Methods: Next-generation sequencing

The mutational background was determined using targeted sequencing of approx. 10 Mb of exonic regions of cancer-associated genes. Genomic DNA libraries were prepared using the KAPA HTP library preparation kit, multiplexed prior to solution-based custom capture (Roche NimbleGen) and sequenced to reach ≥ 100× mean coverage (Illumina HiSeq 1500). All steps were carried out according to original protocols with minor modifications.

Variant discovery included the following steps: quality control of raw fastq, adapter trimming and removal of low quality reads using Trimmomatic [1], read mapping to hg19 using BWA [2], duplication removal, local realignment and quality recalibration using GATK and Picard and variant calling using UnifiedGenotyper, HaplotypeCaller [3] and Freebayes [4].

Variants were filtered using public (NHLBI ESP [5], gnomAD [6]) and internal databases in order to remove common genetic variation. Six bioinformatics methods were used to identify possible protein-damaging variants: CADD [7], PolyPhen2 [8], CHASM [9], SIFT [10], FATHMM [11] and Mutation Taster [12].

Copy-number alterations were detected by sequencing coverage analysis using CNVkit v0.9.5 [13]. CNVkit was run with default settings and a group of 18 control samples was used to create a reference coverage model. Chromosome arm-level heterozygosity alterations were analyzed across covered regions by plotting a chart of the observed absolute shifts of variant allele frequencies from 0.5 (heterozygosity) for all variants with allele frequencies between 0.0001 and 0.95 in gnomAD, called by HaplotypeCaller with coverage larger than 30×.

References

1. 10.1093/bioinformatics/btu170
2. 10.1093/bioinformatics/btp324
3. 10.1002/0471250953.bi1110s43
4. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907 [q-bio. GN] 2012
5. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/)
6. 10.1038/nature19057
7. 10.1038/ng.2892
8. 10.1002/0471142905.hg0720s76
9. 10.1093/bioinformatics/btr357
10. 1038/nprot.2009.86
11. 1002/humu.22225
12. 1038/nmeth.2890
13. 1371/journal.pcbi.1004873

Supplementary Table 1. List of genes targeted by custom design (RefSeq)

Gene name						
1-200	201-400	401-600	601-800	801-1000	1001-1200	1201-1203
AADACL3	CLP1	FOXL2	LRRK2	PHF16	SLC17A4	ZNF835
ABCA10	CLTC	FOXM1	LRRN3	PHF6	SLC22A4	ZRSR2
ABCA12	CLTCL1	FOXO1	LTBP2	PHOX2B	SLC45A3	ZSWIM5
ABCA3	CMAS	FOXO3	LTBP3	PICALM	SLC6A15	
ABCA5	CMC4	FOXO4	LY75-CD302	PIK3AP1	SLIT2	
ABCA7	CNGA4	FOXP1	LYL1	PIK3C2A	SLITRK3	
ABCB1	CNOT1	FOXQ1	LYN	PIK3C2B	SLTM	
ABCC6	CNOT3	FRG2C	MACF1	PIK3C2G	SLX4	
ABCG2	CNTNAP4	FSTL3	MAD2L1BP	PIK3CA	SMAD2	
ABI1	CNTRL	FUBP1	MAF	PIK3CB	SMAD3	
ABL1	COL1A1	FUS	MAFB	PIK3CG	SMAD4	
ABL2	COL27A1	FUT3	MAGEA6	PIK3IP1	SMARCA2	
ACCS	COL4A1	FYB	MAGEC3	PIK3R1	SMARCA4	
ACPL2	COL6A6	FZD8	MAGED1	PIK3R2	SMARCA5	
ACSL3	COL7A1	FZR1	MAGI1	PIK3R3	SMARCB1	
ACSL6	COX6C	GALNT3	MALT1	PIK3R4	SMC1A	
ACTB	CREB1	GAS7	MAML2	PIK3R5	SMC3	
ACTG1	CREB3L1	GATA1	MAP2K1	PIK3R6	SMO	
ACVR1B	CREB3L2	GATA2	MAP2K2	PIM1	SMOX	
ACVR2A	CREBBP	GATA3	MAP2K4	PINK1	SMYD4	
ADAM29	CRIPAK	GATA6	MAP3K1	PKD1L1	SNX29	
ADCK5	CRKL	GBF1	MAP3K8	PKHD1	SOCS1	
AFF1	CRLF2	GBGT1	MAP9	PKHD1L1	SOHLH2	
AFF2	CRTC1	GLMN	MAPK1	PLAG1	SON	
AFF3	CRTC3	GMPS	MAPK3	PLB1	SORCS1	
AFF4	CSF1R	GNA11	MAPK8IP1	PLEC	SORCS2	
AIDA	CSF3R	GNA13	MBD4	PLK1	SOX11	
AIP	CSMD1	GNAQ	MC1R	PLXNA1	SOX17	
AJUBA	CSMD3	GOLGA5	MC2R	PMEPA1	SOX2	
AK8	CSN3	GOLGA6L6	MCL1	PML	SOX9	
AKAP9	CTCF	GOLPH3	MDH1B	PMS1	SPECC1	
AKT1	CTNNA1	GOPC	MDM2	PNLIPRP1	SPEN	
AKT1S1	CTNNA3	GP1BB	MDM4	PNRC1	SPOP	
AKT2	CTNNB1	GPC3	MECOM	POLQ	SPOPL	
AKT3	CTSG	GPHN	MED1	POSTN	SRC	
ALDH2	CTTNBP2	GPM6A	MED12	POT1	SRCAP	
ALK	CUBN	GPR149	MED13	POTEE	SRD5A2	
ALMS1	CUL3	GPR37	MEF2B	POTEF	SRGAP3	
ALOX12B	CXCR4	GPR98	MEN1	POTEH	SRRM2	

Table 1. Cont.

Gene name					
ALOX15	CXCR7	GPRASP2	MET	POU2AF1	SRSF2
ALOX5	CYFIP1	GRB2	MFSD4	POU5F1	SS18
ALS2	CYLD	GRIN3A	MGA	PPARG	SS18L1
AMER1	CYP1B1	GSR	MGAM	PPEF2	SSTR1
AMER3	CYP2B6	GSTM1	MIR5088	PPFIA2	SSX1
ANTXR1	DAAM2	GUCY1A2	MITF	PPP2CA	STAG1
AP5M1	DAXX	H3F3B	MKL1	PPP2R1A	STAG2
APC	DCC	H3F3C	MLF1	PPT1	STAT1
AR	DCDC5	HCN3	MLH1	PRCC	STAT2
ARAF	DDB2	HDAC4	MLLT1	PRDM1	STAT3
ARHGAP30	DDIT3	HERC2	MLLT10	PRDM14	STAT4
ARHGAP35	DDR2	HERPUD1	MLLT3	PRDM2	STAT5A
ARHGEF12	DDX10	HEY1	MLLT4	PRDM9	STAT5B
ARID1A	DDX25	HGF	MLLT6	PREX2	STAT6
ARID1B	DDX3X	HIP1	MLST8	PRF1	STIL
ARID2	DDX41	HIST1H1C	MN1	PRKCD	STK11
ARID3A	DDX43	HIST1H1E	MNX1	PRKD1	STK4
ARID3B	DDX5	HIST1H2BD	MPEG1	PRKDC	STX11
ARID4A	DDX55	HIST1H4A	MPL	PRPF40B	STXBP2
ARID4B	DDX6	HIST1H4B	MPO	PRPF4B	SUFU
ARID5A	DDX60L	HIST1H4C	MRE11A	PRRX1	SULT1A1
ARID5B	DEK	HIST1H4D	MRPL36	PRX	SUZ12
ARNT	DENND1B	HIST1H4E	MRPS18B	PSIP1	SYK
ARSD	DENND5B	HIST1H4F	MS4A3	PTCH1	SYNE1
ASH1L	DEPDC1	HIST1H4H	MSH2	PTEN	SYNE2
ASPSCR1	DFFB	HIST1H4J	MSH6	PTK2	SYNE3
ASTN2	DHX30	HIST1H4K	MSI1	PTK2B	SYT15
ASXL1	DHX58	HIST1H4L	MSI2	PTPN11	TACSTD2
ASXL2	DHX9	HIST2H4A	MSN	PTPN14	TAF1
ASXL3	DICER1	HIST4H4	MST1	PTPN22	TAF11
ATAD2	DIS3	HLA-A	MST1L	PTPRA	TAF15
ATF1	DIS3L2	HLA-B	MTF2	PTPRB	TAF4
ATIC	DLEC1	HLA-C	MTOR	PTPRC	TAL1
ATM	DMD	HLF	MTUS2	PTPRD	TAL2
ATP10B	DNAH7	HMGA1	MUC1	PTPRE	TANK
ATR	DNAH9	HMGA2	MUC16	PTPRF	TBC1D10A
ATRX	DNHD1	HNF1A	MUC4	PTPRG	TBL1XR1
AURKA	DNM2	HNMT	MUC6	PTPRH	TBX3
AURKAIP1	DNMT1	HNRNPA2B1	MUSK	PTPRJ	TBX5
AXIN1	DNMT3A	HNRNPD	MUTYH	PTPRK	TCEA1

Table 1. Cont.

Gene name					
AXIN2	DNMT3B	HNRNPK	MYB	PTPRM	TCEB3B
AXL	DOCK7	HOOK3	MYBPC1	PTPRN	TCF12
B2M	DST	HOXA11	MYC	PTPRN2	TCF3
B4GALT3	DTX1	HOXA13	MYCL1	PTPRO	TCF4
BAG3	DYNC1H1	HOXA9	MYCN	PTPRQ	TCF7L2
BAI3	DYNC2H1	HOXC11	MYD88	PTPRR	TDRP
BAP1	DYSF	HOXD11	MYH11	PTPRS	TECTA
BARD1	EBF1	HOXD13	MYH13	PTPRT	TENM2
BAZ2B	ECM1	HRAS	MYH14	PTPRU	TERT
BCL10	ECT2L	HS3ST5	MYH9	PTPRZ1	TET1
BCL11A	EEF1A1	HS3ST6	MYO10	PYROXD2	TET2
BCL11B	EFCAB4B	HSD17B3	MYO3A	QTRT1	TFE3
BCL2	EGF	HSD3B2	MYOC	R3HDM1	TFEB
BCL2L1	EGFR	HSP90AA1	MYOCD	RABEP1	TFG
BCL3	EGLN1	HSP90AB1	NACA	RAD21	TFPT
BCL6	EGR1	HTR1A	NAIP	RAD51B	TFRC
BCL7A	EGR2	HTR3A	NALCN	RAD51C	TGFBR2
BCL7C	EGR3	HTRA4	NARG2	RADIL	THRAP3
BCL9	EHBP1	HVCN1	NAV3	RAF1	TLN2
BCOR	EHMT1	HYDIN	NBN	RALGDS	TLR2
BCORL1	EIF4A1	ID3	NCKAP5	RANBP17	TLR4
BCR	ELAC1	IDH1	NCKIPSD	RANBP2	TLX1
BIRC2	ELAC2	IDH2	NCOA1	RAP1GDS1	TLX3
BIRC3	ELANE	IDO2	NCOA2	RAP2B	TMEM127
BIRC5	ELF3	IFI30	NCOA3	RARA	TMEM132D
BIRC6	ELF4	IGDCC4	NCOA4	RASA1	TMEM42
BIRC7	ELK4	IGF1R	NCOR1	RAVER1	TMPRSS2
BIVM-ERCC5	ELL	IGLL5	NDRG1	RB1	TMSB4X
BLM	ELN	IKBKE	NDST4	RBL1	TMTC1
BMPR1A	EML4	IKZF1	NECAP1	RBM10	TNFAIP3
BNC2	ENPP1	IKZF2	NEFH	RBM15	TNFRSF14
BPTF	ENPP2	IKZF3	NF1	RBM34	TNFRSF17
BRAF	EP300	IL12RB1	NF2	RECQL4	TOP1
BRCA1	EPB41L3	IL1RAPL1	NFE2L2	REL	TOP2A
BRCA2	EPCAM	IL2	NFE2L3	RET	TP53
BRCC3	EPHA10	IL21R	NFIB	REV3L	TP63
BRD3	EPHA2	IL2RG	NIN	RFX7	TPM3
BRD4	EPHA3	IL6	NKX2-1	RGS17	TPM4
BRD7	EPHA4	IL6R	NONO	RGS7	TPRX1
BRD9	EPHA5	IL6ST	NOTCH1	RGS8	TRAF3
BRIP1	EPHA6	IL7R	NOTCH2	RHBDF2	TRIM24

Table 1. Cont.

Gene name					
BSN	EPHA7	INO80D	NOTCH3	RHEB	TRIM27
BTG1	EPHB6	IRF2	NOTCH4	RHOA	TRIM33
BTG2	EPPK1	IRF6	NOV	RHOH	TRIM37
BTK	EPS15	IRS2	NPM1	RICTOR	TRIM42
BTNL9	ERBB2	ITK	NR4A3	RIMBP2	TRIO
BUB1B	ERBB3	ITPKB	NR5A1	RIMS2	TRIP11
C11orf42	ERBB4	JAK1	NRAP	RIPK1	TROAP
C1orf127	ERCC1	JAK2	NRAS	RIPK4	TSC1
C7	ERCC2	JAK3	NRD1	RIT1	TSC2
CACNA1A	ERCC3	JAZF1	NRK	RMI2	TSHR
CACNA1D	ERCC4	JMJD1C	NRXN2	RNASEL	TSHZ2
CACNA1S	ERCC5	JPH2	NSD1	RNF17	TSHZ3
CAD	ERG	JUN	NTHL1	RNF2	TTC3
CALD1	ESRRB	KANK4	NTRK1	RNF207	TTC37
CALR	ETNK1	KARS	NTRK2	RNF213	TTL
CANT1	ETV1	KAT6A	NTRK3	RNF43	TTN
CARD11	ETV4	KAT6B	NUMA1	ROBO1	TUFM
CARS	ETV5	KCNH3	NUP160	ROBO2	TULP4
CASC5	ETV6	KCNH5	NUP214	ROBO3	TXNIP
CASK	EVPL	KCNJ3	NUP98	ROS1	TYK2
CASP8	EWSR1	KDM4B	NUTM1	RPA1	U2AF2
CBFA2T3	EXT1	KDM4C	NUTM2A	RPL10	UBE2A
CBFB	EXT2	KDM5A	NUTM2B	RPL22	UBE4A
CBL	EYA4	KDM5C	OBSCN	RPL5	UGT2B4
CBLB	EZH2	KDM6A	OCRL	RPN1	UHRF1BP1
CBLC	F8A2	KDR	OLIG2	RPS6KA3	ULK1
CCDC147	FAM105A	KDSR	OMD	RPTOR	UNC13C
CCDC168	FAM133A	KEAP1	OPTN	RREB1	UNC13D
CCDC169-SOHLH2	FAM135B	KIAA0195	OR10A2	RRM1	USH2A
CCDC6	FAM46C	KIAA0430	OR12D3	RTL1	USP24
CCDC88A	FAM5B	KIAA1109	OR2T33	RTP2	USP36
CCKBR	FAM5C	KIAA1468	OR4C16	RUNX1	USP6
CCNB1IP1	FAM71B	KIAA1549	OR4C3	RUNX1T1	USP9X
CCND1	FAM89A	KIAA1731	OR4D6	RUNX2	UTRN
CCND2	FANCA	KIF1B	OR4P4	RUNX3	UTY
CCNE1	FANCB	KIT	OR51G2	RXFP3	VCAM1
CCR7	FANCC	KL	OR51I1	RYR1	VCX3B
CD274	FANCD2	KLF5	OR52A1	RYR2	VEZF1
CD302	FANCE	KLHL24	OR5M8	RYR3	VHL
CD3EAP	FANCF	KLHL6	OSBPL6	SAMHD1	VPREB1

Table 1. Cont.

Gene name					
CD70	FANCG	KLK2	OTOF	SARDH	VPS13B
CD79A	FANCI	KMT2A	OTX2	SATB2	VRK2
CD79B	FANCL	KMT2B	P2RY8	SBDS	VTI1A
CDA	FANCM	KMT2C	PABPC3	SCAF1	WAS
CDC27	FASN	KMT2D	PAFAH1B2	SCAI	WDR36
CDC42BPA	FAT1	KRAS	PAIP1	SCN1A	WHSC1
CDC73	FAT2	KRT15	PAK7	SDHA	WHSC1L1
CDH1	FAT3	KRT17	PALB2	SDHAF2	WIF1
CDH11	FAT4	KRT33B	PALLD	SDHB	WRB
CDH6	FBN1	KRT35	PAPPA	SDHC	WRN
CDH9	FBXO10	KRT4	PARD3B	SEMA5A	WT1
CDK12	FBXO11	KRT7	PARP4	SENP3	XIAP
CDK4	FBXW7	KRTAP9-4	PATZ1	SEPT6	XIRP2
CDK6	FCGR2B	KTN1	PAX2	SEPT9	XPA
CDK8	FCRL4	L1CAM	PAX3	SET	XPC
CDKN1A	FEV	LAMA1	PAX5	SETBP1	XPO1
CDKN1B	FGF23	LAMA2	PAX6	SETD1A	XRCC1
CDKN1C	FGFR1	LASP1	PAX7	SETD1B	YWHAE
CDKN2A	FGFR1OP	LCK	PAX8	SETD2	ZBTB16
CDKN2C	FGFR2	LCP1	PBRM1	SETDB1	ZCCHC8
CDX2	FGFR3	LHFP	PBX1	SETDB2	ZEB1
CEBPA	FGFR4	LIFR	PCBP1	SF1	ZFHX4
CEL	FH	LILRA2	PCDHB13	SF3A1	ZFP36L1
CELSR3	FHIT	LILRB1	PCDHB3	SF3B1	ZIM2
CEP112	FIP1L1	LMCD1	PCDHGB3	SF3B2	ZMYM1
CEP57	FKBP1B	LMO1	PCDHGC5	SFPQ	ZMYM2
CGNL1	FKBP9	LMO2	PCLO	SFRP4	ZMYM3
CHD3	FLCN	LOC100862671	PCM1	SGCG	ZNF169
CHD4	FLG	LOC440243	PCSK7	SGK1	ZNF175
CHD5	FLNC	LOC728637	PDCD1LG2	SGK223	ZNF217
CHD6	FLT1	LPA	PDE4DIP	SH3BP1	ZNF276
CHEK1	FLT3	LPHN2	PDGFB	SH3GL1	ZNF331
CHEK2	FLT4	LPHN3	PDGFRA	SHC1	ZNF384
CHIC2	FN1	LPPR2	PDGFRB	SHC2	ZNF460
CHN1	FNBP1	LPPR3	PDPK1	SHC3	ZNF521
CHPF2	FOLR1	LRBA	PDZD2	SI	ZNF668
CHST10	FOXA1	LRCOL1	PER1	SIN3A	ZNF681
CIC	FOXA2	LRP1B	PGBD1	SKA3	ZNF716
CIITA	FOXC1	LRP4	PGK2	SLAIN1	ZNF717
CILP2	FOXF2	LRP5	PHEX	SLC16A4	ZNF827