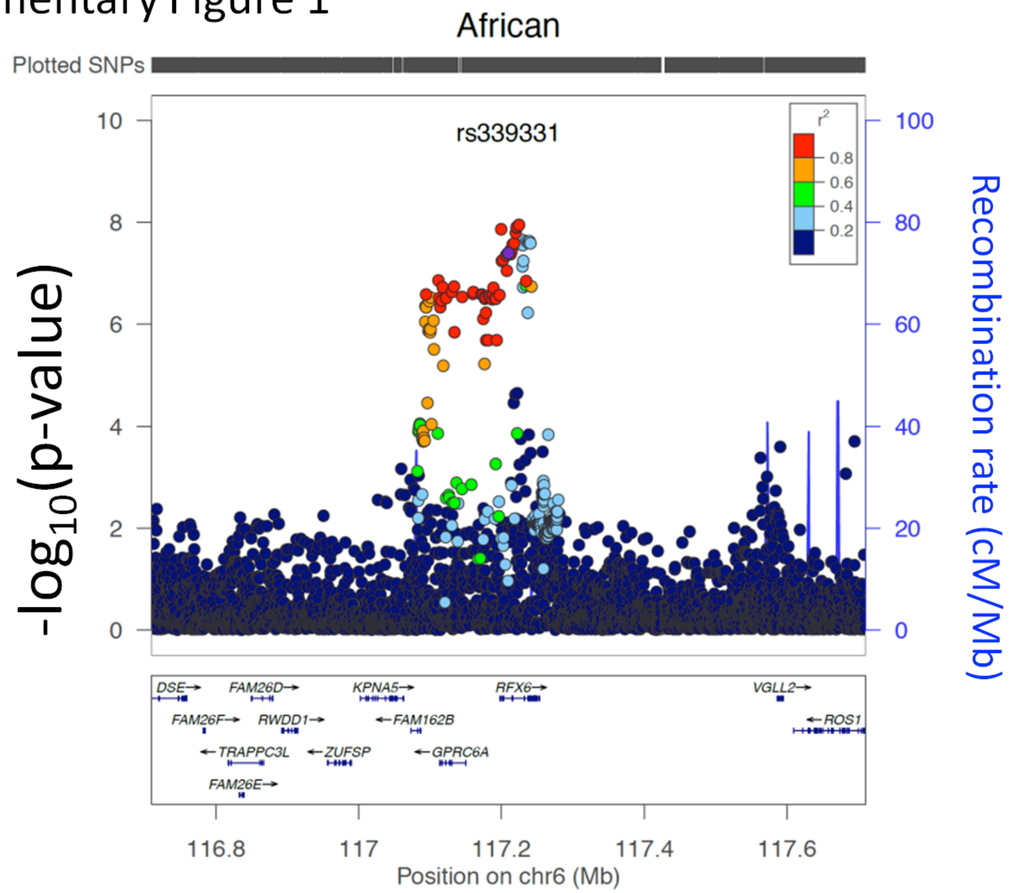
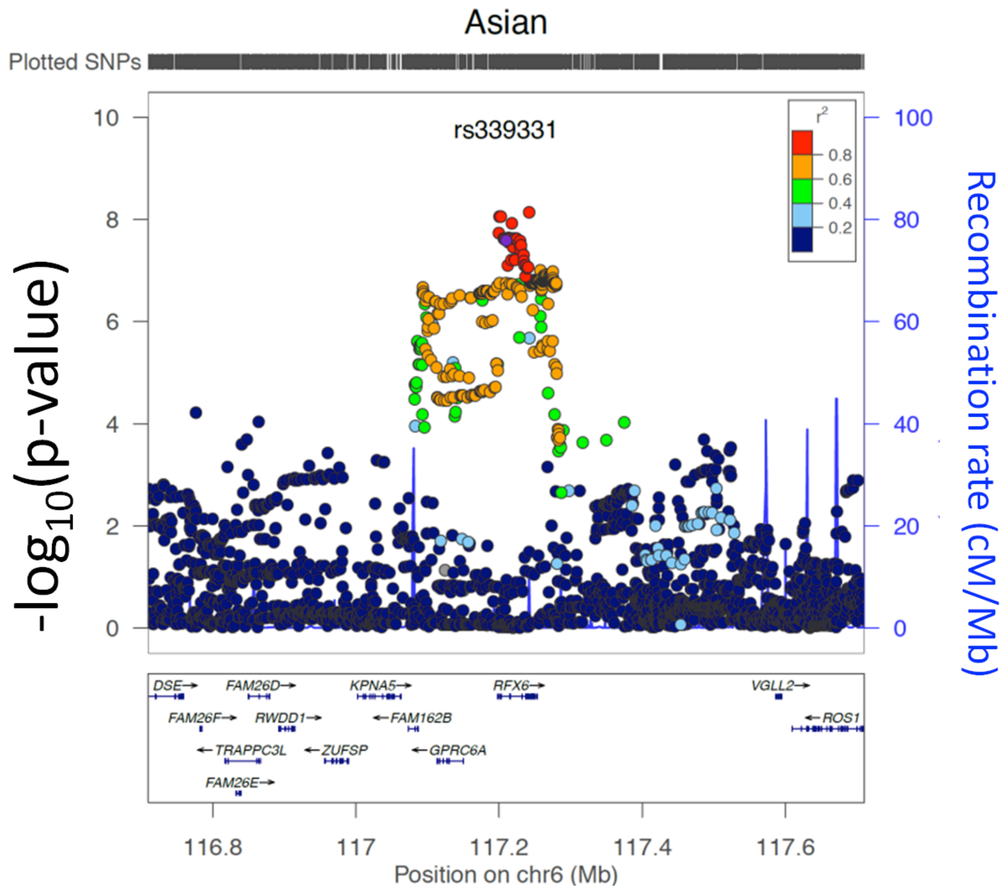
Supplementary Figure 1

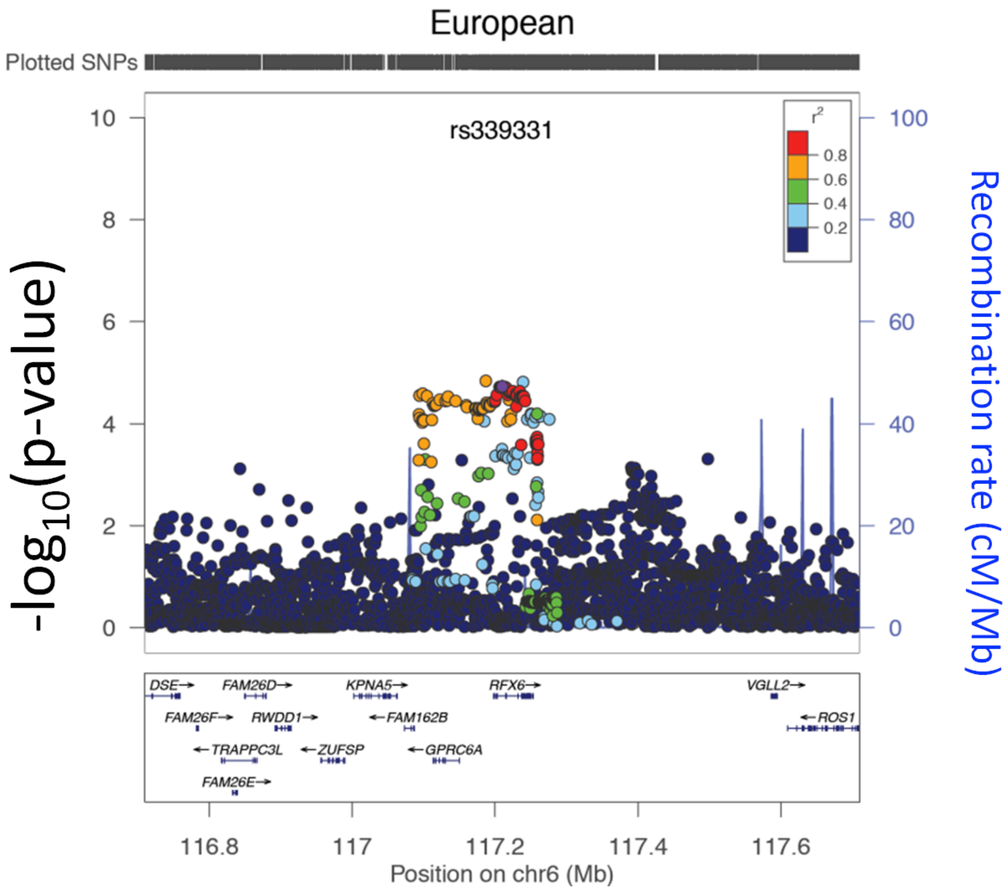
a



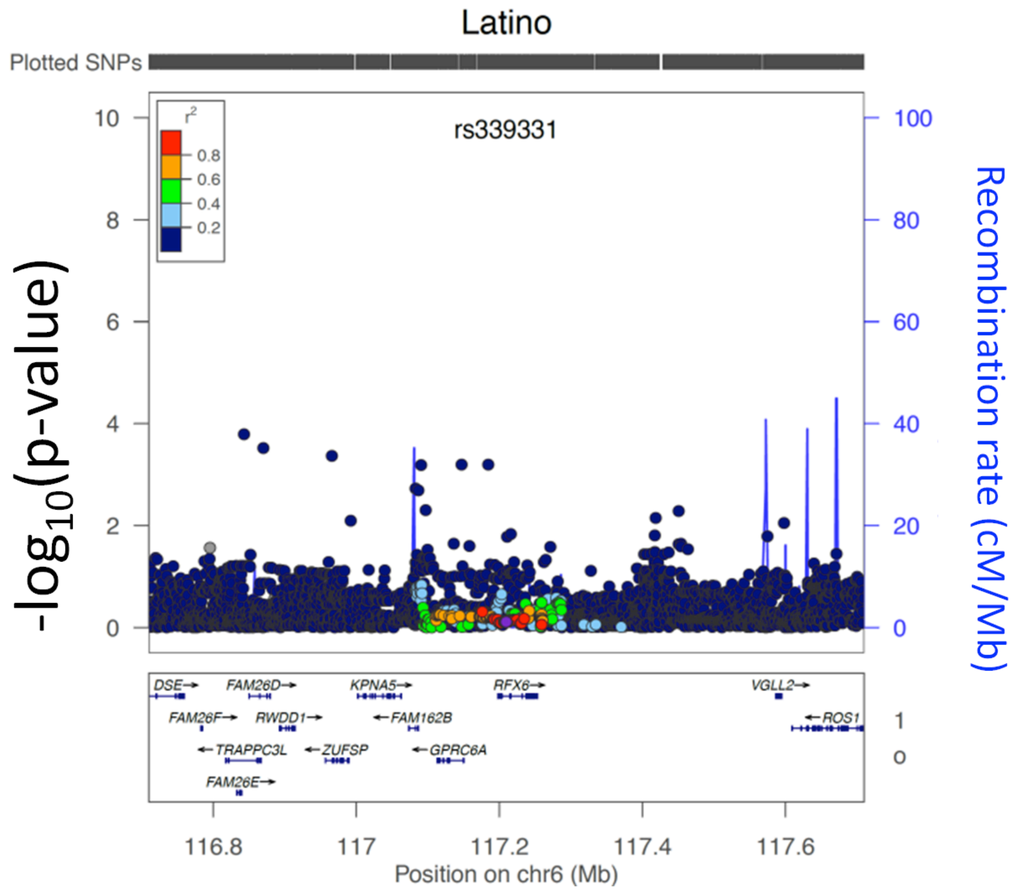
b



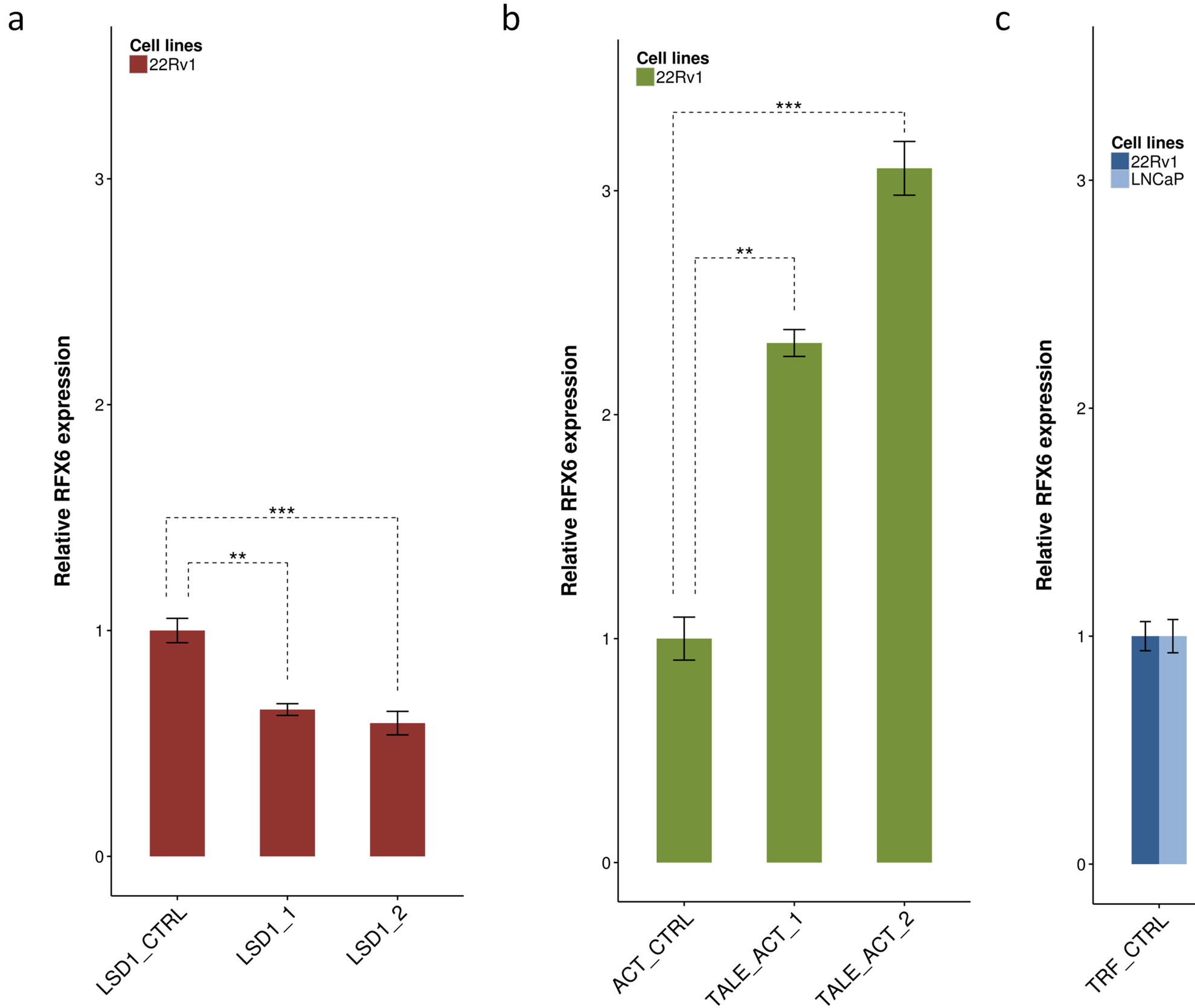
c



d



Supplementary Figure 2



Supplementary Figure 3

GAACTCTCTCTCCCCAGTTTCATGAGGTTTATCTTTAGTG

GAACTCTCTCTCCCCAGTTTTATGAGGTTTATCTTTAGTG

parental

a

Left TALEN

gDNA

TGAGATGAAAAATTTATGTACACTTTGCATGAACTCTCTCTCCCCAGTTTTATGAGGTTTATCTTTAGTGACTAGGAGGACAGAAAGCT

*

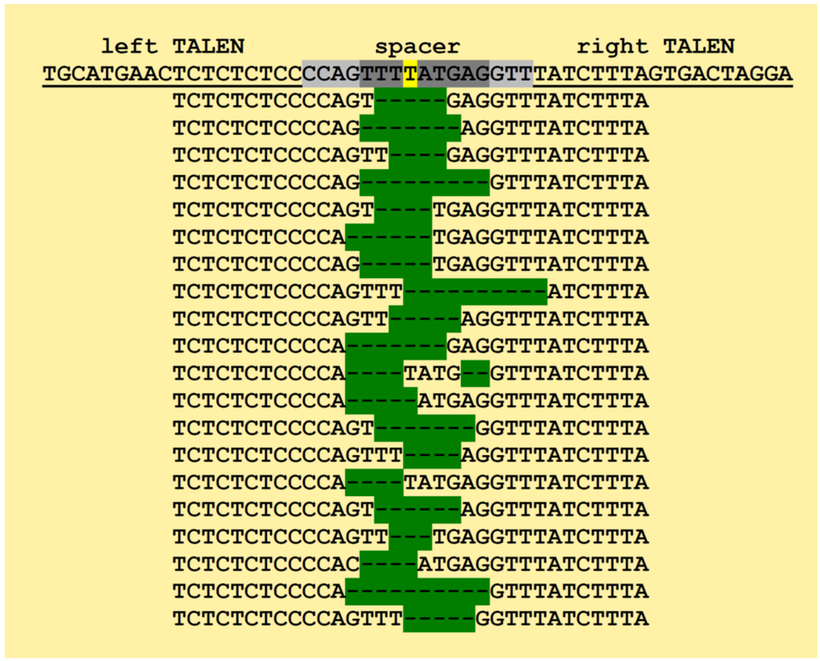
Right TALEN

X

HDR template oligo 200 bp

...AATTTATGTACACTTTGCATGAACTCTCTCTCCCCAGTTTCATGAGGTTTATCTTTAGTGACTAGGAGGACA...

NHEJ



HDR

wild type

GAACTCTCTCTCCCCAGTTTCATGAGGTTTATCTTTAGTG

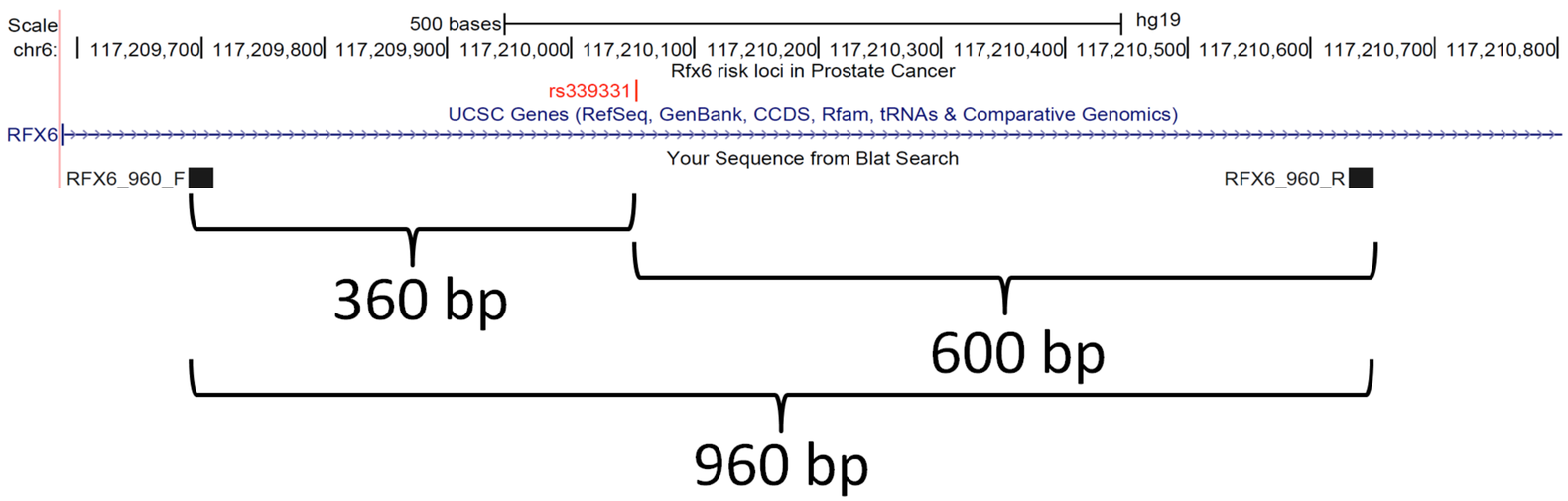
GAACTCTCTCTCCCCAGTTTCATGAGGTTTATCTTTAGTG

mutant

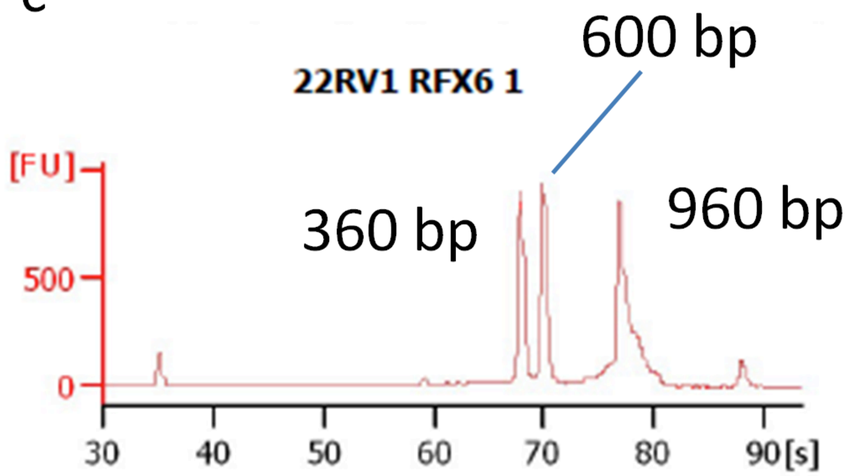
GAACTCTCTCTCCCCAGTTTTATGAGGTTTATCTTTAGTG

GAACTCTCTCTCCCCAGTTTTATGAGGTTTATCTTTAGTG

b



c



Supplementary Figure 4

a

PCR1

gDNA

b

PCR2

Gene specific forward primer with barcode compatible adapter 1

PCR1 product

c

PCR3

Illumina compatible barcode sequences

Gene specific reverse primer with barcode compatible adapter 2

PCR2 product

Gene specific amplicon with adapter sequences

d

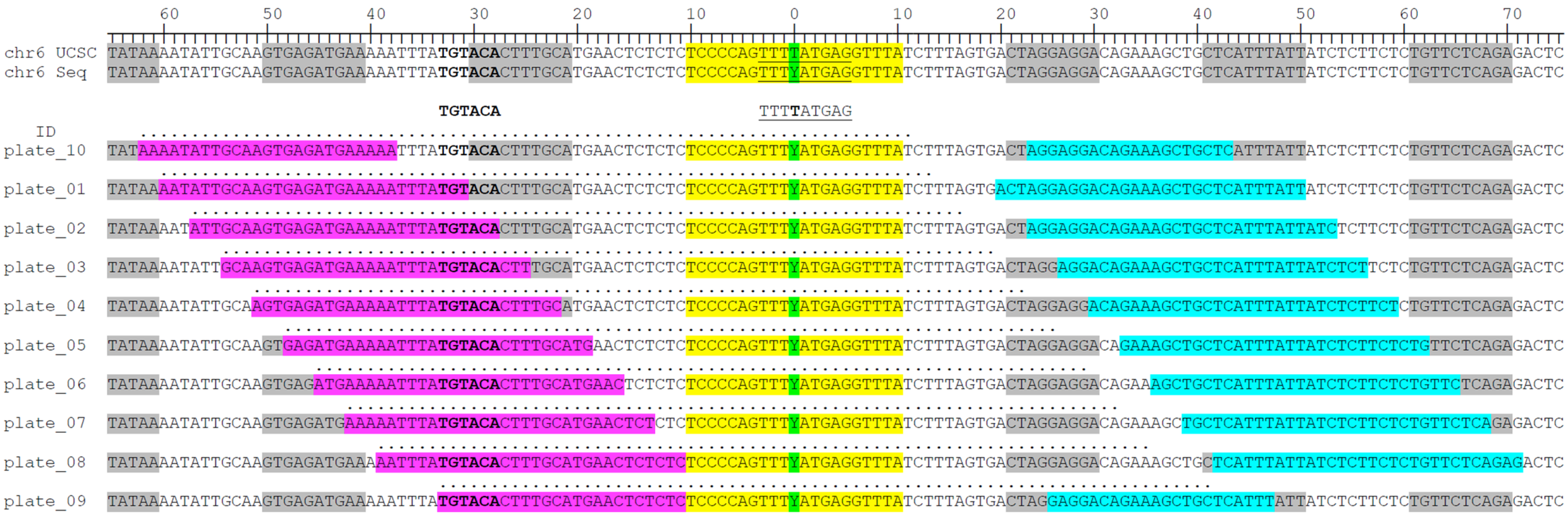
BC AD GS amplicon GS AD BC

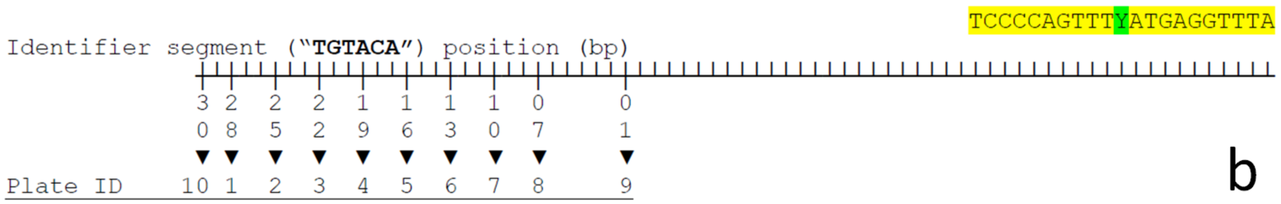
a

Supplementary Figure 5

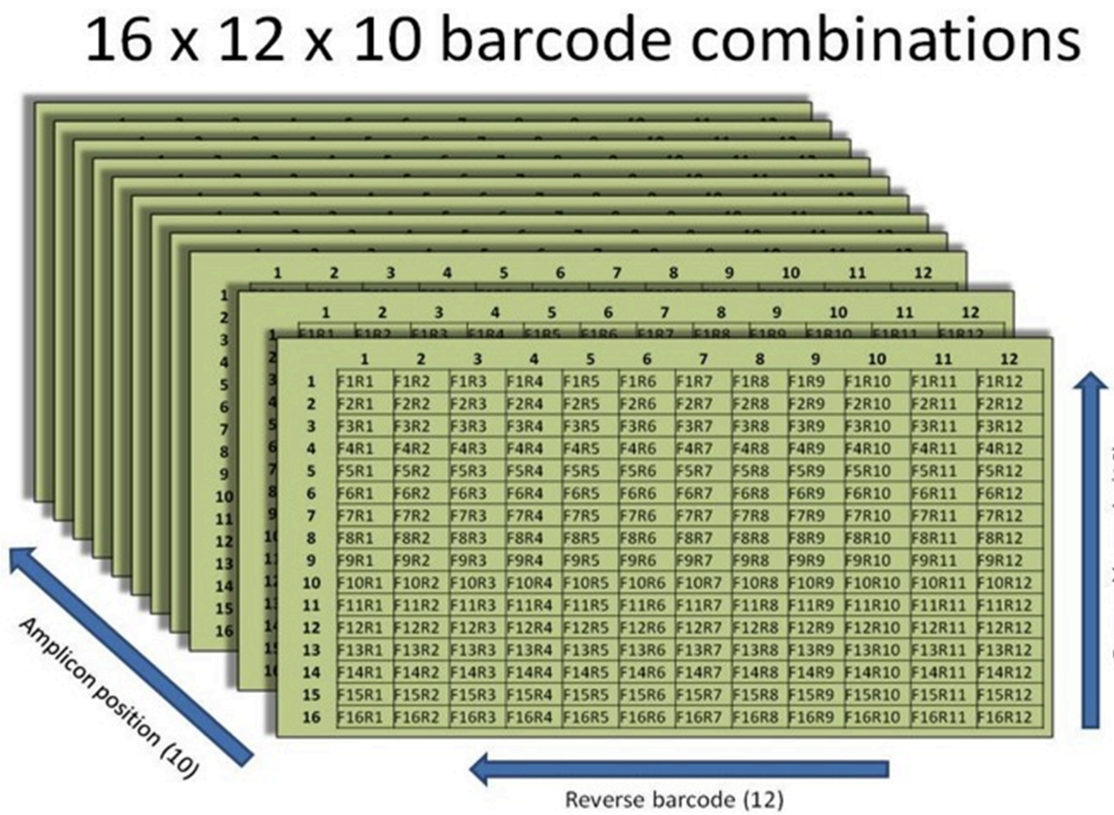
Interesting SNP: rs339331
Interesting SNP location: RFX6 gene 4th intron
Genomic position of SNP (GRCh37/hg19)chr6:117,210,052-117,210,052
Genomic position of selected 141bp refseq (GRCh37/hg19): chr6:117,209,986-117,210,126

Rfx6_amplicon_refseq_141bp





b



TTTTATGAG = **HoxB13 binding site**

Y **interesting SNP position**

Y = **C or T according to the UPCAC nomenclature**

NNNNN = **forward primer binding site**

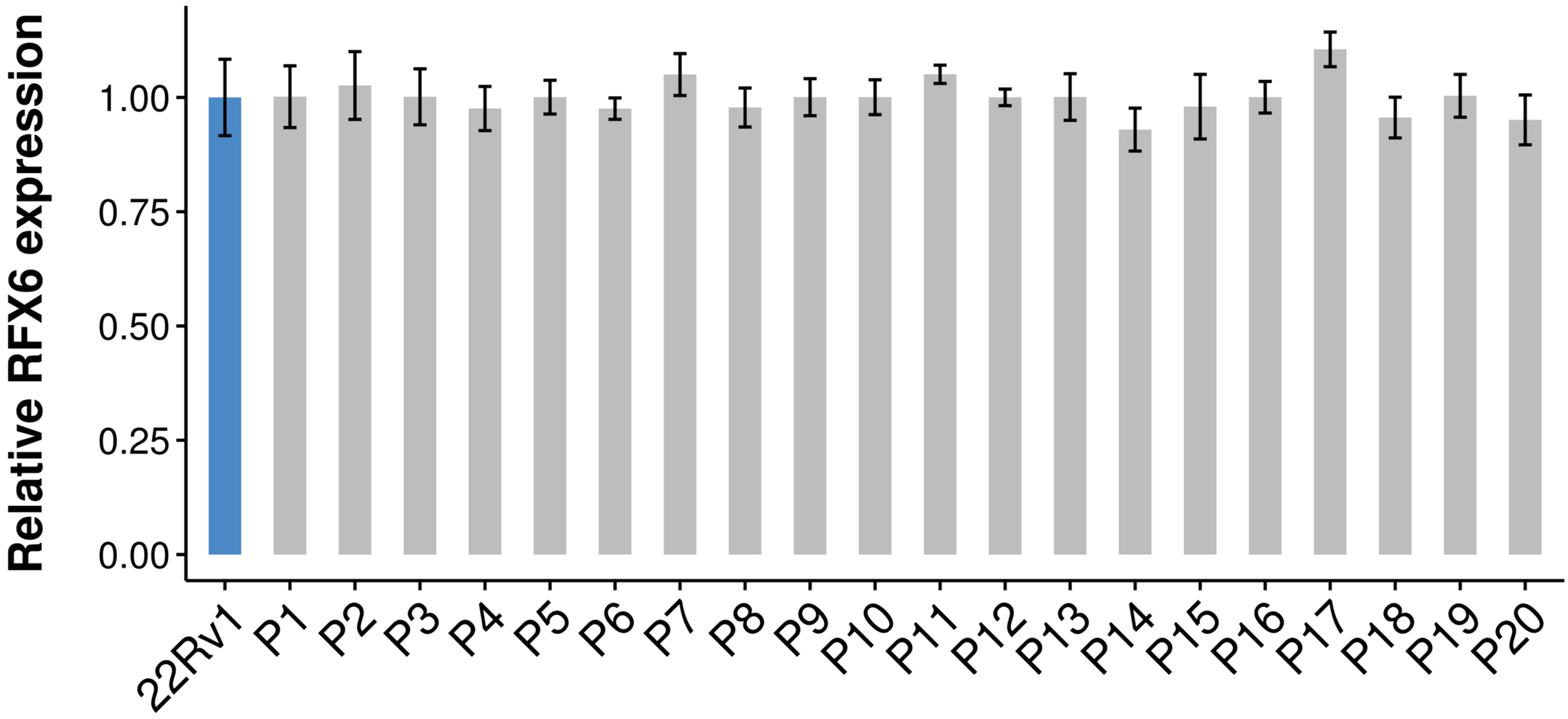
NNNNN = **reverse primer binding site**

..... = **position of the 75 bp forward reads**

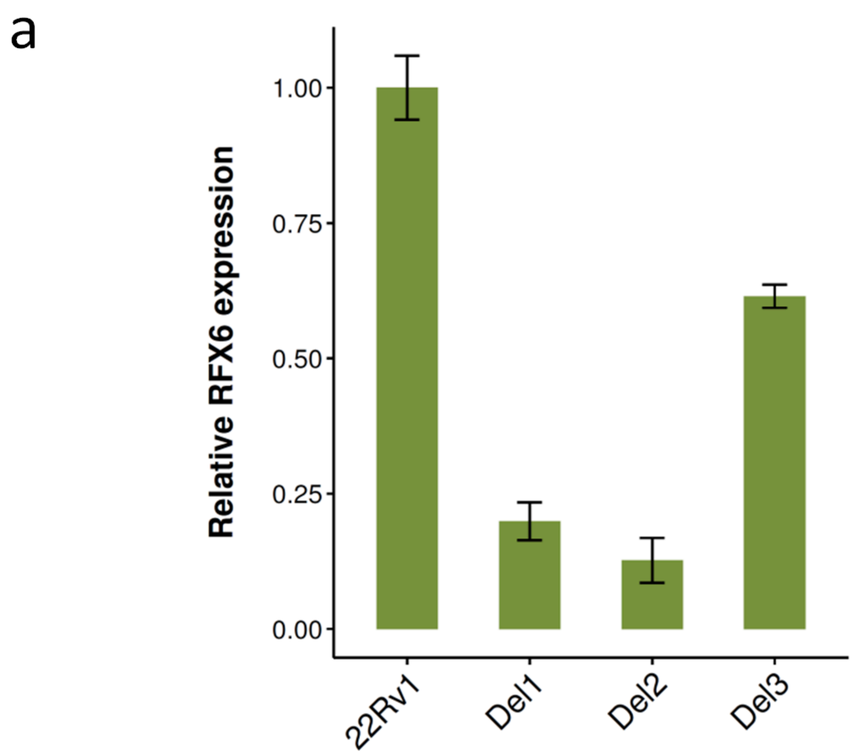
TGTACA = identifier segment – the position encoding the late ID

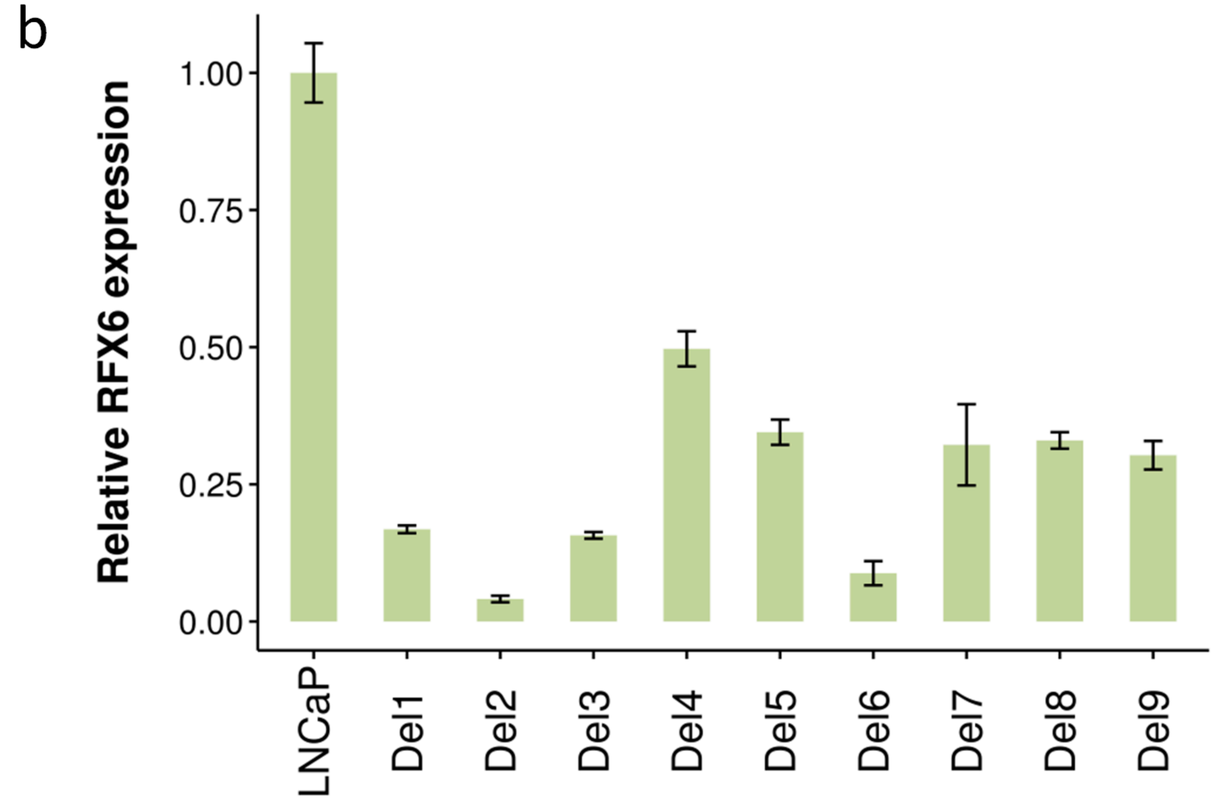
TCCCCAGTTTYATGAGGTTTA = **21 bp consensus core region for genotyping and insertion/deletion analysis**

Supplementary Figure 6

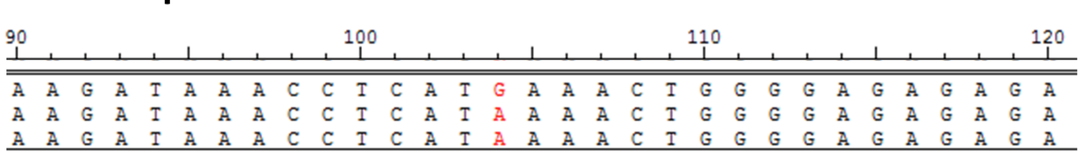


Supplementary Figure 7

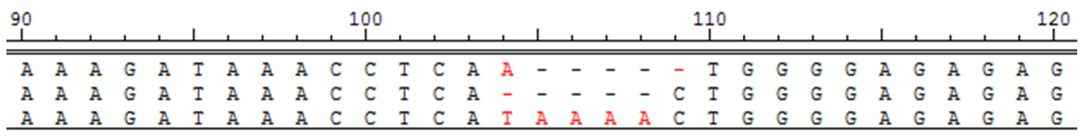




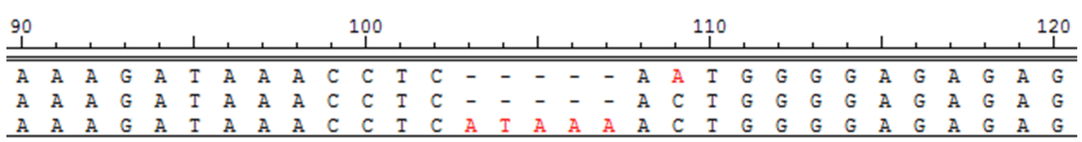
22Rv1 - parental



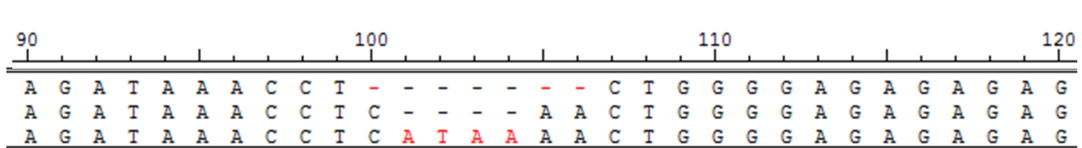
Del1

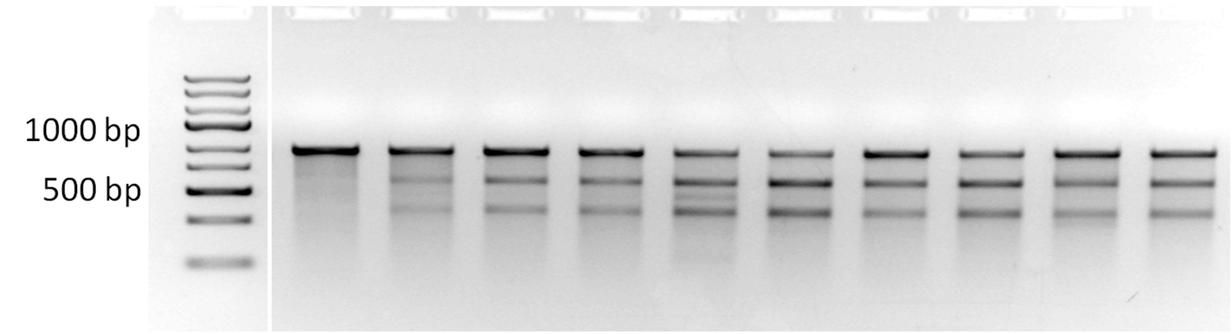


Del2



Del3





Supplementary Figure 8

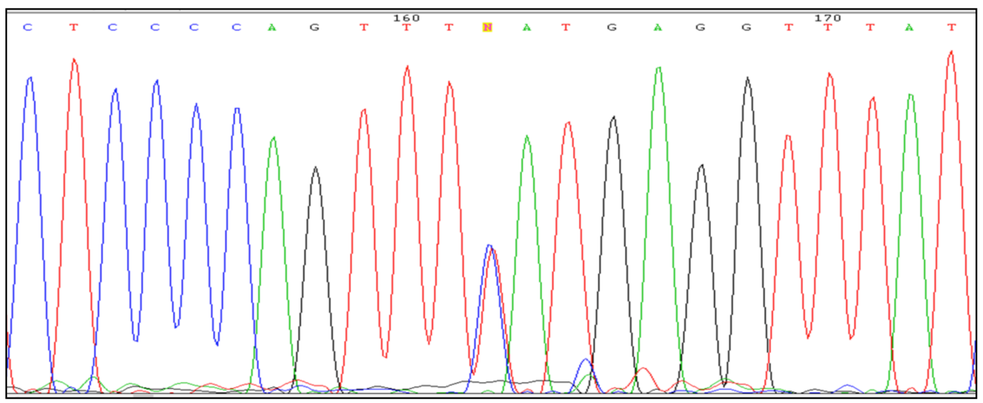
a

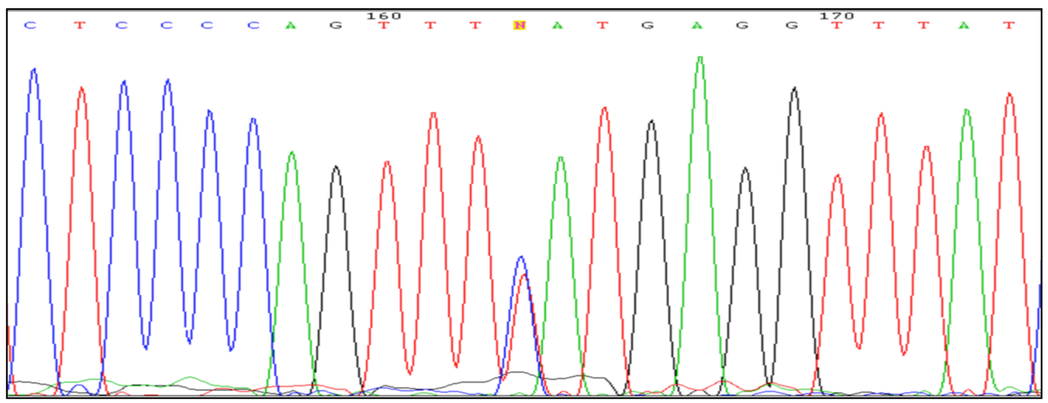
b

HoxB13

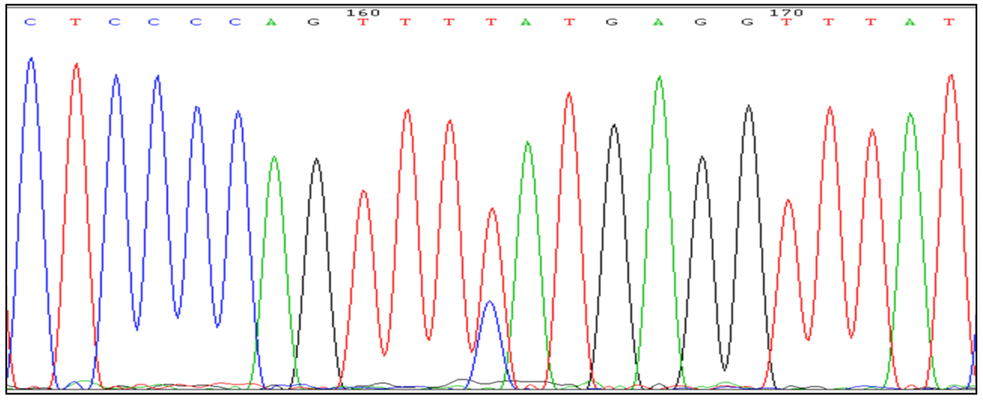
H3K4Me2

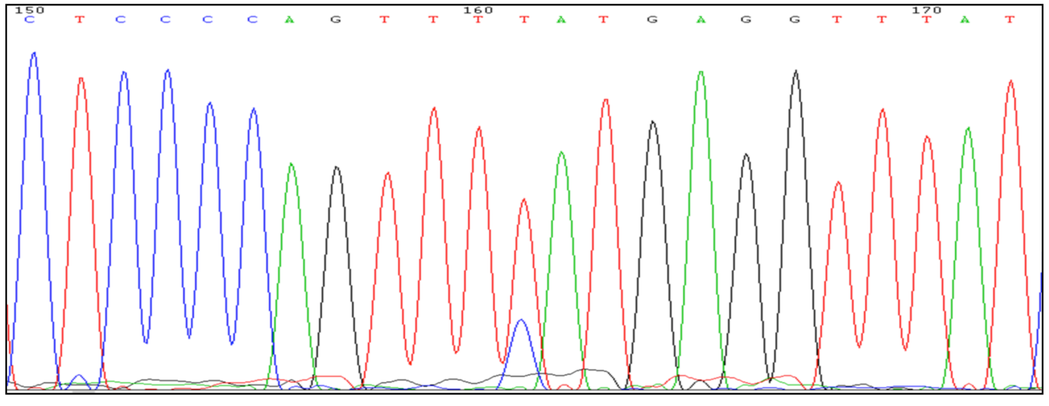
INPUT





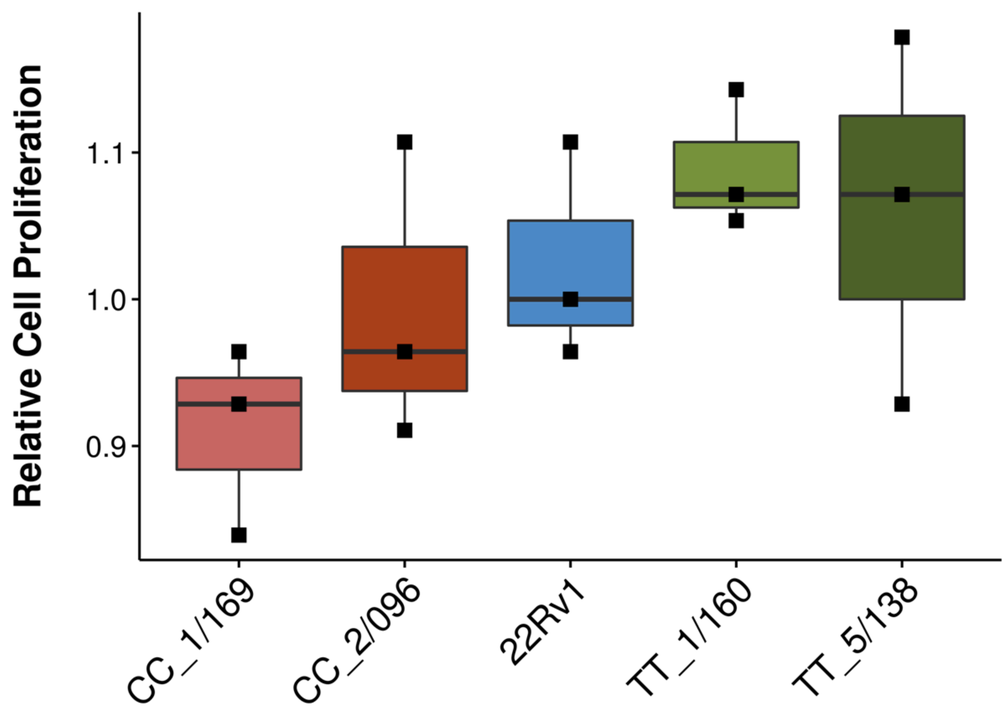
CHIP



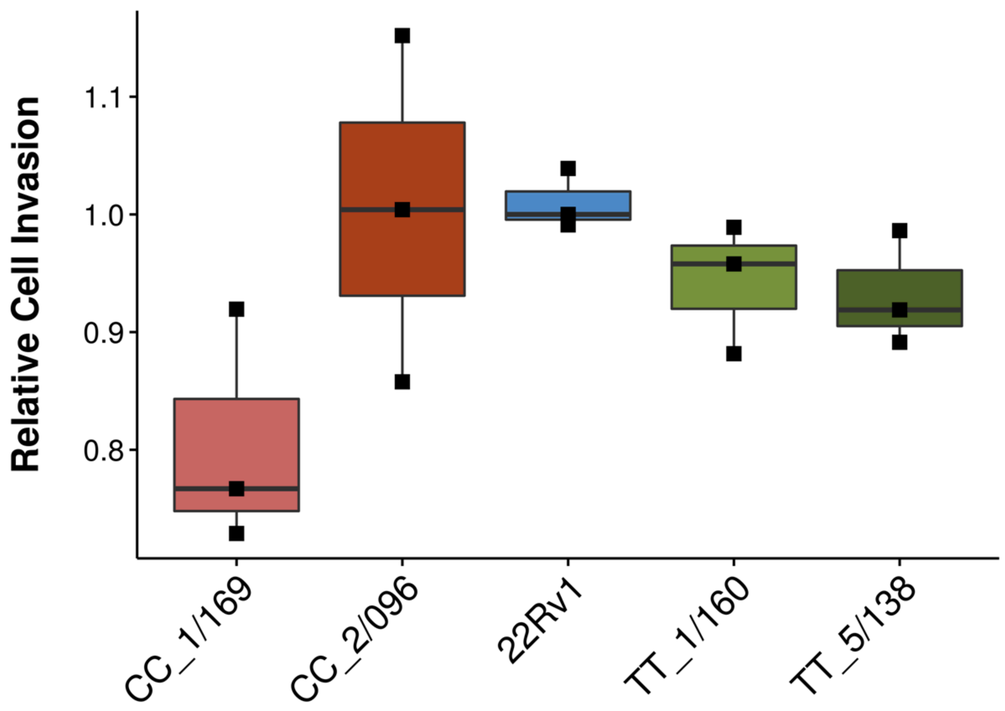


Supplementary Figure 9

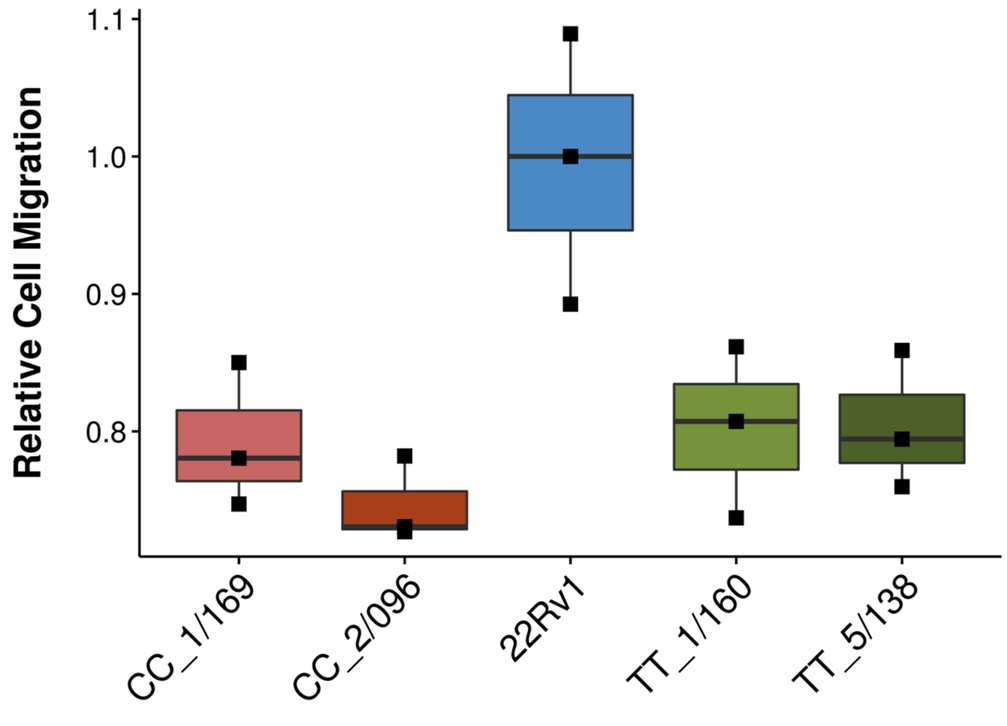
a



b

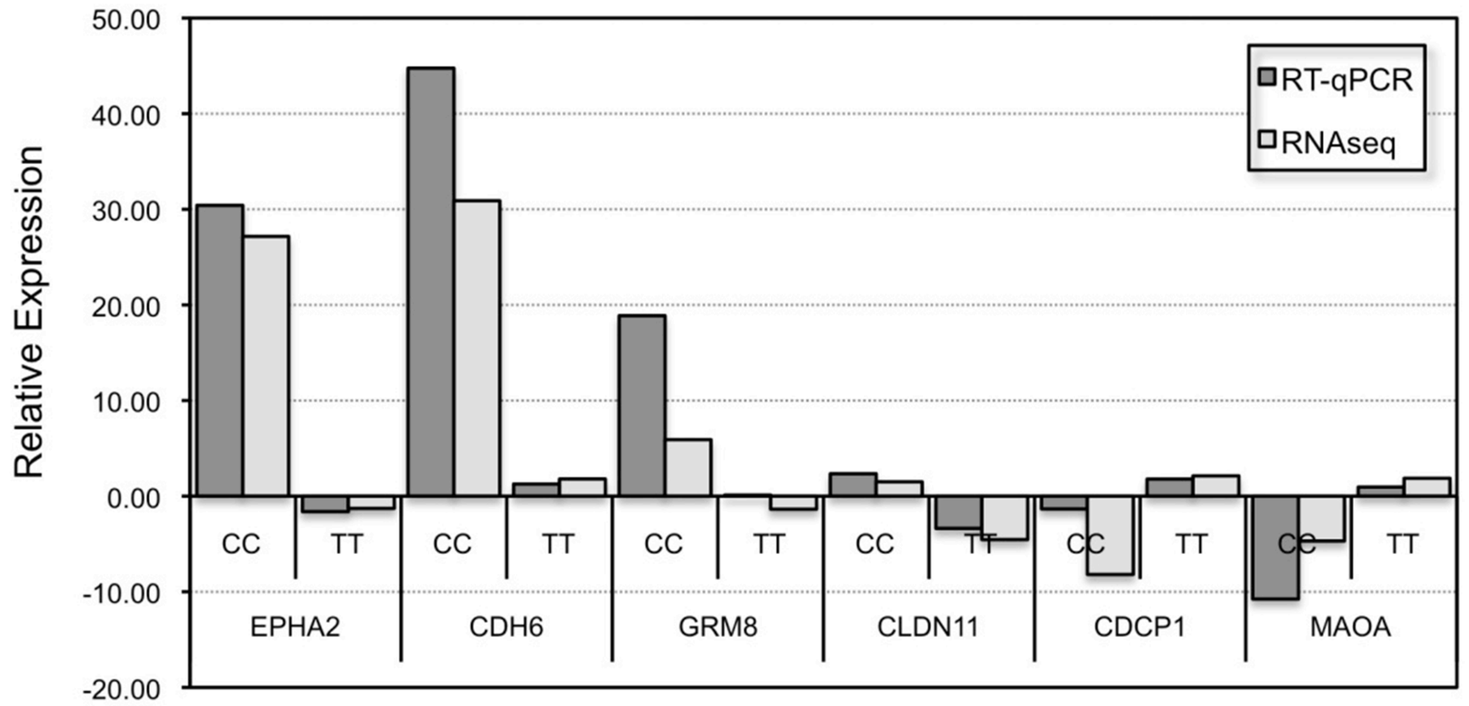


c

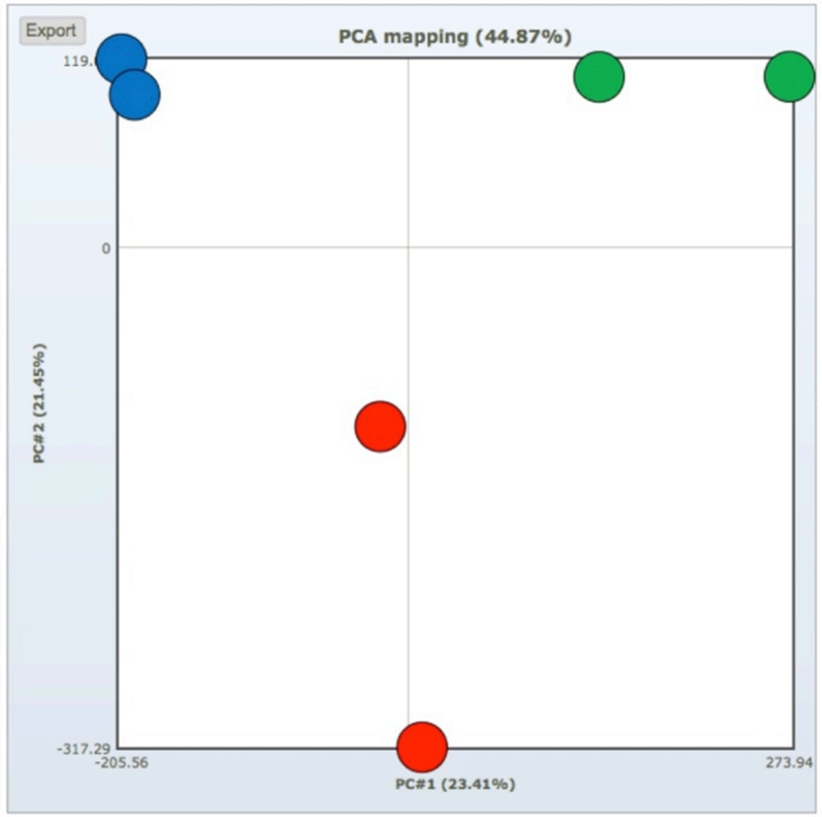


Supplementary Figure 10

a



b



type

Color	Value
	CC
	parental
	TT

Supplementary Figure Legends

Supplementary Figure 1: SNPs associated with prostate cancer cases across four populations

Plots of all SNPs in a 1 megabase interval centered on rs339331 and their significance (-log(p-value)) in 4 populations (**a-d**). Rs339331 is denoted in purple. Colors denote the degree of LD between rs339331 and other variants.

Supplementary Figure 2: Epigenome editing of the HOXB13 binding site influences *RFX6* gene expression changes in the 22Rv1 cell line

a. TALE-LSD1 reagents decrease RFX6 expression. The LSD1_CTRL is a control using a vector containing only the LSD1 enzyme. **b.** TALE-VP64 increases RFX6 expression The ACT_CTRL control contains only the VP-64 domain. **c.** Control experiments demonstrating no significant alterations in RFX6 levels in both cell lines; TRF_CTRL is a blank transfection without any vector construction, DBD_1 and DBD_2 are vectors that contain only the TALE DNA Binding Domain (without the LSD1 or VP64 domains), which were used for targeting the HOXB13 binding domain. *** $p<0.001$ and **$p<0.01$

Supplementary Figure 3: TALEN design against the rs339331 position

a. The ZiFit program was used to design a TALEN pair (http://zifit.partners.org/ZiFiT) against our sequence of interest[25,26]. The output demonstrates the repeat variable domain (RVD) structure and sequence localizations of the TALEN pairs to generate double stranded breaks for HDR or insertion/deletion by NHEJ in the rs339331 position (*). **b.** The genomic location of the RFX6_960F and RFX6_960R primers, which amplify the 960 bp segment of the intron 4 bearing the rs339331 for the cleavage assay. The T7E1 assay is predicted to be produce 600 bp and 360 bp DNA fragments. **c.** The result of the T7E1 cleavage assay shows roughly 50% cutting efficiency.

Supplementary Figure 4: Barcoding 192 clones within a plate.

a. 1st round PCR, amplicon generation using gene specific primers. **b.** 2nd round PCR, creates an amplicon that has generic adapters (red) that serve as binding sites for the barcoded (green) adapters in the 3rd round PCR **c.** The 16 forward barcodes

(green) and 12 reverse (green) barcodes can result in 16 x 12 = 192 unique identifiers. **d.** Anatomy of the amplicon, BC=barcode, AD=adapter, GS=gene specific primer.

Supplementary Figure 5: Shifted amplicon sequencing to uniquely identify plates allows the identification of 1,920 clones

a. A 141 bp reference sequence was selected in the region of interest on chromosome 6 to design 30 bp long gene specific forward (purple) and reverse (light blue) primers to amplify the interesting SNP (green) and the surrounding ± 10 bp area (yellow). *Each plate has its own gene specific primer with a unique starting position thereby allowing a unique barcode for each plate*. After sequencing, each plate can be identified using the starting *position* of the six nucleotide "TGTACA" string and then can be associated with the plate ID shown in the ruler under "Primer starting position". For example, sequences where the identifier segment starts at the position 25 belonging to the plate 2.

b. Three dimensional barcoding strategy. Within the plates, the forward and reverse barcodes can be multiplied and are considered as the first two dimensions. The third dimension, plate identification, employs distinct amplicons (each containing the region of interest), to distinguish each plate. Thus, the same set of forward and reverse barcode combinations can be used across amplicons. For example, well A1 across all plates used the same forward and reverse barcodes and the plates are distinguished by the position of the "TGTACA" nucleotide sequence as described above.

Supplementary Figure 6: RFX6 expression levels do not significantly differ across 20 clones

Twenty independent clones were derived from the parental 22Rv1 prostate cell line and *RFX6* levels were measured by qRT-PCR.

Supplementary Figure 7: TALEN-induced deletions at the HOXB13 locus in the 22Rv1 and LNCaP prostate cancer cell lines decrease RFX6 expression.

a. For 22Rv1, three independent clones were analyzed. From these clones, a 960 bp amplicon was cloned into a vector and after bacterial transformation, 12 colonies

were Sanger sequenced for each 22Rv1 clone. One representative allele was aligned against the RFX6 reference sequence. Each sample has three tracks (sequences); allele 1, allele 2 and the reference sequence. The Del1 sample shows two 5-bp deletions. The Del2 sample shows two 5-bp deletions. The Del3 sample has a 7 and a 4 base pair deletion. **b.** For LNCaP, transfections were performed with a TALEN pair and nine clones were selected. The T7E1 assay was used to screen for deletions and then *RFX6* levels were measured in clones that were positive by the T7E1 assay.

Supplementary Figure 8: Effect of the C and T alleles on the HOXB13 binding and histone (H3K4Me2) enrichment

ChIP-Sanger sequencing in the parental C/T heterozygote DNA at the rs339331 position. The T allele binds more avidly to HOXB13 and H3K4me2 than the C allele.

Supplementary Figure 9: Proliferation, Migration, and Invasion

Proliferation (a) assays: 1000 cells/well were seeded into 96 well plates and incubated at 37°C for 48 hours. MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) (VWR) was added at a final concentration of 500 ng ml^{-1}. Cells were incubated for 2 hours, media removed and cells washed with PBS. Cells were lysed in 100% DMSO and absorbance read at 570 nm on a Varioskan Flash platereader. Three technical replicates were performed for each sample. Data were normalized to the parental 22Rv1 cell line.

Migration (b) and Invasion (c) Assays: 96-well plate format migration and invasion assays (Trevigen) were performed according to the manufacturers instructions. Briefly, $1x10^6$ cells were starved for 24 hours, then harvested and washed in serum free medium. Cell suspensions were normalized to $1x10^6$ cells ml^{-1}, and 50 µl seeded into migration chambers, or for invasion assays, onto chambers that had been pre-coated with a 1:10 dilution of basement membrane extract. 10% serum was used as a chemoattractant and cells were incubated for 24 hours before disassembly of the transwell plate inserts, washing of the membranes, disassociation and detection of the invaded/migratory cells, using Calcein-AM diluted in cell dissociation buffer.

Relative fluorescence units were read on a Varioskan Flash platereader (485 nm excitation, 520 nm emission). Data were normalized to the parental 22Rv1 cell line. No significant differences were observed (P>0.05, two-tailed paired T-test) for any of the tested assays (proliferation, migration, and invasion).

Supplementary Figure 10: RNAseq profiling of isogenic models

a. Validation of selected genes by RT-qPCR. The average change in expression for two biological replicates is shown relative to the average expression in two independent RNA samples from the parental line. **b.** Principal component analysis (PCA) plot. Two clones were analyzed per genotype, plus two independent RNA isolates from the parental cell line. Biological replicate samples cluster together.