Frequent CHD1 deletions in prostate cancers of African American men is associated with rapid disease progression

Supplementary Note

Authors:

Miklos Diossy, Viktoria Tisza, Hua Li, Pranshu Sahgal, Jia Zhou, Zsofia Sztupinszki, Denise Young, Darryl Nousome, Claire Kuo, Jiji Jiang, Yongmei Chen, Reinhard Ebner, Isabell A. Sesterhenn, Joel T. Moncur, Gregory T. Chesnut, Gyorgy Petrovics, Gregory T. Klus, Gabor Valcz, Pier Vitale Nuzzo, Dezso Ribli, Judit Börcsök, Aurel Prosz, Marcin Krzystanek, Thomas Ried, David Szuts, Kinza Rizwan, Salma Kaochar, Shailja Pathania, Alan DAndrea, Istvan Csabai, Shiv Srivastava, Matthew L Freedman, Albert Dobi, Sandor Spisak, Zoltan Szallasi

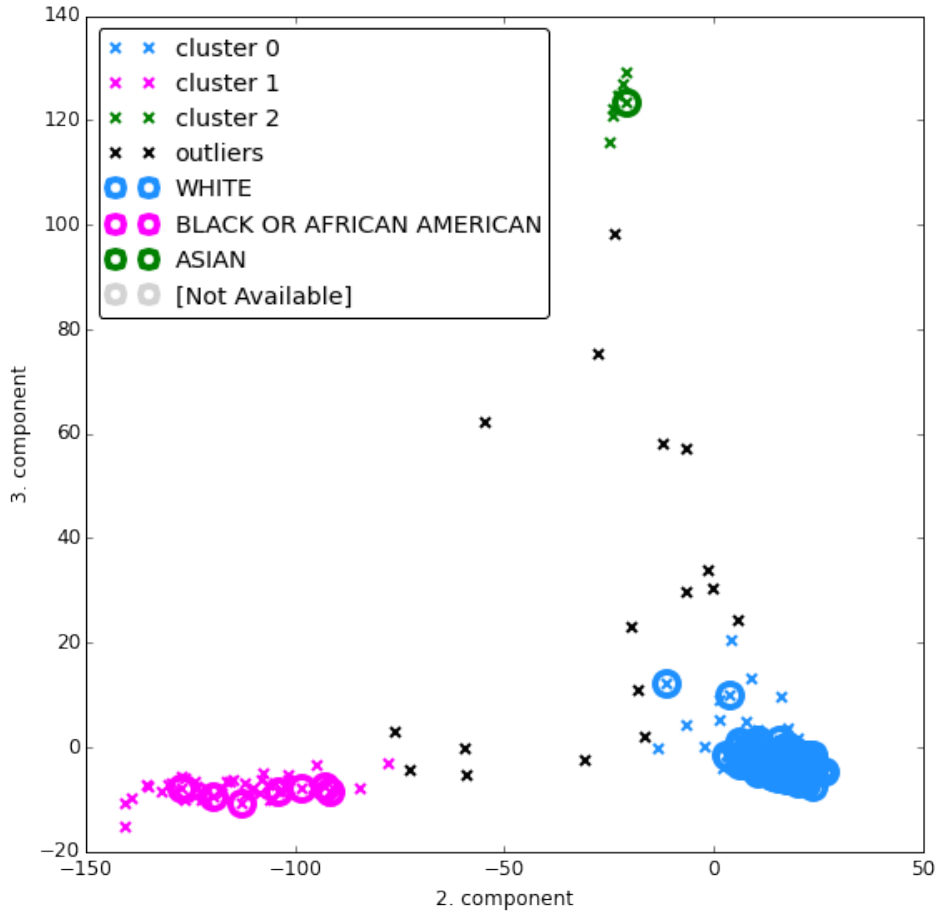
List of Supplementary Figures

List of Supplementary Tables

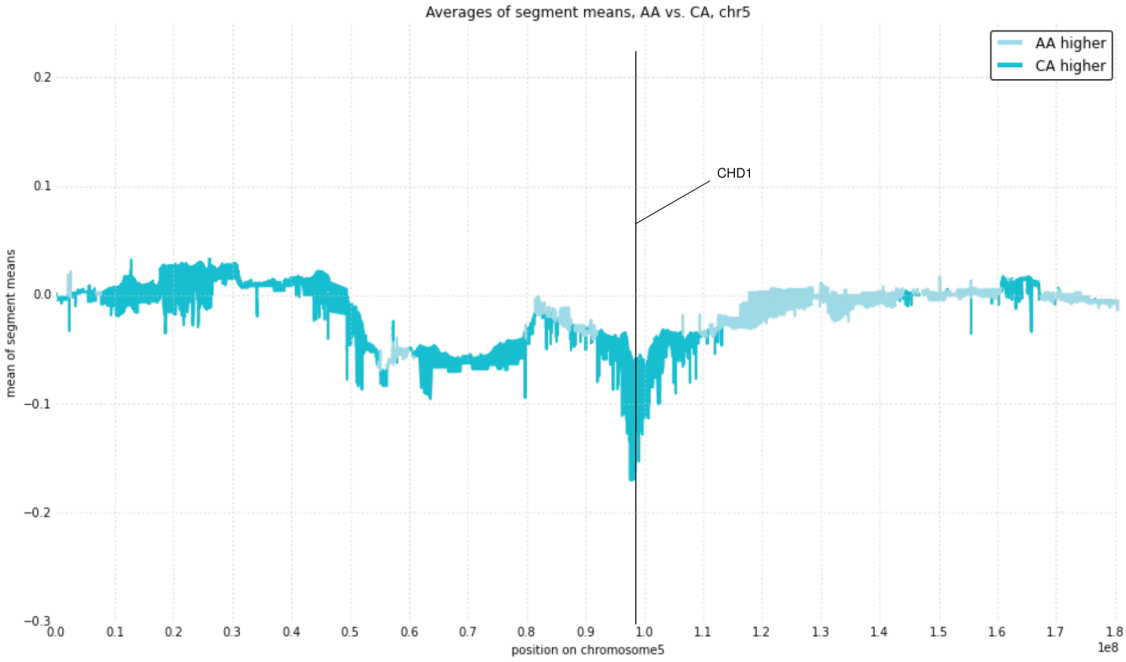
Supplementary DATA

- **Supplementary Data 1**: Subclonal CHD1 loss prediction in the WGS cohorts (available separetely in xlsx format)
- **Supplementary Data 2**: Subclonal CHD1 loss prediction in the WES cohorts (available separately in xlsx format)

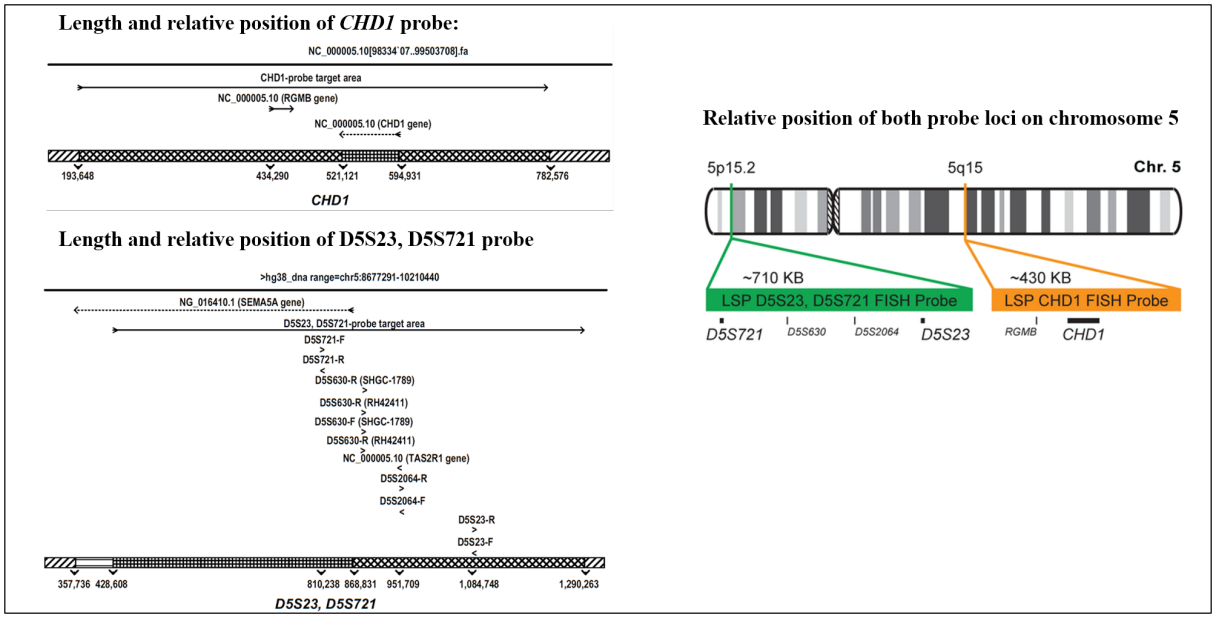
1 Supplementary Material



Supplementary Figure 1: Non-declared ancestry classification on SNP-array data. Circles indicate samples with self declared ancestry, while colored crosses indicate the density-estimated clusters, according to the DBSCAN classification algorithm.



Supplementary Figure 2: The average raw copynumbers of Caucasian (CA) and African American (AA) individuals on chromosome 5. The plotting style is fill between the lines. Darker shades indicate regions where the average copynumber of CA patients was higher, lighter color correspond to regions where AA mean segments were higher. A vertical line indicates the location of CHD1 on the chromosome. The fitted segment mean of the African American samples ($\langle s_{AA} \rangle$ = -0.057, 95% CI: [-0.059, -0.055]) was almost three times lower than of the fitted mean of the European Americans ($\langle s_{CA} \rangle$ = -0.167, 95% CI: [-0.176, -0.158])



Supplementary Figure 3: The design strategy of FISH probes for CHD1 and chromosome 5 short arm.

a

	African American (N=91)	European American (N=109)	P value
Age at Dignosis			0.020
Mean (SD)	59.7 (6.6)	61.9 (6.2)	
Range	44.8-73.2	45.7-74.6	
PSA at Diagnosis (ng/ml)			0.435
<4.0	7 (8.1%)	16 (15.5%)	
4.0-9.0	54 (62.8%)	63 (61.2%)	
10.0-20.0	21 (24.4%)	20 (19.4%)	
>20.0	4 (4.7%)	4 (3.9%)	
Pathological T Stage			0.631
pT2	44 (48.4%)	49 (45.0%)	
pT3-4	47 (51.6%)	60 (55.0%)	
Pathological Gleason Score			0.601
3+3	33 (36.2%)	41 (37.7%)	
3+4	29 (31.9%)	35 (32.1%)	
4+3	10 (11.0%)	7 (6.4%)	
8 to 10	16 (17.6%)	18 (16.5%)	
Treatment	3 (3.3%)	8 (7.3%)	
Grade Group			0.406
GG1-GG2	62 (68.1%)	76 (69.7%)	
GG3	10 (11.0%)	7 (6.4%)	
GG4-GG5	16 (17.6%)	18 (16.5%)	
Treatment	3 (3.3%)	8 (7.3%)	
Margin Status			0.297
Negative	51 (56.0%)	69 (63.3%)	
Positive	40 (44.0%)	40 (36.7%)	
BCR			0.324
No	58 (63.7%)	62 (56.9%)	
Yes	33 (36.3%)	32 (29.4%)	
Metastasis			0.440
No	81 (89.0%)	93 (85.3%)	
Yes	10 (11.0%)	16 (14.7%)	

b

Patient	TMA Block	TMA Row	CORE a	CORE b	CORE c	CORE d	CORE e	CORE f	CORE g	CORE h	CORE i	CORE j	CORE k	CORE l
1	1	1	Normal	Normal	PIN	PIN	T1A	T1A	T2A	T2A	T3A	T3A	T4A	T4A
2	1	2	Normal	Normal	PIN	PIN	T1A	T1A	T1B	T2A	T2A	T3A		
3	1	3	Normal	Normal	T1A	T1A	T1B	T1B						
4	1	4	Normal	Normal	PIN	PIN	T1A	T1A	T1B	T1B	T2A	T2A		
5	1	5	Normal	Normal	PIN	PIN	T1A	T1A	T1B	T1B				
6	1	6	Normal	Normal	PIN	PIN	T1A	T1A	T1B	T1B	T1C	T1C		
7	1	7	Normal	Normal	T1A	T1A	T2A	T2A	T4A	T4A	T7A	T7A		
8	1	8	Normal	Normal	T1A	T1A	T1B	T1B	T2A	T2A	T2B	T2B		
9	1	9	Normal	Normal	PIN	PIN	T1A	T1A	T2A	T2A	T3A	T3A	T4A	T4A

c

Race	Patients	Focal Tumors	Tumor Cores
AA (African American)	91	162	385
EA (European American)	109	189	478

d

	CHD1 Deletion (N=41)	No *CHD1* Deletion (N=159)	P value
Pathologic T Stage			0.043
pT2	13 (31.7%)	81 (50.9%)	
pT3-pT4	28 (68.3%)	78 (49.1%)	
Pathologic Gleason Score			<0.001
3+3	3 (7.3%)	70 (44.0%)	
3+4	17 (41.5%)	47 (29.6%)	
4+3	7 (17.1%)	10 (6.3%)	
8 to 10	11 (26.8%)	25 (15.7%)	
Treatment	3 (7.3%)	7 (4.4%)	
Grade Group			0.024
GG1-GG2	20 (48.8%)	117 (7.36%)	
GG3	7 (17.1%)	10 (6.3%)	
GG4-GG5	11 (26.8%)	25 (15.7%)	
Treatment	3 (7.3%)	7 (4.4%)	

e

Gene Defects	AA (N=42)	EA (N=59)	P value
PTEN status			<0.0001
Deletion	8 (19.0%)	38 (64.4%)	
No-deletion	34 (81.0%)	21 (35.6%)	
ERG status			0.000525
Positive	9 (21.4%)	33 (55.9%)	
Negative	33 (78.6%)	26 (44.1%)	

Suppementary Table 1: **a**: Distributions of clinico-pathological features in combined prostate cancer cohort (N=200).; **b**: The grid map of a representative TMA block construction (PIN=prostatic intraepithelial neoplasia, T1= index tumor; T2=secondary tumor and so on) **c**: Multiple tumor samples of 91 AA and 109 EA patients derived tumor TMA; **d**:CHD1 deletion frequency correlations with pathological T-stage, pathological Gleason score and Grade group in prostate cancer patients (N=200) **e**: Frequency of PTEN deletion and expression of ERG protein in a subset of the cohort (N=101, AA=42 and CA=59).

Supplementary Figure 4: a: The discordance map of CHD1 deletion vs. PTEN deletion and ERG expression in TMA cores carrying CHD1 deletion (purple, AA cases 1 to 12 and 6 EA cases 43-48). Samples from benign tissue are marked in CORE a and b. T1-IN indicates index, T2-NIN marks secondary tumors and so on, in TMA cores c to i. Patient codes are shown in the Case column **b:** The discordance map of CHD1 deletion, PTEN deletion and ERG expression at patient level. Patient codes are shown in the Case column.

a

	Case	Race	CORE a	CORE b	CORE c	CORE d	CORE e	CORE f	CORE g	CORE h	CORE i	CORE j	CORE k	CORE i	BCR	Met
PTEN	1	AA					T1-IN	T1-IN			T2-NIN	T2-NIN			Y	N
ERG	1	AA					T1-IN	T1-IN			T2-NIN	T2-NIN			Y	N
CHD1	1	AA					T1-IN	T1-IN							Y	N
PTEN	2	AA						T1-IN							Y	N
ERG	2	AA													Y	N
CHD1	2	AA					T1-IN		T2-NIN						Y	N
PTEN	3	AA					T1-IN	T1-IN							Y	N
ERG	3	AA													Y	N
CHD1	3	AA					T1-IN	T1-IN			T2-NIN	T2-NIN	T2-NIN		Y	N
PTEN	4	AA					T1-IN	T1-IN	T2-NIN						Y	N
ERG	4	AA													Y	N
CHD1	4	AA									T3-NIN				Y	N
PTEN	5	AA													Y	Y
ERG	5	AA							T1-IN						Y	Y
CHD1	5	AA							T1-IN	T1-IN					Y	Y
PTEN	6	AA													Y	Y
ERG	6	AA													Y	Y
CHD1	6	AA							T1-IN	T1-IN					Y	Y
PTEN	7	AA													Y	N
ERG	7	AA													Y	N
CHD1	7	AA				T1-IN						T1-IN			Y	N
PTEN	8	AA													Y	N
ERG	8	AA													Y	N
CHD1	8	AA								T2-NIN					Y	N
PTEN	9	AA													Y	N
ERG	9	AA													Y	N
CHD1	9	AA					T1-IN	T1-IN							Y	N
PTEN	10	AA													Y	N
ERG	10	AA													Y	N
CHD1	10	AA							T1-IN						Y	N
PTEN	11	AA													Y	N
ERG	11	AA													N	N
CHD1	11	AA						T1-IN	T1-IN	T1-IN					N	N
PTEN	12	AA													N	N
ERG	12	AA													N	N
CHD1	12	AA							T1-IN	T1-IN					N	N

	Case	Race	CORE a	CORE b	CORE c	CORE d	CORE e	CORE f	CORE g	CORE h	CORE i	CORE j	CORE k	CORE i	BCR	Met
PTEN	43	EA				T1-IN	T1-IN	T1-IN	T1-IN	T1-IN	T2-NIN	T2-NIN			Y	Y
ERG	43	EA				T1-IN	T1-IN	T1-IN	T1-IN	T1-IN	T2-NIN	T2-NIN			Y	Y
CHD1	43	EA					T1-IN								Y	Y
PTEN	44	EA			T1-IN	T1-IN		T1-IN		T1-IN					Y	Y
ERG	44	EA			T1-IN	T1-IN	T1-IN	T1-IN	T1-IN	T1-IN					Y	Y
CHD1	44	EA						T1-IN	T1-IN						Y	Y
PTEN	45	EA					T1-IN	T1-IN		T1-IN					Y	N
ERG	45	EA					T1-IN	T1-IN							Y	N
CHD1	45	EA						T1-IN		T1-IN					Y	N
PTEN	46	EA							T2-NIN						N	N
ERG	46	EA				T1-IN									N	N
CHD1	46	EA							T2-NIN	T2-NIN					N	N
PTEN	47	EA					T1-IN	T1-IN	T1-IN	T1-IN	T2-NIN				Y	N
ERG	47	EA													Y	N
CHD1	47	EA							T1-IN						Y	N
PTEN	48	EA													N	N
ERG	48	EA													N	N
CHD1	48	EA			T1-IN		T1-IN	T1-IN	T1-IN						N	N

	PTEN deletion
	ERG expression
	CHD1 deletion
	No defect
	Benign
T1-IN	Index Tumor
T2-NIN	Secondary Tumor
T3-NIN	Tertial Tumor
Y	Yes, BCR
Y	Yes, Met
N	No BCR/Met

b

Case	Race	CHD1	PTEN	ERG	BCR	Met
1	AA				Y	N
2	AA				Y	N
3	AA				Y	N
4	AA				Y	N
5	AA				Y	Y
6	AA				Y	Y
7	AA				Y	N
8	AA				Y	N
9	AA				Y	N
10	AA				Y	N
11	AA				N	N
12	AA				N	N
13	AA				Y	N
14	AA				N	N
15	AA				Y	N
16	AA				N	N
17	AA				Y	Y
18	AA				Y	Y
19	AA				Y	N
20	AA				Y	N
21	AA				Y	N
22	AA				Y	Y
23	AA				Y	Y
24	AA				Y	N
25	AA				Y	N
26	AA				Y	N
27	AA				Y	N
28	AA				Y	N
29	AA				N	N
30	AA				N	N
31	AA				N	N
32	AA				N	N
33	AA				N	N
34	AA				N	N
35	AA				N	N
36	AA				N	N
37	AA				N	N
38	AA				N	N
39	AA				N	N
40	AA				N	N
41	AA				N	N
42	AA				N	N

Case	Race	CHD1	PTEN	ERG	BCR	Met
43	EA				Y	Y
44	EA				Y	Y
45	EA				Y	N
46	EA				N	N
47	EA				Y	N
48	EA				N	N
49	EA				Y	Y
50	EA				Y	Y
51	EA				Y	Y
52	EA				Y	Y
53	EA				Y	N
54	EA				Y	N
55	EA				Y	N
56	EA				Y	N
57	EA				Y	N
58	EA				N	N
59	EA				N	N
60	EA				N	N
61	EA				N	N
62	EA				N	N
63	EA				N	N
64	EA				N	N
65	EA				N	N
66	EA				N	N
67	EA				N	N
68	EA				Y	N
69	EA				Y	N
70	EA				Y	N
71	EA				Y	N
72	EA				Y	N
73	EA				Y	N
74	EA				Y	N
75	EA				Y	N
76	EA				Y	N
77	EA				N	Y
78	EA				N	N
79	EA				N	N
80	EA				N	N
81	EA				N	N
82	EA				Y	N
83	EA				Y	N
84	EA				N	N
85	EA				N	N
86	EA				N	N
87	EA				N	N
88	EA				N	N
89	EA				N	N
90	EA				N	N
91	EA				N	N
92	EA				Y	Y
93	EA				Y	N
94	EA				Y	N
95	EA				Y	N
96	EA				Y	N
97	EA				Y	N
98	EA				Y	N
99	EA				N	N
100	EA				N	N
101	EA				N	N

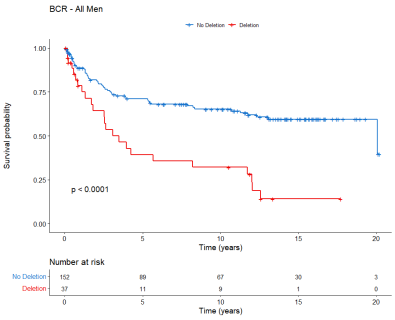
	PTEN deletion
	ERG expression
	CHD1 deletion
	No defect
	Benign
Y	Yes, BCR
Y	Yes, Met
N	No BCR/Met

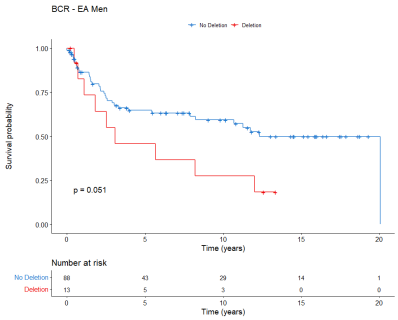
Supplementary Figure 5: **a:** Univariable survival analysis of the clinical features associated with BCR (N=189, excluding 11 patients receiving neo-adjuvant therapy). **b:** Association between CHD1 deletion and BCR in all (AA and EA) and EA only prostate cancer patients. **c:** Univariable survival analysis of the clinical features associated with metastasis (N=189, excluding 11 patients receiving neo-adjuvant therapy). **d:** Association between CHD1 deletion and metastasis in all and EA prostate cancer patients.

a

				Time_RP_BCR		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P value**	**Assumption P value**	**Log-rank P value**
***CHD1* deletion**	**Deletion**	37	2.80 (1.72-4.57)	**<0.001**	0.193	**<0.001**
	No Deletion	152				
Race	**European American**	101	1.38 (0.88-2.19)	0.164	0.285	0.162
	African American	88				
Path T Stage	**pT3-pT4**	104	5.58 (3.23-9.62)	**<0.001**	0.218	**<0.001**
	pT2	85				
GG cat	**GG4-GG5**	40	2.30 (1.38-3.83)	**0.001**	0.421	**<0.001**
	GG1-GG3	149				
Path Gleason	**8 to 10**	34	3.83 (1.92-7.66)	**<0.001**	0.449	**<0.001**
	4+3	17	4.86 (2.29-10.34)	**<0.001**		
	3+4	64	2.41 (1.35-4.30)	**0.003**		
	3+3	74				
Surgical Margin	**Positive**	76	3.80 (2.39-6.04)	**<0.001**	0.313	**<0.001**
	Negative	113				
Age at Diagnosis		189	1.03 (0.99-1.06)	0.142	0.154	
PSA at Diagnosis		178	1.42 (0.99-2.03)	0.059	0.433	

b





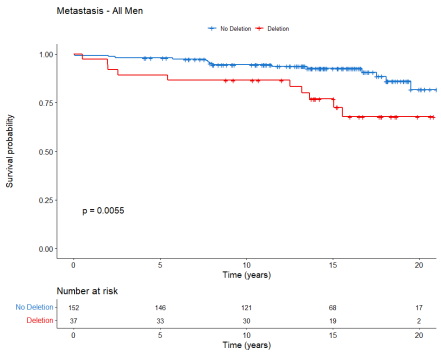
	HR	P value	CI. Lower. HR	CI. upper. HR
***CHD1* Deletion vs No Deletion**	2.1	**0.012**	1.2	3.7
Age at Diagnosis	1.0	0.992	1.0	1.0
Race EA vs AA	1.4	0.194	0.8	2.4
Diagnosis PSA	1.4	0.106	0.9	2.0
Pathological Stage pT3-pT4 vs pT2	3.0	0.003	1.4	6.1
GG4-GG5 vs GG1-GG3	1.3	0.350	0.7	2.4
Surgical Margin (positive vs negative)	1.4	0.261	0.8	2.6

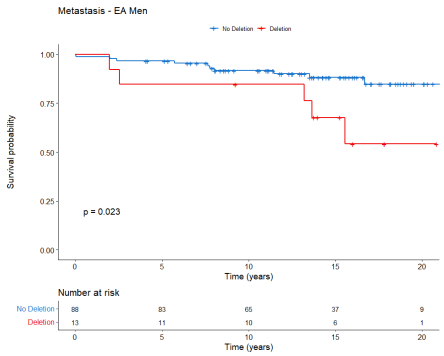
	HR	P value	CI. Lower. HR	CI. upper. HR
***CHD1* Deletion vs No Deletion**	1.9	**0.032**	1.1	3.5
Age at Diagnosis	1.0	0.801	1.0	1.0
Race EA vs AA	1.4	0.260	0.8	2.3
Diagnosis PSA	1.4	0.084	1.0	2.1
Pathological Stage pT3-pT4 vs pT2	3.1	0.002	1.5	6.5
Pathological Gleason 3+4 vs 3+3	1.2	0.629	0.6	2.3
Pathological Gleason 4+3 vs 3+3	2.6	0.029	1.1	5.9
Pathological Gleason 8-10 vs 3+3	2	0.092	0.9	4.3
Surgical Margin (positive vs negative)	1.4	0.314	0.7	2.5

c

				Time_Dx_Mets		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P value**	**Assumption P value**	**Log-rank P value**
***CHD1* deletion**	**Deletion**	37	2.99 (1.33-6.75)	**0.008**	0.466	**0.005**
	No Deletion	152				
Race	**European American**	101	1.45 (0.65-3.22)	0.367	0.199	0.364
	African American	88				
Path T Stage	**pT3-pT4**	104	4.48 (1.53-13.05)	**0.006**	0.405	**0.003**
	pT2	85				
GG cat	**GG4-GG5**	40	3.04 (1.36-6.82)	**0.007**	0.252	**0.005**
	GG1-GG3	149				
Path Gleason	**8 to 10**	34	5.30 (1.83-15.34)	**0.002**	0.286	**0.002**
	4+3	17	1.90 (0.37-9.85)	0.443		
	3+4	64	1.40 (0.44-4.40)	0.57		
	3+3	74				
Surgical Margin	**Positive**	76	2.76 (1.22-6.26)	**0.015**	0.375	**0.011**
	Negative	113				
Age at Diagnosis		189	1.06 (0.99-1.13)	0.094	0.213	
PSA at Diagnosis		178	1.03 (0.54-1.97)	0.923	0.785	

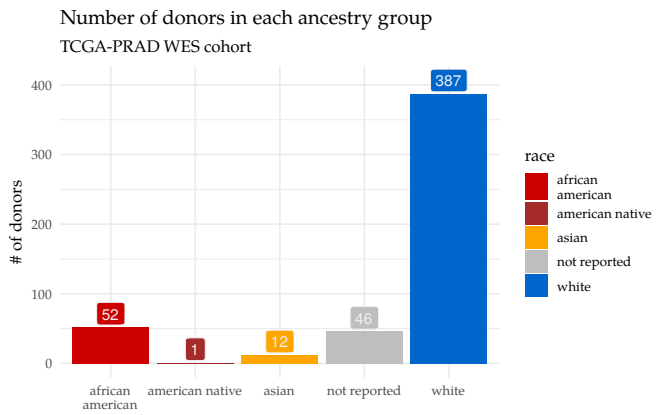
d



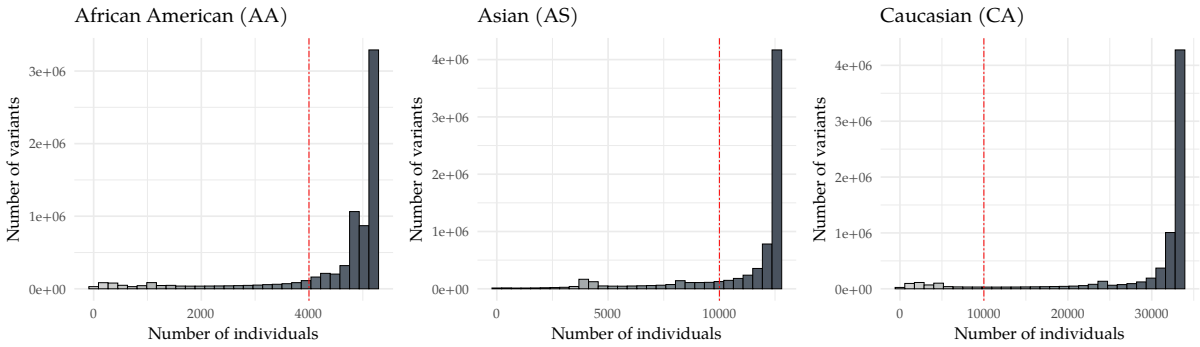


	HR	P value	CI. Lower. HR	CI. upper. HR
***CHD1* Deletion vs No Deletion**	2.8	**0.032**	1.1	7.2
Age at Diagnosis	1.0	0.340	1.0	1.1
Race EA vs AA	1.6	0.372	0.6	4.2
Diagnosis PSA	0.8	0.452	0.4	1.6
Pathological Stage pT3-pT4 vs pT2	1.9	0.328	0.5	7.1
GG4-GG5 vs GG1-GG3	1.6	0.295	0.6	4.3
Surgical Margin (positive vs negative)	2.2	0.157	0.7	6.4

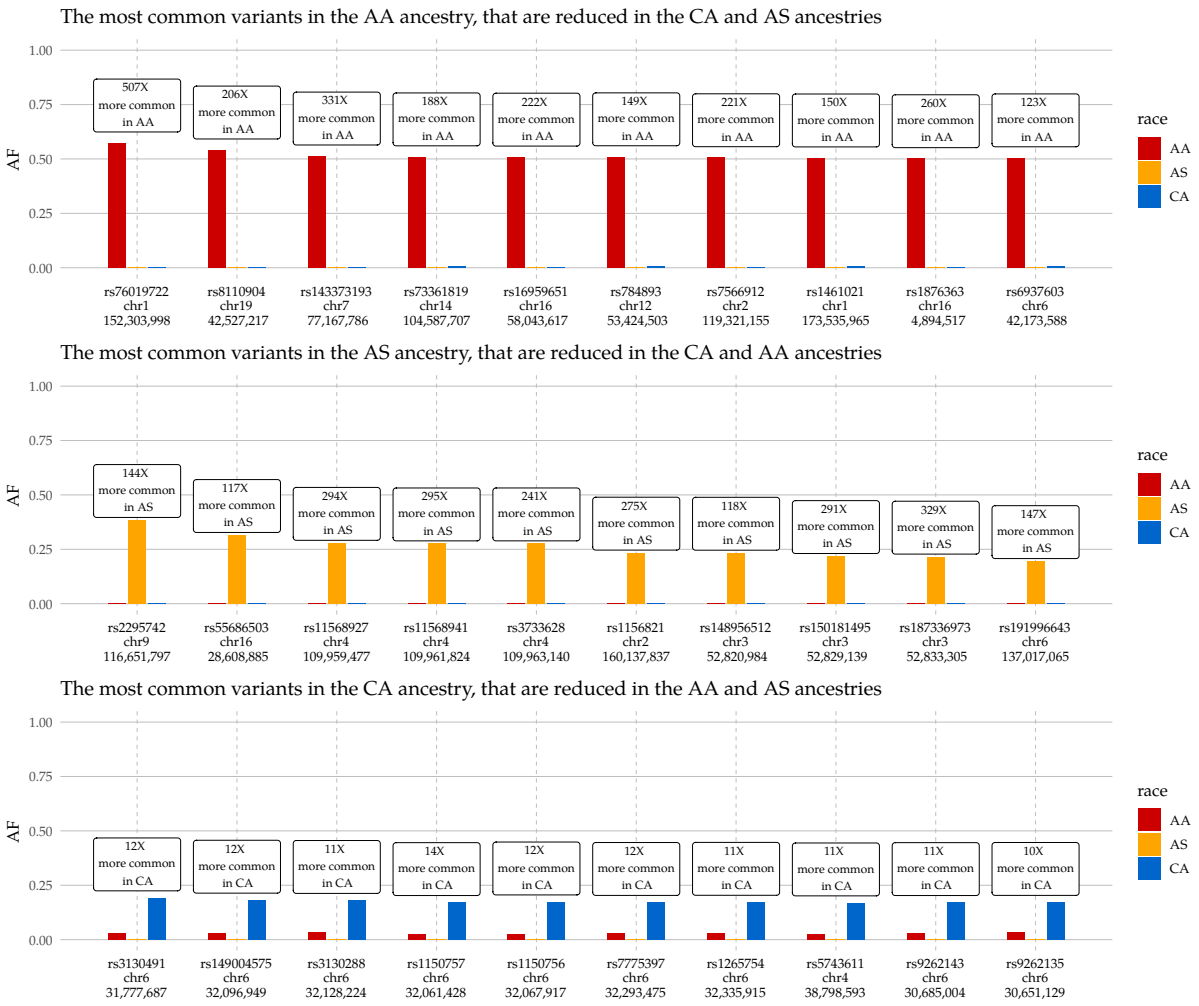
	HR	P value	CI. Lower. HR	CI. upper. HR
***CHD1* Deletion vs No Deletion**	2.6	**0.048**	1.0	6.9
Age at Diagnosis	1.0	0.341	1.0	1.1
Race CA vs AA	1.2	0.671	0.5	3.4
Diagnosis PSA	0.7	0.306	0.3	1.4
Pathological Stage pT3-pT4 vs pT2	1.8	0.380	0.5	6.9
Pathological Gleason 3+4 vs 3+3	0.9	0.849	0.2	3.4
Pathological Gleason 4+3 vs 3+3	0.9	0.911	0.1	5.5
Pathological Gleason 8-10 vs 3+3	2.7	0.125	0.8	10.1
Surgical Margin (positive vs negative)	1.9	0.239	0.6	5.6



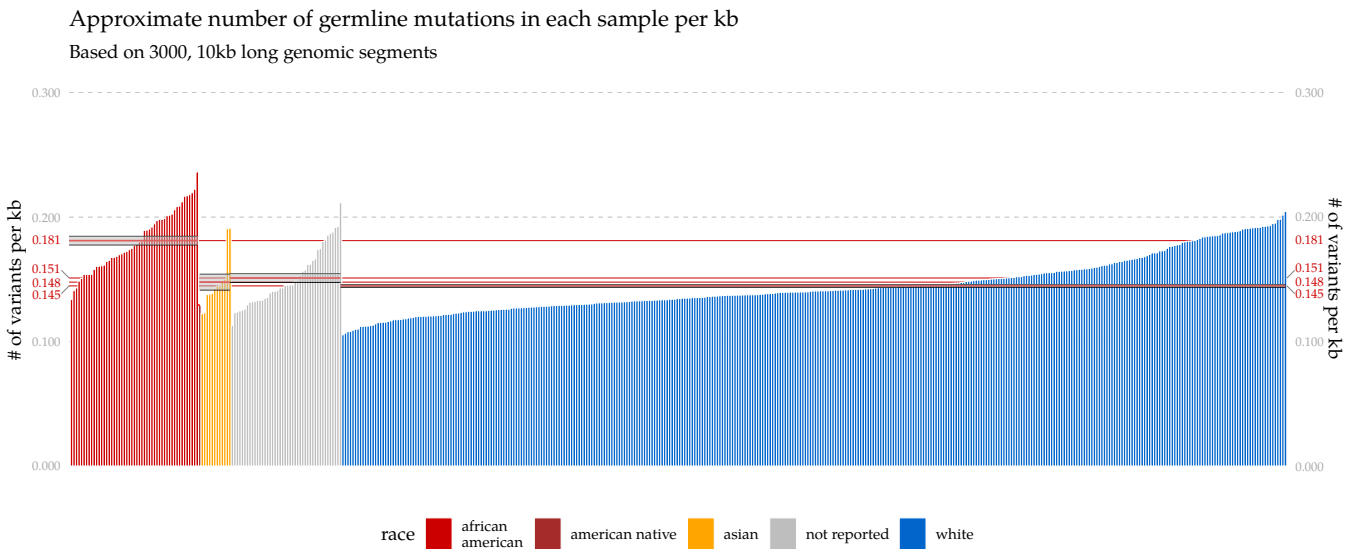
Supplementary Figure 6: Distribution of the self-declared ancestries within the TCGA WES dataset.



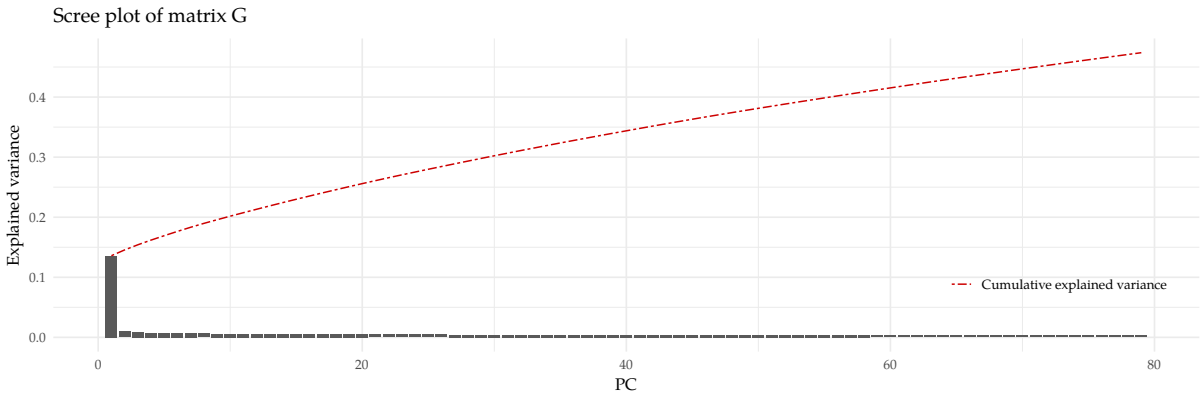
Supplementary Figure 7: Distributions of the number of individuals supporting SNPs in the ExAC database in each ancestry group. The red dashed vertical lines represent a threshold, below which the variants were excluded from the search.



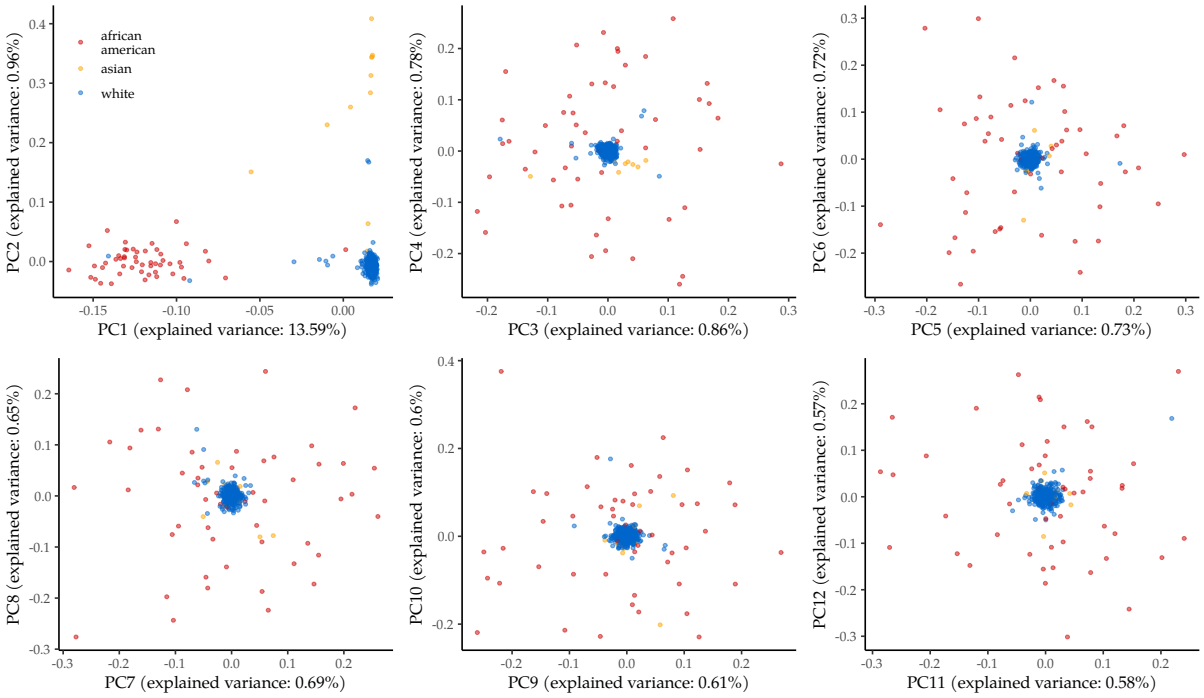
Supplementary Figure 8: The ten most ancestry-specific SNPs of the three most frequent WES ancestry groups. The relative commonnes of each SNP was compared to the mean of the population allele frequencies (AF) of the other two ancestries.



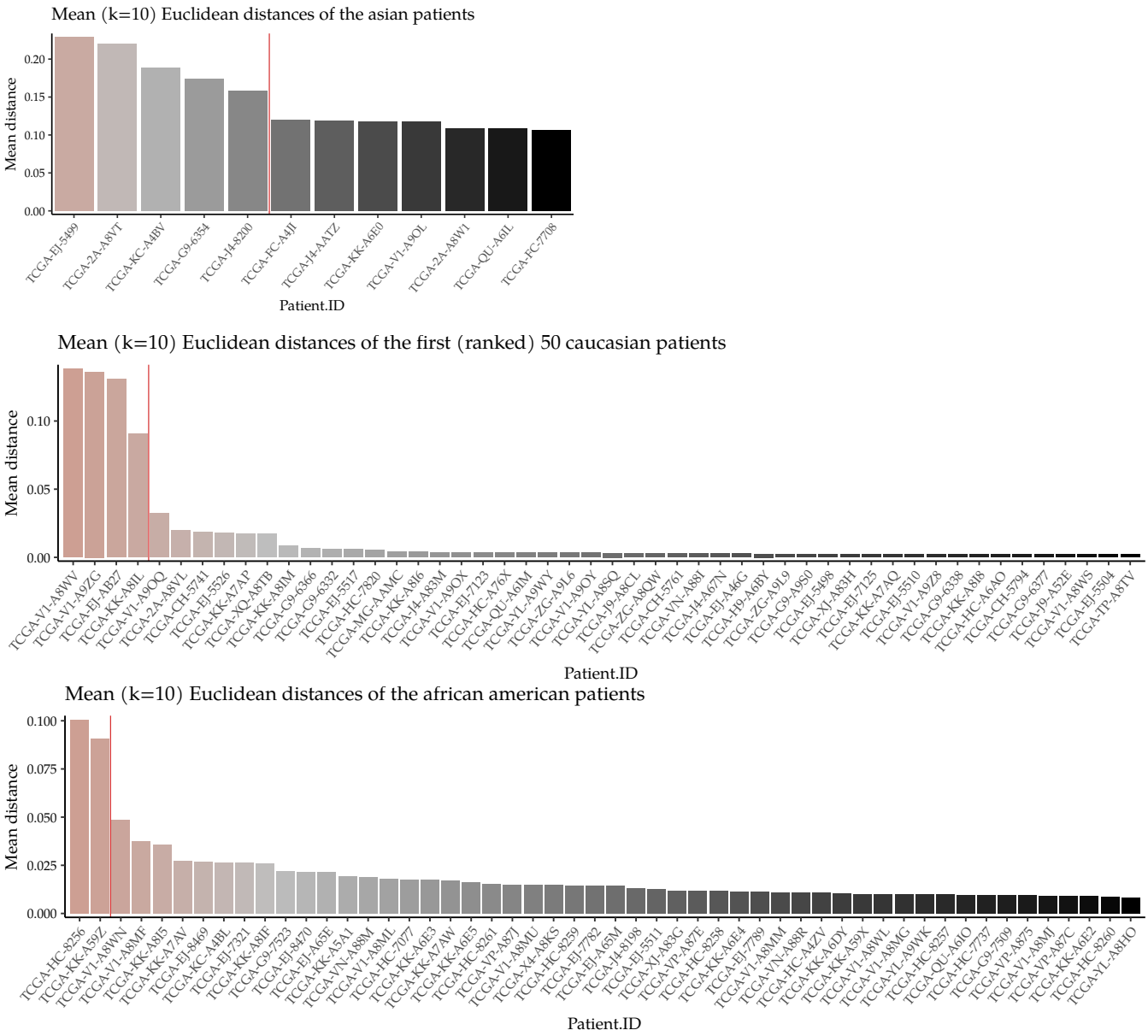
Supplementary Figure 9: Frequencies of the exonic germline mutations within the PRAD-US whole exomes, grouped by the self-reported or not-reported ancestries.



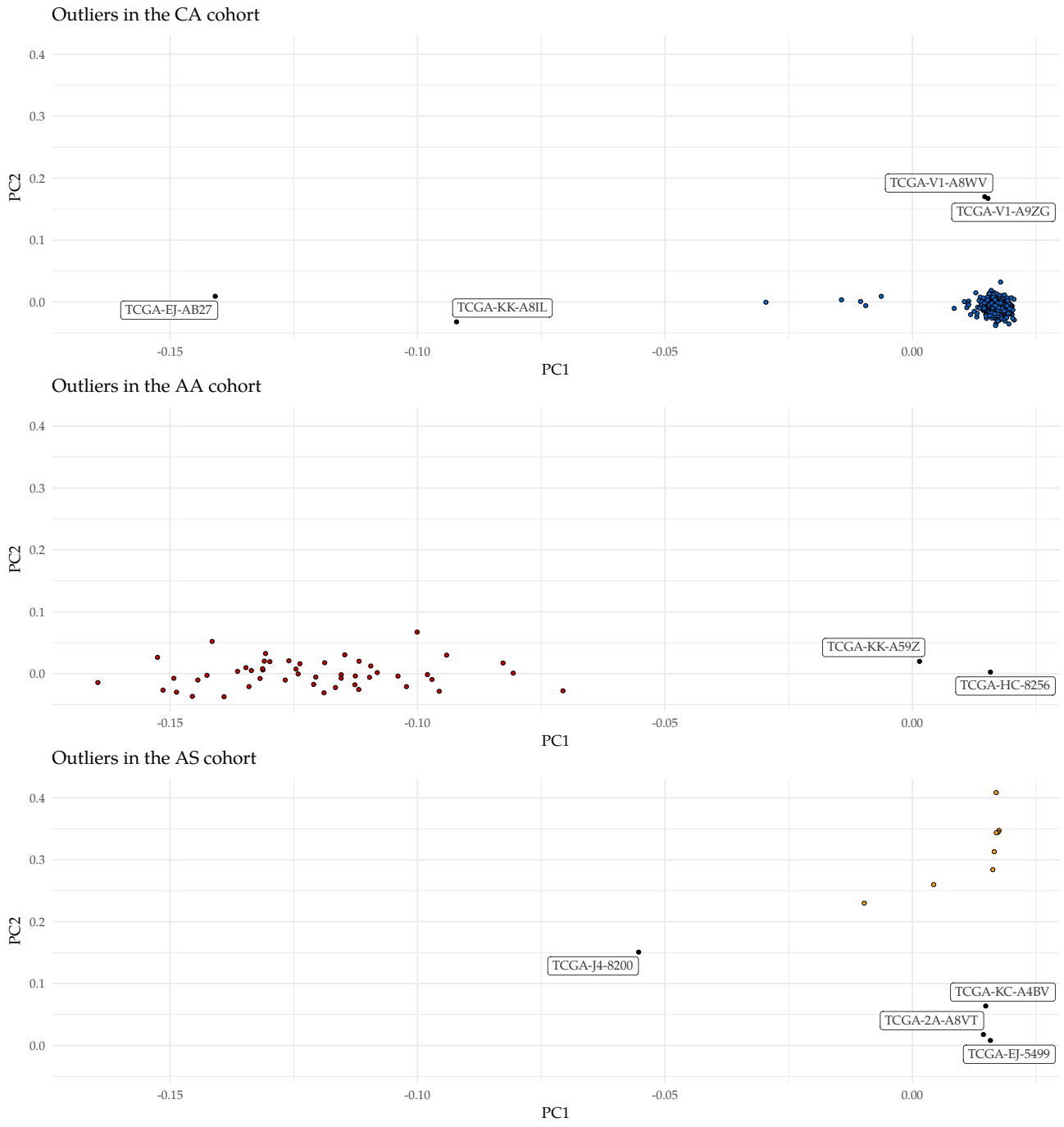
Supplementary Figure 10: Scree plot of the ancestry-genotype matrix **G**



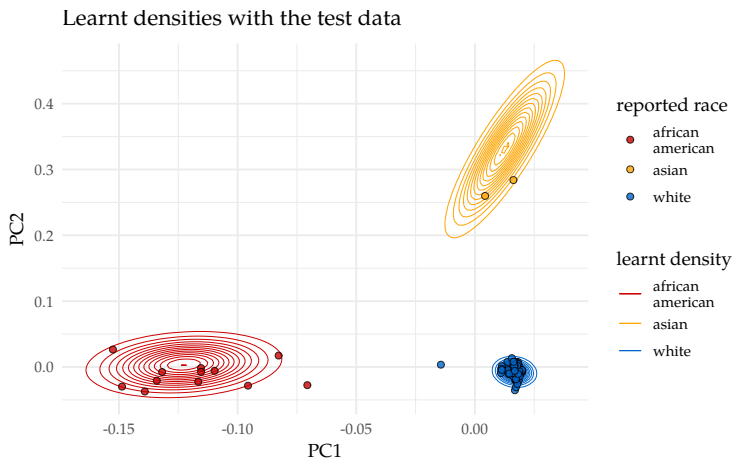
Supplementary Figure 11: Projection of the ancestry genotype matrix onto its first few principal components.



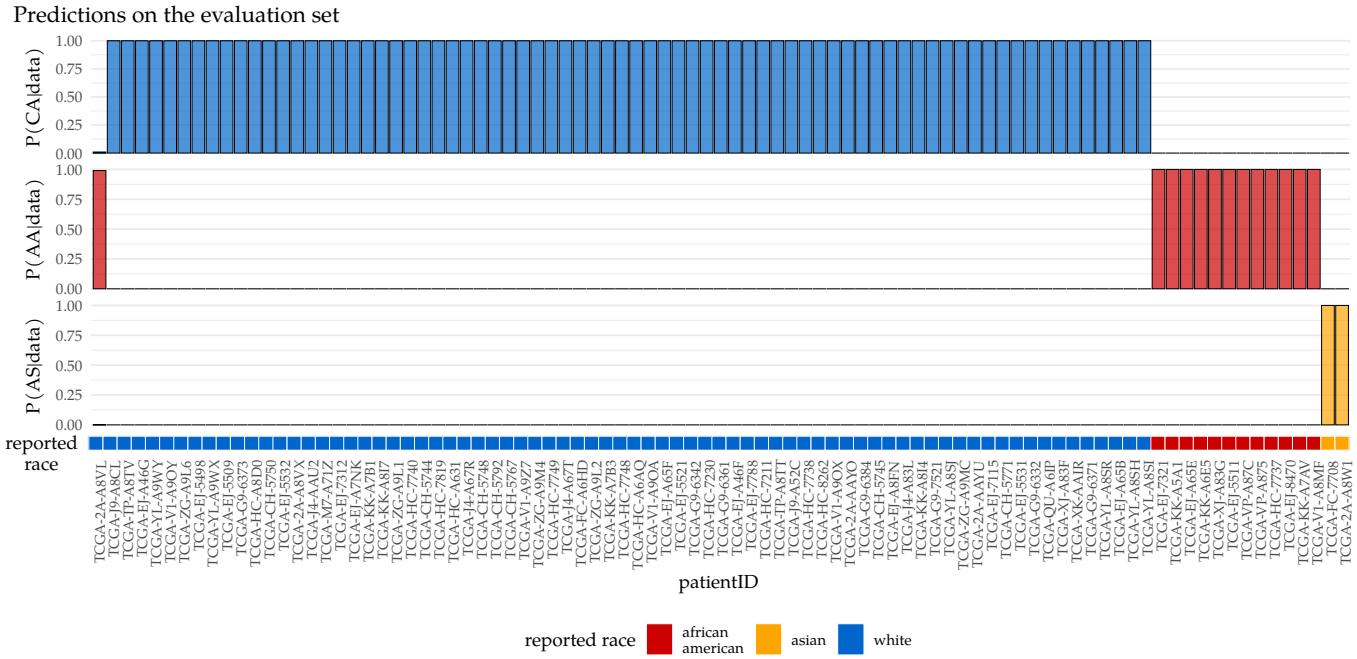
Supplementary Figure 12: Euclidean mean distances in the PC1-PC2 space. Large mean distances indicate, that the sample is located further from its cluster than the rest of the members of that group. Red vertical lines indicate the thresholds that were used for outlier detection. The ancestry of samples left from the threshold were reclassified as "outliers". These samples were excluded from the training process introduced in the next section.



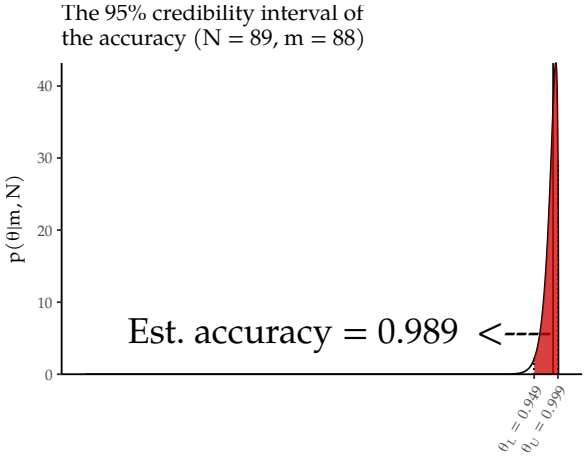
Supplementary Figure 13: Outliers of the three ancestries. From top panel to the bottom, European American, African American, and Asian American ancestries. Samples that have not deviated from their main cluster significantly are colored according to the color-code of their ancestry, while clear outliers are colored black. The TCGA submitter IDs of these samples are also indicated on the figure.



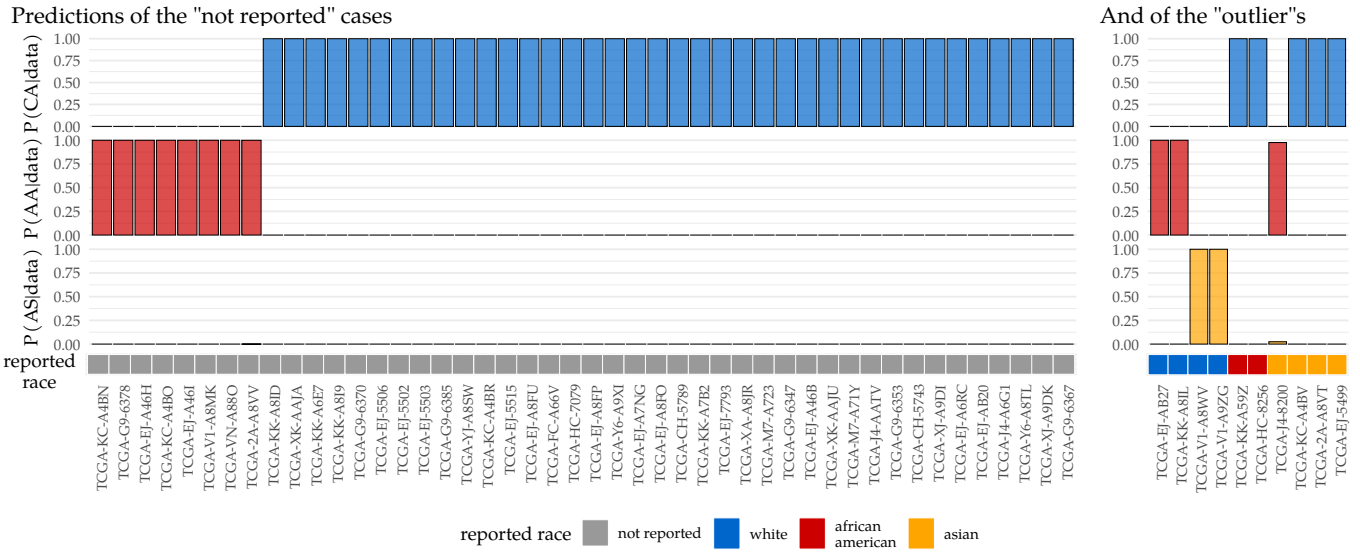
Supplementary Figure 14: Learnt distributions of the ancestries in the PC1-PC2 space of the genotype matrix. Samples of the test set are also represented in the figure, colored by their original, self-declared ancestries.



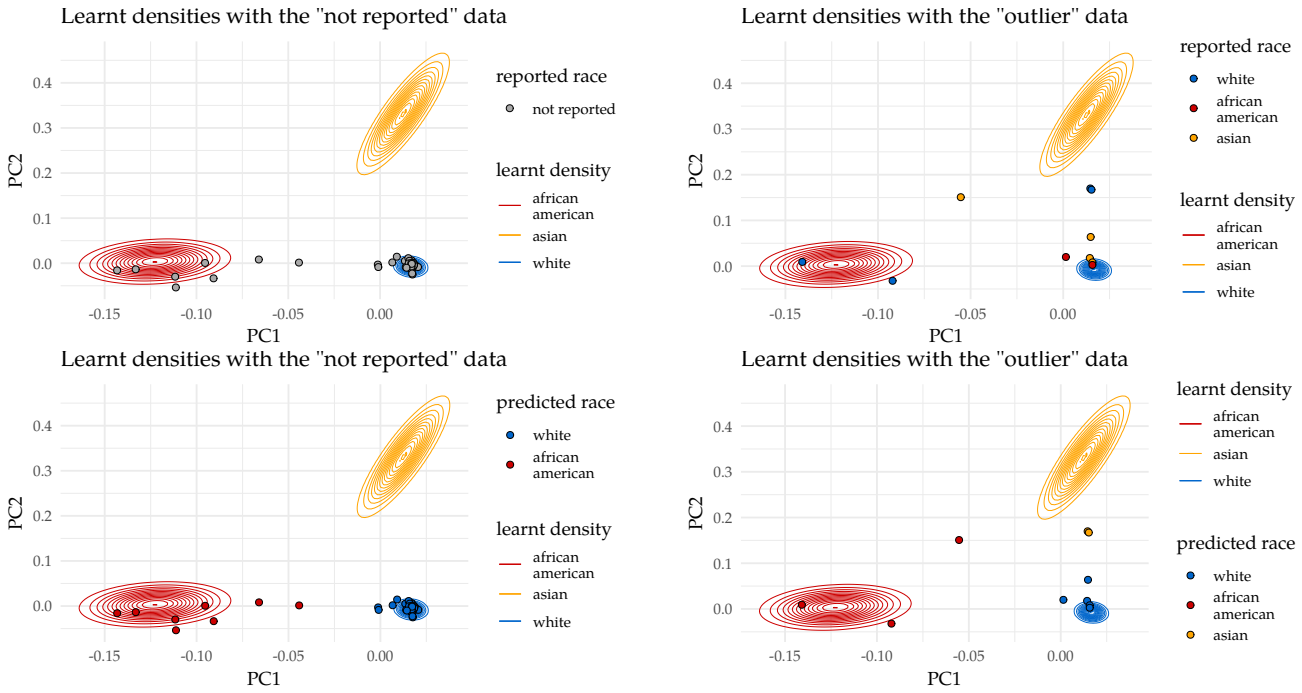
Supplementary Figure 15: Predicted probabilities of samples of the test set. The reported, i.e. self-declared ancestries are also illustrated under the bar plots.



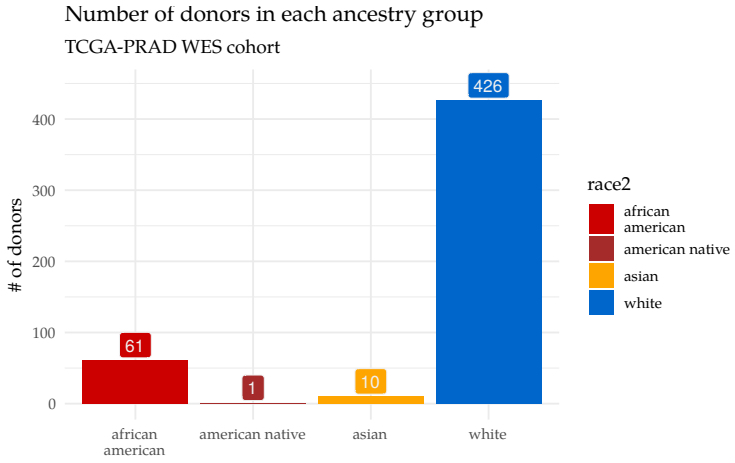
Supplementary Figure 16: Estimated Accuracy (θ) of the Bayes classifier. The 95% credibility interval was estimated using a binomial distribution of $N = 89$ trials and $m = 88$ successes.



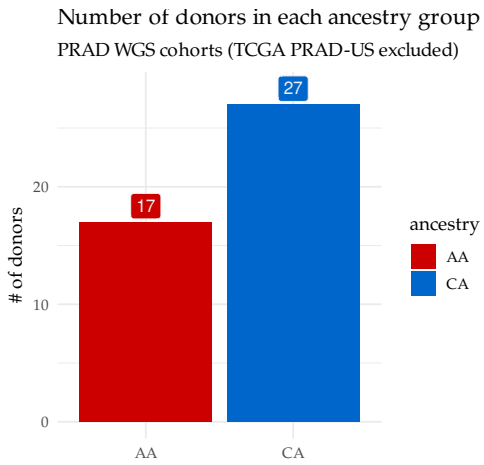
Supplementary Figure 17: Predicted probabilities of samples of the "outlier" and "not reported" samples. The reported, i.e. self-declared ancestries are also illustrated under the bar plots.



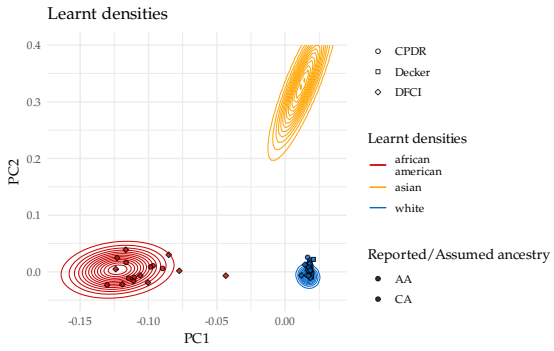
Supplementary Figure 18: Learnt distributions of the ancestries in the PC1-PC2 space of the genotype matrix. Samples of the "not reported" and "outlier" sets are also represented in the figures, in the top two figures they are colored according to their original, self-declared ancestries, while in the bottom two figures they are colored according to their reclassified ancestries.



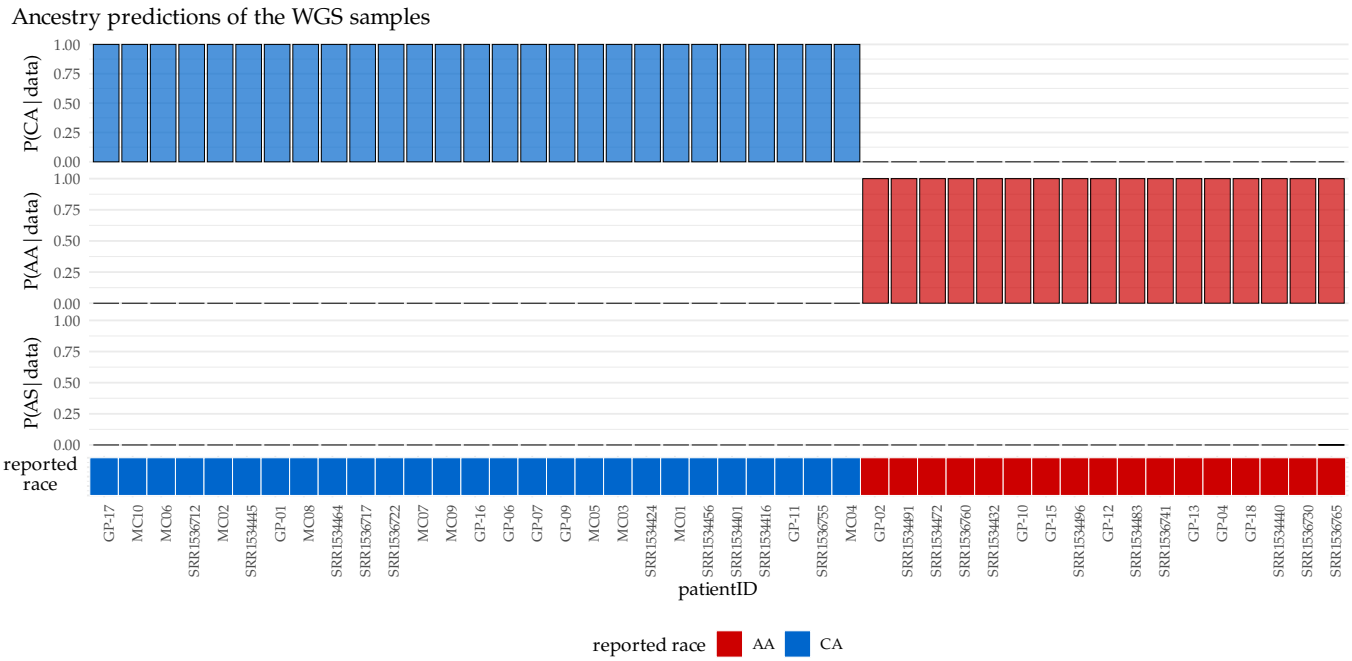
Supplementary Figure 19: Final distribution of the validated ancestries in the TCGA WES dataset



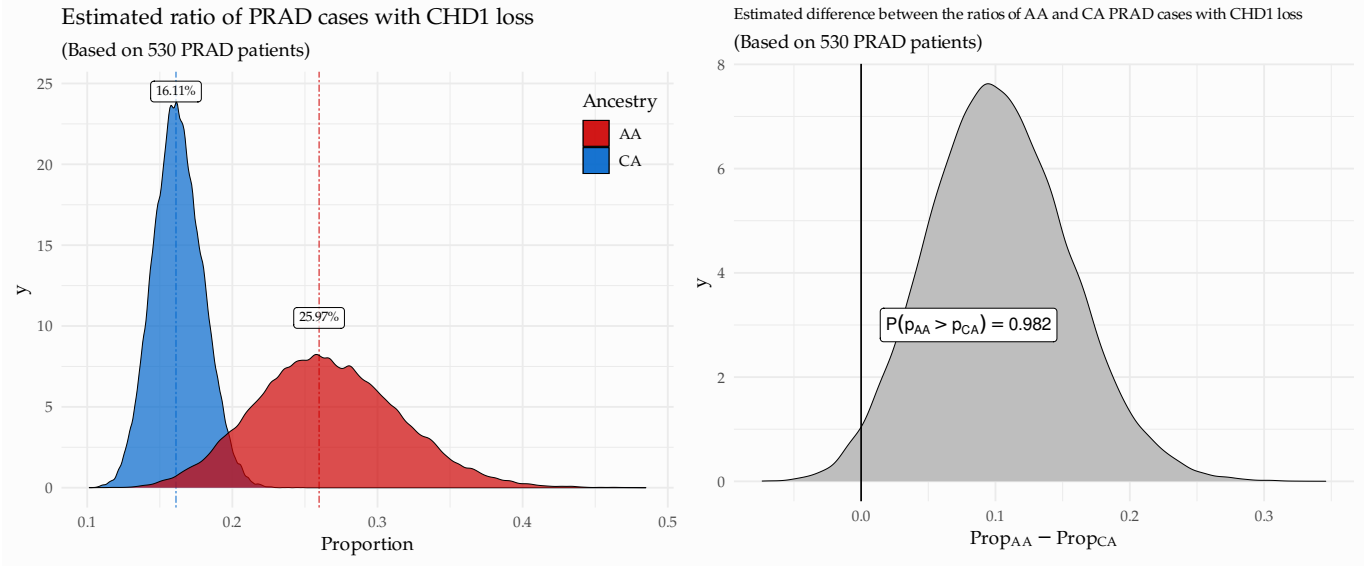
Supplementary Figure 20: Number of patients with AA/CA self-declared ancestries among the PRAD WGS samples. As the 20 TCGA donors have been analyzed in the whole exome section, the were not involved in the WGS section.



Supplementary Figure 21: Distribution of the self-declared ancestries of the PRAD WGS samples. As the 20 TCGA donors have been already analyzed in the whole exome section, they were not included in the WGS section.

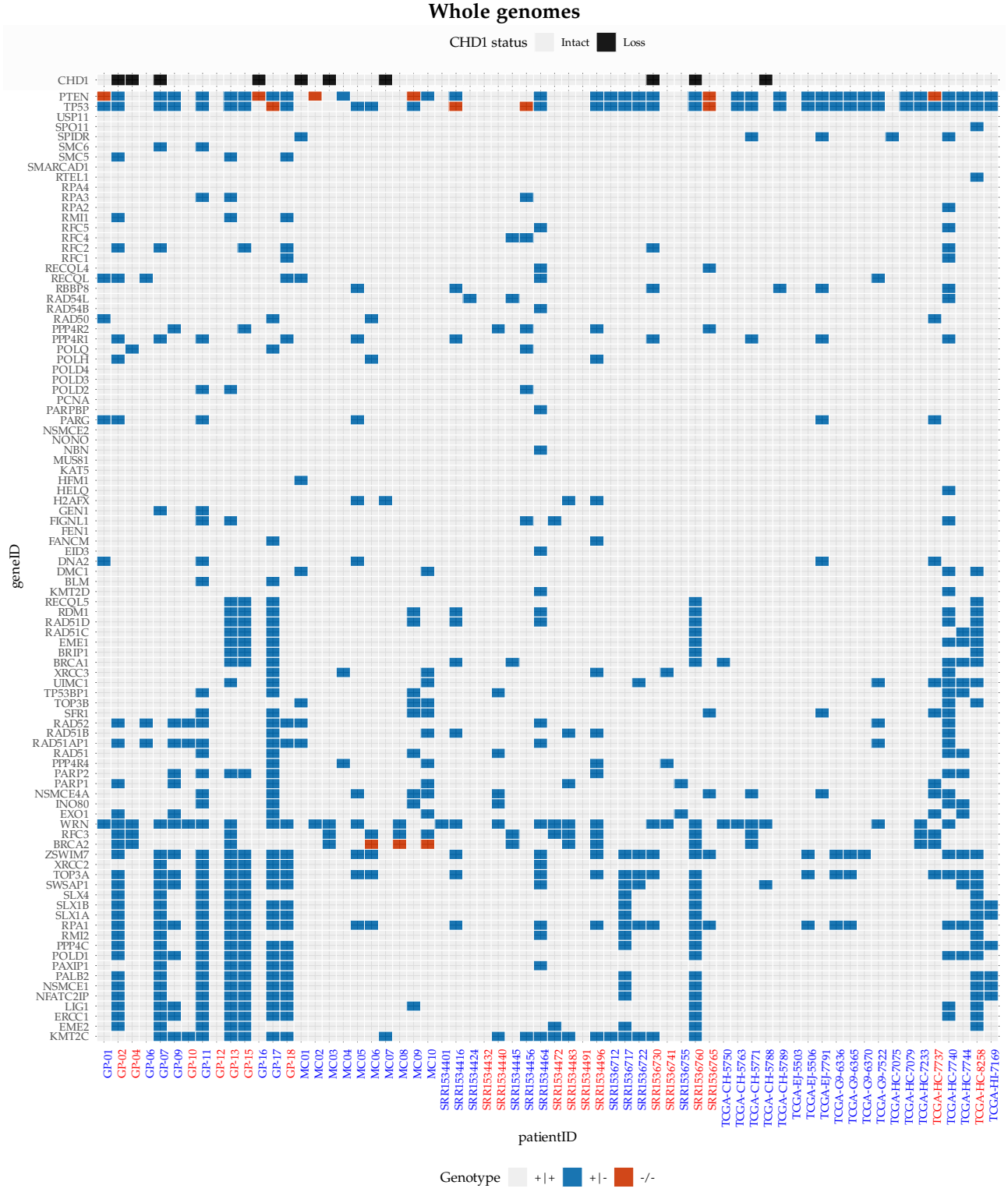


Supplementary Figure 22: Predicted probabilities of samples of the "outlier" and "not reported" samples. The reported, i.e. self-declared ancestries are also illustrated under the bar plots.

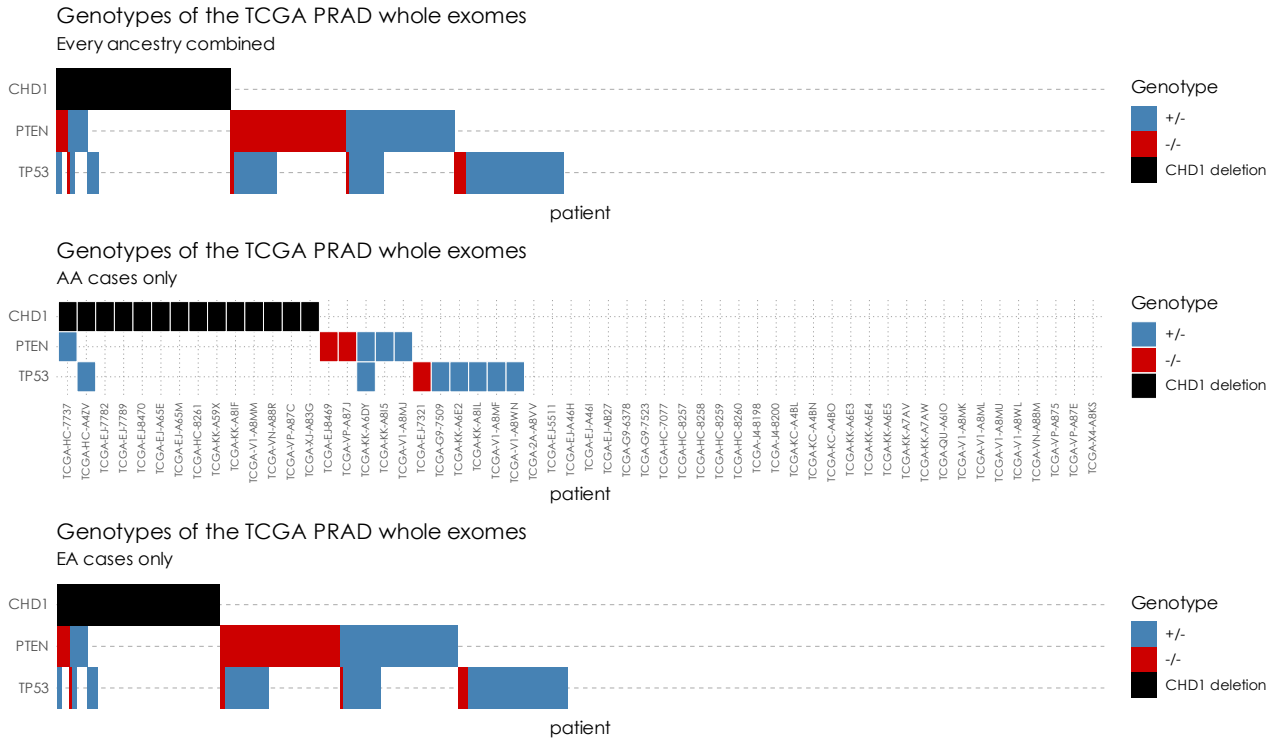


Supplementary Figure 23: On the left: The estimated distributions of the prevalence of CHD1 loss cases in AA and CA PRAD cases. The figure illustrates the likely range of the true ratio of CHD1-loss in AA and CA tumors in the greater population.
On the right: The difference of the two posteriors $Prop_{AA} - PropCA$.

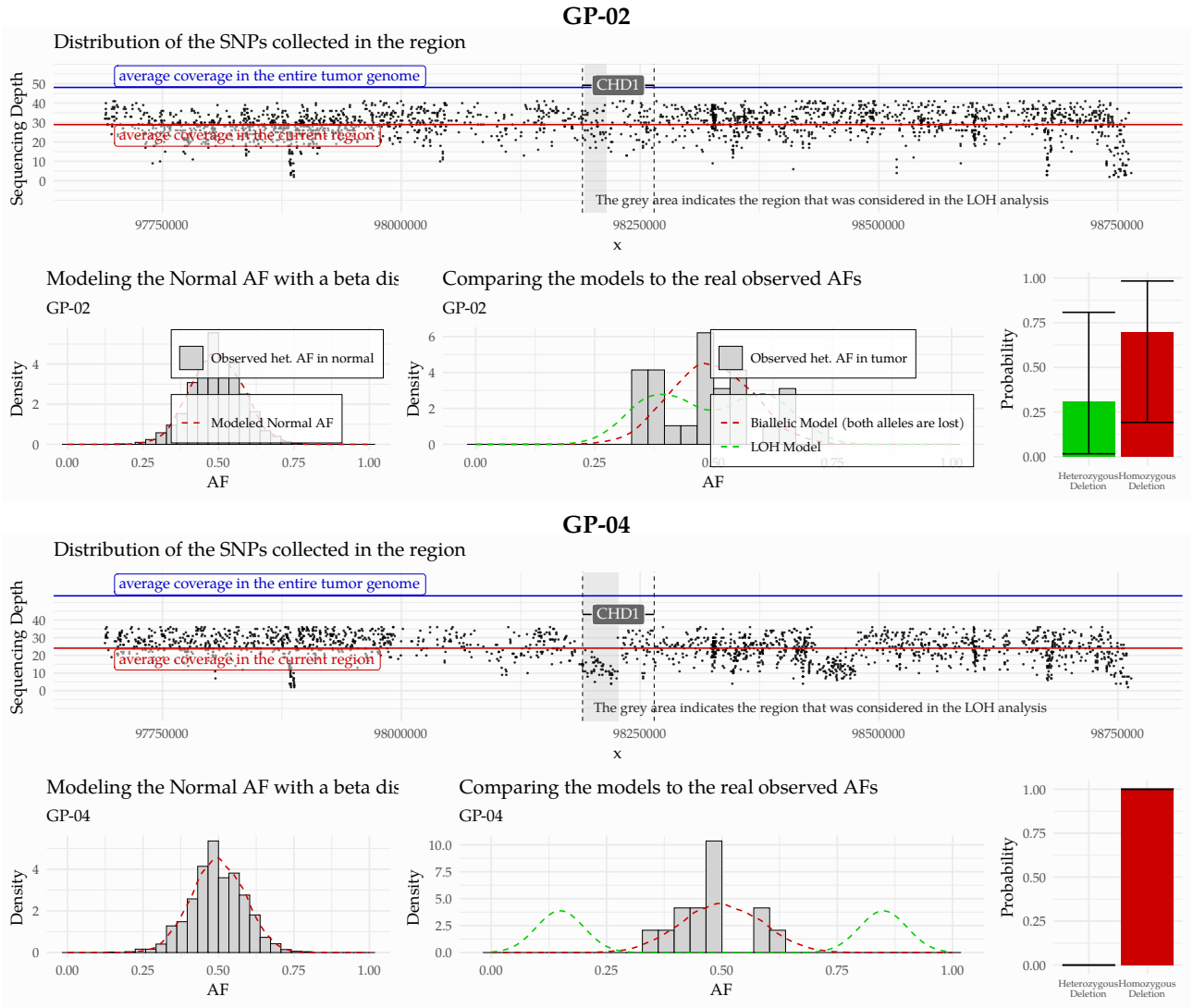
Supplementary Figure 24: Genotyping of genes involved in DNA repair in the NGS dataset.



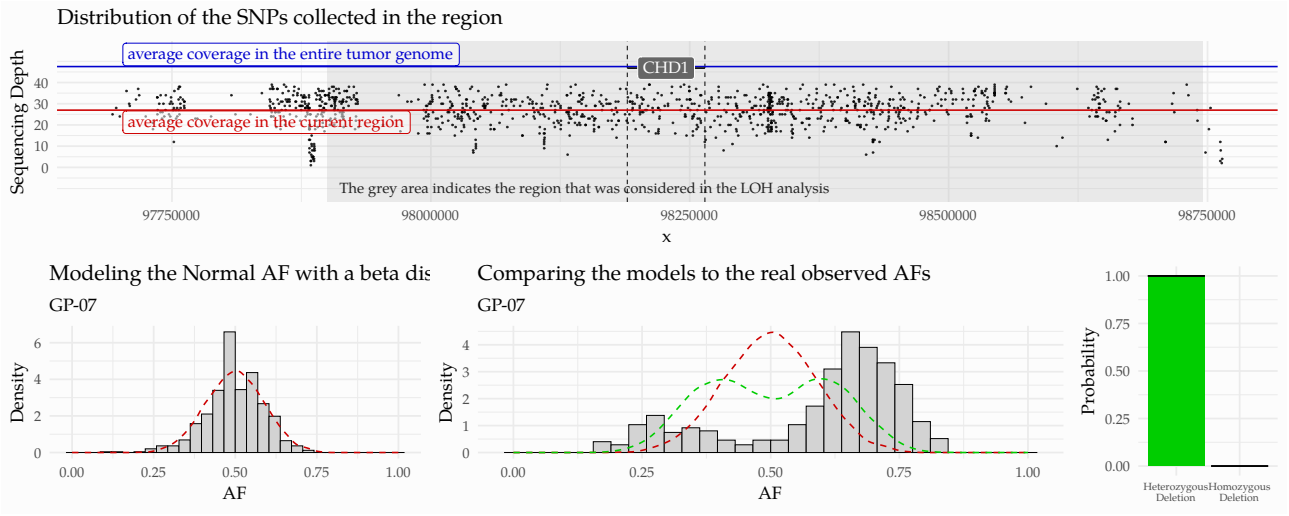
TCGA whole exomes



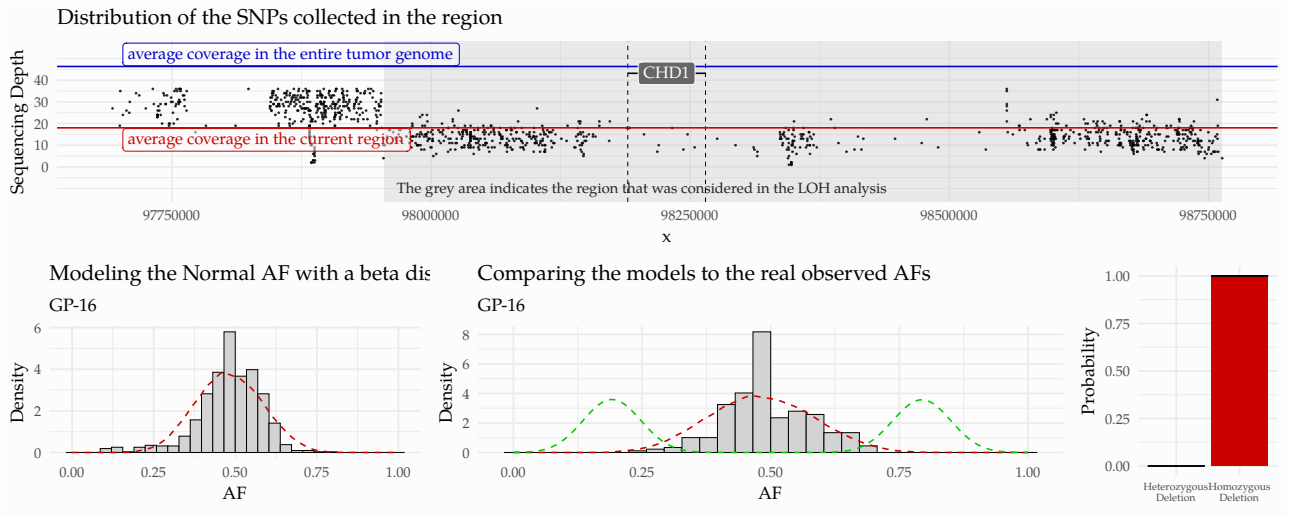
Supplementary Figure 25: Illustrations of the LOH calling procedure on the CHD1-loss whole genomes. Each sample has its own figure set: the top panel shows the coverage and spatial distribution of the heterozygous SNPs around CHD1 in the sample, with a grey box around those that were included in the LOH analysis. The figure also indicates the mean coverage over the entire tumor genome (blue line) and the average coverage in the displayed region (red line). The bottom panel shows the various observed AF-distribution as histograms, and the theoretical distributions that are expected if the deletion is homozygous or heterozygous. The final probabilities that infer whether the observed AF-distribution in the tumor above and around CHD1 result from an LOH or not are also displayed. The error bars indicate the 95% percentile intervals of 1000 bootstrap iterations.



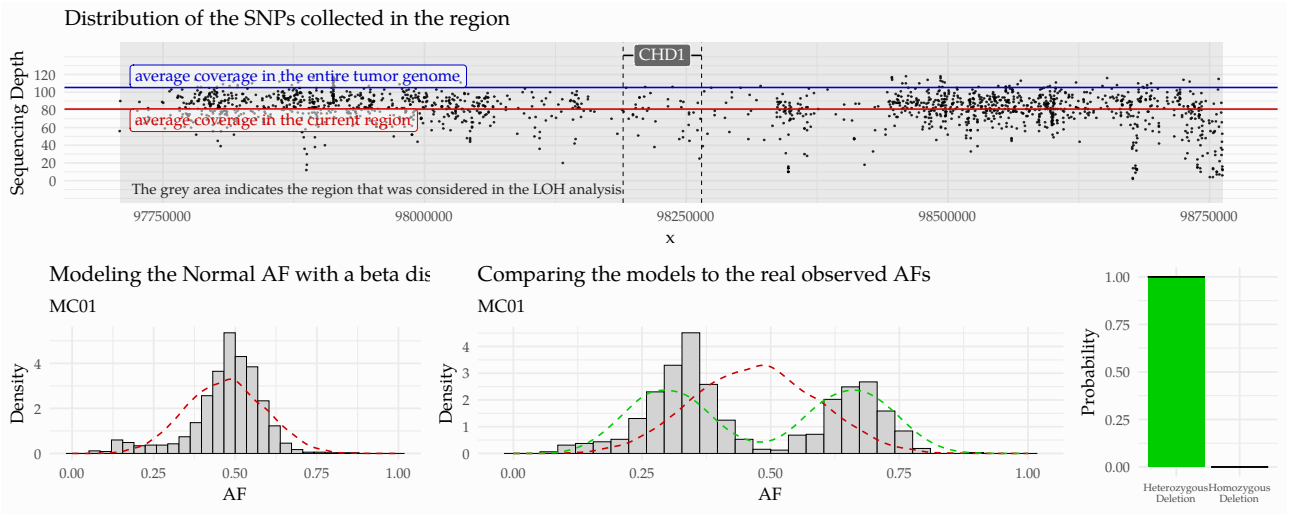
GP-07



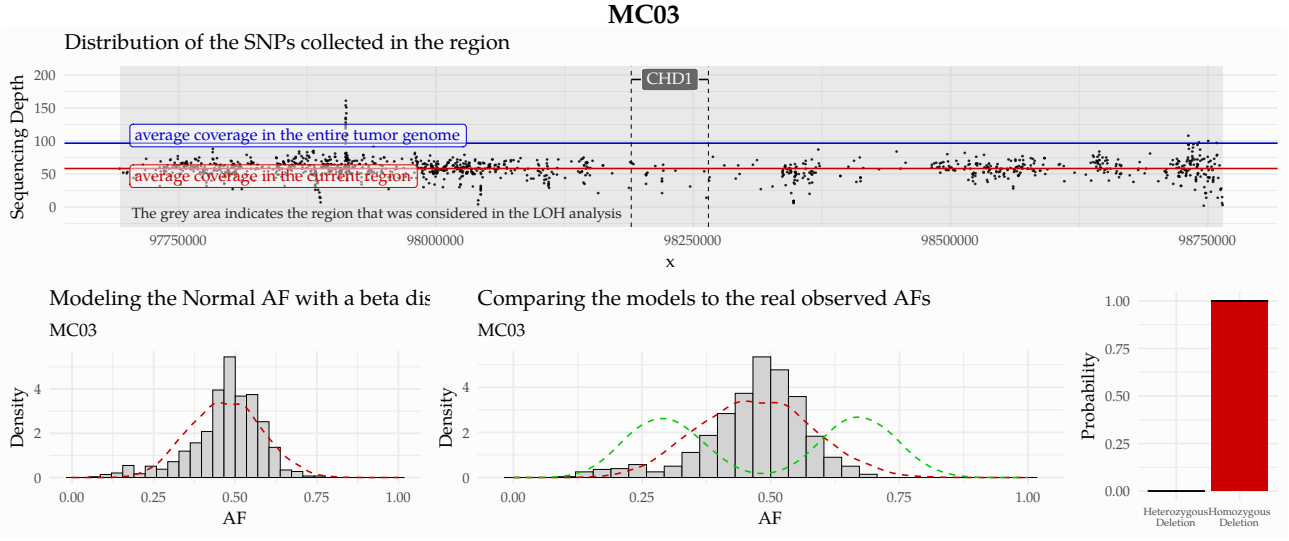
GP-16



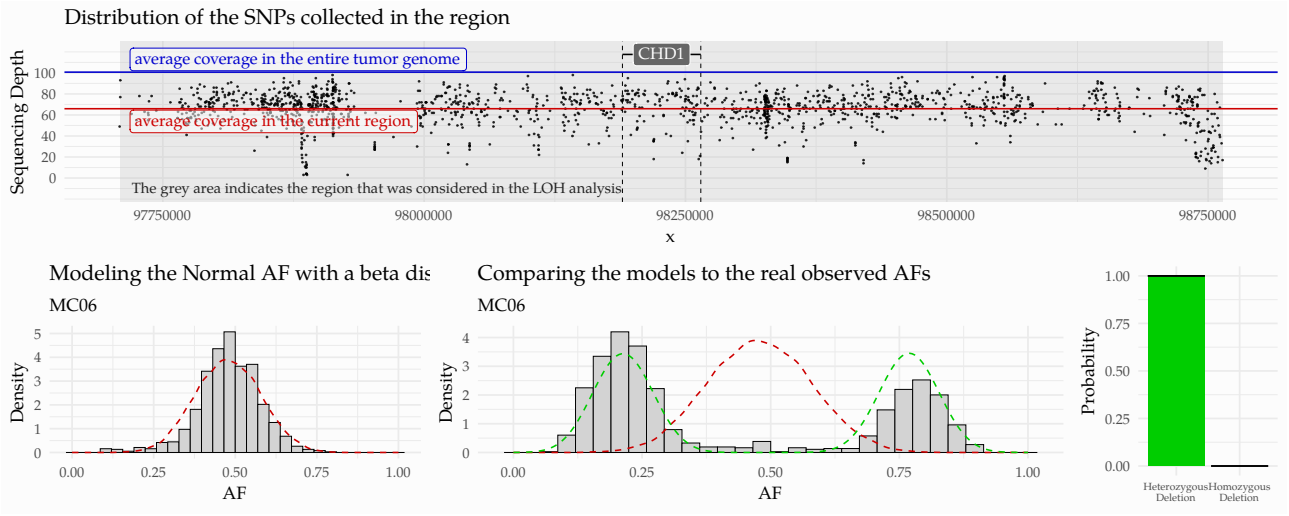
MC01



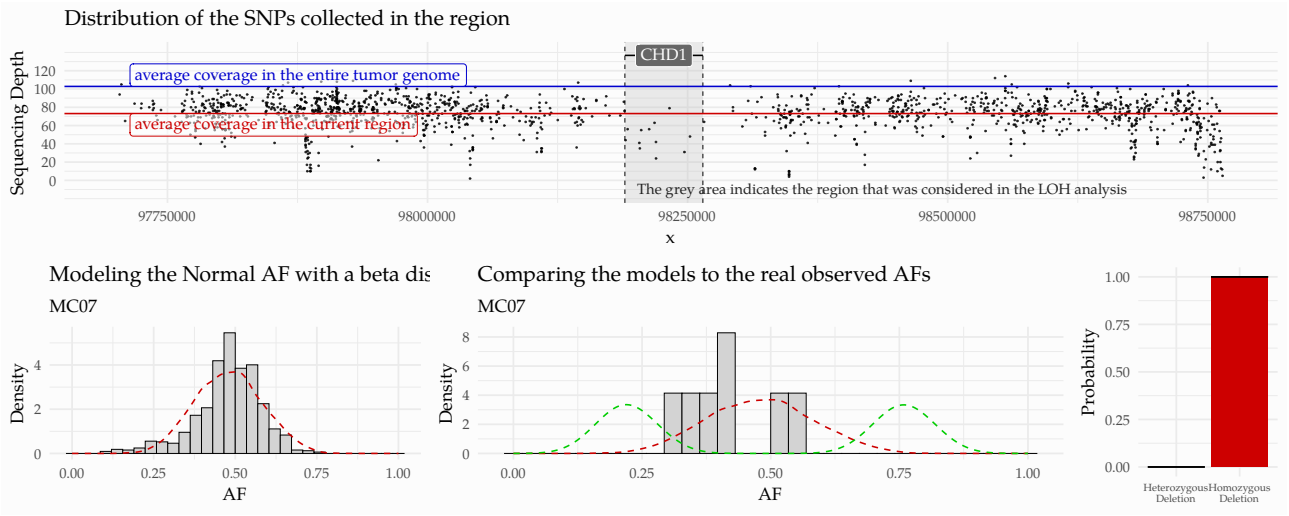
MC03



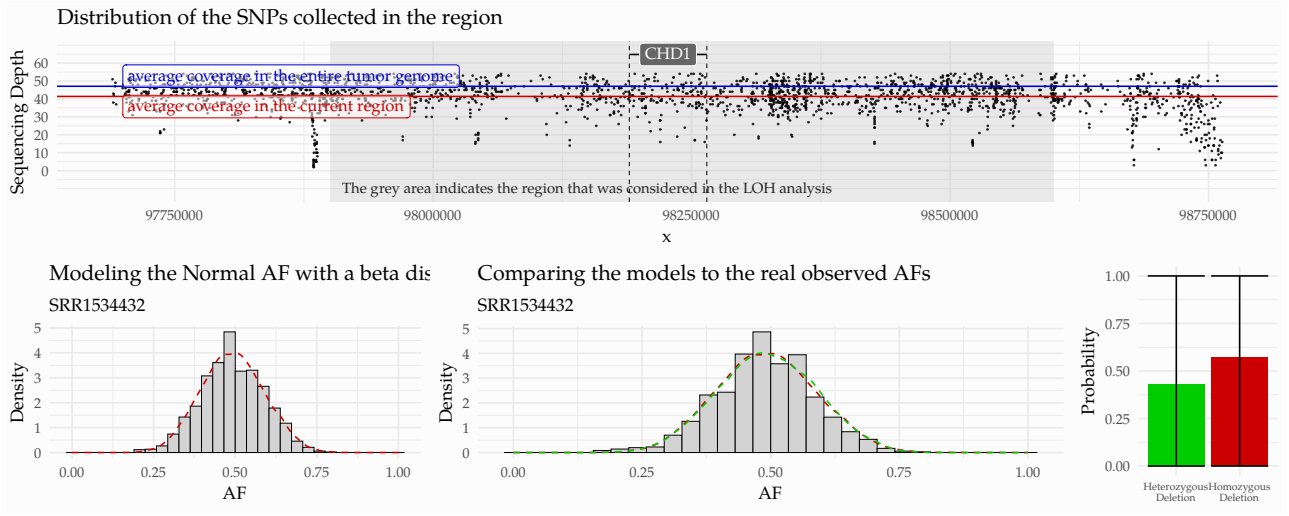
MC06



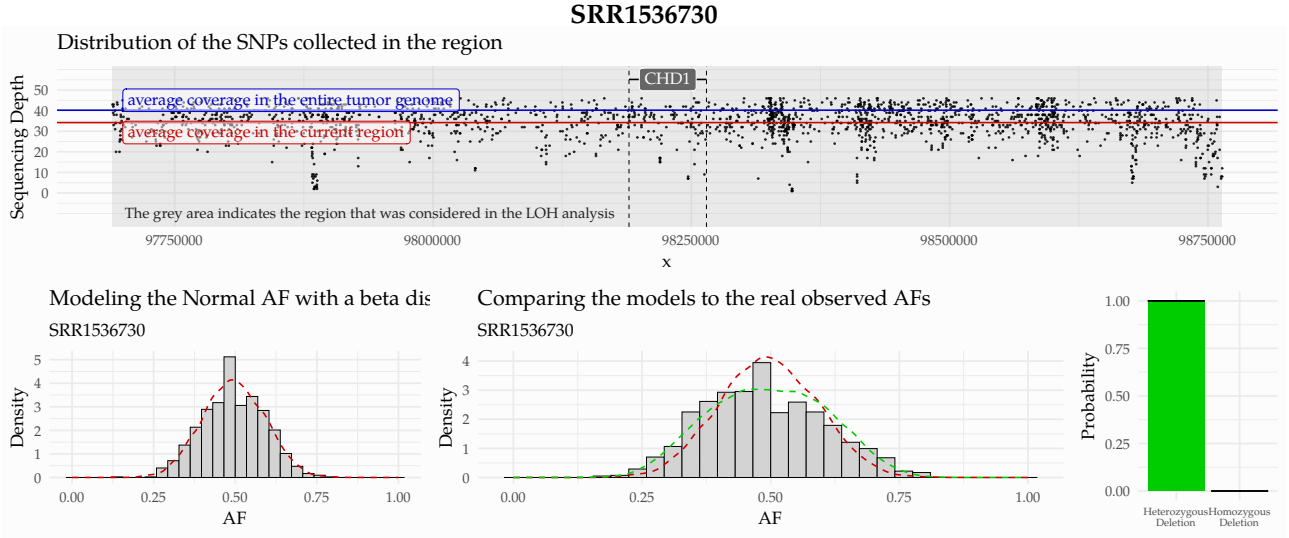
MC07



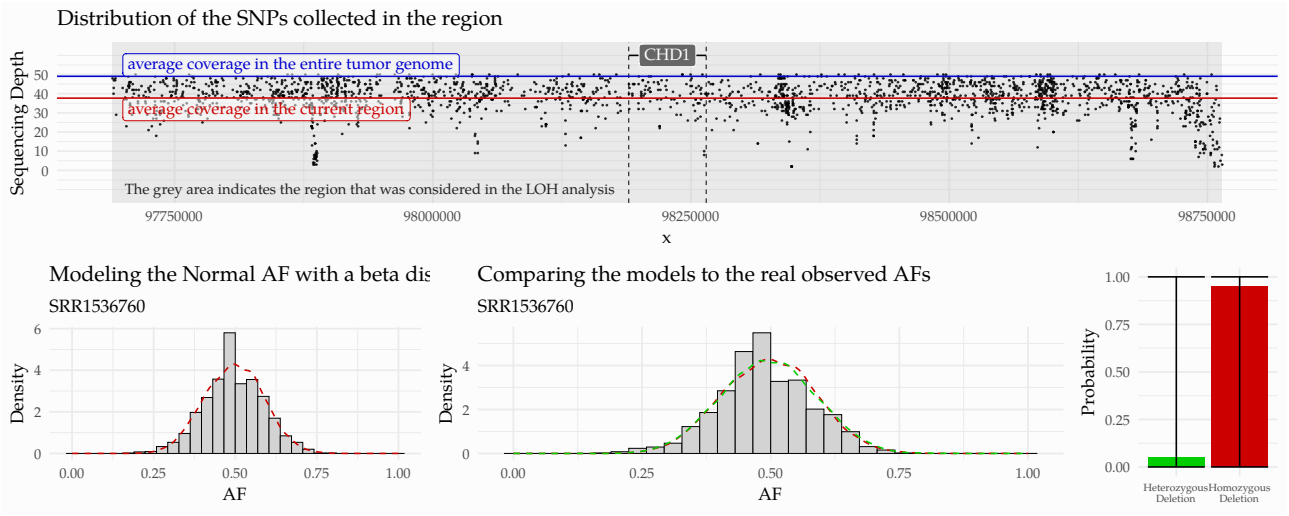
SRR1534432



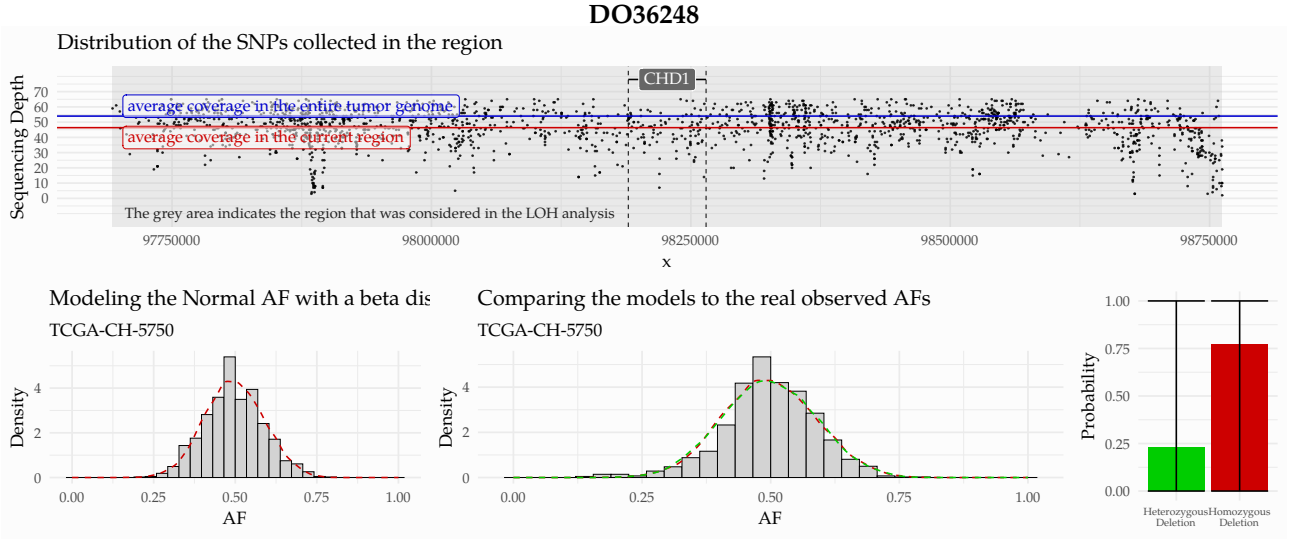
SRR1536730



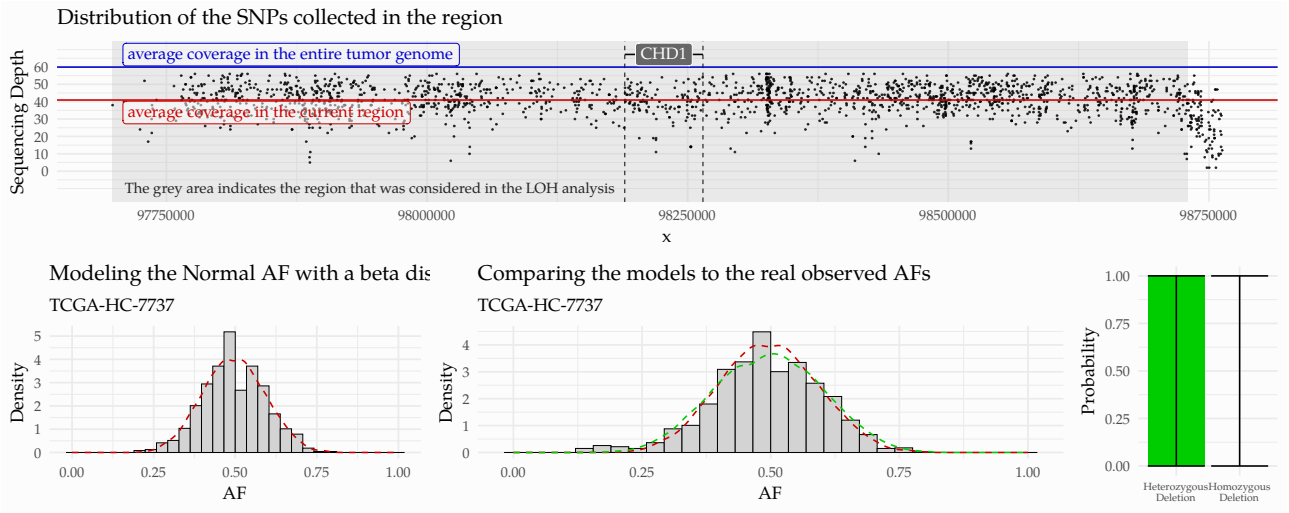
SRR1536760



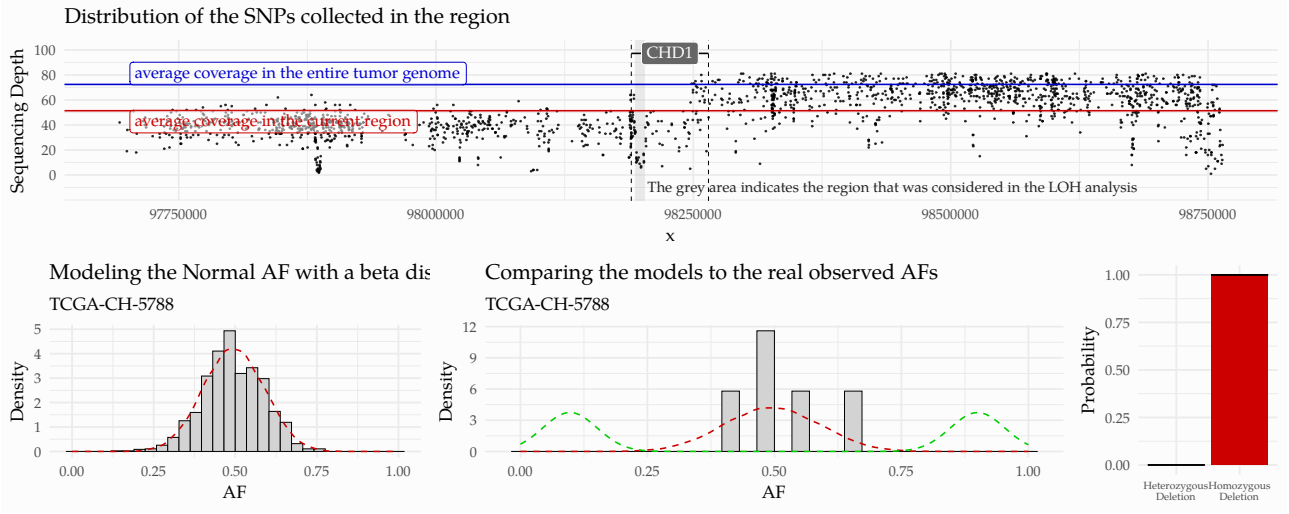
DO36248



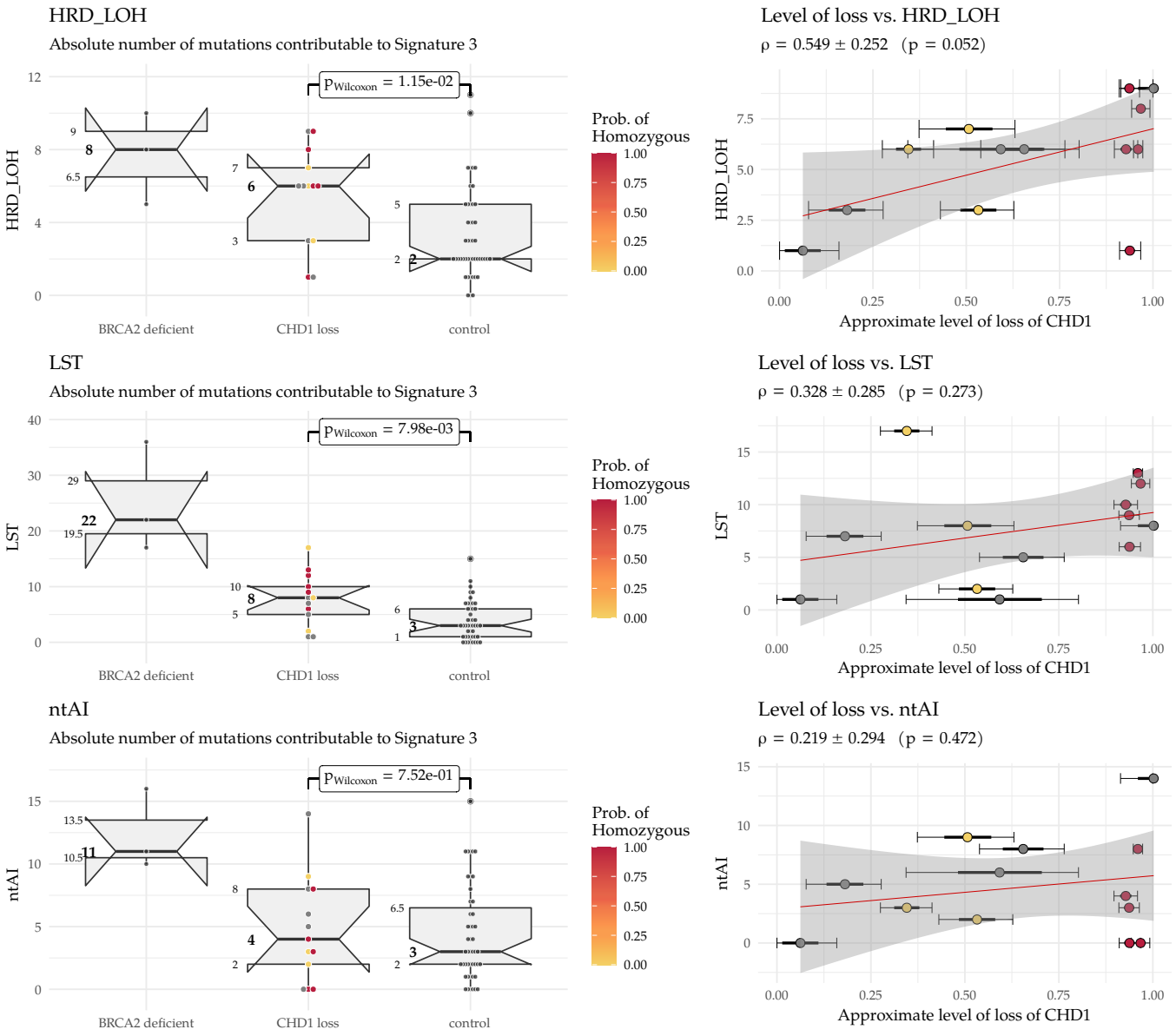
DO36285



DO36253

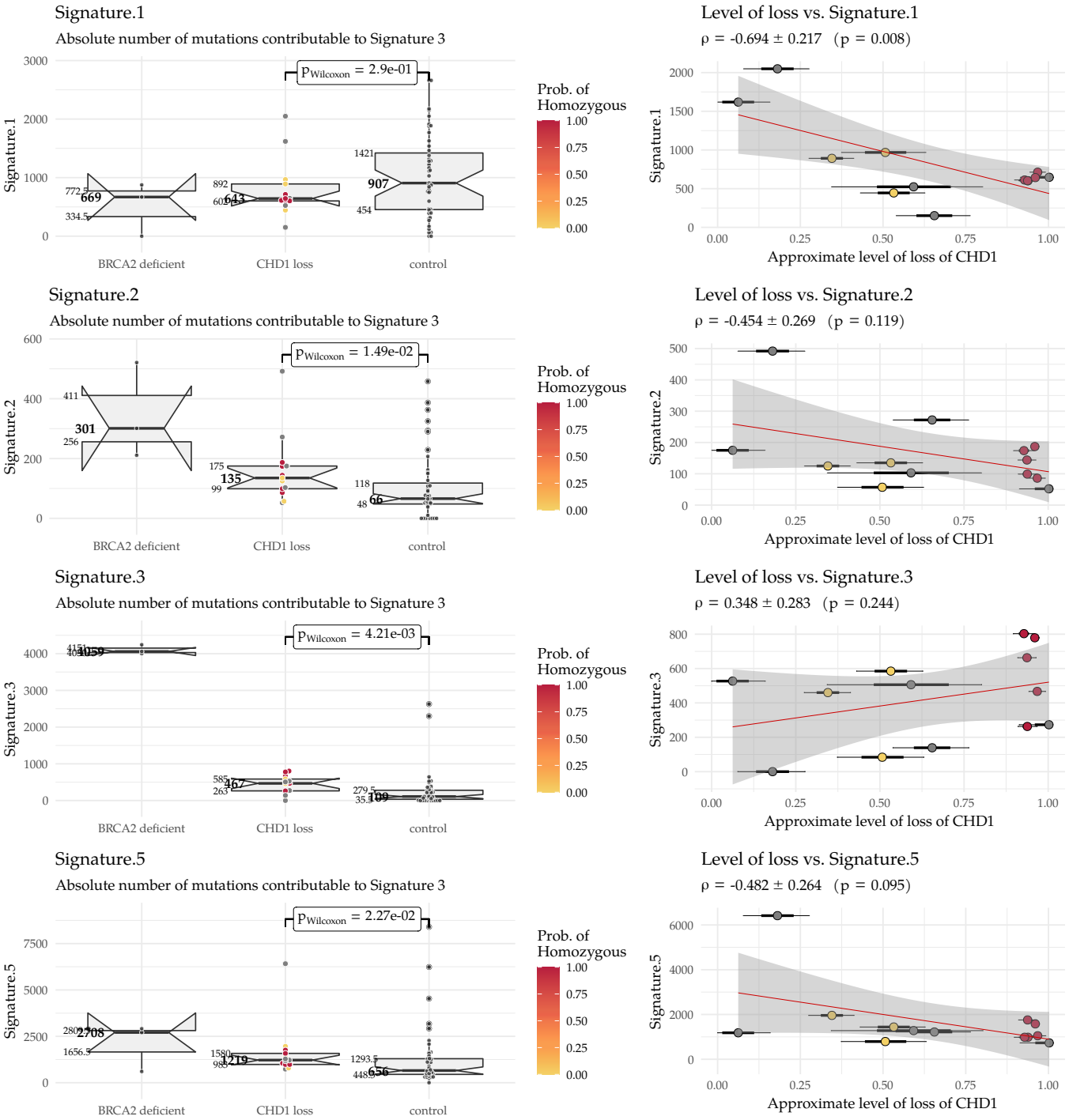


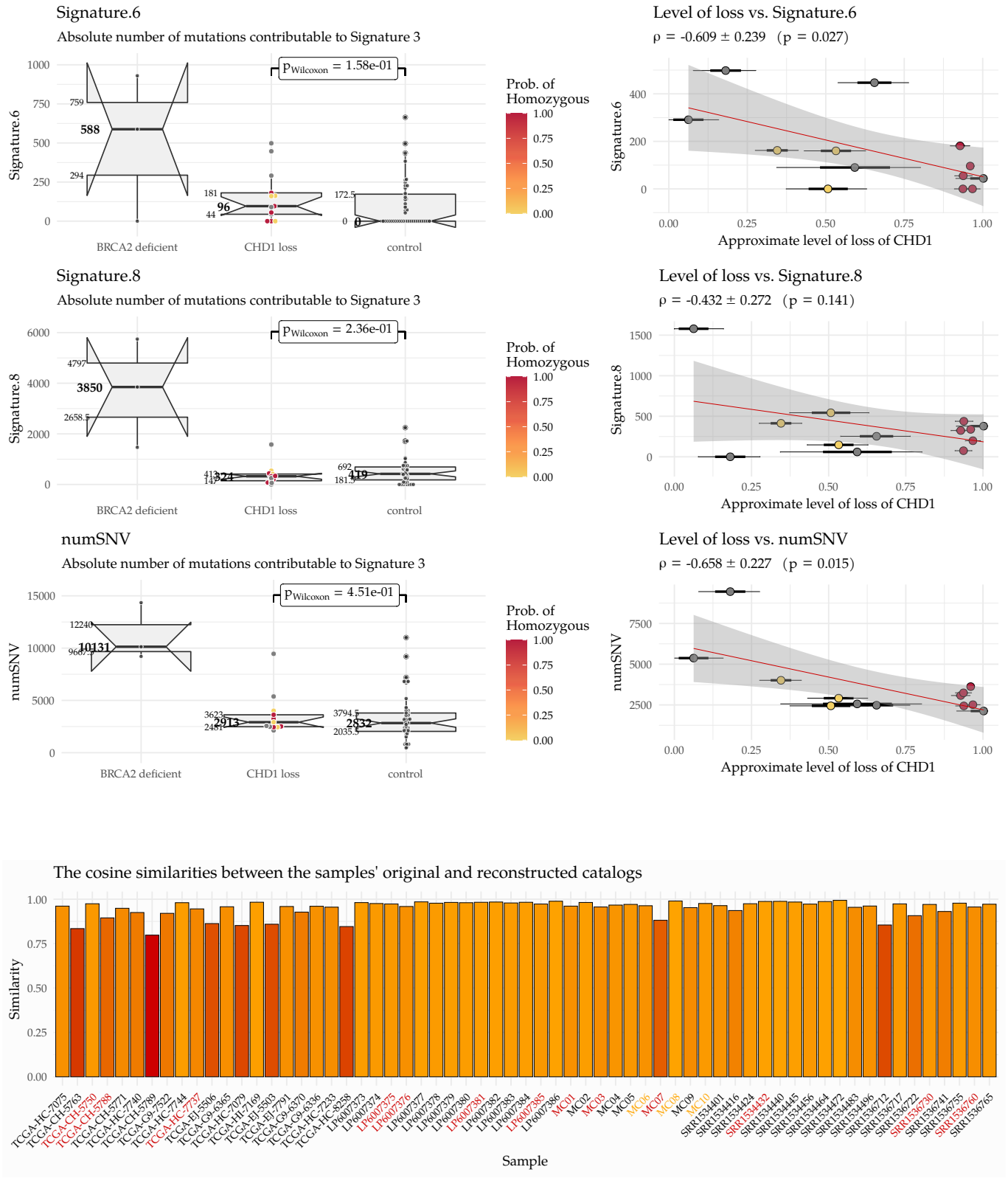
Genomic features of the whole genomes



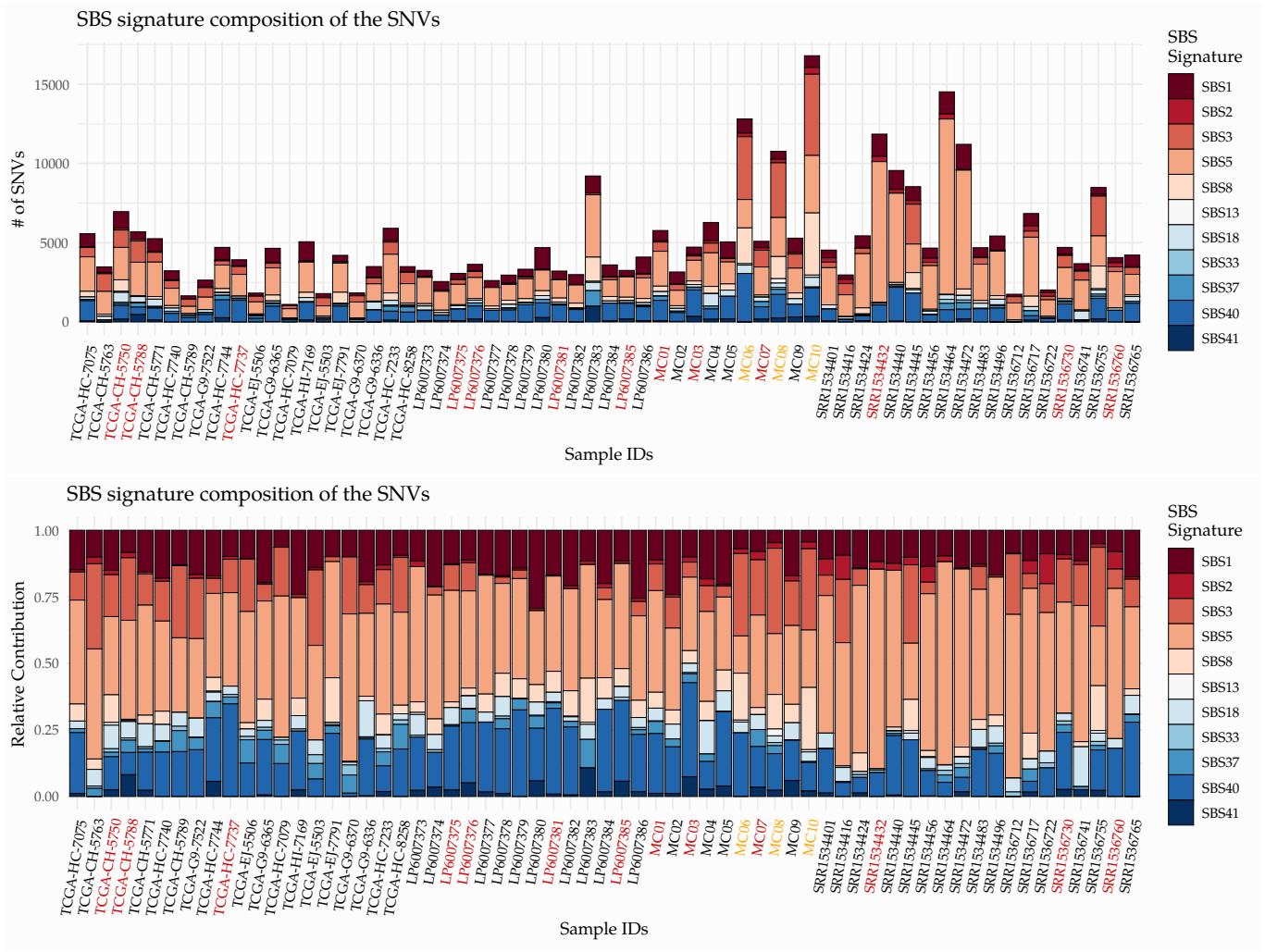
Supplementary Figure 26: HRD-related genomic scars in the whole genomes. **HRD_LOH**: HRD Loss of Heterozygosity, **LST**: Large Scale Transition, **ntAI**: number of telomeric Allelic Imbalance

Supplementary Figure 27: Number of variants contributable ot the 2nd generation single nucleotide signatures. Signatures 1,2,3,5,6,8 and the total number of SNVs in the sample (numSNV)



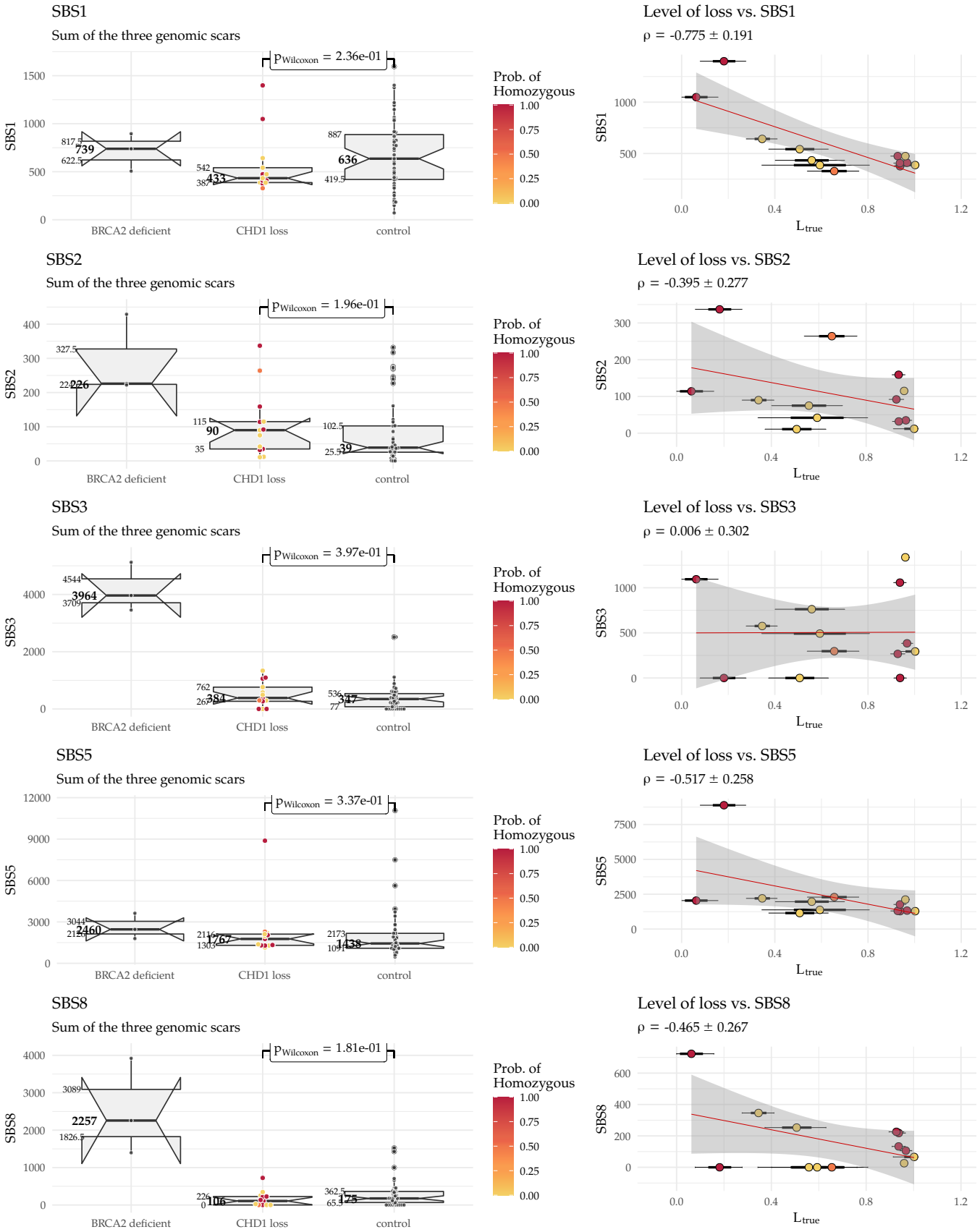


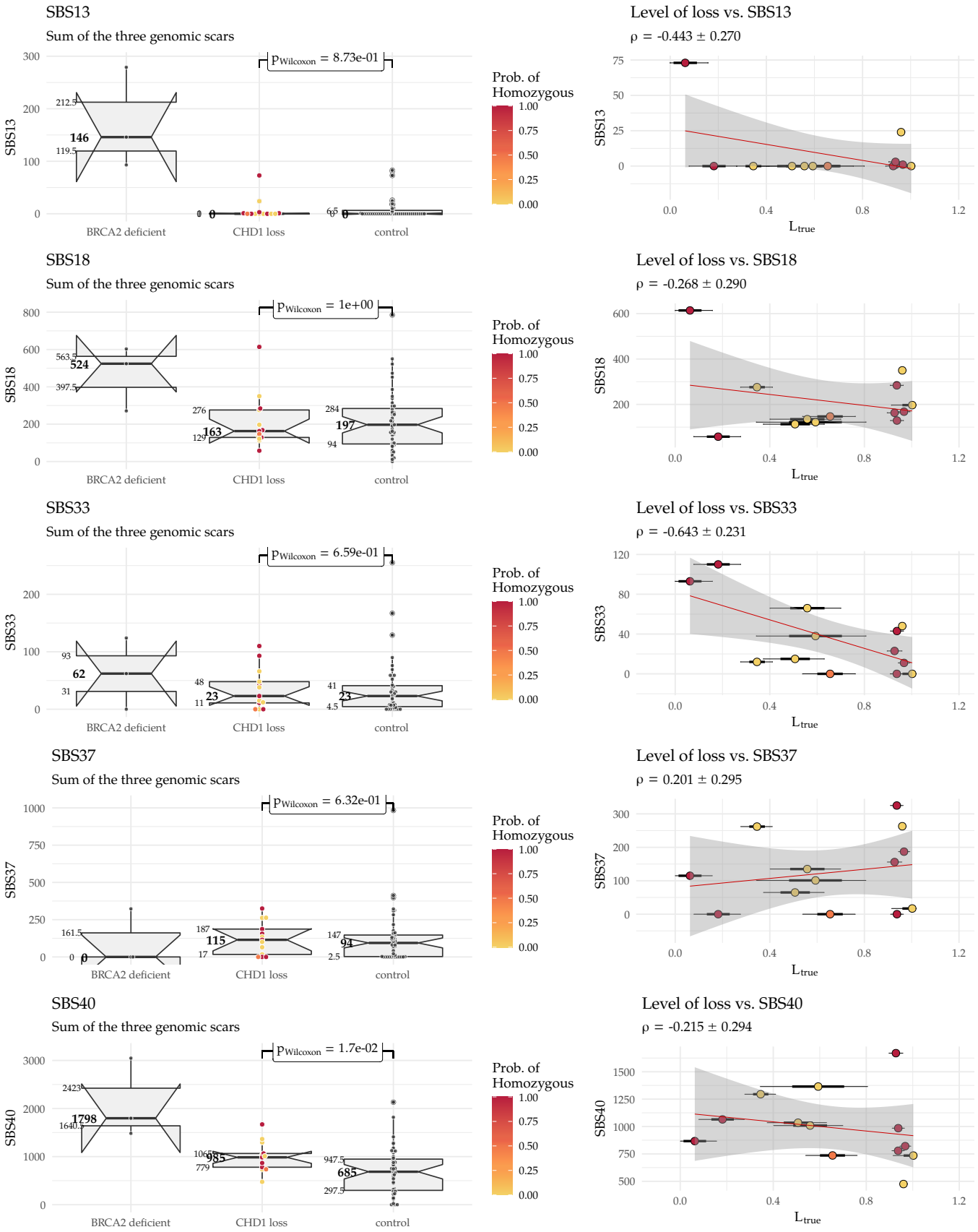
Supplementary Figure 28: The cosine similarity between the original and reconstructed mutational SBS spectra in the whole genomes.

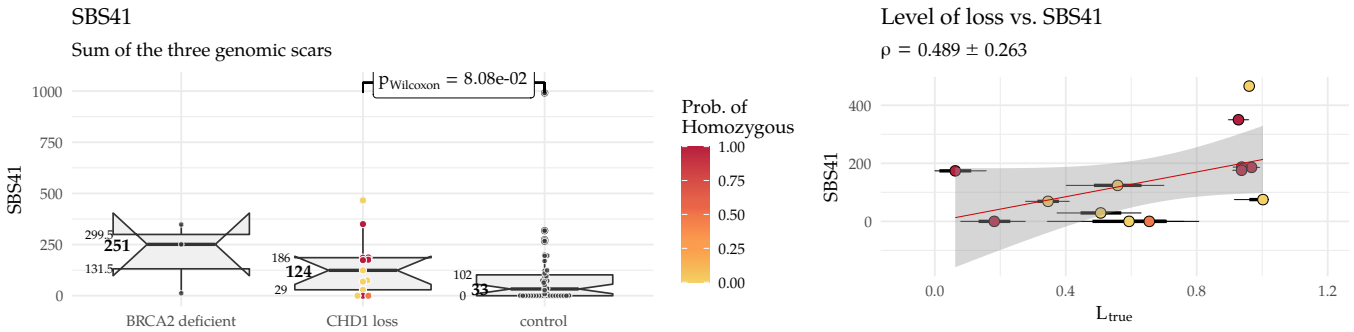


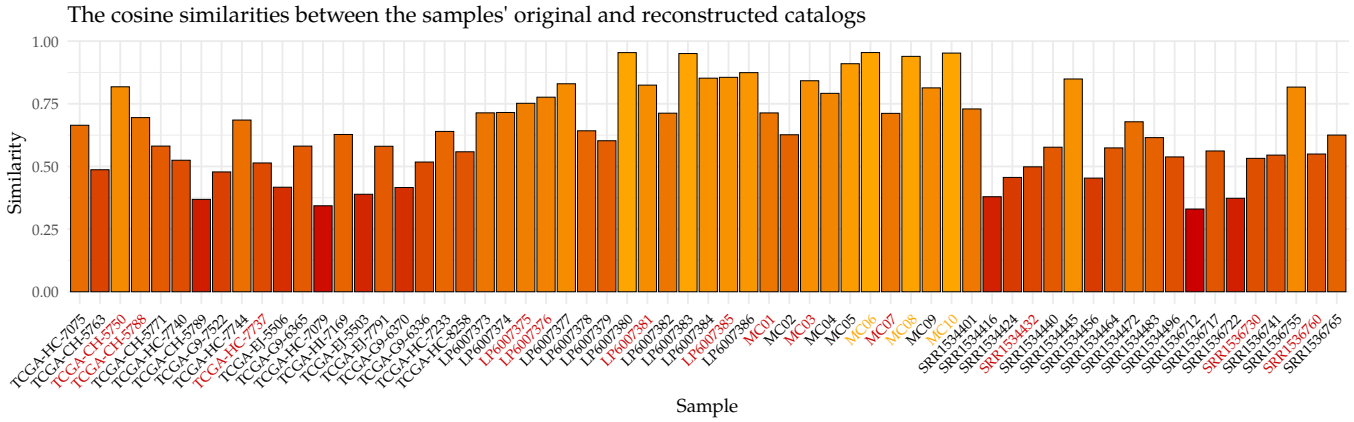
Supplementary Figure 29: The SBS signature compositon of the whole genomes. Above are the absolute numbers, below the relative ratios.

Supplementary Figure 30: Third generation single nucleotide (SBS) signatures.

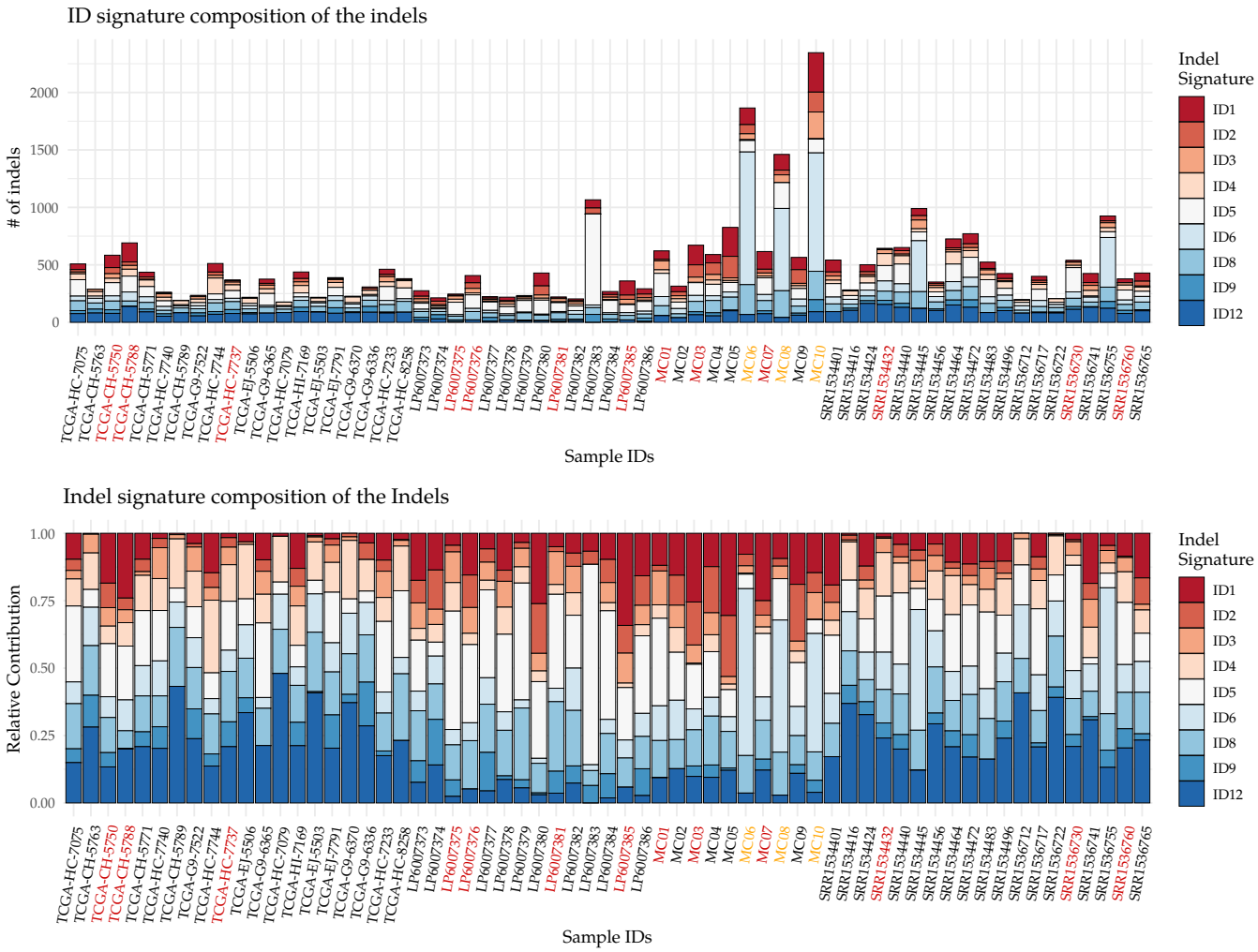



SBS13
Sum of the three genomic scars
PWilcoxon = 8.73e-01
SBS13
BRCA2 deficient
CHD1 loss
control
Prob. of Homozygous
Level of loss vs. SBS13
ρ = -0.443 ± 0.270
Ltrue
SBS18
Sum of the three genomic scars
PWilcoxon = 1e+00
SBS18
Level of loss vs. SBS18
ρ = -0.268 ± 0.290
SBS33
Sum of the three genomic scars
PWilcoxon = 6.59e-01
SBS33
Level of loss vs. SBS33
ρ = -0.643 ± 0.231
SBS37
Sum of the three genomic scars
PWilcoxon = 6.32e-01
SBS37
Level of loss vs. SBS37
ρ = 0.201 ± 0.295
SBS40
Sum of the three genomic scars
PWilcoxon = 1.7e-02
SBS40
Level of loss vs. SBS40
ρ = -0.215 ± 0.294



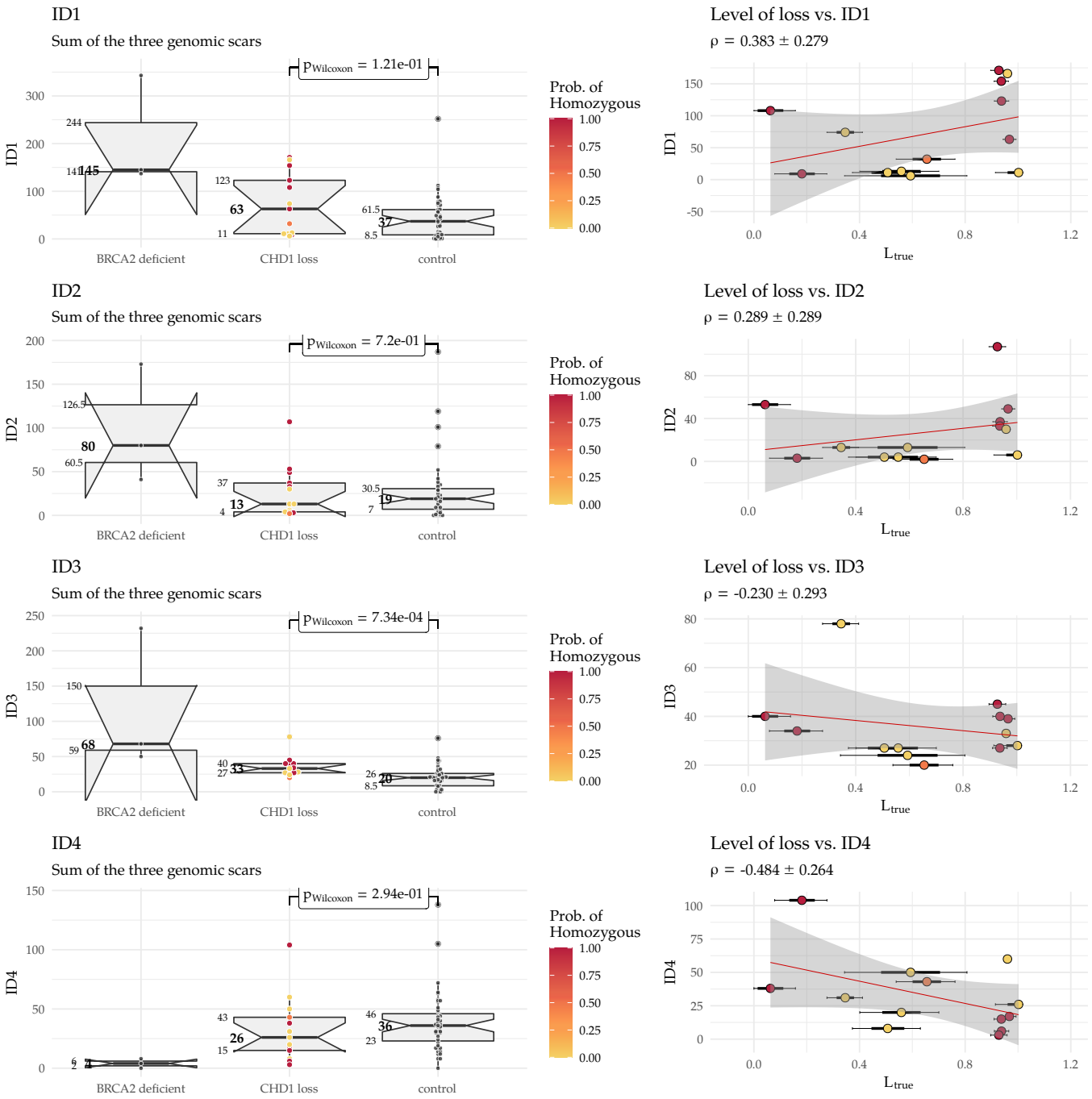


Supplementary Figure 31: The cosine similarity between the original and reconstructed mutational ID spectra in the whole genomes.



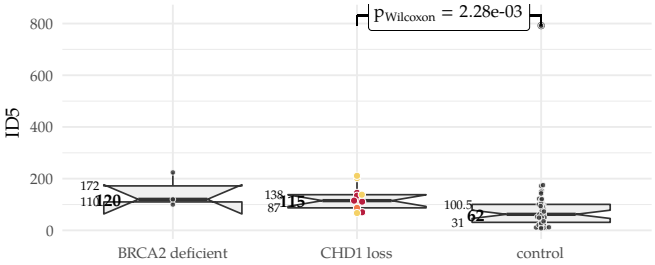
Supplementary Figure 32: The ID signature compositon of the whole genomes. Above are the absolute numbers, below the relative ratios.

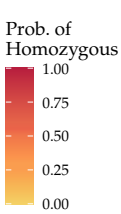
Supplementary Figure 33: ↓ Number of variants contributable to the the third generation indel (ID) signatures.



ID5

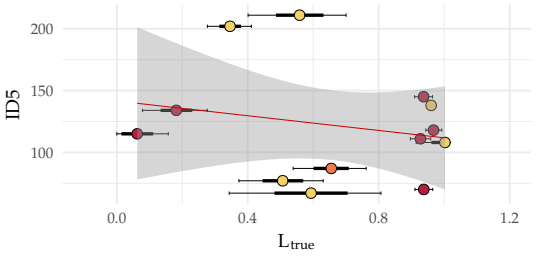
Sum of the three genomic scars





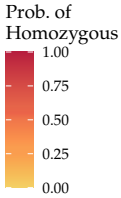
Level of loss vs. ID5

ρ = -0.212 ± 0.295



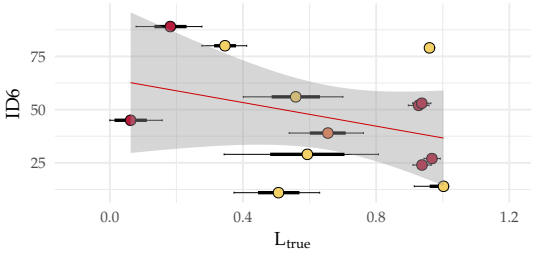
ID6

Sum of the three genomic scars



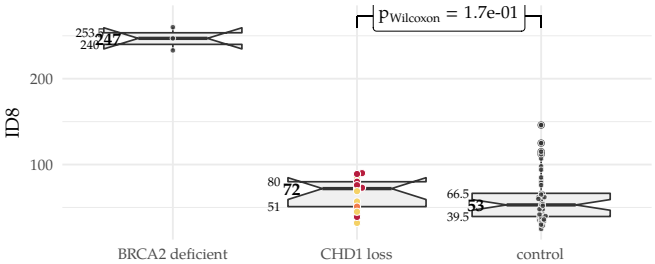
Level of loss vs. ID6

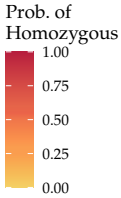
ρ = -0.352 ± 0.282



ID8

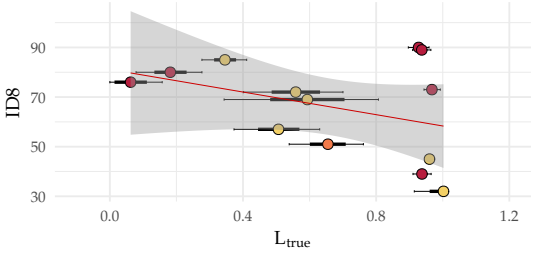
Sum of the three genomic scars





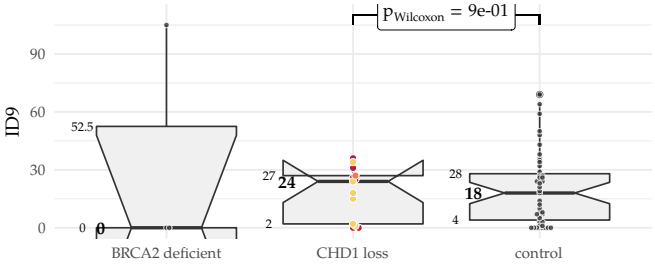
Level of loss vs. ID8

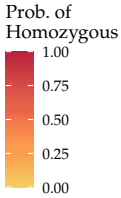
ρ = -0.380 ± 0.279



ID9

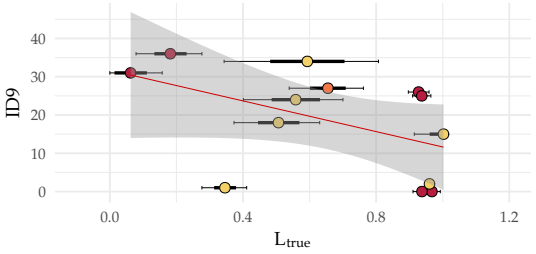
Sum of the three genomic scars





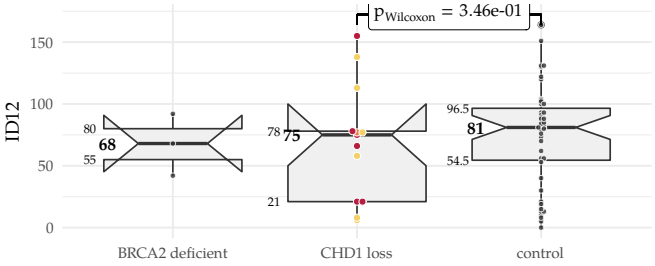
Level of loss vs. ID9

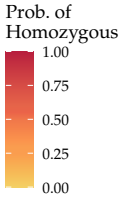
ρ = -0.480 ± 0.264



ID12

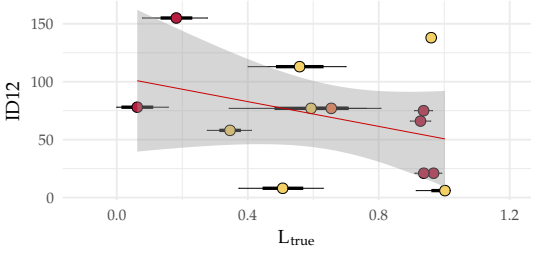
Sum of the three genomic scars

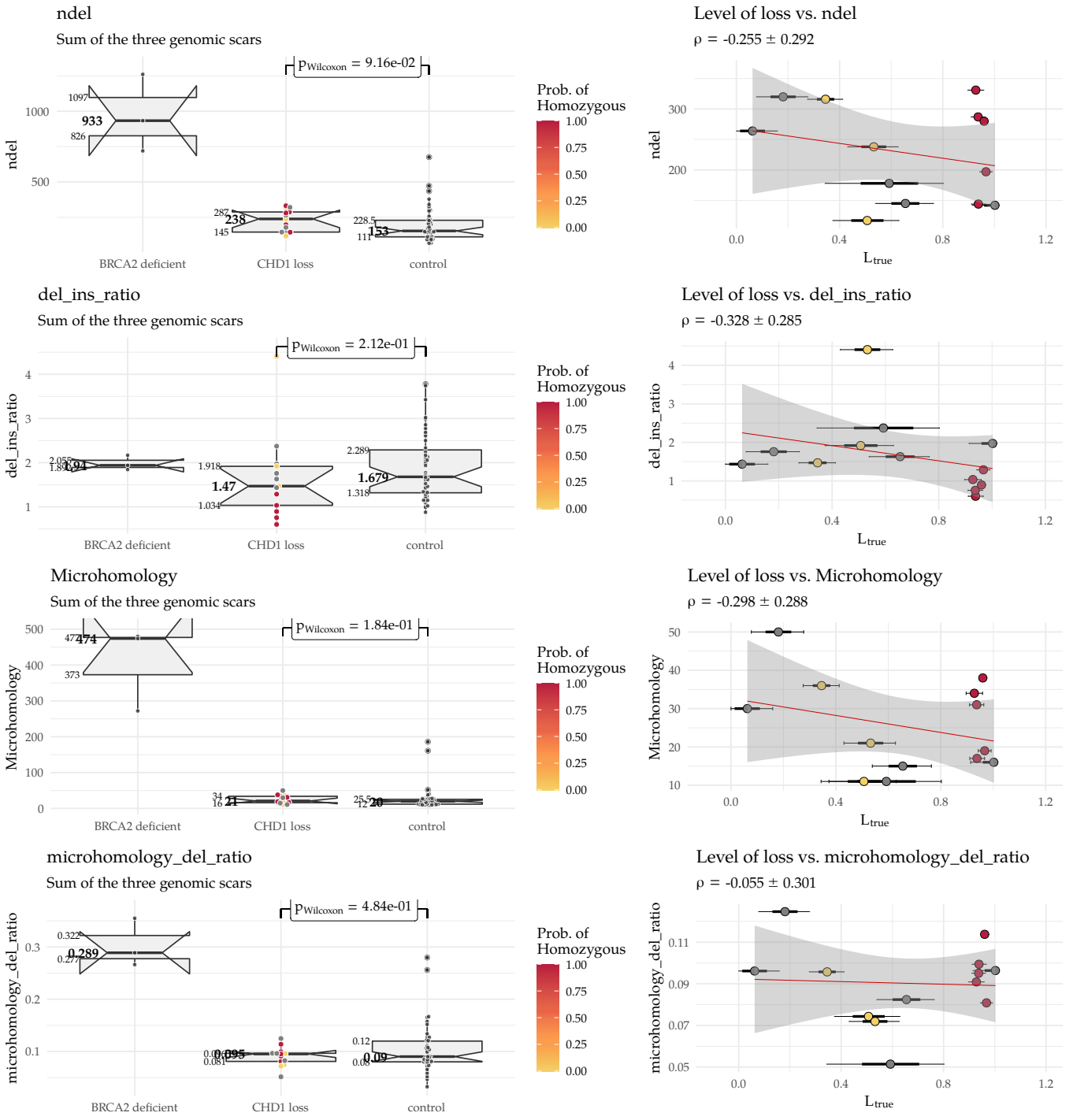




Level of loss vs. ID12

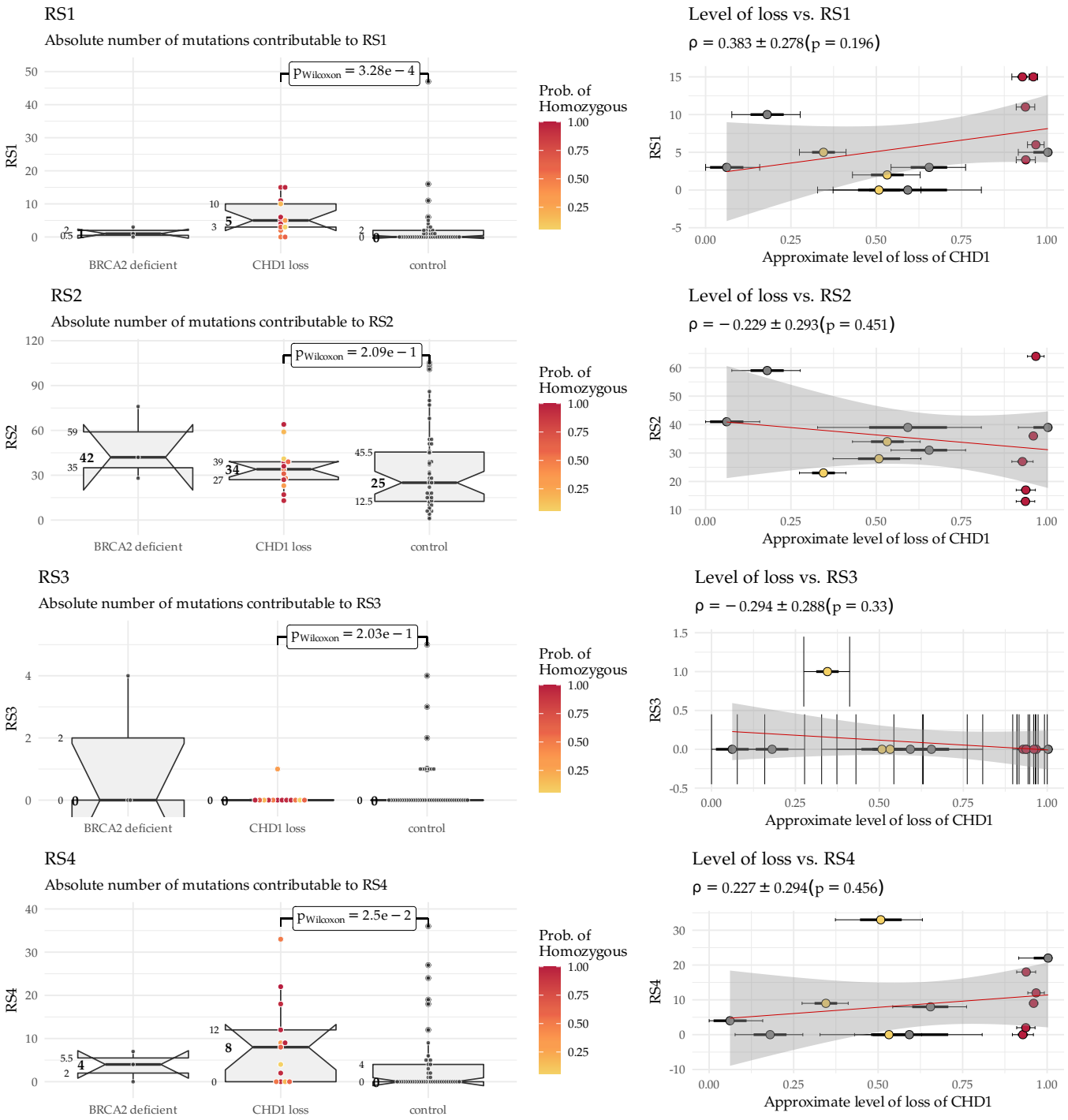
ρ = -0.365 ± 0.281

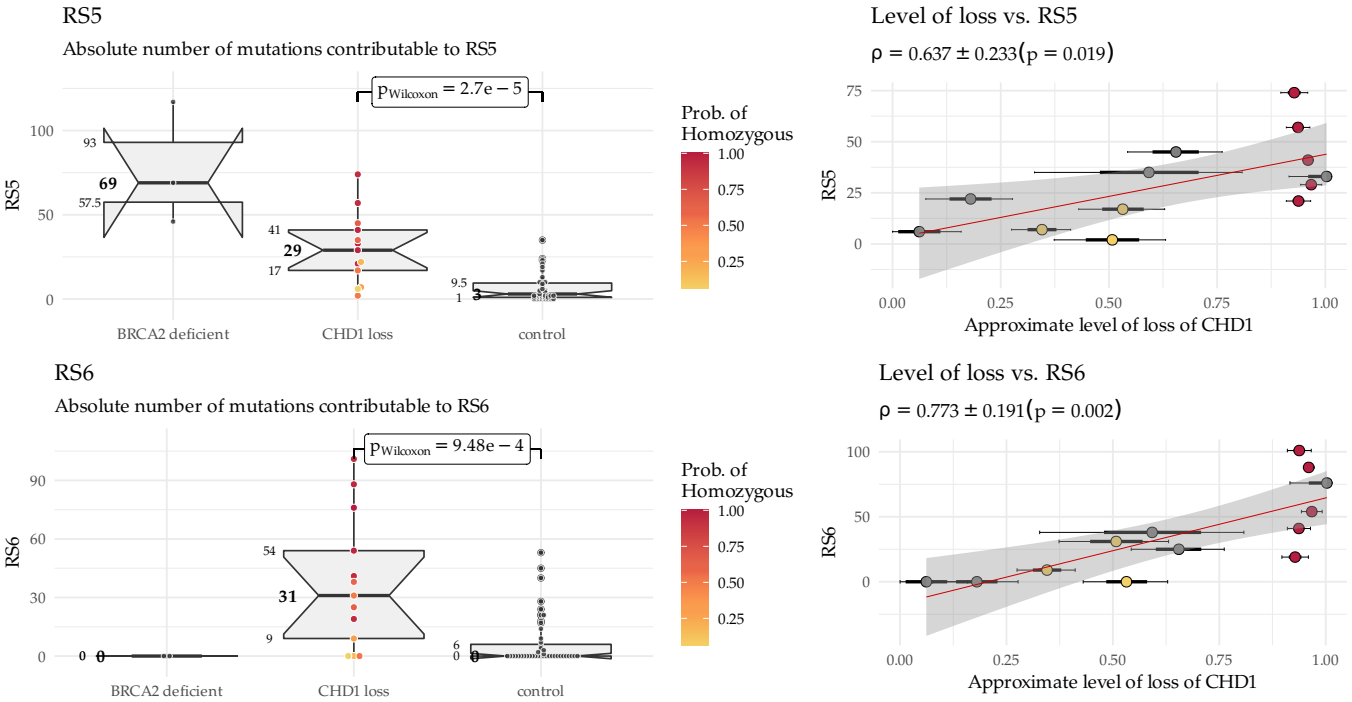


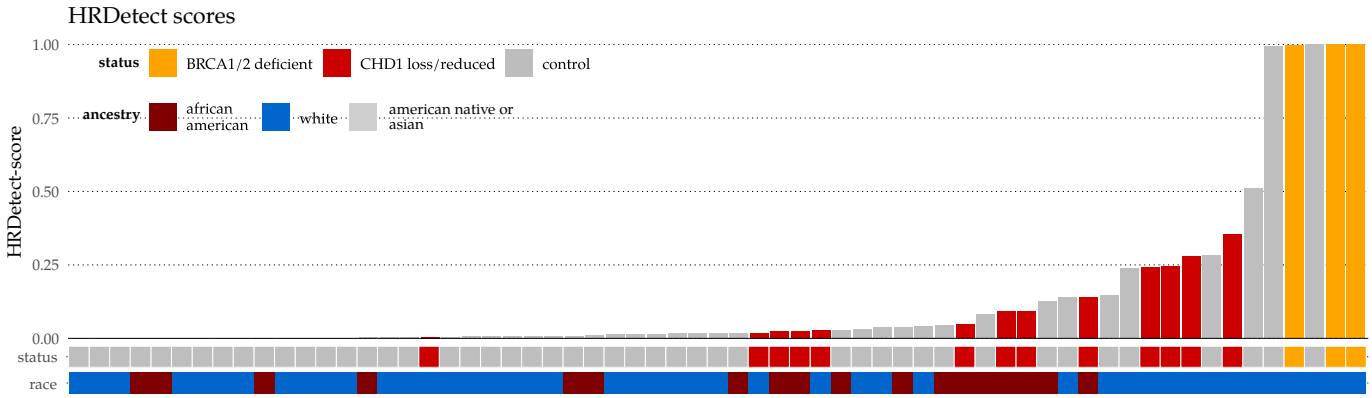


Supplementary Figure 34: The "traditional" classification of deletions. ndel: number of deletions, del_ins_ratio: deletion insertion ratio, microhomology: number of microhomology mediated deletions, microhomology_del_ratio: ratio of microhomology mediated deletions relative to the number of deletions

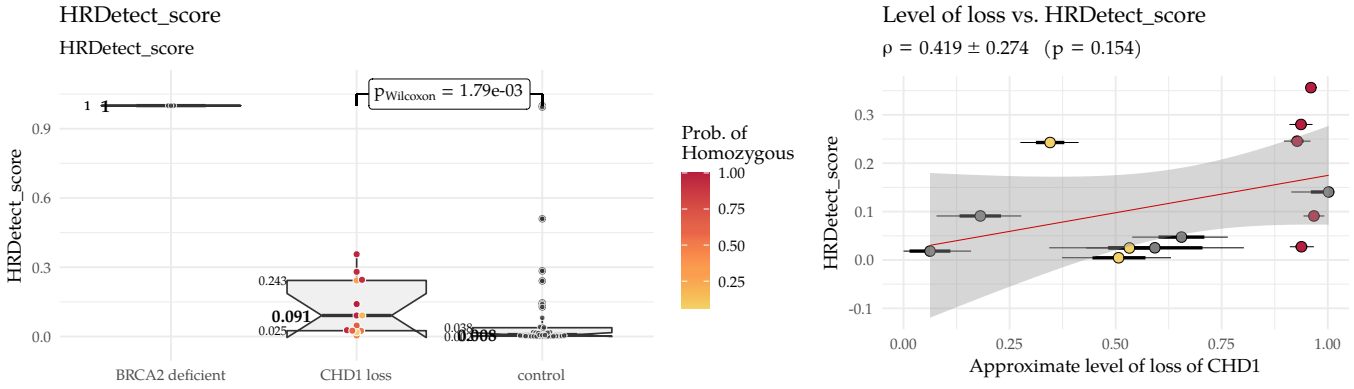
Supplementary Figure 35: The six structural-variant-based rearrangement signatures



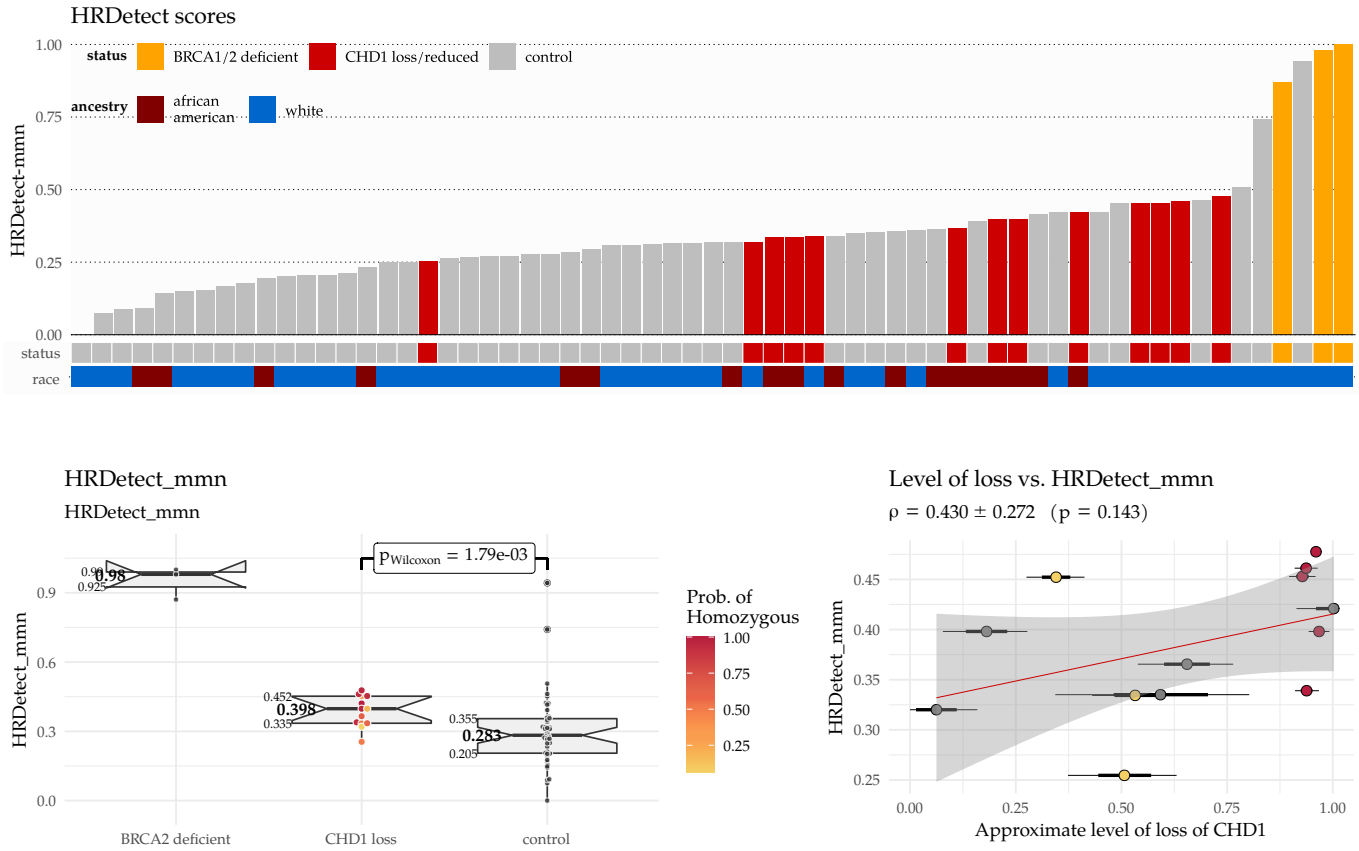

RS5
Absolute number of mutations contributable to RS5
$p_{Wilcoxon} = 2.7e-5$
RS5
100
50
0
93
69
57.5
41
29
17
9.5
1
BRCA2 deficient
CHD1 loss
control
Prob. of Homozygous
1.00
0.75
0.50
0.25
Level of loss vs. RS5
$\rho = 0.637 \pm 0.233 (p = 0.019)$
RS5
75
50
25
0
0.00
0.25
0.50
0.75
1.00
Approximate level of loss of CHD1
RS6
Absolute number of mutations contributable to RS6
$p_{Wilcoxon} = 9.48e-4$
RS6
90
60
30
0
54
31
9
6
0
BRCA2 deficient
CHD1 loss
control
Prob. of Homozygous
1.00
0.75
0.50
0.25
Level of loss vs. RS6
$\rho = 0.773 \pm 0.191 (p = 0.002)$
RS6
100
50
0
0.00
0.25
0.50
0.75
1.00
Approximate level of loss of CHD1



Supplementary Figure 36: HRDetect score summary - WGS. The weights of the model from the original HRDetect publication from Davies et. al, in ascending order: intercept = -3.3642, Signature.8 = 0.09062, HRD-LOH = 0.6666, RS5 = 0.8467, RS3 = 1.1532, Signature.3 = 1.6114, mhm.del.ratio = 2.3977

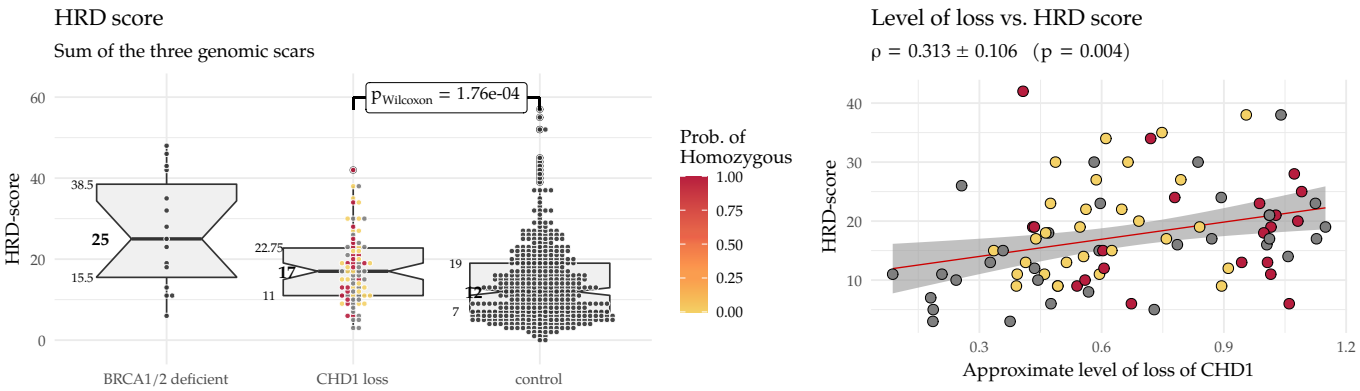


Supplementary Figure 37: HRDetect scores in samples with CHD1 loss vs. CHD1 intact samples

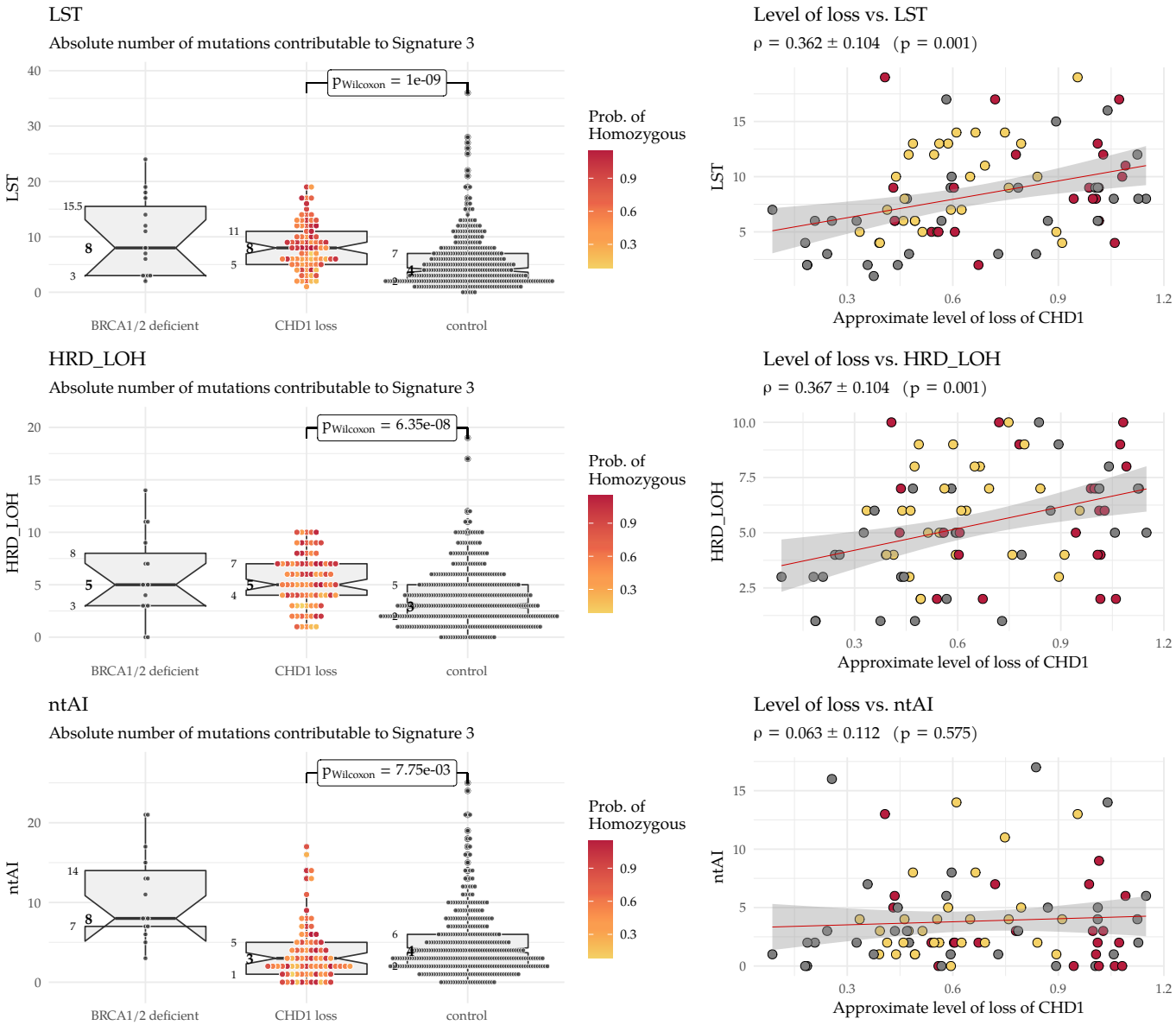


Supplementary Figure 38: The HRDetect model is a logistic regression model, in which the ordinary linear regression scores ($\mathbf{x}$) are transformed into the range of [0,1] by putting them through a nonlinear logistic function: $\text{HRDetect}_{\text{probability}} = \mathbf{s} = \frac{1}{1+e^{-\mathbf{x}}}$. The linearized, non-probabilistic "raw" scores are extracted from this formula: $\mathbf{x} = \ln\left(\frac{\mathbf{s}}{1-\mathbf{s}}\right)$. The scores were min-max normalized (mmn) for visualization purposes.

Genomic Features of the Whole Exomes

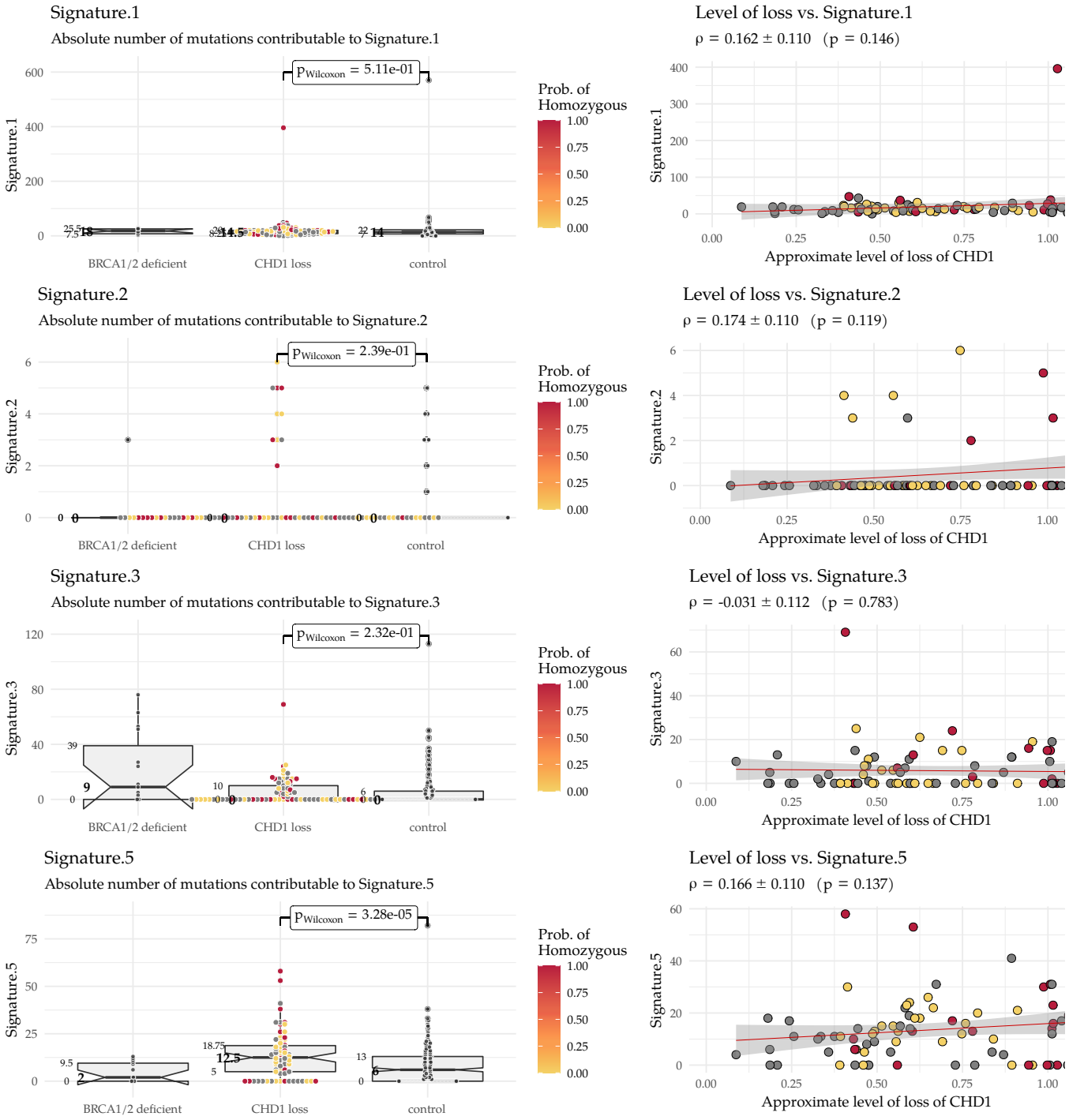


Supplementary Figure 39: HRD-scores (sum of the three genomic scars) of the whole exomes



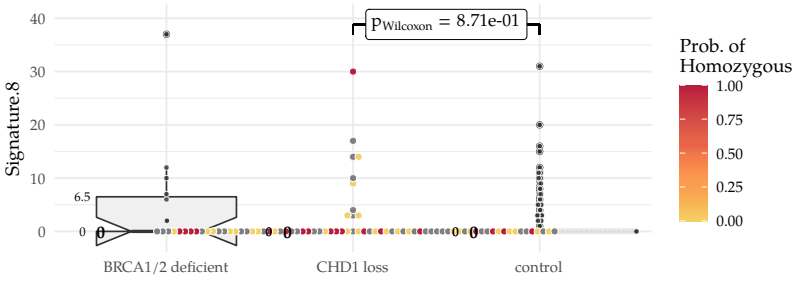
Supplementary Figure 40: HRD-related genomic scars in the whole exomes. **HRD_LOH**: HRD Loss of Heterozygosity, **LST**: Large Scale Transition, **ntAI**: number of telomeric Allelic Imbalance

Supplementary Figure 41: Number of variants contributable ot the 2nd generation single nucleotide signatures – WES. Signatures 1,2,3,5,6,8 and the total number of SNVs in the sample (numSNV)



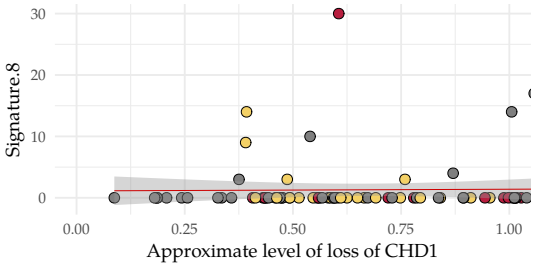
Signature.8

Absolute number of mutations contributable to Signature.8



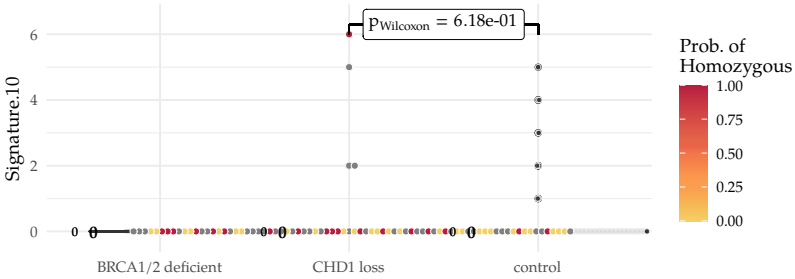
Level of loss vs. Signature.8

$\rho = 0.017 \pm 0.112$ ($p = 0.881$)



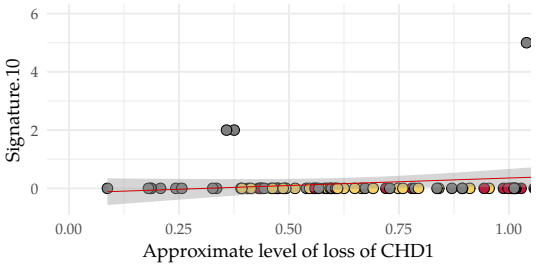
Signature.10

Absolute number of mutations contributable to Signature.10



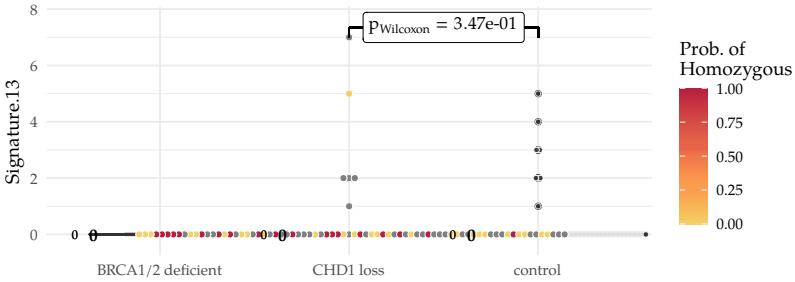
Level of loss vs. Signature.10

$\rho = 0.158 \pm 0.110$ ($p = 0.156$)



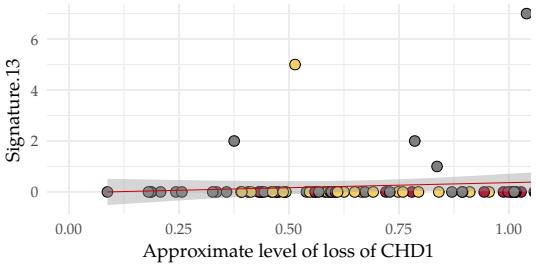
Signature.13

Absolute number of mutations contributable to Signature.13



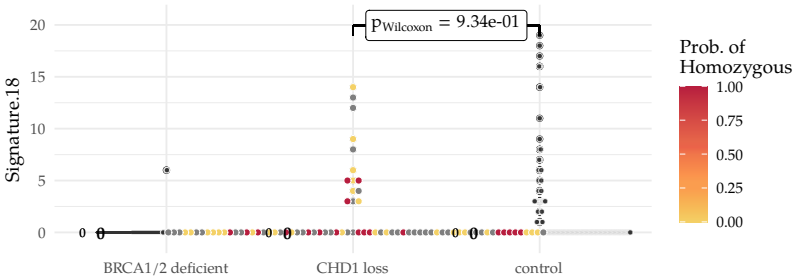
Level of loss vs. Signature.13

$\rho = 0.110 \pm 0.111$ ($p = 0.326$)



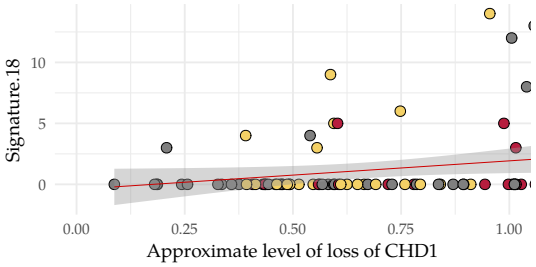
Signature.18

Absolute number of mutations contributable to Signature.18



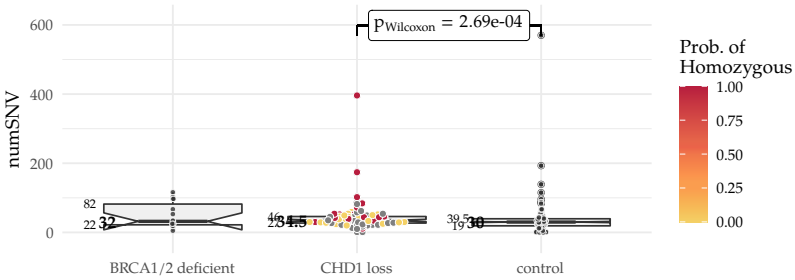
Level of loss vs. Signature.18

$\rho = 0.219 \pm 0.109$ ($p = 0.048$)



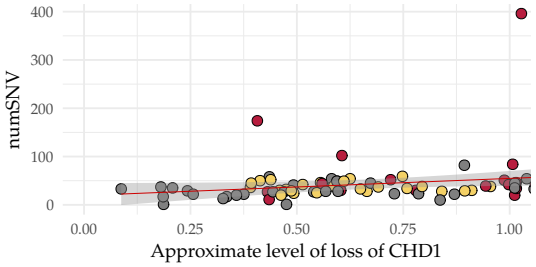
numSNV

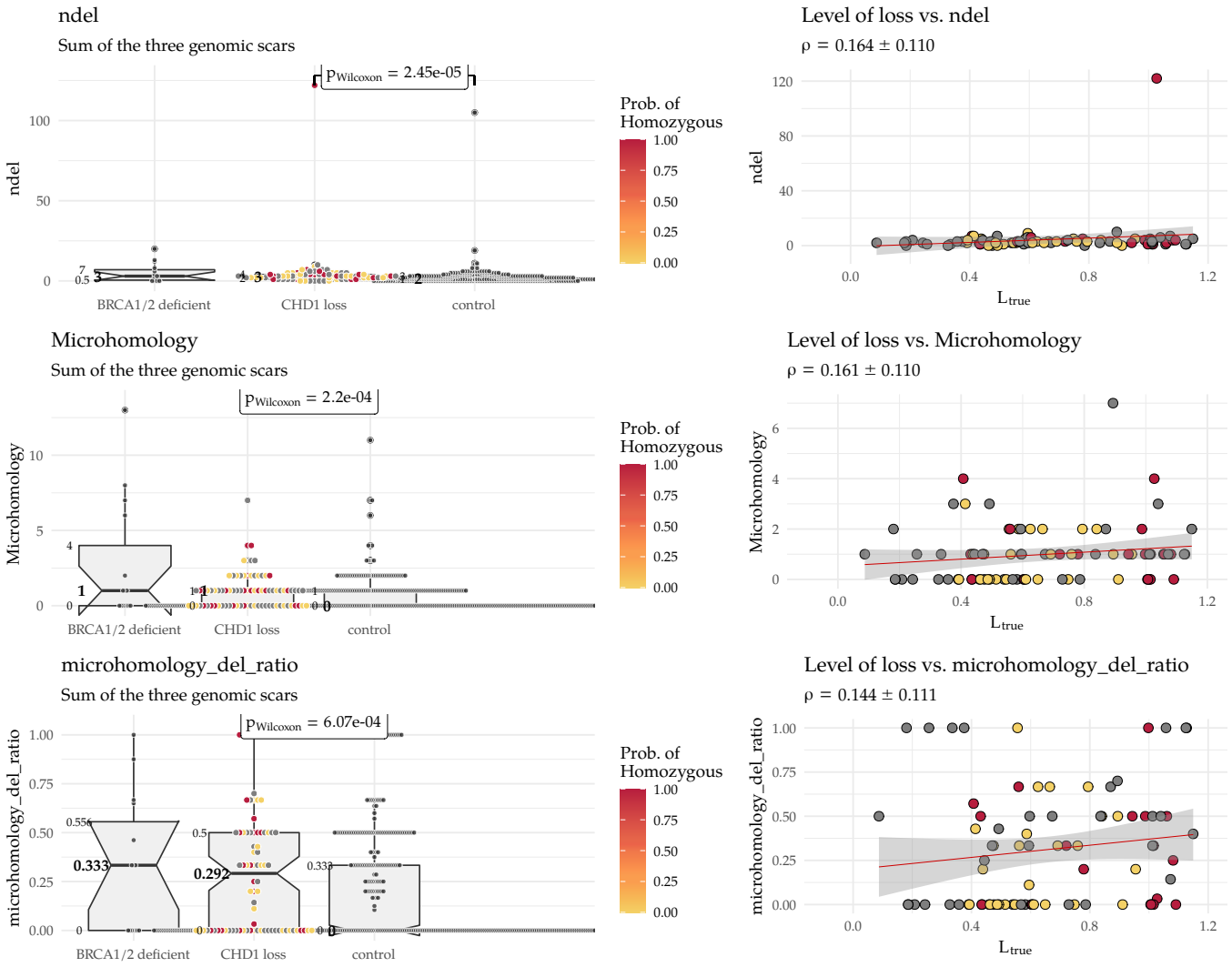
Absolute number of mutations contributable to numSNV



Level of loss vs. numSNV

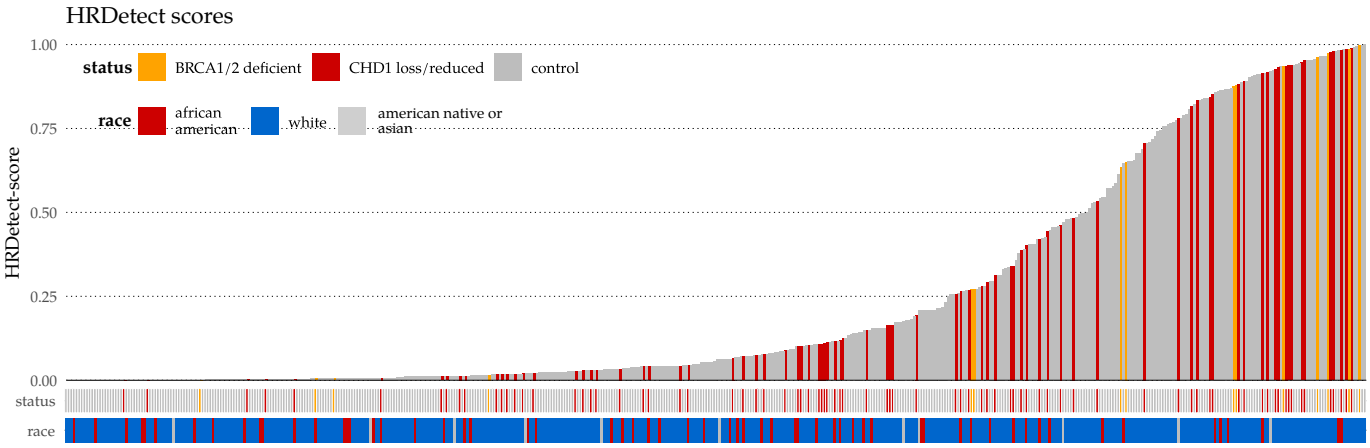
$\rho = 0.216 \pm 0.109$ ($p = 0.051$)



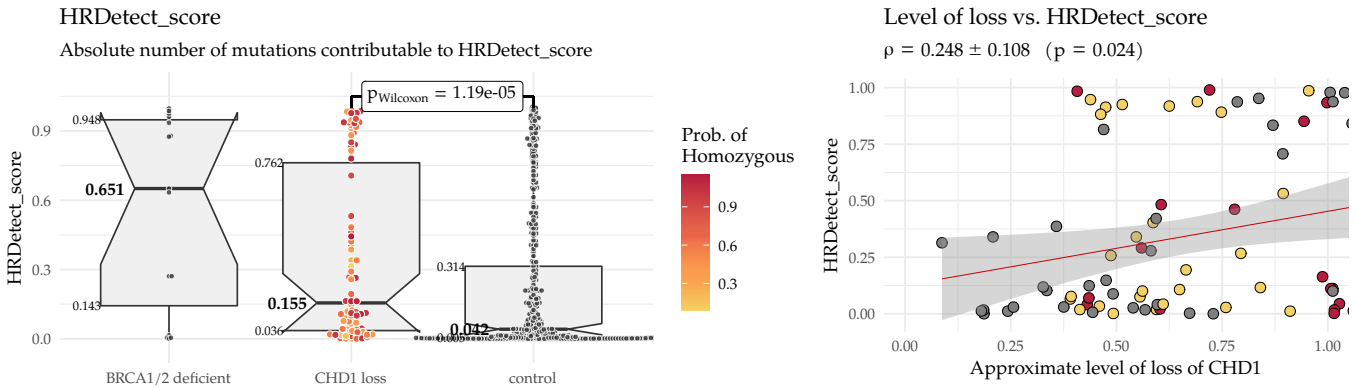


Supplementary Figure 42: The "traditional" classification of deletions. ndel: number of deletions, del_ins_ratio: deletion insertion ratio, microhomology: number of microhomology mediated deletions, microhomology_del_ratio: ratio of microhomology mediated deletions relative to the number of deletions – WES

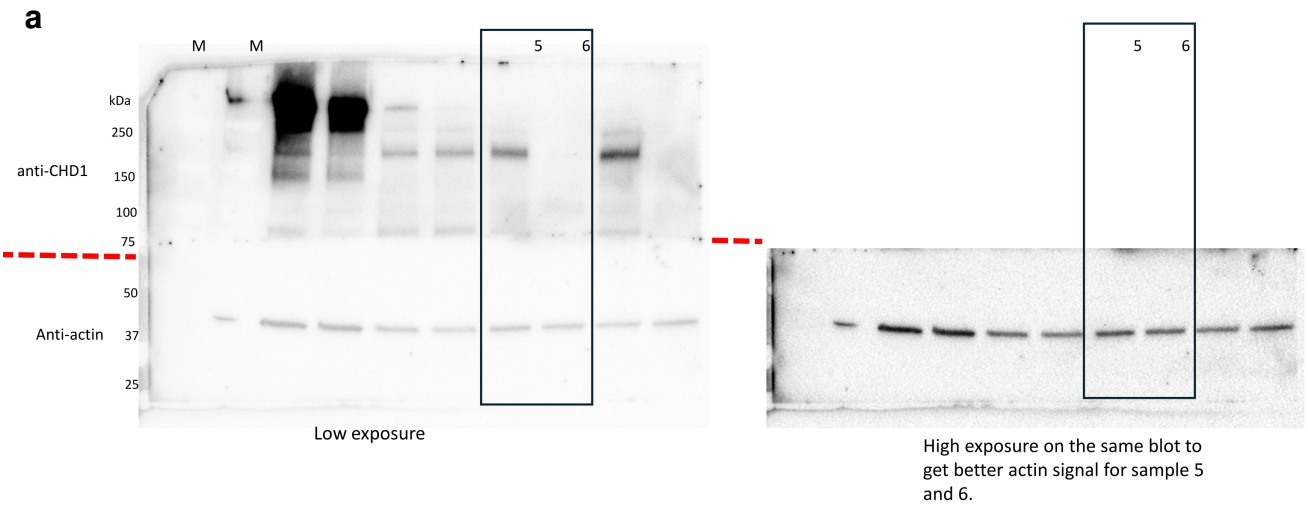
1.1 HRDetect

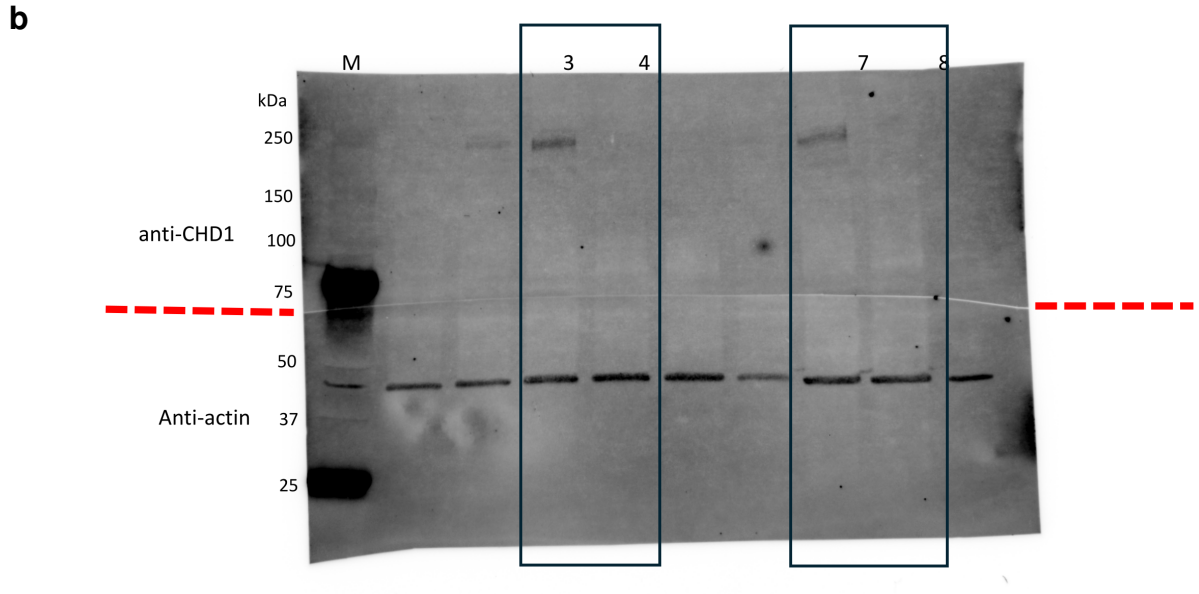


Supplementary Figure 43: HRDetect score summary – WES: For the whole exomes, we have used a different model, trained on 560 artificially derived (from whole genomes) breast cancer whole exomes: intercept = -2.619, Signature.17 = 0.0671, Signature.20 = 0.09409, Signature.26 = 0.1617, Signature.6 = 0.31, Signature.18 = 0.312, mhm.del.ratio = 0.314, Signature.8 = 0.615, Signature.13 = 0.8302, Signature.3 = 2.01, HRD-LOH = 2.387.

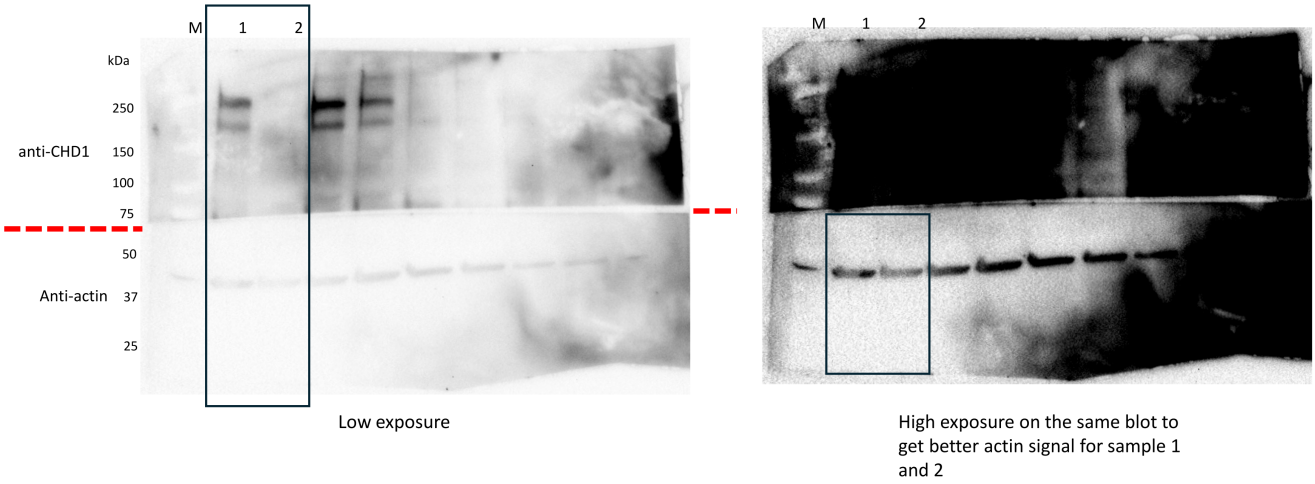


Supplementary Figure 44: HRDetect scores in samples with CHD1 loss vs. CHD1 intact samples

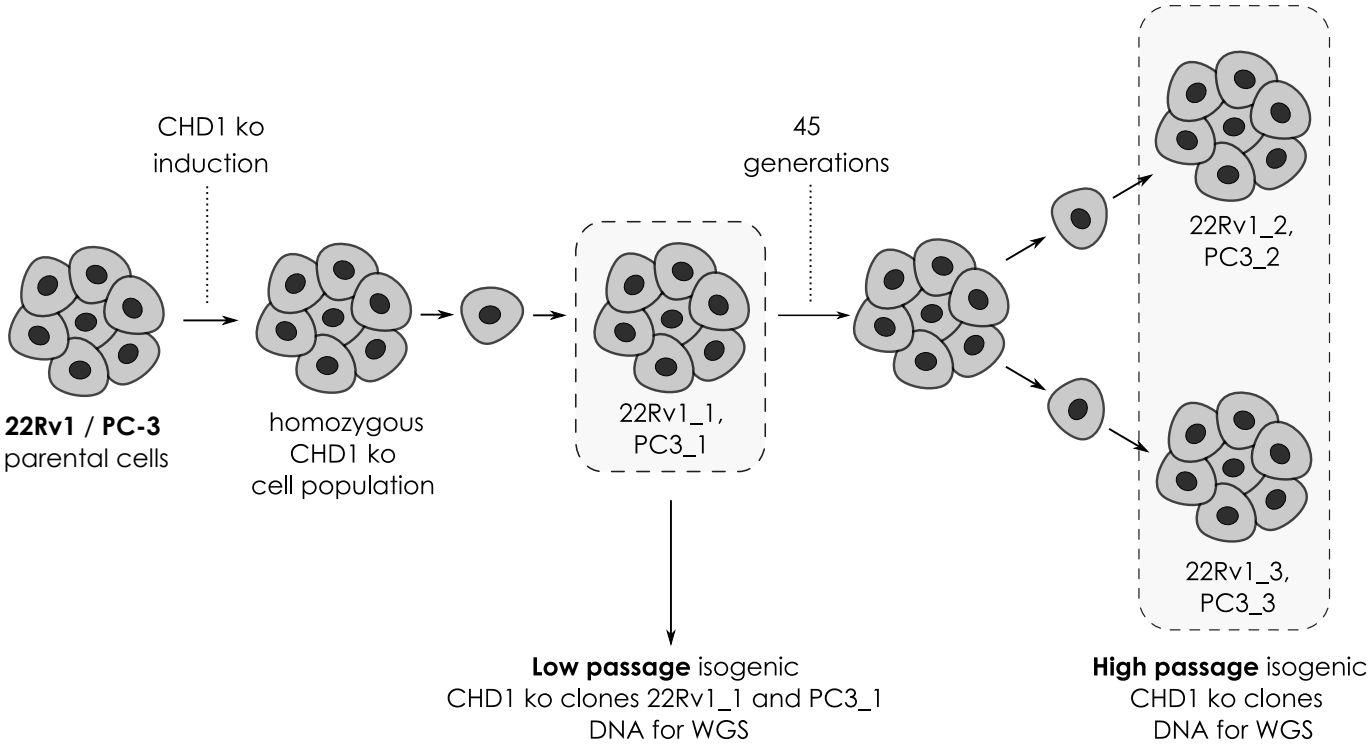




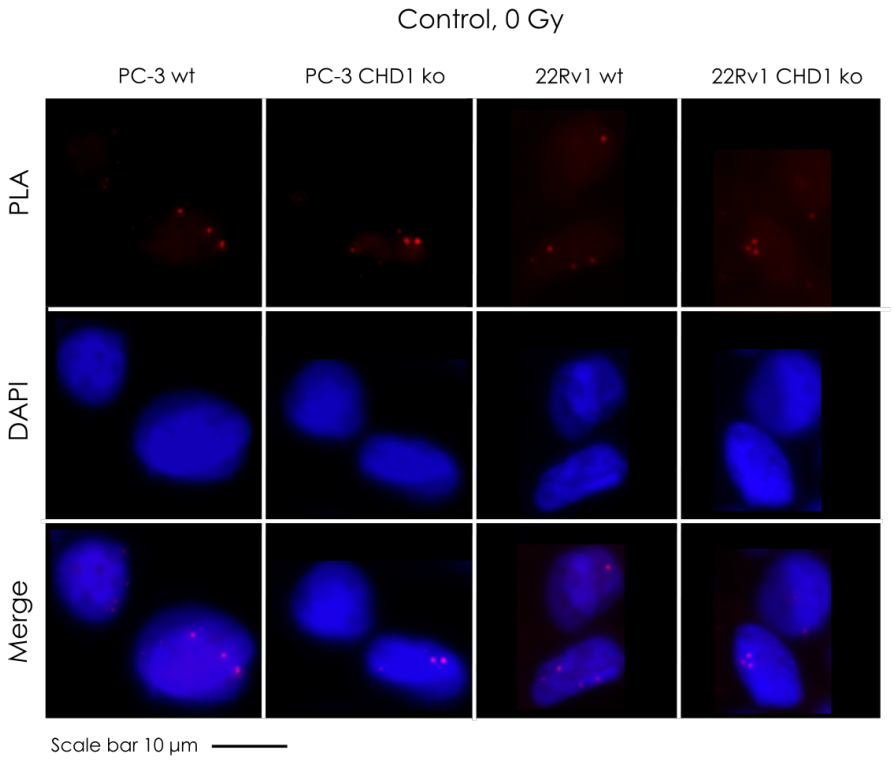
Supplementary Figure 45: a: Membrane cut between 75 and 50 KDa (red line) and incubated in anti-CHD1 (top half) and anti-Actin (bottom half) antibodies. Protein ladder (Precision Plus Protein Dual Color Standard (BioRad #1610374) marked as M. Samples 5 and 6 are presented in Figure 4g. **b:** Membrane cut between 75 and 50 KDa (red line) and incubated in anti-CHD1 (top half) and anti-Actin (bottom half) antibodies. Protein ladder (Precision Plus Protein Dual Color Standard (BioRad #1610374) marked as M. Samples 3 and 4 are presented in Figure 4h. Samples 7 and 8 are presented in Figure 4m.



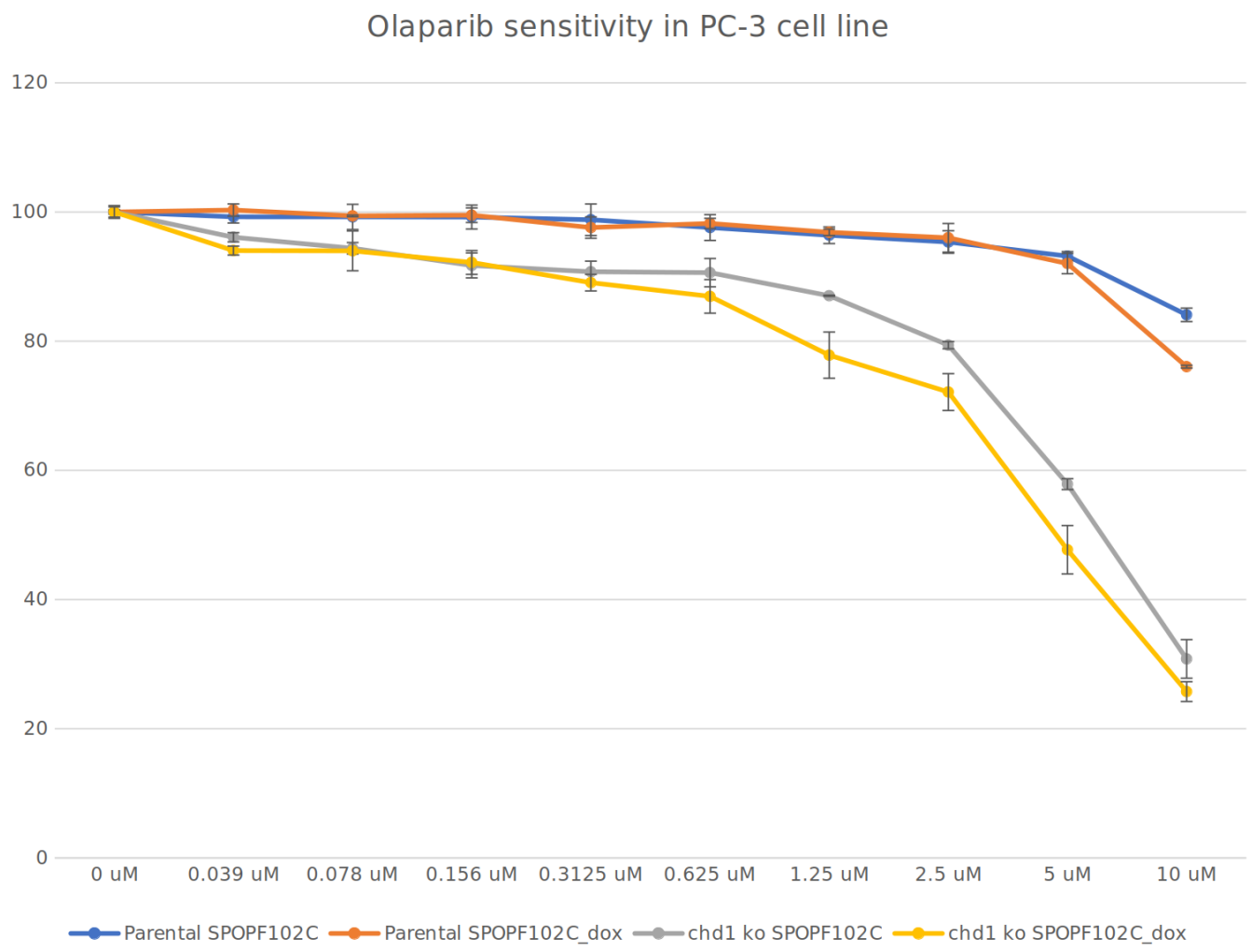
Supplementary Figure 46: Membrane cut between 75 and 50 KDa (red line) and incubated in anti-CHD1 (top half) and anti-Actin (bottom half) antibodies. Protein ladder (Precision Plus Protein Dual Color Standard (BioRad #1610374) marked as M. Samples 1 and 2 are presented in Figure 4n.



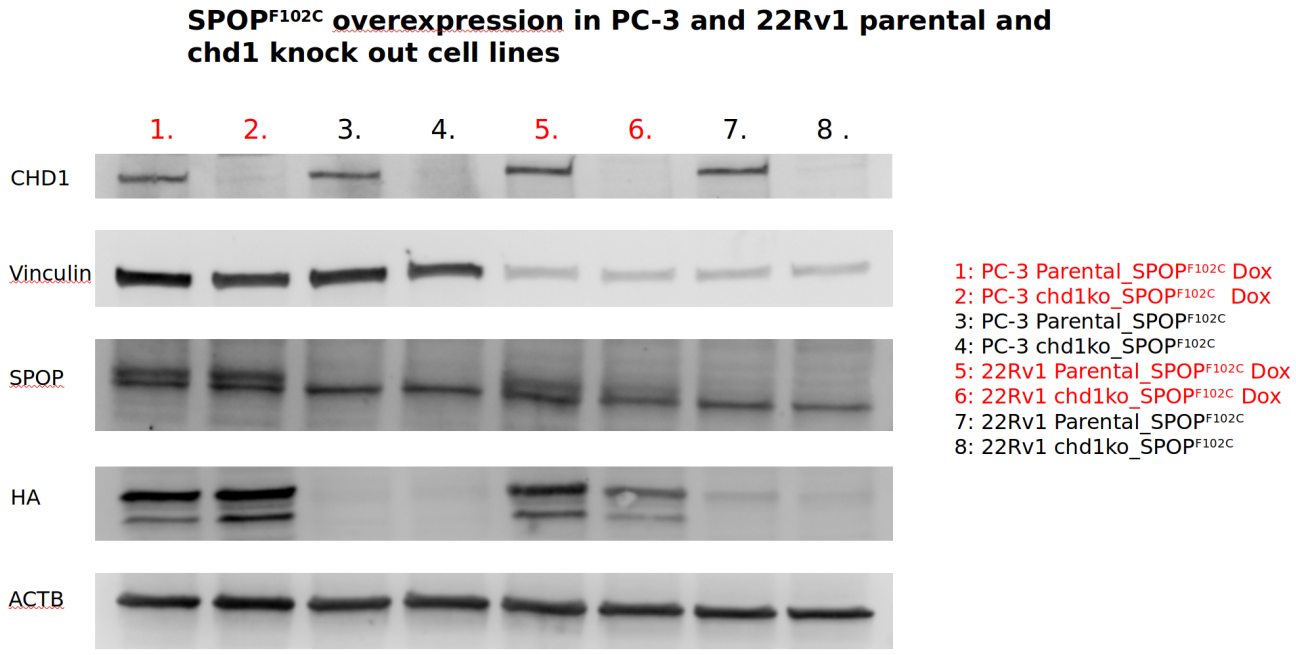
Supplementary Figure 47: Illustration of the PC-3 and 22Rv1 CHD1 ko isogenic clone generation for WGS. CHD1 knock out (ko) was induced in the parental PC-3 and 22Rv1 cell lines. The CHD1 ko cell populations were single cell cloned. Isogenic cell lines displaying homozygous CHD1 ko were identified. DNA was extracted directly from the regenerated population (22Rv1_1 and PC3_1, low passage stage). Cells were further propagated through 45 generations, then the high passage cell populations were single cell cloned. DNA was extracted from two isogenic CHD1 ko clones for each cell line (22Rv1_2, 22Rv1_3, and PC3_2, PC3_3 high passage stage) after propagation. The process is illustrated in Suppl. Figure 47. Normal references were downloaded from the Sequence Read Archive (SRA, 22Rv1: SRX5437595, PC-3: SRX5466646)



Supplementary Figure 48: RAD51 foci formation. Examples of the most common staining patterns in WT and CHD1 ko 22Rv1 and PC-3 cell lines. Control cells were fixed by 4% PFA without irradiation (IR=0Gy). PLA was carried out using antibodies against γH2Ax and RAD51 proteins.

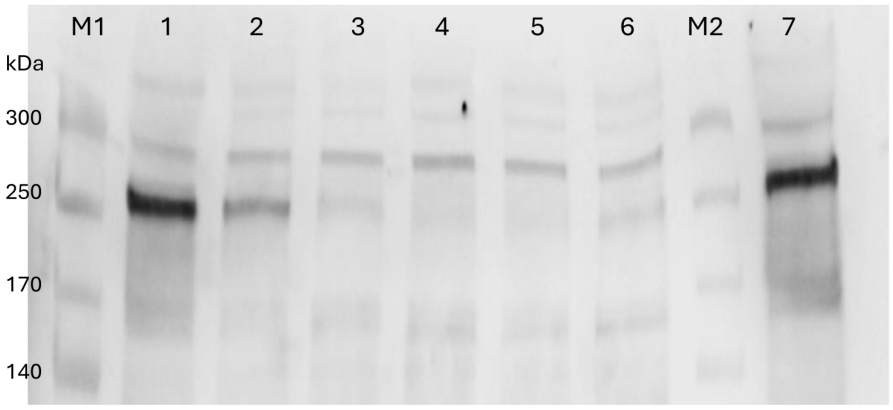


Supplementary Figure 49: Olaparib sensitivity in PC3 cell line

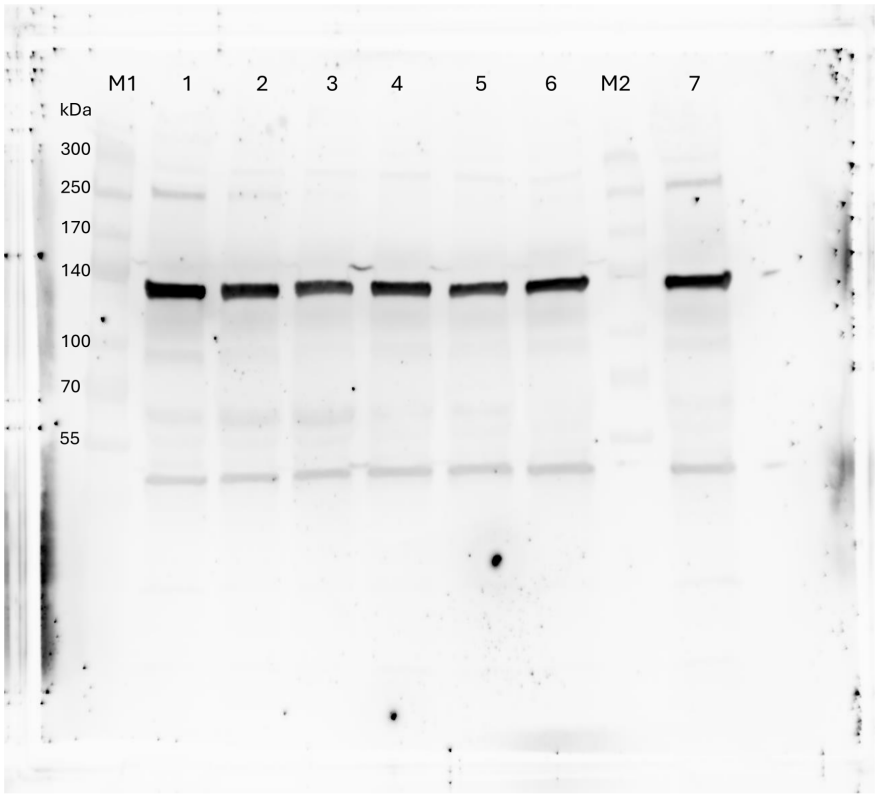


Supplementary Figure 50: *SPOP*^{F102C} overexpression in PC-3 and 22Rv1 parental and CHD1 knock out cell lines

a



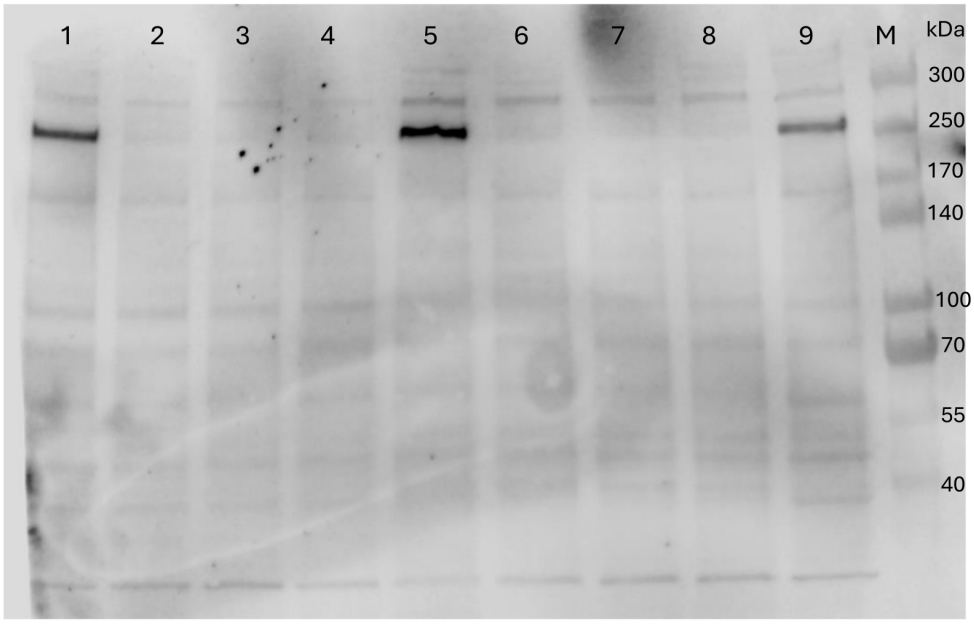
b



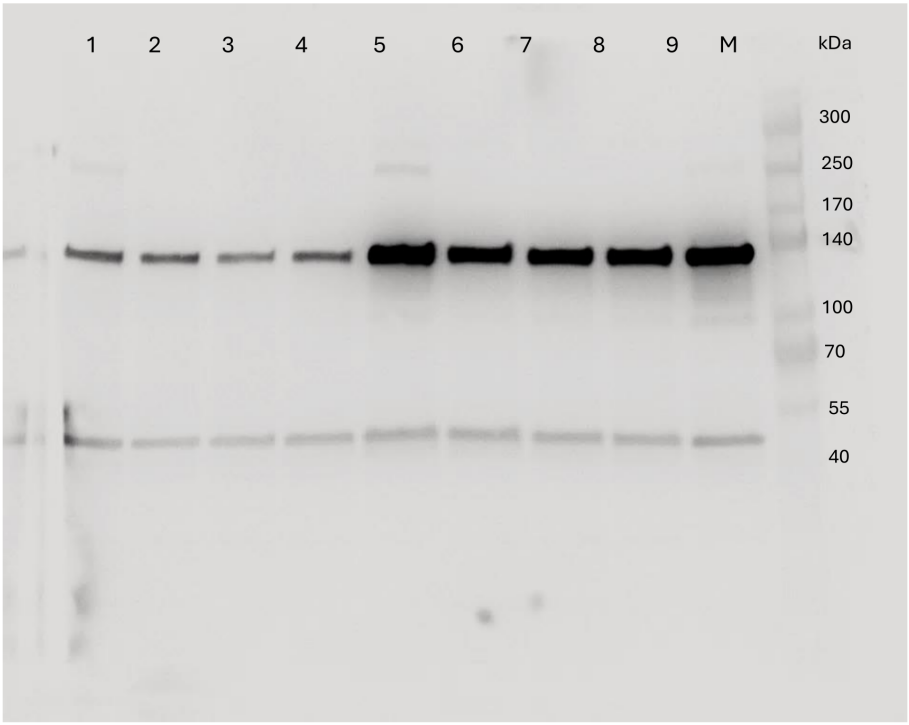
1: PC-3 Cas9 expressor wt parental
2: PC-3 del51 chd1 ko
3: PC-3 del6 chd1 ko
4: PC-3 del35 chd1 ko
5: PC-3 del62 chd1 ko
6: PC-3 del71 chd1 ko
7: PC-3 wt population
M1,2: ProSieve QuadColor Protein Marker (Lonza)

Supplementary Figure 51: a: Uncropped CHD1 (240 kDa) immunoblot Samples 1, 4 and 6 are presented in Figure 4a. **b:** Uncropped Vinculin (upper bands, 125 kDa) and Actin (lower bands, 42kDa) immunoblot. Samples 1, 4 and 6 are presented in Figure 4a.

a

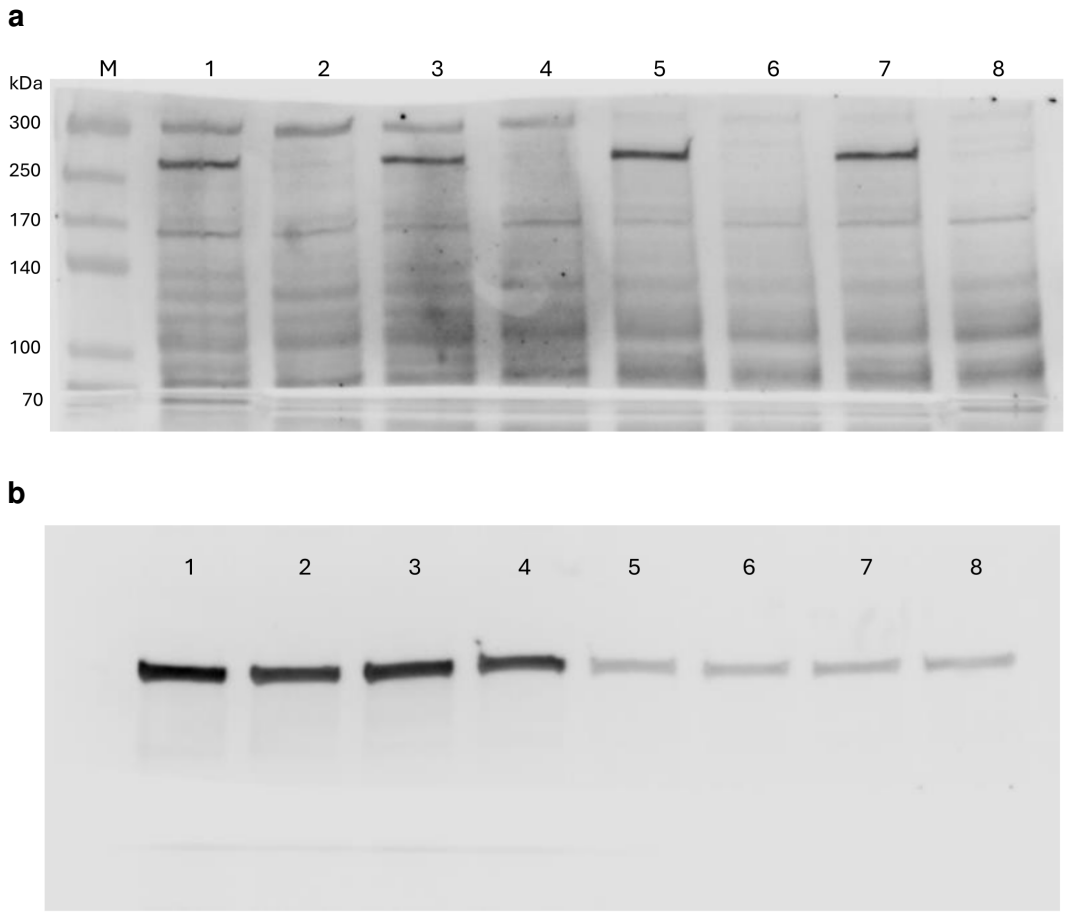


b



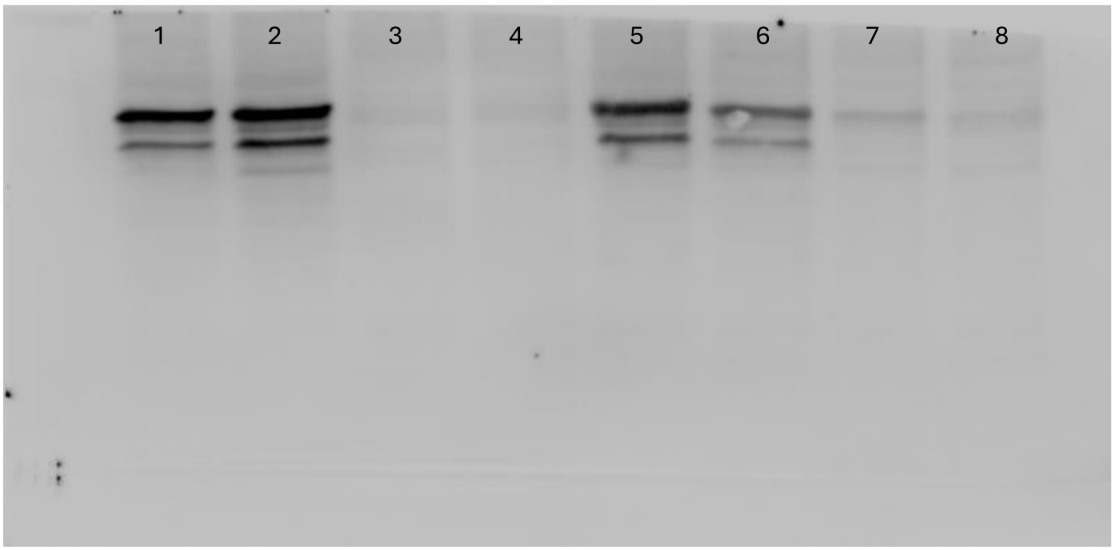
1: 22Rv1 wt
2: 22Rv1_chd1 ko P early
3: 22Rv1_chd1 ko P45_1
4: 22Rv1_chd1 ko P45_1
5: PC3/Cas9 Clone1
6: PC3_chd1 ko P early
7: PC3_chd1 ko P30_1
8: PC3_chd1 ko P30_2
9: RPE1/Cas9 P45_pop
M: ProSieve QuadColor Protein Marker (Lonza)

Supplementary Figure 52: a: Uncropped CHD1 (240 kDa) immunoblot. Samples 1 and 2 are presented in Figure 4b. **b:** Uncropped Vinculin (upper bands, 125 kDa) and Actin (lower bands, 42 kDa) immunoblot. Samples 1 and 2 are presented in Figure 4b.



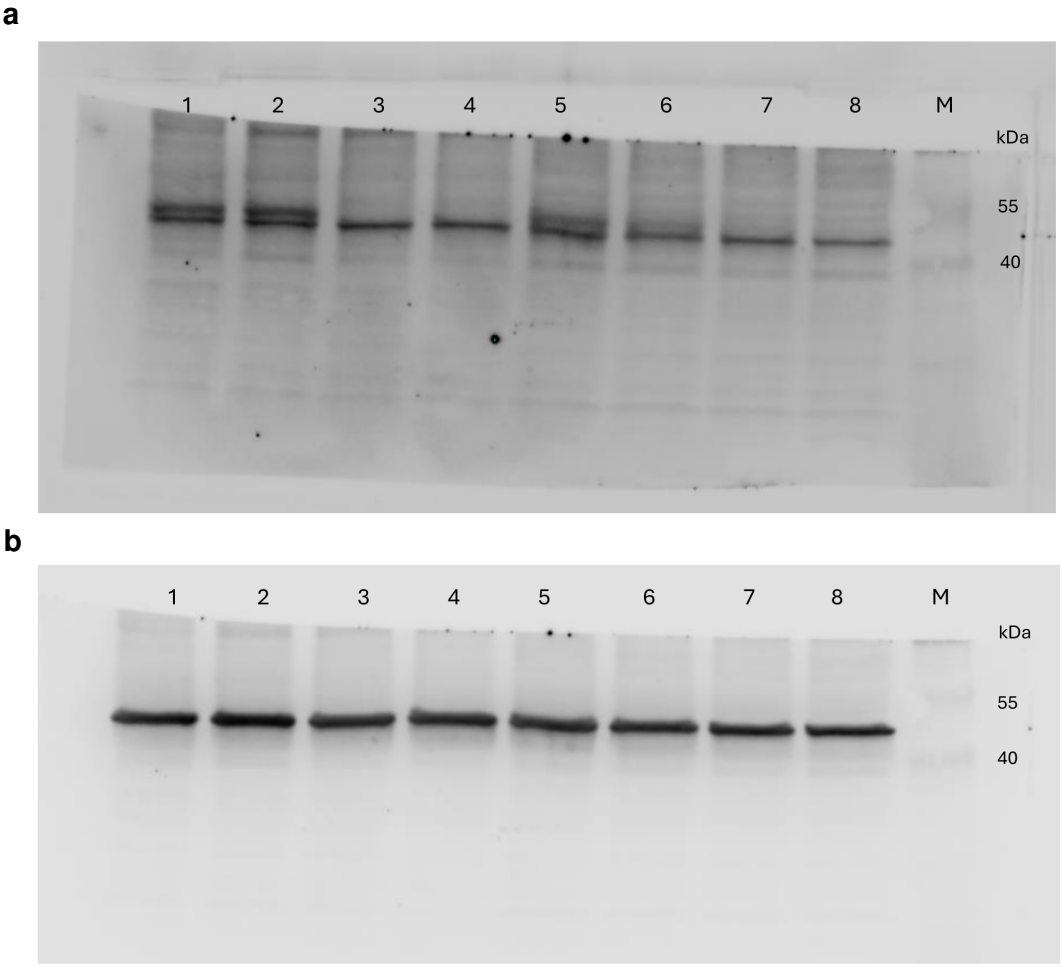
1: PC-3 Parental_SPOP^{F102C} Dox
2: PC-3 chd1ko_SPOP^{F102C} Dox
3: PC-3 Parental_SPOP^{F102C}
4: PC-3 chd1ko_SPOP^{F102C}
5: 22Rv1 Parental_SPOP^{F102C} Dox
6: 22Rv1 chd1ko_SPOP^{F102C} Dox
7: 22Rv1 Parental_SPOP^{F102C}
8: 22Rv1 chd1ko_SPOP^{F102C}
M: ProSieve QuadColor Protein Marker (Lonza)

Supplementary Figure 53: a: Uncropped CHD1 (240 kDa, panel **a**) and Vinculin (125 kDa, panel **b**) immunoblot. Samples 1-8 are presented in Supplementary Figure 48.



1: PC-3 Parental_SPOP^{F102C} Dox
2: PC-3 chd1ko_SPOP^{F102C} Dox
3: PC-3 Parental_SPOP^{F102C}
4: PC-3 chd1ko_SPOP^{F102C}
5: 22Rv1 Parental_SPOP^{F102C} Dox
6: 22Rv1 chd1ko_SPOP^{F102C} Dox
7: 22Rv1 Parental_SPOP^{F102C}
8: 22Rv1 chd1ko_SPOP^{F102C}

Supplementary Figure 54: Uncropped HA (SPOP+HA cca 43 kDa) immunoblot. Samples 1-8 are presented in Supplementary Figure 48.



1: PC-3 Parental_SPOP^{F102C} Dox
2: PC-3 chd1ko_SPOP^{F102C} Dox
3: PC-3 Parental_SPOP^{F102C}
4: PC-3 chd1ko_SPOP^{F102C}
5: 22Rv1 Parental_SPOP^{F102C} Dox
6: 22Rv1 chd1ko_SPOP^{F102C} Dox
7: 22Rv1 Parental_SPOP^{F102C}
8: 22Rv1 chd1ko_SPOP^{F102C}
M: ProSieve QuadColor Protein Marker (Lonza)

Supplementary Figure 55: Uncropped SPOP (42 kDa, upper) and Actin (42 kDa, lower) immunoblot. Samples 1-8 are presented in Supplementary Figure 48.