## Supplementary Fig. S1. Toca 511 detection in RNA isolated from patient plasma samples from Study 11.

Heatmap representation of Toca 511 RNA signal in plasma as a function of time for the 22 patients of 56 in Study 11 with detectable signal.

## Supplementary Fig. S2. Cartoon illustration of integration site identification.

1. Genomic DNA (green) from four cells that contain integrated Toca 511 (red).
2. Isolated genomic DNA is randomly fragmented to mode size of ~300 bp.
3. A partially double stranded adaptor is ligated to DNA ends, PCR is performed with a Toca 511-specific primer and adaptor primer.
4. Nested PCR is performed with a Toca 511-specific primer and adaptor primer. Primers contain Illumina adaptors and barcodes for pooled sequencing.
5. Data processing following sequencing: e.g., raw sequencing reads are filtered to remove sequences that do not contain the Toca 511 integration sequence, then mapped to the human genome, and finally reads with the same integration site and ligation/break site are collapsed. Note that integration events 1 and 4 are not combined because they have different ligation sites and therefore originated from different cells.
6. Final results.

## Supplementary Fig. S3. Validation of Toca 511 integration profiling workflow.

1. Barplot representation of qPCR results showing fold-enrichment of serially diluted genomic DNA from Toca 511 infected U-87MG cells into noninfected DNA isolated from human brain. qPCR was performed with primers used for the second round of PCR for integration site sequencing library preparation.
2. Barplot representation of the total number of integration events recovered from sequencing libraries generated from the serially diluted control samples.
3. Cumulative fraction plots showing the fraction of all mapped sequencing reads accounted for by the 100 most represented sites (due to PCR duplication).
4. Histogram showing characteristic peak of integration sites near TSSs. For this analysis, sites from the four positive control samples were pooled.

## Supplementary Fig. S4. Barcoded ligation strategy did not perform as expected.

1. An example of a single integration site and fragmentation site represented by tens of thousands of sequence reads due to PCR duplication. The x-axis shows the distance of fragmentation site reads from integration site reads.
2. Distribution of the 100 top ranking 7mer barcodes from (A). Based on (A), we expect all reads originate from a single ligation event and thus should all share the same 7mer barcode.
3. Similar to (A), except for a different sample
4. Similar to (B) for sample in (C).
5. Similar to (A), except for a different sample.
6. Similar to (B) for sample in (E).

## Supplementary Fig. S5. Summary of Toca 511 integration profile results from patient samples.

1. Barplot showing the number of integration events identified in each sample.
2. Scatterplot between the total number of reads that mapped to the human genome (x-axis) versus the total number of integration events for each sample (y-axis).

## Supplementary Fig. S6. Integration site upstream of CAMLG.

1. Follow-up analyses on patient 11\_33 blood samples show there is no overlap in integration sites identified from technical replicates and different time points, suggesting that the apparent multiple sites obtained from sample 4A are artefactual.
2. Genome browser view of read pairs that mapped to a region immediately upstream of CAMLG from blood sample 4A from patient 11\_33. The paired reads map to an ALU element adjacent to the integration site and most contain multiple mismatches with the reference genome sequence (color coded).
3. Simulation showing that based on replicate experiment in which 60 unique sites were obtained from blood sample visit 4A from patient 11\_33 the likelihood that one site is actually present at various percentages (x-axis) given 100, 1000 or 10000 actual sites in total.

## Supplementary Fig. S7. PCR amplification of Toca 511 genomes.

Cartoon showing genome organization of Toca 511 and PCR amplicons used for Toca 511 amplification and sequencing.

## Supplementary Fig. S8. Summary of indels identified in Toca 511 genomes.

1. Boxplot showing the total number of indels from different classes of samples at a frequency threshold of 3%: Toca 511 containing plasmid and cell lines (black, n=4)**,** blood (red, n=3), brain tumors re-resected after further progression of disease after Toca 511 delivery (magenta, n=4), brain tumors following IV dosing prior to Toca FC treatment (blue, n=6).
2. Barplot showing the fraction of indels across all patients from whom data could be obtained, with frequency 3-10 % (light green), 10-30% (green) and greater than 30% (dark green).

## Supplementary Fig. S9. Cytosine deaminase multiple sequence alignment.

Alignment of yCD2, its *S. cerevisiae* parent and representative orthologs from other fungal species. Conservation is indicated on a white to grey scale and also shown below the alignment as a barplot, where the color code and numbers indicate sequence identity/conservation. W10 in yCD2 is at position 10, enclosed in a green box.

##  Supplementary Fig. S10. Inactivating mutations in yCD2.

1. Table showing the percentage frequencies of three mutations that are expected to inactivate yCD2. The first set of columns corresponds to values obtained from targeted yCD2 sequencing and the second set of columns corresponds to values obtain from Toca 511 sequencing. The “functional” column is 100 minus the maximum of the three mutations and does not take into account that mutations may occur independently.
2. Scatterplot comparing mutation percentage of the three inactivating mutations according to yCD2 sequencing (x-axis) versus Toca 511 sequencing (y-axis).

## Supplementary Fig. S11. yCD2 sequencing technical replicates match.

1. Scatterplot comparing mutation frequencies across yCD2 for two technical replicates of 11\_06 visit 5 blood.
2. Same as (A) except for 11\_33 visit 4 blood.
3. Same as (A) except for 11\_33 visit 7 blood.
4. Same as (A) except for 13\_08 piece 14 tumor.

## Supplementary Fig. S12. Cytosine deaminase expression in tumor samples measured by IHC.

1. IHC staining with anti-CD antibody on resected tumor sections after Toca 511 delivery and multiple cycles of Toca FC. Positive staining of CD protein in a patient 11\_02 tumor section after Toca 511 treatment and multiple cycles to Toca FC. CD protein is stained brown; nuclei are in blue.
2. Positive staining of CD protein in a patient 11\_31 tumor section after Toca 511 treatment and a single cycle to Toca FC. CD protein is stained brown; nuclei are in blue.

## Supplementary Table S1. Tg 511-11-01 (Study 11) dosing table.

## Supplementary Table S2. Tg 511-13-01 (Study 13) dosing table.

## Supplementary Table S3. Toca 511 signal in tumor and nonneoplastic brain samples taken at time of autopsy from a Study 8 patient.

## Dataset S1. Toca 511 qPCR data.

## Dataset S2. Locations of Toca 511 Integration Sites.