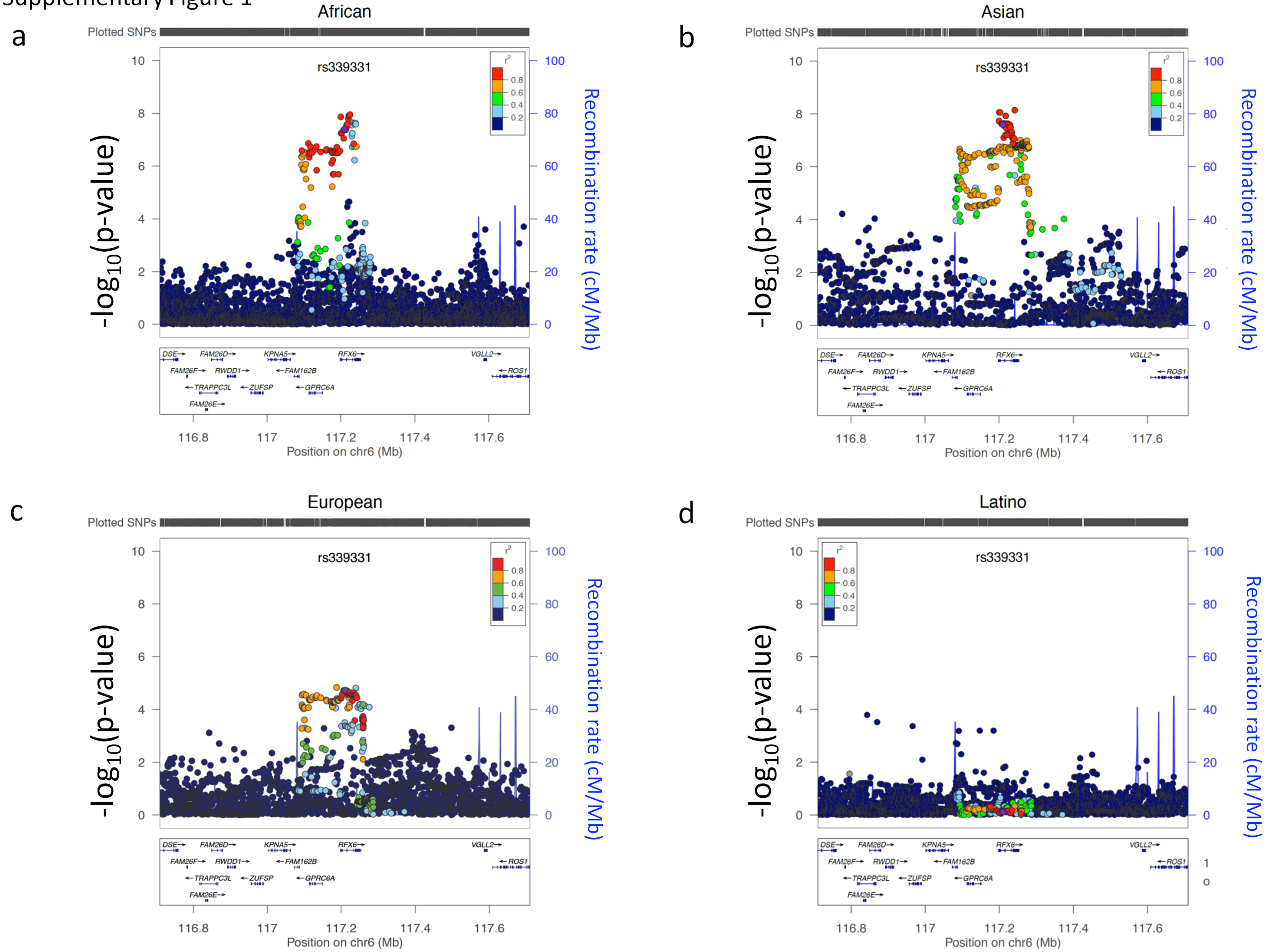
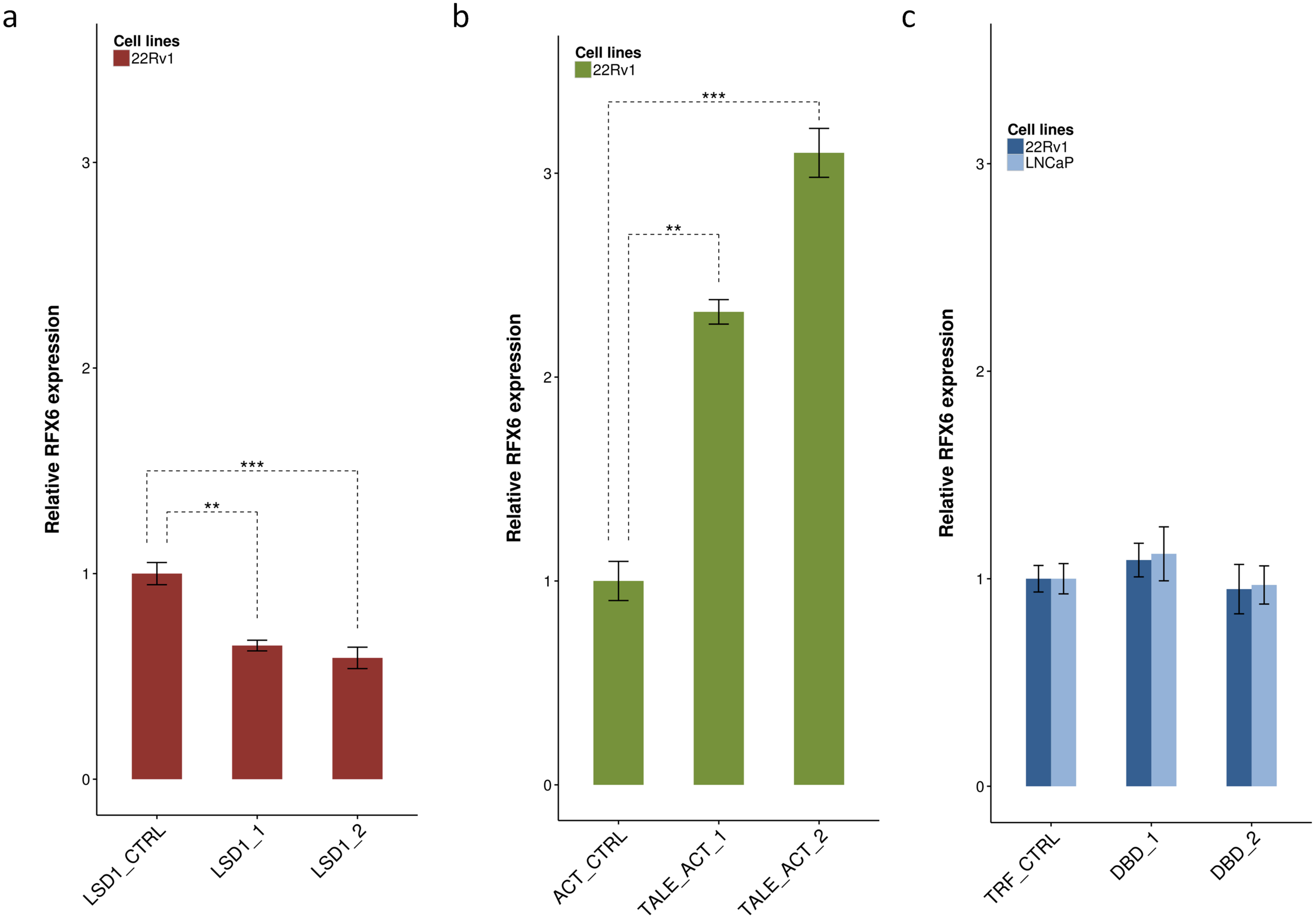
Supplementary Figure 1



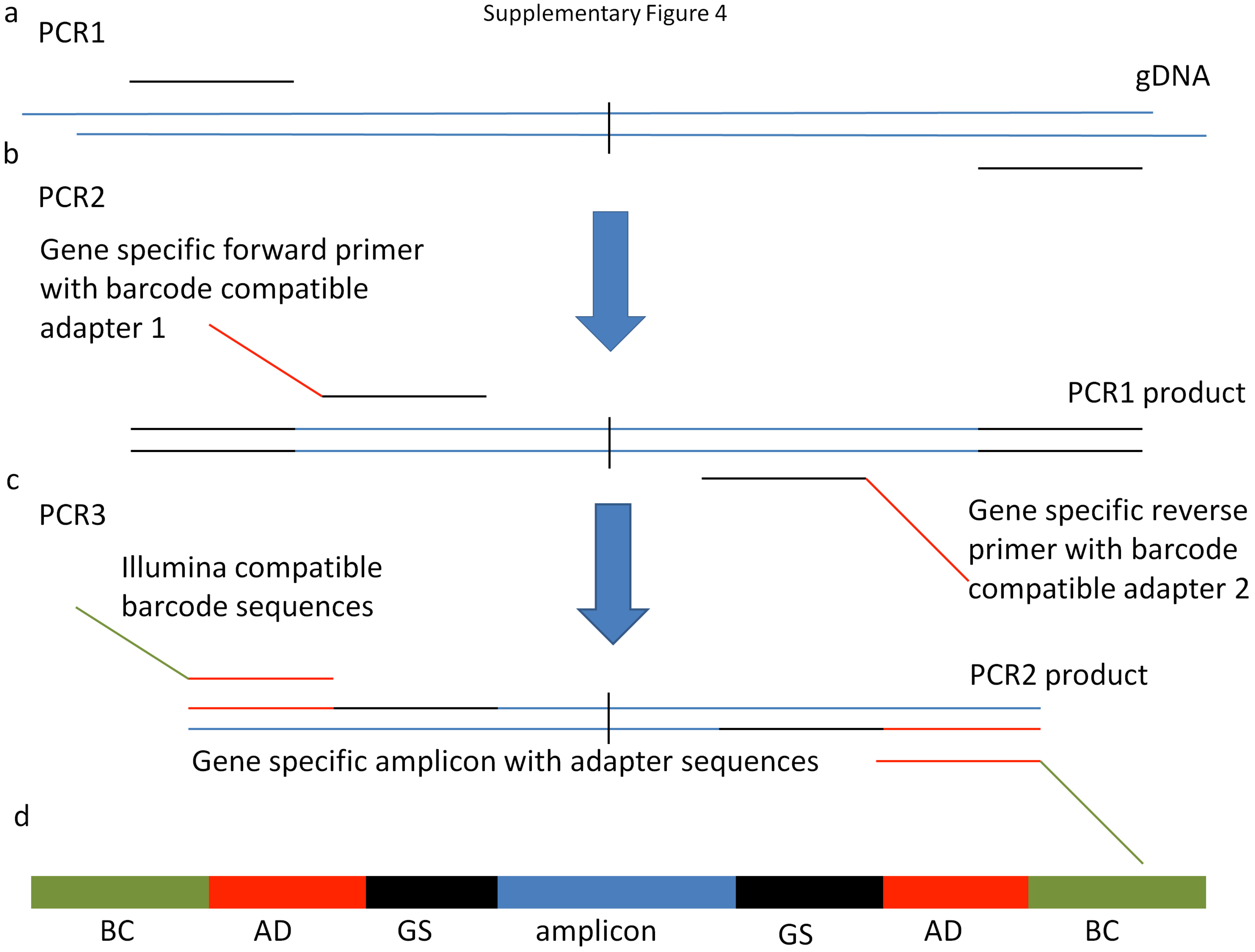
Supplementary Figure 2



Supplementary Figure 3

GAACTCTCTCTCCCCAGTTTCATGAGGTTTATCTTTAGTG
GAACTCTCTCTCCCCAGTTTTATGAGGTTTATCTTTAGTG
parental
a
Left TALEN
N H N N N N N H N H N H N H N H
N D I G N I I D G D G D G D G D H D
tGCATGAACTCTCTCTCC
gDNA
*
TGAGATGAAAAATTTATGTACACTTTGCATGAACTCTCTCTCCCCAGTTTTATGAGGTTTATCTTTAGTGACTAGGAGGACAGAAAGCT
Right TALEN
X
HDR template oligo 200 bp
...AATTTATGTACACTTTGCATGAACTCTCTCTCCCCAGTTTCATGAGGTTTATCTTTAGTGACTAGGAGGACA...
NHEJ
left TALEN
spacer
right TALEN
TGCATGAACTCTCTCTCCCCAGTTTTATGAGGTTTATCTTTAGTGACTAGGA
TCTCTCTCCCCAGT-----GAGGTTTATCTTTA
TCTCTCTCCCCAG-------AGGTTTATCTTTA
TCTCTCTCCCCAGTT----GAGGTTTATCTTTA
TCTCTCTCCCCAG---------GTTTATCTTTA
TCTCTCTCCCCAGT----TGAGGTTTATCTTTA
TCTCTCTCCCCA------TGAGGTTTATCTTTA
TCTCTCTCCCCAG-----TGAGGTTTATCTTTA
TCTCTCTCCCCAGTTT----------ATCTTTA
TCTCTCTCCCCAGTT-----AGGTTTATCTTTA
TCTCTCTCCCCA-------GAGGTTTATCTTTA
TCTCTCTCCCCA----TATG--GTTTATCTTTA
TCTCTCTCCCCA-----ATGAGGTTTATCTTTA
TCTCTCTCCCCAGT-------GGTTTATCTTTA
TCTCTCTCCCCAGTTT----AGGTTTATCTTTA
TCTCTCTCCCCA----TATGAGGTTTATCTTTA
TCTCTCTCCCCAGT------AGGTTTATCTTTA
TCTCTCTCCCCAGTT---TGAGGTTTATCTTTA
TCTCTCTCCCCAC----ATGAGGTTTATCTTTA
TCTCTCTCCCCA----------GTTTATCTTTA
TCTCTCTCCCCAGTTT-----GGTTTATCTTTA
HDR
wild type
GAACTCTCTCTCCCCAGTTTCATGAGGTTTATCTTTAGTG
GAACTCTCTCTCCCCAGTTTCATGAGGTTTATCTTTAGTG
mutant
GAACTCTCTCTCCCCAGTTTTATGAGGTTTATCTTTAGTG
GAACTCTCTCTCCCCAGTTTTATGAGGTTTATCTTTAGTG
b
Scale
500 bases
hg19
chr6:
117,209,700
117,209,800
117,209,900
117,210,000
117,210,100
117,210,200
117,210,300
117,210,400
117,210,500
117,210,600
117,210,700
117,210,800
Rfx6 risk loci in Prostate Cancer
rs339331
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
RFX6
Your Sequence from Blat Search
RFX6_960_F
RFX6_960_R
360 bp
600 bp
960 bp
c
22RV1 RFX6 1
[FU]
500
0
30
40
50
60
70
80
90 [s]
600 bp
360 bp
960 bp

Supplementary Figure 4


a
PCR1
gDNA
b
PCR2
Gene specific forward primer with barcode compatible adapter 1
PCR1 product
c
PCR3
Gene specific reverse primer with barcode compatible adapter 2
Illumina compatible barcode sequences
PCR2 product
Gene specific amplicon with adapter sequences
d
BC
AD
GS
amplicon
GS
AD
BC

Supplementary Figure 5

a

```
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

               60        50        40        30        20        10         0         10        20        30        40        50        60        70
chr6 UCSC TATAAAATATTGCAAGTGAGATGAAAAATTTATGTACACTTTGCATGAACTCTCTCTCCCCAGTTTTATGAGGTTTATCTTTAGTGACTAGGAGGACAGAAAGCTGCTCATTTATTATCTCTTCTCTGTTCTCAGAGACTC
chr6 Seq  TATAAAATATTGCAAGTGAGATGAAAAATTTATGTACACTTTGCATGAACTCTCTCTCCCCAGTTTYATGAGGTTTATCTTTAGTGACTAGGAGGACAGAAAGCTGCTCATTTATTATCTCTTCTCTGTTCTCAGAGACTC

                                          TGTACA                          TTTTATGAG
   ID
             ...........................................................................
plate_10  TATAAAATATTGCAAGTGAGATGAAAAATTTATGTACACTTTGCATGAACTCTCTCTCCCCAGTTTYATGAGGTTTATCTTTAGTGACTAGGAGGACAGAAAGCTGCTCATTTATTATCTCTTCTCTGTTCTCAGAGACTC
               ...........................................................................
plate_01  TATAAAATATTGCAAGTGAGATGAAAAATTTATGTACACTTTGCATGAACTCTCTCTCCCCAGTTTYATGAGGTTTATCTTTAGTGACTAGGAGGACAGAAAGCTGCTCATTTATTATCTCTTCTCTGTTCTCAGAGACTC
                  ...........................................................................
plate_02  TATAAAATATTGCAAGTGAGATGAAAAATTTATGTACACTTTGCATGAACTCTCTCTCCCCAGTTTYATGAGGTTTATCTTTAGTGACTAGGAGGACAGAAAGCTGCTCATTTATTATCTCTTCTCTGTTCTCAGAGACTC
                     ...........................................................................
plate_03  TATAAAATATTGCAAGTGAGATGAAAAATTTATGTACACTTTGCATGAACTCTCTCTCCCCAGTTTYATGAGGTTTATCTTTAGTGACTAGGAGGACAGAAAGCTGCTCATTTATTATCTCTTCTCTGTTCTCAGAGACTC
                        ...........................................................................
plate_04  TATAAAATATTGCAAGTGAGATGAAAAATTTATGTACACTTTGCATGAACTCTCTCTCCCCAGTTTYATGAGGTTTATCTTTAGTGACTAGGAGGACAGAAAGCTGCTCATTTATTATCTCTTCTCTGTTCTCAGAGACTC
                           ...........................................................................
plate_05  TATAAAATATTGCAAGTGAGATGAAAAATTTATGTACACTTTGCATGAACTCTCTCTCCCCAGTTTYATGAGGTTTATCTTTAGTGACTAGGAGGACAGAAAGCTGCTCATTTATTATCTCTTCTCTGTTCTCAGAGACTC
                              ...........................................................................
plate_06  TATAAAATATTGCAAGTGAGATGAAAAATTTATGTACACTTTGCATGAACTCTCTCTCCCCAGTTTYATGAGGTTTATCTTTAGTGACTAGGAGGACAGAAAGCTGCTCATTTATTATCTCTTCTCTGTTCTCAGAGACTC
                                 ...........................................................................
plate_07  TATAAAATATTGCAAGTGAGATGAAAAATTTATGTACACTTTGCATGAACTCTCTCTCCCCAGTTTYATGAGGTTTATCTTTAGTGACTAGGAGGACAGAAAGCTGCTCATTTATTATCTCTTCTCTGTTCTCAGAGACTC
                                    ...........................................................................
plate_08  TATAAAATATTGCAAGTGAGATGAAAAATTTATGTACACTTTGCATGAACTCTCTCTCCCCAGTTTYATGAGGTTTATCTTTAGTGACTAGGAGGACAGAAAGCTGCTCATTTATTATCTCTTCTCTGTTCTCAGAGACTC
                                          ...........................................................................
plate_09  TATAAAATATTGCAAGTGAGATGAAAAATTTATGTACACTTTGCATGAACTCTCTCTCCCCAGTTTYATGAGGTTTATCTTTAGTGACTAGGAGGACAGAAAGCTGCTCATTTATTATCTCTTCTCTGTTCTCAGAGACTC

                                                         TCCCCAGTTTYATGAGGTTTA
```

Identifier segment ("TGTACA") position (bp)	30	28	25	22	19	16	13	10	07	01
	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼
Plate ID	10	1	2	3	4	5	6	7	8	9

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

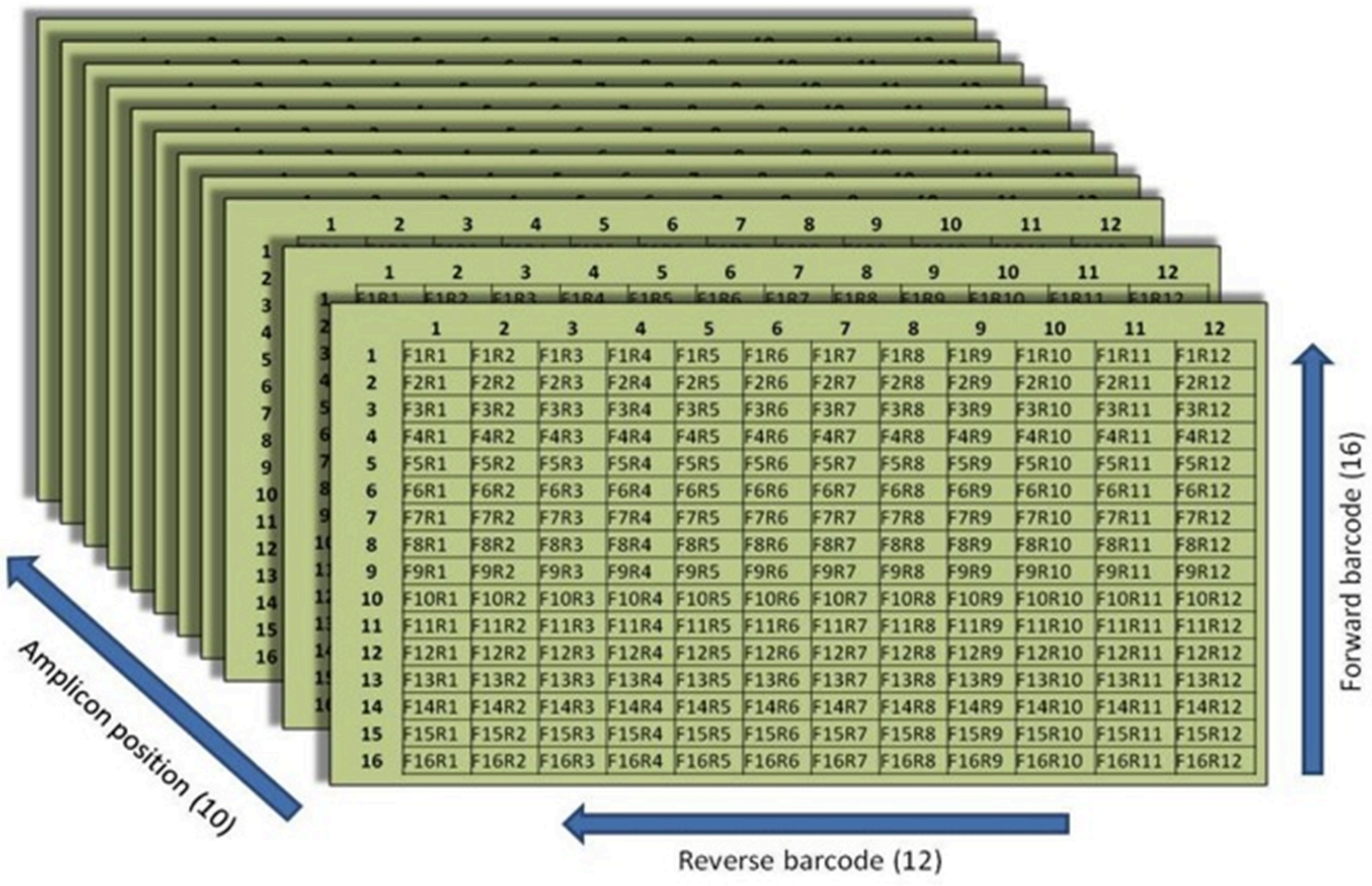
..... = position of the 75 bp forward reads

TGTACA = identifier segment – the position encoding the late ID

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

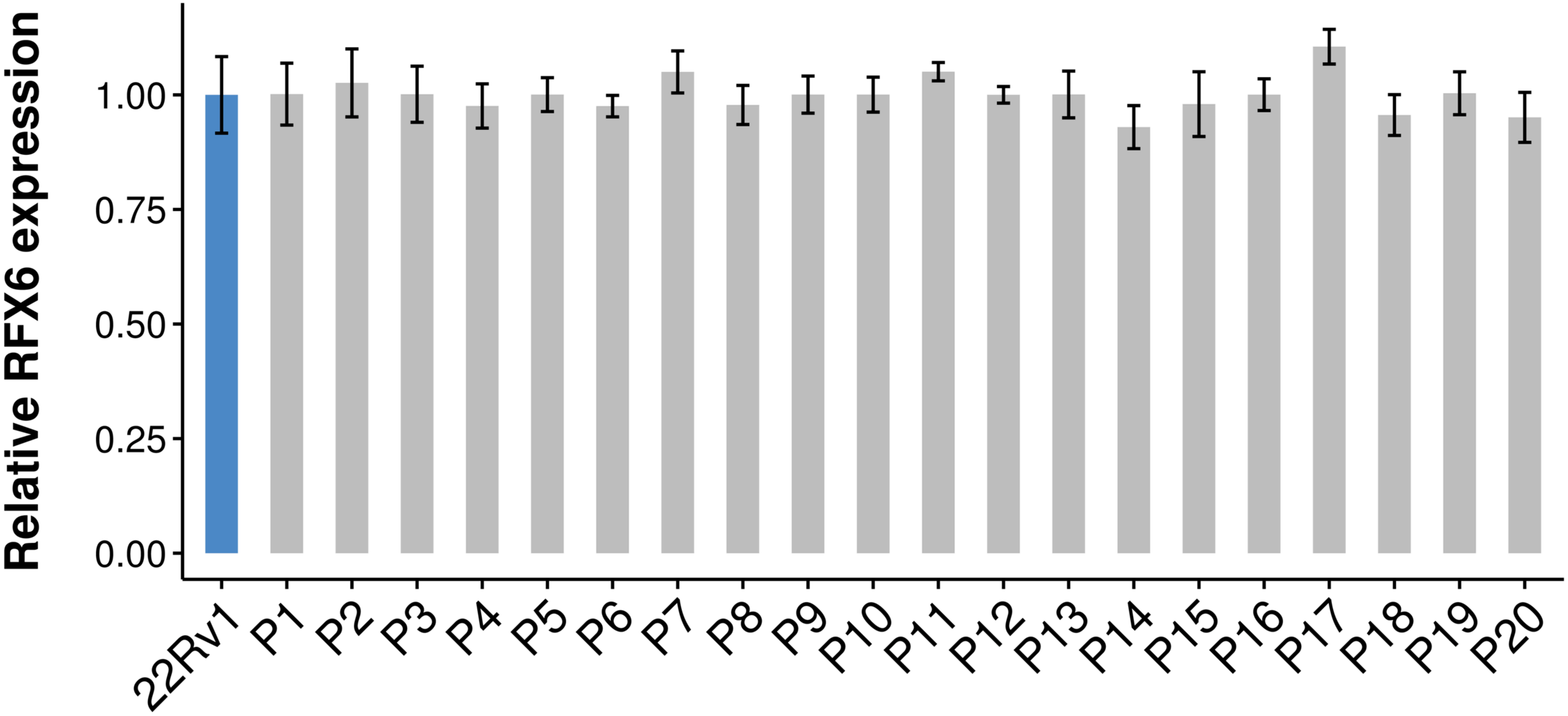
b

16 x 12 x 10 barcode combinations

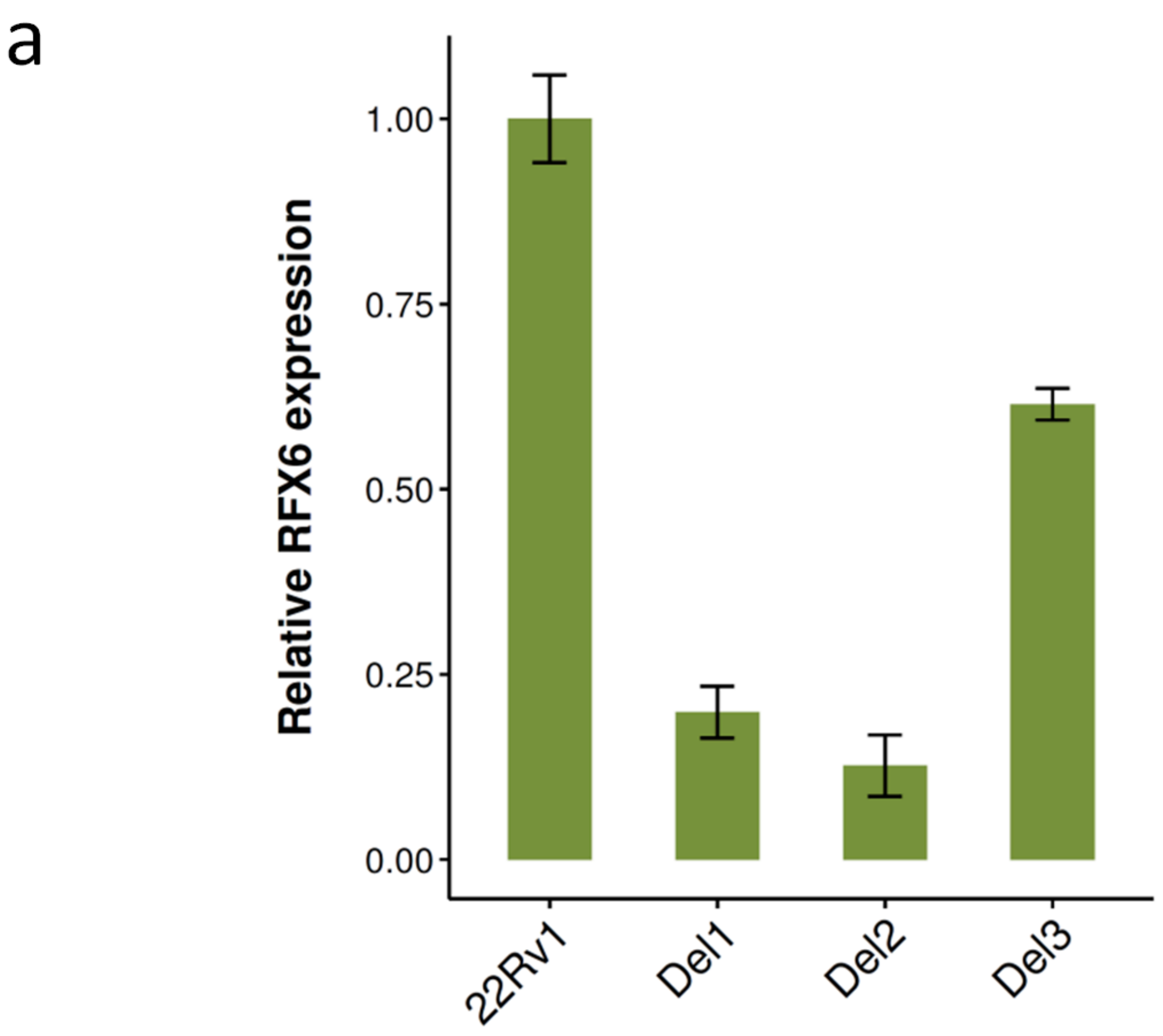


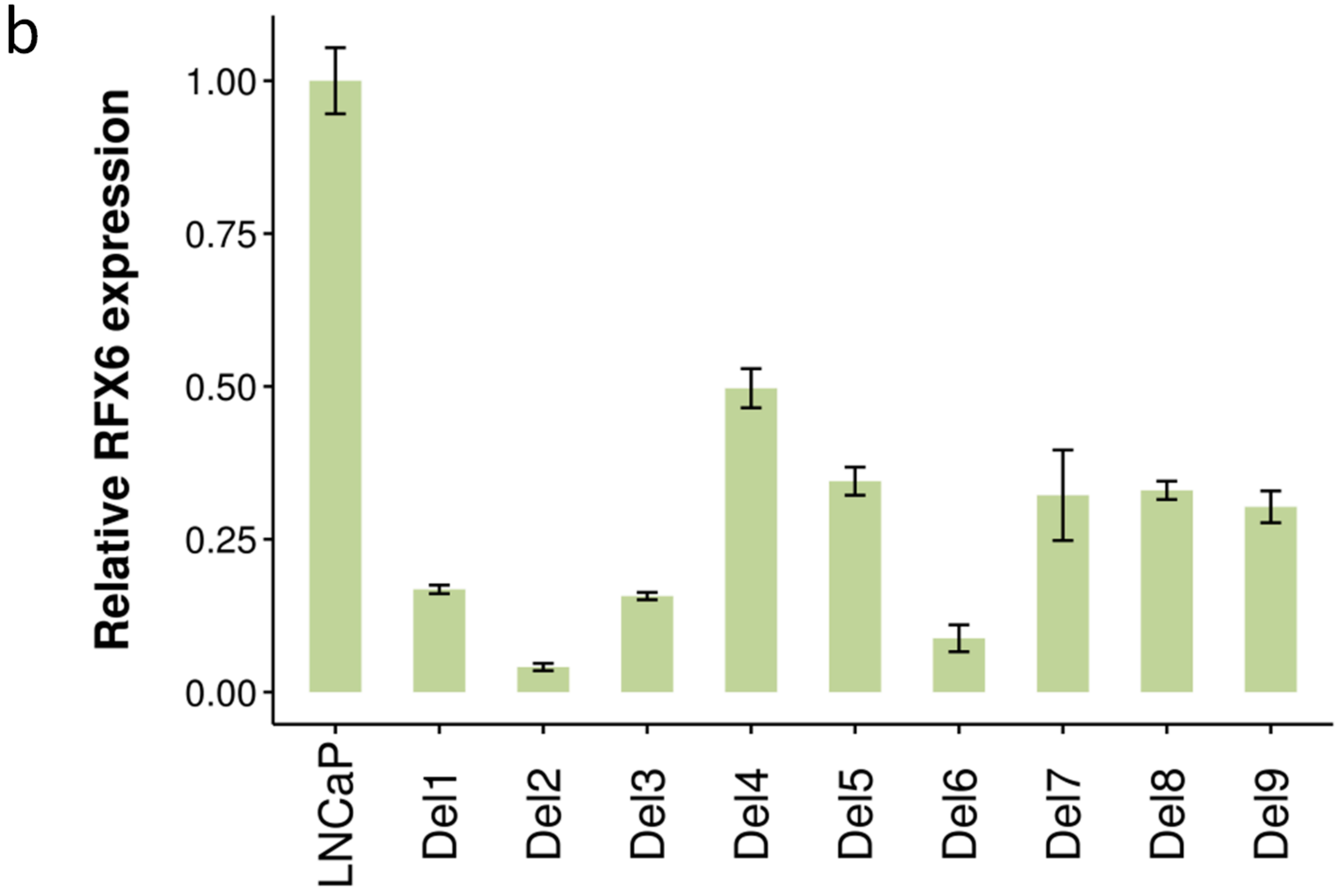
	1	2	3	4	5	6	7	8	9	10	11	12
1	F1R1	F1R2	F1R3	F1R4	F1R5	F1R6	F1R7	F1R8	F1R9	F1R10	F1R11	F1R12
2	F2R1	F2R2	F2R3	F2R4	F2R5	F2R6	F2R7	F2R8	F2R9	F2R10	F2R11	F2R12
3	F3R1	F3R2	F3R3	F3R4	F3R5	F3R6	F3R7	F3R8	F3R9	F3R10	F3R11	F3R12
4	F4R1	F4R2	F4R3	F4R4	F4R5	F4R6	F4R7	F4R8	F4R9	F4R10	F4R11	F4R12
5	F5R1	F5R2	F5R3	F5R4	F5R5	F5R6	F5R7	F5R8	F5R9	F5R10	F5R11	F5R12
6	F6R1	F6R2	F6R3	F6R4	F6R5	F6R6	F6R7	F6R8	F6R9	F6R10	F6R11	F6R12
7	F7R1	F7R2	F7R3	F7R4	F7R5	F7R6	F7R7	F7R8	F7R9	F7R10	F7R11	F7R12
8	F8R1	F8R2	F8R3	F8R4	F8R5	F8R6	F8R7	F8R8	F8R9	F8R10	F8R11	F8R12
9	F9R1	F9R2	F9R3	F9R4	F9R5	F9R6	F9R7	F9R8	F9R9	F9R10	F9R11	F9R12
10	F10R1	F10R2	F10R3	F10R4	F10R5	F10R6	F10R7	F10R8	F10R9	F10R10	F10R11	F10R12
11	F11R1	F11R2	F11R3	F11R4	F11R5	F11R6	F11R7	F11R8	F11R9	F11R10	F11R11	F11R12
12	F12R1	F12R2	F12R3	F12R4	F12R5	F12R6	F12R7	F12R8	F12R9	F12R10	F12R11	F12R12
13	F13R1	F13R2	F13R3	F13R4	F13R5	F13R6	F13R7	F13R8	F13R9	F13R10	F13R11	F13R12
14	F14R1	F14R2	F14R3	F14R4	F14R5	F14R6	F14R7	F14R8	F14R9	F14R10	F14R11	F14R12
15	F15R1	F15R2	F15R3	F15R4	F15R5	F15R6	F15R7	F15R8	F15R9	F15R10	F15R11	F15R12
16	F16R1	F16R2	F16R3	F16R4	F16R5	F16R6	F16R7	F16R8	F16R9	F16R10	F16R11	F16R12

Supplementary Figure 6



Supplementary Figure 7





22Rv1 - parental

```
90                     100                      110                      120
A A G A T A A A C C T C A T G A A A C T G G G G A G A G A G A
A A G A T A A A C C T C A T A A A A C T G G G G A G A G A G A
A A G A T A A A C C T C A T A A A A C T G G G G A G A G A G A
```

Del1

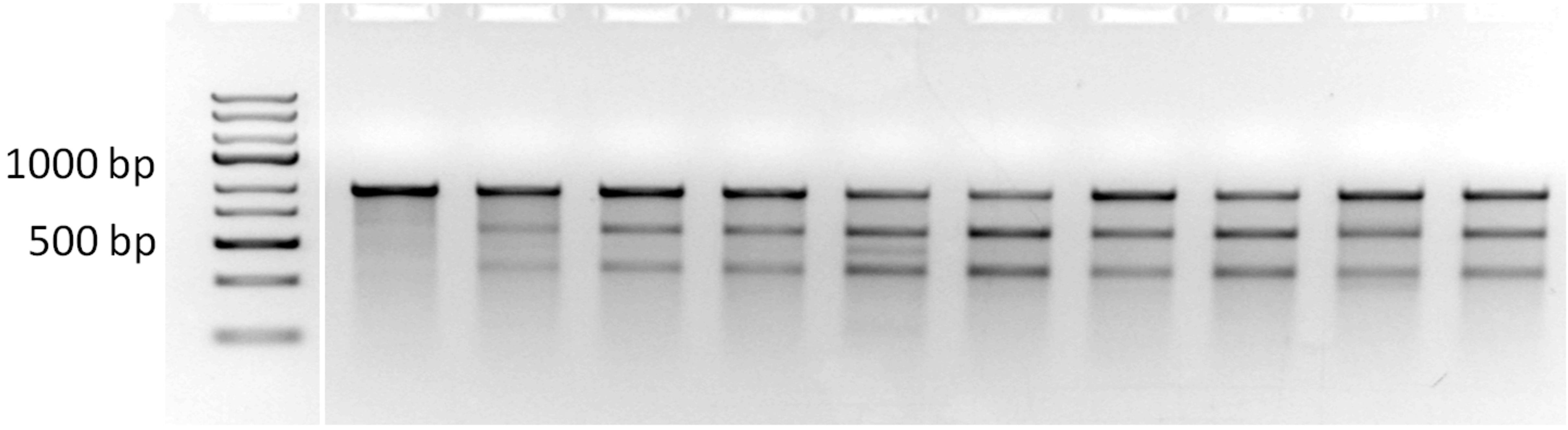
```
90                     100                      110                      120
A A A G A T A A A C C T C A A - - - - - T G G G G A G A G A G
A A A G A T A A A C C T C A - - - - - C T G G G G A G A G A G
A A A G A T A A A C C T C A T A A A A C T G G G G A G A G A G
```

Del2

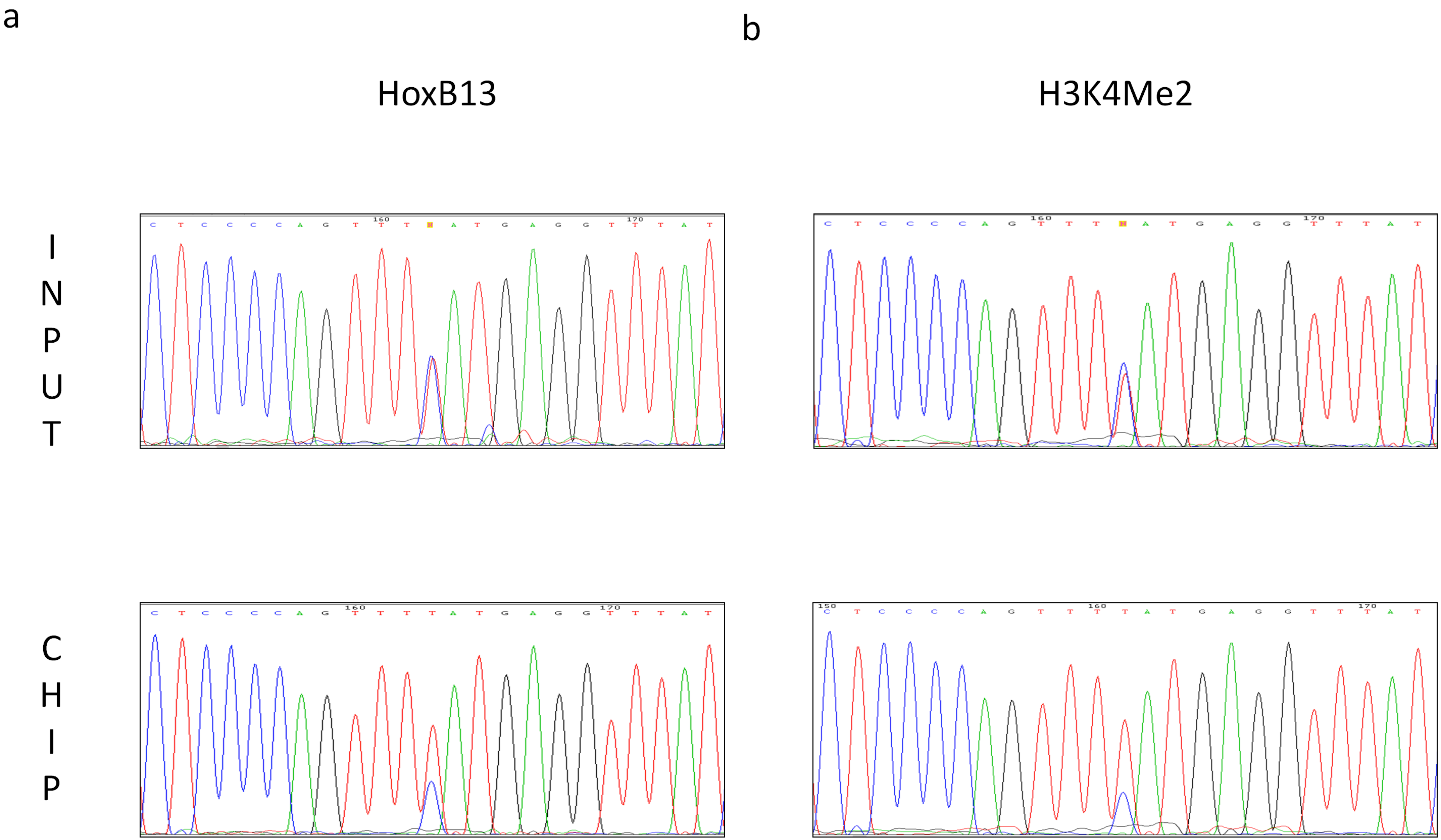
```
90                     100                      110                      120
A A A G A T A A A C C T C - - - - - A A T G G G G A G A G A G
A A A G A T A A A C C T C - - - - - A C T G G G G A G A G A G
A A A G A T A A A C C T C A T A A A A C T G G G G A G A G A G
```

Del3

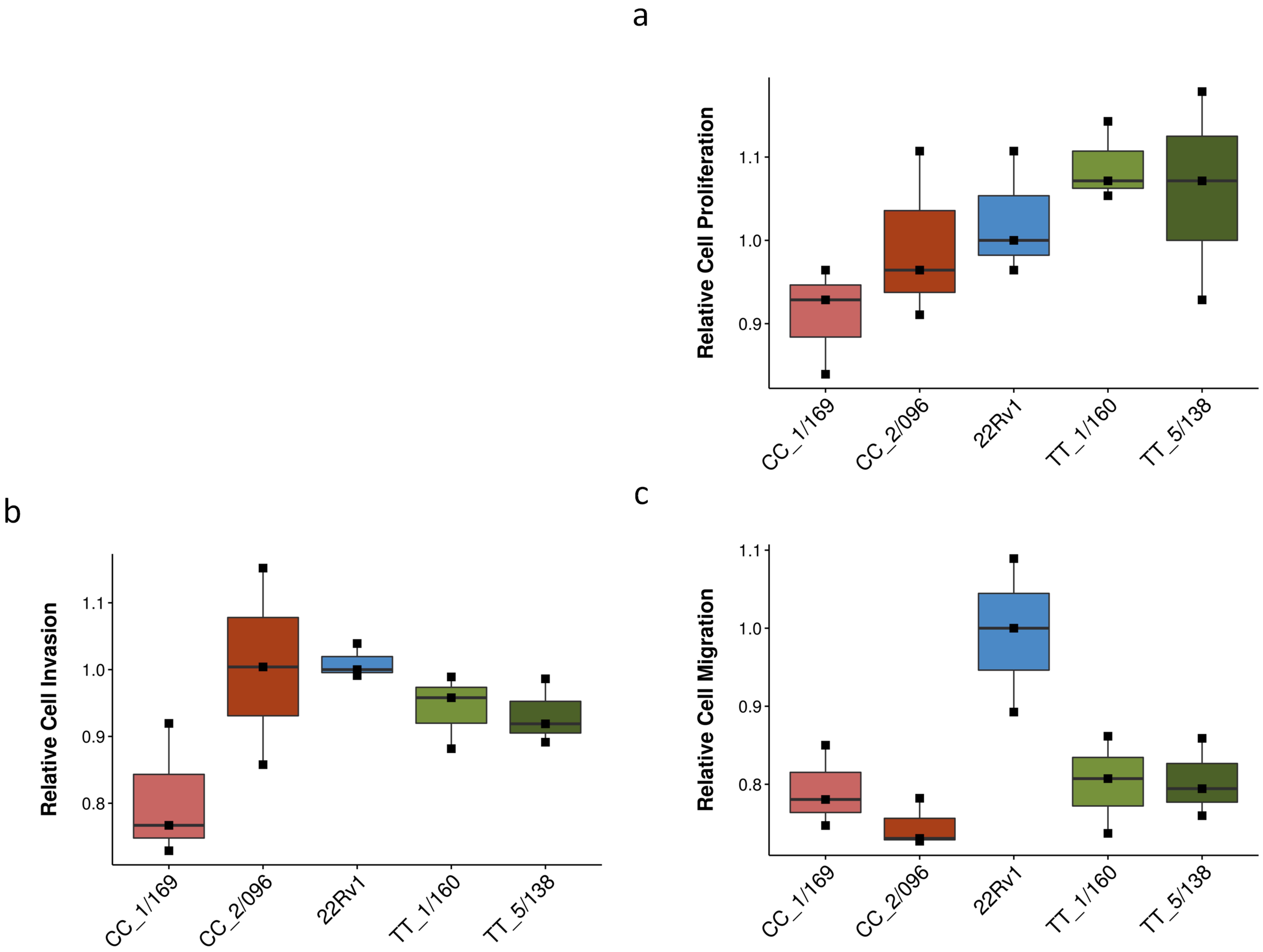
```
90                     100                      110                      120
A G A T A A A C C T - - - - - - - C T G G G G A G A G A G A G
A G A T A A A C C T C - - - - A A C T G G G G A G A G A G A G
A G A T A A A C C T C A T A A A A C T G G G G A G A G A G A G
```



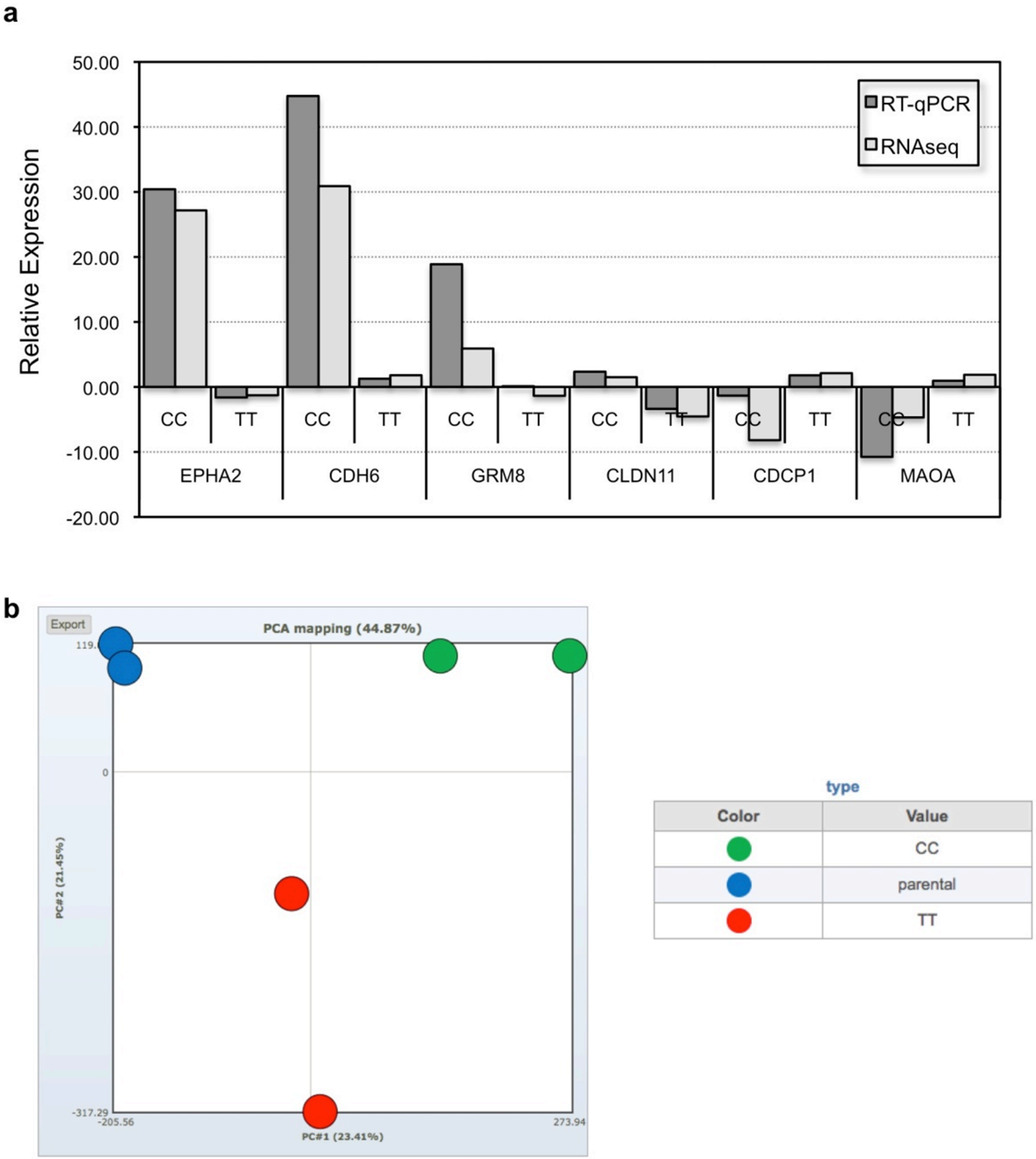
Supplementary Figure 8



Supplementary Figure 9



Supplementary Figure 10



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.