Characterization of gene mutations and copy number changes in acute myeloid leukemia using a rapid target enrichment protocol

Niccolò Bolli,[1,2,3] Nicla Manes,[3,4] Thomas McKerrell,[4] Jianxiang Chi,[5] Naomi Park,[6] Gunes Gundem,[1] Michael A. Quail,[6] Vijitha Sathiaseelan,[1] Bram Herman,[7] Charles Crawley,[3] Jenny I. O. Craig,[3] Natalie Conte,[4,8] Carolyn Grove,[4] Elli Papaemmanuil,[1] Peter J. Campbell,[1] Ignacio Varela,[9] Paul Costeas,[5,10] and George S. Vassiliou[4]

[1]Cancer Genome Project, Wellcome Trust Sanger Institute, Cambridge, UK; [2]Department of Haematology, University of Cambridge, UK; [3]Department of Haematology, Addenbrookes Hospital, Cambridge, UK; [4]Haematological Cancer Genetics, Wellcome Trust Sanger Institute, Cambridge, UK; [5]The Center for the Study of Haematological Malignancies, Nicosia, Cyprus; [6]Sequencing Research and Development, Wellcome Trust Sanger Institute, Cambridge, UK; [7]Agilent Technologies, Agilent Technologies LDA UK Ltd., Cheadle, UK; [8]EMBL-European Bioinformatics Institute, Cambridge, UK; [9]Instituto de Biomedicina y Biotecnología de Cantabria (CSIC-UC-Sodercan), Departamento de Biología Molecular, Universidad de Cantabria, Santander, Spain; and [10]Molecular Haematology and Immunogenetics Center, The Karaiskakio Foundation, Nicosia, Cyprus

©2015 Ferrata Storti Foundation. This is an open-access paper. *doi:10.3324/haematol.2014.113381*
Manuscript received on July 8, 2014. Manuscript accepted on November 4, 2014.
Correspondence: nb8@sanger.ac.uk or gsv20@sanger.ac.uk

Supplementary Methods

Sequence variant detection and filtering criteria

Base substitutions and small insertions or deletions were identified by comparison of 42 MDS samples against unmatched normal samples using established bioinformatics algorithms[1-3]. To account for the absence of matched control a bespoke variant selection pipeline was developed. Each putative variant was annotated using the following resources:

1. Known constitutional polymorphisms using known human variation databases, Ensembl GRCh37.5, 1000 genomes release 2.2.2 and ESP6500[4 5].
2. Known somatic variation in myeloid and other common malignancies as reported in COSMIC v67[6].
3. Exome or whole genome sequencing data derived from 317 constitutional DNA samples analyzed in CGP (CGP normal panel).
4. Sequence context 5' and 3' to the reported sequence change highlighting regions of homopolymer sequences that are prone to PCR slippage and artifacts altering the last base of the homopolymer or inserting the same base as the homopolymer at +1, +2 of the track.
5. Variant specific metrics to include protein annotation, sequence depth and % of reads reporting the variant allele.

To enrich for high-confidence somatic variants that impact on protein function further filtering was conducted using the following criteria:

1. Removal of all variants with a predicted effect of a silent amino acid change on all transcripts corresponding to each gene.
2. Removal of known polymorphisms present in either of the human variation databases at a population frequency > 0.0014 (reflecting the population incidence of myeloid disease and potentially rare variants that could be associated with myeloid malignancies) unless variant is present as confirmed somatic mutation in COSMIC.
3. Removal of known polymorphisms present in human variation databases at a population frequency < 0.0014 and also represented in the extended normal CGP panel, available form in house exome and whole genome sequencing projects.
4. Retention of all variants present in human variation databases at a population frequency < 0.0014 and also present in COSMIC as confirmed somatic in Haematopoietic tissue.
5. Removal of all sequence variants that were represented in at least 2 normal individuals in the CGP normal panel with a minimum variant allele proportion of 10%.
6. Removal of variants present within regions prone to sequence context specific artifacts, including regions of high depth, enriched for reads of low mapping quality that harbor multiple mismatches.
7. Removal of all 1bp insertions or deletions present adjacent to regions of more than 5 homopolymer bases (i.e insA adjacent to AAAAA) and a variant allele proportion of < 12% and evidence of occurrence in CGP normal panel;

Once low confidence or likely polymorphisms were removed from the dataset, likely oncogenic were annotated and selected for the study among the shortlist of high confidence variants in accordance to prior evidence in the literature. To reflect the confidence that one would use these as diagnostic biomarkers in the clinic, variants were annotated conservatively, so that we only reported known oncogenic variants previously reported in the literature, or novel variants that cluster with known somatic variants in cancer driver genes, or truncating variants (nonsense mutations, essential splice mutations or frameshift indels) in genes implicated in myeloid malignancies through acquisition of loss of function mutations.

Validation

Copy number alterations of KRAS and BCOR were validated on genomic DNA with SYBRgreen quantitative PCR using the ACTB gene as endogenous control, and applying the $\Delta\Delta$CT method to perform a relative quantification[7]. Furthermore, the copy number pattern identified by NGS in sample PD17946a was validated using the Agilent SurePrint G3 ISCA CGH+SNP Microarray. MLL-PTDs were validated by long range PCR as described in[8]. FLT3-ITDs were assessed on genomic DNA by PCR followed by either agarose gel electrophoresis or Bioanalyzer using a high sensitivity analysis kit (Agilent Technologies) for 40 samples. NPM1 exon 12 mutations were validated in 33 samples using genomic DNA PCR followed by capillary sequencing. All primer sequences are provided in Supplementary Table 1.

Supplementary References

1. Bolli N, Avet-Loiseau H, Wedge DC, Van Loo P, Alexandrov LB, Martincorena I, et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nature Communications*. 2014;5:2997.
2. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood*. 2013;122(22):3616-27.
3. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009;25(21):2865-71.
4. Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2014. *Nucleic Acids Res*. 2014;42(Database issue):D749-55.
5. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56-65.
6. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res*. 2011;39(Database issue):D945-50.
7. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. 2001;25(4):402-8.
8. Schnittger S, Kinkelin U, Schoch C, Heinecke A, Haase D, Haferlach T, et al. Screening for MLL tandem duplication in 387 unselected patients with AML identify a prognostically unfavorable subset of AML. *Leukemia*. 2000;14(5):796-804.

Supplementary Table S1. Primers used for PCR validation

name	Sequence (5'-3')
BCOR_ex1_F	TTTAGCACAGTCCTCCACCCCA
BCOR_ex1_R	CATTCCGTTCAAACCCAGCAGC
BCOR_ex4_F	CGGAAGACAGCGGTTCAAGACA
BCOR_ex4_R	GTATCGCCCAGTCCAATGCCTT
ACTB_ex3_F	GGAAGGAAGGCTGGAAGAGTGC
ACTB_ex3_R	TGTGCTATCCCTGTACGCCTCT
KRAS_ex3_F	CACTACCGATGCAGTCTGGAGC
KRAS_ex3_R	GGACTGGGGAGGGCTTTCTTTG
NPM1_F	ATTGGCCATATGGGTCTCTG
NPM1_R	AACACGGTAGGGAAAGTTCTCA
FLT3-ITD_F	GCAATTTAGGTATGAAAGCCAGC
FLT3-ITD_R	CTTTCAGCATTTTGACGGAACC
MLL-6.1	GTCCAGAGCAGAGCAAACAG
MLL-2.0	CGCACTCTGACTTCTTCATC

Supplementary Table S2. Variants identified in the study and their validation

Algorithm	Sample	CHR	START	END	Gene	Transcript	Protein	Effect	Validation method	Validation outcome
Pindel	PD17929a	13	28608280	28608281	FLT3	CCDS31953.1	p.D600_L601insDFREYEYD	frameshift	PCR + agarose gel	validated
Pindel	PD17929a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Pindel	PD17929a	13	28608290	28608298	FLT3	CCDS31953.1	p.E598_Y599insNEYFYVDFREYE	frameshift	PCR + agarose gel	validated
Caveman	PD17929a	20	31022938	31022938	ASXL1	CCDS13201.1	p.P808H	missense	MiSeq	validated
Caveman	PD17929a	2	209113112	209113112	IDH1	CCDS2381.1	p.R132L	missense	MiSeq	validated
Pindel	PD17930a	2	25463299	25463300	DNMT3A	CCDS33157.1	p.E733fs*1	frameshift	MiSeq	validated
Pindel	PD17930a	13	28608271	28608272	FLT3	CCDS31953.1	p.K602_W603insEYEYDLK	frameshift	PCR + agarose gel	validated
Pindel	PD17930a	13	28608274	28608275	FLT3	CCDS31953.1	p.E608_N609insYEYDLKWEFPRE	frameshift	PCR + agarose gel	validated
Caveman	PD17930a	2	209113113	209113113	IDH1	CCDS2381.1	p.R132S	missense	MiSeq	validated
Pindel	PD17930a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Pindel	PD17931a	13	28608286	28608287	FLT3	CCDS31953.1	p.Y597_E598insDYVDFREY	frameshift	PCR + agarose gel	validated
Caveman	PD17931a	2	25457242	25457242	DNMT3A	CCDS33157.1	p.R882H	missense	MiSeq	validated
Caveman	PD17931a	4	106180928	106180928	TET2	CCDS47120.1	p.?	ess splice	MiSeq	validated
	PD17931a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Pindel	PD17932a	13	28608280	28608281	FLT3	CCDS31953.1	p.D600_L601insFREYEYD	frameshift	PCR + agarose gel	validated
Caveman	PD17932a	2	25467449	25467449	DNMT3A	CCDS33157.1	p.G543C	missense	MiSeq	validated
Caveman	PD17932a	4	106156570	106156570	TET2	CCDS47120.1	p.Q491K	missense	MiSeq	No coverage
	PD17932a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Pindel	PD17932c	13	28608280	28608281	FLT3	CCDS31953.1	p.D600_L601insFREYEYD	frameshift	PCR + agarose gel	validated
Caveman	PD17932c	2	25467449	25467449	DNMT3A	CCDS33157.1	p.G543C	missense	MiSeq	validated
	PD17932c	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	not confirmed
Caveman	PD17933a	2	25463182	25463182	DNMT3A	CCDS33157.1	p.R771*	nonsense	MiSeq	validated
Caveman	PD17933a	19	33792981	33792981	CEBPA	ENST00000498907	p.G114C	missense	MiSeq	validated

	PD17933a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Caveman	PD17934b	2	25457243	25457243	DNMT3A	CCDS33157.1	p.R882C	missense	SureSelect + NGS	validated
Caveman	PD17934b	13	28592642	28592642	FLT3	CCDS31953.1	p.D835Y	missense	SureSelect + NGS	validated
	PD17934b	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Caveman	PD17935a	4	106158455	106158455	TET2	CCDS47120.1	p.L1119*	nonsense	SureSelect + NGS	validated
Caveman	PD17935a	12	25378647	25378647	KRAS	CCDS8703.1	p.K117N	missense	SureSelect + NGS	validated
Caveman	PD17936a	15	90631934	90631934	IDH2	CCDS10359.1	p.R140Q	missense	SureSelect + NGS	validated
Pindel	PD17936a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	PCR failure
Pindel	PD17936c	13	28608280	28608281	FLT3	CCDS31953.1	p.D600_L601insFREYEYD	frameshift	PCR + agarose gel	not confirmed
Caveman	PD17936c	15	90631934	90631934	IDH2	CCDS10359.1	p.R140Q	missense	SureSelect + NGS	validated
Caveman	PD17936c	13	28602340	28602340	FLT3	CCDS31953.1	p.N676K	missense	SureSelect + NGS	validated
Pindel	PD17936c	13	28608288	28608302	FLT3	CCDS31953.1	p.E608_N609insDNEYFYVDFREYEYDLKWEFPRE	frameshift	PCR + agarose gel	not confirmed
Pindel	PD17936c	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Pindel	PD17937c	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Caveman	PD17937c	2	209113112	209113112	IDH1	CCDS2381.1	p.R132H	missense	SureSelect + NGS	validated
Caveman	PD17937c	13	28592642	28592642	FLT3	CCDS31953.1	p.D835Y	missense	SureSelect + NGS	validated
Caveman	PD17937c	1	115258748	115258748	NRAS	CCDS877.1	p.G12S	missense	SureSelect + NGS	validated
Pindel	PD17938a	13	28608308	28608309	FLT3	CCDS31953.1	p.E598_Y599insCRSSDNEYFYVDFREYE	frameshift	PCR + agarose gel	validated
Caveman	PD17938a	2	25457242	25457242	DNMT3A	CCDS33157.1	p.R882H	missense	MiSeq	validated
Caveman	PD17938a	12	112940014	112940014	PTPN11	CCDS9163.1	p.D556Y	missense	MiSeq	No coverage
	PD17938a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Pindel	PD17939a	19	33792393	33792394	CEBPA	ENST00000498907	p.E309_T310insN	inframe	SureSelect + NGS	validated
Pindel	PD17939a	11	32456252	32456254	WT1	CCDS7878.2	p.N214fs*36	frameshift	SureSelect + NGS	validated
Pindel	PD17939a	13	28608288	28608291	FLT3	CCDS31953.1	p.Y597_E598insDFYVDFREY	frameshift	PCR + agarose gel	validated
Caveman	PD17939a	1	115258747	115258747	NRAS	CCDS877.1	p.G12D	missense	SureSelect + NGS	validated
	PD17939a	19	33792982	33792983	CEBPA	ENST00000498907_r69	p.A111fs*56	frameshift	SureSelect + NGS	validated
Caveman	PD17940a	20	31022297	31022297	ASXL1	CCDS13201.1	p.C594*	nonsense	SureSelect + NGS	validated

Caveman	PD17940a	12	25398284	25398284	KRAS	CCDS8703.1	p.G12V	missense	SureSelect + NGS	validated
Pindel	PD17941a	2	25467105	25467109	DNMT3A	CCDS33157.1	p.G590fs*61	frameshift	SureSelect + NGS	validated
Caveman	PD17941a	15	90631838	90631838	IDH2	CCDS10359.1	p.R172K	missense	SureSelect + NGS	validated
Pindel	PD17942a	13	28608286	28608287	FLT3	CCDS31953.1	p.Y597_E598insDYVDFREY	frameshift	PCR + agarose gel	validated
Caveman	PD17942a	2	25457242	25457242	DNMT3A	CCDS33157.1	p.R882H	missense	MiSeq	validated
Caveman	PD17942a	2	209113113	209113113	IDH1	CCDS2381.1	p.R132C	missense	MiSeq	validated
	PD17942a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Caveman	PD17943a	15	90631934	90631934	IDH2	CCDS10359.1	p.R140Q	missense		
	PD17943a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Pindel	PD17944a	20	31022412	31022413	ASXL1	CCDS13201.1	p.H633fs*2	frameshift	MiSeq	validated
Caveman	PD17945a	1	115256536	115256536	NRAS	CCDS877.1	p.A59S	missense	MiSeq	not confirmed
Caveman	PD17945a	7	148508721	148508721	EZH2	CCDS5891.1	p.G648V	missense	MiSeq	No coverage
	PD17945a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*	frameshift	PCR + capillary sequencing	validated
Pindel	PD17946a	4	106156658	106156660	TET2	CCDS47120.1	p.S521fs*1	frameshift		
Caveman	PD17946a	21	36231774	36231774	RUNX1	CCDS13639.1	p.R204*	nonsense	MiSeq	validated
Caveman	PD17946a	20	31022625	31022625	ASXL1	CCDS13201.1	p.G704R	missense	MiSeq	No coverage
Caveman	PD17946a	4	106164773	106164773	TET2	CCDS47120.1	p.R1214Q	missense	MiSeq	not confirmed
Caveman	PD17947a	17	29557890	29557890	NF1	CCDS42292.1	p.W1048C	missense	MiSeq	not confirmed
Caveman	PD17948a	X	44937750	44937750	KDM6A	CCDS14265.1	p.D980Y	missense	MiSeq	validated
Caveman	PD17949a	2	25458661	25458661	DNMT3A	CCDS33157.1	p.N838D	missense	MiSeq	validated
Pindel	PD17950a	2	25463567	25463568	DNMT3A	CCDS33157.1	p.I705fs*8	frameshift		
Caveman	PD17950a	9	5073770	5073770	JAK2	CCDS6457.1	p.V617F	missense		
Pindel	PD17951a	11	119149254	119149274	CBL	CCDS8418.1	p.I423_E427delIKGTE	inframe	MiSeq	validated
Caveman	PD17951a	21	36231774	36231774	RUNX1	CCDS13639.1	p.R204*	nonsense	MiSeq	validated
Caveman	PD17951a	7	148508719	148508719	EZH2	CCDS5891.1	p.E649*	nonsense	MiSeq	No coverage
Pindel	PD17952a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Pindel	PD17953a	4	106157954	106157955	TET2	CCDS47120.1	p.R953fs*19	frameshift	MiSeq	validated

Pindel	PD17953a	4	106156406	106156418	TET2	CCDS47120.1	p.Y437fs*7	frameshift	MiSeq	validated
Pindel	PD17953a	13	28608316	28608317	FLT3	CCDS31953.1	p.E598_Y599insWVTGSSDNEYFYVDFREYE	frameshift	PCR + agarose gel	validated
Caveman	PD17953a	X	44938474	44938474	KDM6A	CCDS14265.1	p.A1008S	missense	MiSeq	validated
	PD17953a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Caveman	PD17954a	15	90631934	90631934	IDH2	CCDS10359.1	p.R140Q	missense	MiSeq	validated
	PD17954a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Pindel	PD17955a	4	106157812	106157817	TET2	CCDS47120.1	p.M906fs*17	frameshift	MiSeq	validated
Caveman	PD17955a	4	106164741	106164741	TET2	CCDS47120.1	p.S1203R	missense	MiSeq	not confirmed
	PD17955a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Caveman	PD17956a	4	106156348	106156348	TET2	CCDS47120.1	p.Q417*	nonsense	MiSeq	validated
	PD17956a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Caveman	PD17957a	20	31025013	31025013	ASXL1	CCDS13201.1	p.E1500*	nonsense	MiSeq	validated
Caveman	PD17957a	2	209113113	209113113	IDH1	CCDS2381.1	p.R132S	missense	MiSeq	validated
Caveman	PD17957a	4	106182965	106182965	TET2	CCDS47120.1	p.P1335Q	missense	MiSeq	No coverage
Caveman	PD17957a	4	106164878	106164878	TET2	CCDS47120.1	p.T1249N	missense	MiSeq	validated
Caveman	PD17959a	12	25398281	25398281	KRAS	CCDS8703.1	p.G13D	missense	MiSeq	validated
	PD17959a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Pindel	PD17960a	13	28608304	28608305	FLT3	CCDS31953.1	p.D600_L601insSDNEYFYVDFREYEYD	frameshift	PCR + agarose gel	validated
Caveman	PD17960a	2	25467449	25467449	DNMT3A	CCDS33157.1	p.G543C	missense	MiSeq	validated
Caveman	PD17960a	X	44938447	44938447	KDM6A	CCDS14265.1	p.E999*	nonsense	MiSeq	validated
	PD17960a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Pindel	PD17961a	13	28608216	28608219	FLT3	CCDS31953.1	p.?	frameshift	PCR + agarose gel	validated
Pindel	PD17962a	13	28608274	28608275	FLT3	CCDS31953.1	p.K602_W603insCREYEYDLK	frameshift	PCR + agarose gel	validated
Pindel	PD17962a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Caveman	PD17963a	4	106157573	106157573	TET2	CCDS47120.1	p.S825*	nonsense	MiSeq	No coverage
Caveman	PD17964a	21	36259280	36259280	RUNX1	CCDS13639.1	p.L71M	missense	MiSeq	No coverage
	PD17964a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	MiSeq	No coverage

Pindel	PD17965a	20	31022545	31022546	ASXL1	CCDS13201.1	p.R678fs*40	frameshift	MiSeq	No coverage
Caveman	PD17965a	Y	15417990	15417990	UTY	CCDS14783.1	p.S1018Y	missense	MiSeq	No coverage
	PD17965a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Caveman	PD17966a	7	148511205	148511205	EZH2	CCDS5891.1	p.R566L	missense	MiSeq	validated
Caveman	PD17966a	2	25469038	25469038	DNMT3A	CCDS33157.1	p.R474S	missense	MiSeq	validated
Caveman	PD17966a	13	28592641	28592641	FLT3	CCDS31953.1	p.D835V	missense	MiSeq	validated
Caveman	PD17966a	4	106157119	106157119	TET2	CCDS47120.1	p.Q674K	missense	MiSeq	No coverage
	PD17966a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	PCR + capillary sequencing	validated
Pindel	PD17967a	13	28608286	28608287	FLT3	CCDS31953.1	p.Y597_E598insDYVDFREY	frameshift	PCR + agarose gel	validated
Caveman	PD17967a	1	115258744	115258744	NRAS	CCDS877.1	p.G13D	missense	MiSeq	validated
Caveman	PD17967a	X	39932971	39932971	BCOR	CCDS48093.1	p.S543*	nonsense	MiSeq	validated
	PD17967a	5	170837547	170837548	NPM1	CCDS4376.1	p.W288fs*12	frameshift	MiSeq	validated
Pindel	PD17968a	2	25463235	25463244	DNMT3A	CCDS33157.1	p.F752delF	inframe	MiSeq	validated
Caveman	PD17968a	1	115251178	115251178	NRAS	CCDS877.1	p.G183V	missense	MiSeq	validated
Caveman	PD17968a	17	29554540	29554540	NF1	CCDS42292.1	p.?	ess splice	MiSeq	validated

Supplementary Figure Legends

Legend to Supplementary Figure 1

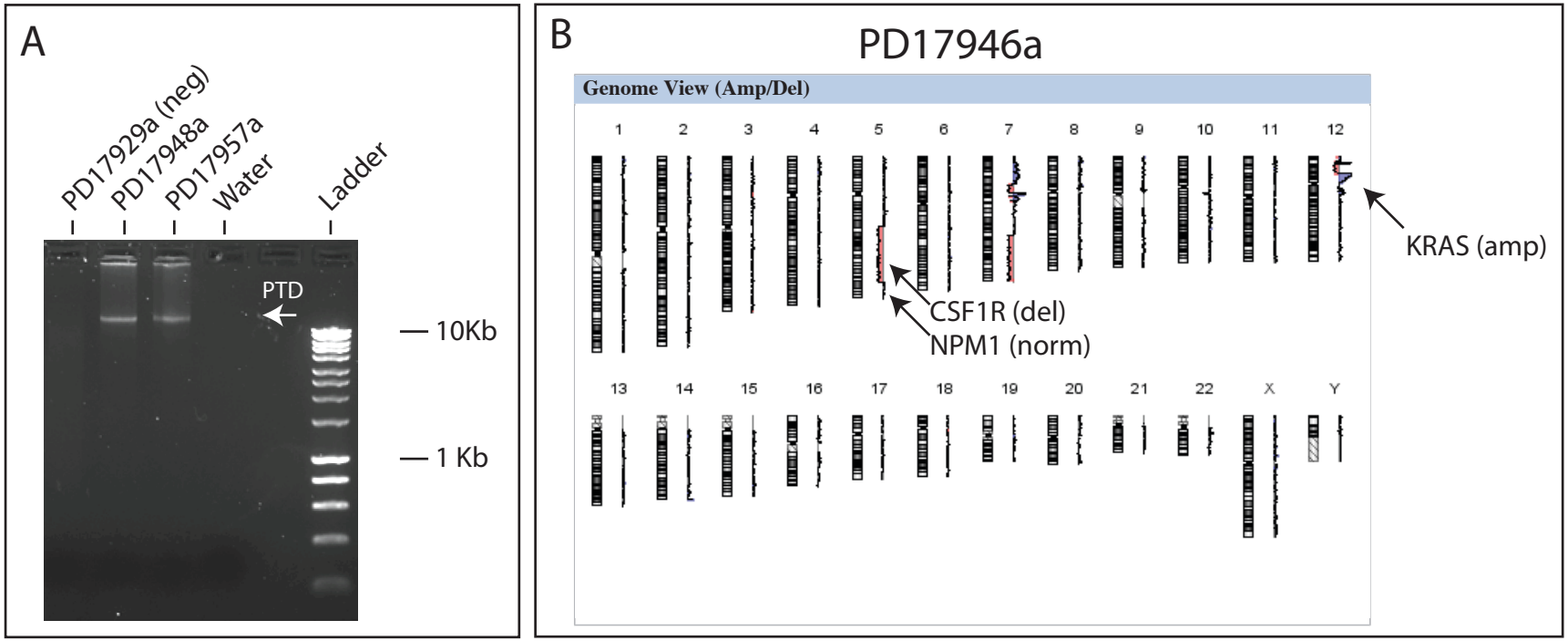
A) Long-range PCR on genomic DNA was performed on samples PD17948a and PD17957a to check for the presence of a MLL-PTD. Sample PD17929a was used as a negative control along with water. The white arrows shows a band at >10 Kb suggestive of an MLL-PTD in the two test samples.

B) Array CHG analysis of sample PD17946a confirms a KRAS amplification in chromosome 12p, and a 5q deletion that involved CSF1R but not NPM1 in keeping with the copy number pattern shown in Figure 3A (red bar).

Legend to Supplementary Figure 2

For 90 single nucleotide polymorphisms (SNPs) covered by the study design, the allelic fraction of the major allele (defined as the most prevalent in the general population) is plotted in the Y-axis. Samples are plotted in the X-axis. Note that 84.6% of SNP calls fall close to the 50% mark for heterozygous SNPs, indicating quantitative value of the allelic fraction of single nucleotide substitutions in Haloplex data.

Supplementary Figure 1



Supplementary Figure 2