforward for RnBeads users to analyze their own DNA methylation data in the context of publicly available reference epigenome maps.