Targeted resequencing analysis of 25 genes commonly mutated in myeloid disorders in del(5q) myelodysplastic syndromes

Marta Fernandez-Mercado, Adam Burns, Andrea Pellagatti, Aristoteles Giagounidis, Ulrich Germing, Xabier Agirre, Felipe Prosper, Carlo Aul, Sally Killick, James S Wainscoat, Anna Schuh and Jacqueline Boultwood

Supplementary Information

Targeted re-sequencing

We designed a TruSeq Custom Amplicon panel (TSCA, Illumina), targeting 25 genes mutated in various myeloid malignancies (Table 2). The panel was developed using the online DesignStudio pipeline (http://designstudio.illumina.com, Illumina), and covers a total of 46,604bp with 322 amplicons. In genes with well-defined mutational hotspots only these regions were targeted; otherwise the entire coding sequence of the gene was sequenced.

Dual-barcoded TSCA libraries were created from 250ng of genomic DNA, in accordance with the manufacturer's instructions, before undergoing 2x150bp paired-end sequencing on the Illumina MiSeq platform. The initial alignment and variant calling analysis was performed with the BaseSpace online analysis tool (https://basespace.illumina.com, Illumina). In order to screen for larger insertions and deletions, the data was also was run through the Stampy[1] and Platypus[2] pipelines, which uses a different algorithm to map sequencing reads to a reference genome. All variants called were visually inspected in IGV.

All candidate sequence variations that passed the internal Illumina integrity filters, and with a quality score greater than Q60, were taken forward for further analysis. All variations were confirmed visually and then checked against dbSNP build 135 (NCBI, National Center for Biotechnology Information, USA) and COSMIC (Catalog of Somatic Mutations In Cancer, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK) databases, to assess whether the variations found were reported polymorphisms or annotated mutations, respectively.

Assay sensitivity

To evaluate the sensitivity of the assay, we used two different approaches: (1) comparison with absolute real-time PCR quantification for a specific mutation and (2) definition of general background noise across all amplicons.

1. Comparison with real-time PCR

We determined the variant allele frequency (VAF) in 7 *JAK2*^{V617F} positive samples by real-time PCR, and compared them with the VAF values from the targeted sequencing assay.

We performed real-time PCR using the commercially available *JAK2* Muta*Quant*™ kit (Ipsogen, Luminy Biotech, Marseille, France), which distinguishes between *JAK2* wild-type and V617F alleles through Taqman allelic discrimination. Allele specific probes, labelled with 5' reporter and 3' quencher dyes, for both wild-type and V617F alleles are used to amplify the region of interest. The $JAK2^{V617F}$ percentage can be calculated from the fluorescent levels of each assay.

DNA samples were quantified using a BioPhotometer (Eppendorf, Hamburg, Germany) and normalised to a working concentration of 5ng/µl in nuclease free water. RT-PCR reactions were setup in a 100 well rotor by a CAS1200 liquid handling instrument (Qiagen, Hilden, Germany). Each reaction contained 6.25µl 2x Taqman Universal PCR Master Mix (Applied Biosystems, Life Technologies, Carlsbad, CA,), 0.5µl 25x primer/probe mix (Ipsogen, Luminy Biotech, Marseille, France), 3.25µl nuclease free water and 2.5µl 5ng/µl sample DNA. 4-point duplicate standard curves were included with each run, amplified from standard plasmids included in the kit. Positive (>99.9% V617F) and negative (<0.1% V617F) controls were also included in each run. Each sample was processed in duplicate for both the wild type and V617F alleles.

The reactions were amplified on a Rotor-Gene 6000 instrument (Qiagen, Hilden, Germany) with the following PCR conditions: 50°C for 2 minutes, 95°C for 10 minutes followed by 50 cycles of 95°C for 15 seconds and 62°C for 1 minute, with acquisition of FAM fluorescence during the 62°C step.

Analysis of the raw data was performed using the Rotor-Gene Q software package (Qiagen, Hilden, Germany). The cycle threshold was set at 0.03 with the slope corrected, as per the manufacturer's guidelines (Ipsogen, Luminy Biotech, Marseille, France). Raw data tables for both Wild-Type and V617F assays were exported into Excel (Microsoft, Redmond, WA) to facilitate further analysis. The standard curves were plotted (y = mean ct, x = Log_{10} CN, where CN is gene copy number/5µl) for both the wild-type and V617F standard samples, and the Y and R^2 values were extracted. The copy number for $JAK2^{V617F}$ was calculated as: (mean $Ct_{JAK2V617F}$ – Standard Curve $Intercept_{JAK2V617F}$)/Standard Curve $Slope_{JAK2V617F}$. JAK2 wild-type copy number was calculated as: (mean Ct_{JAK2WT} – Standard Curve $Intercept_{JAK2WT}$)/Standard Curve $Slope_{JAK2WT}$. Final results were determined as a percentage of $JAK2^{V617F}$ allele load, calculated by: Copy $Number_{JAK2V617F}$/(Copy $Number_{JAK2V617F}$ + Copy $Number_{JAK2WT}$) x 100.

The variant allele frequency of the $JAK2^{V617F}$ positive samples, as determined by the real-time PCR assay, ranged from 1-24% (Table S2). All mutations with a VAF >3% (6/7, 86%) were successfully aligned and called as the $JAK2^{V617F}$ variant. The remaining mutation (1% VAF)

was present in the sequencing reads, but was below the detection limit of the variant calling software.

2. Background noise

We determined the background noise level of our assay by investigating the sequencing read composition at 31 SNP loci over 14 chromosomes in 15 samples. The SNPs were all initially identified by our data analysis pipeline, are bi-allelic and are all recorded in dbSNP135 as being non-pathogenic. At each locus (465 total), we measured the level of background noise by calculating the percentage of sequencing reads containing any of the alternate nucleotides (3 in the case of homozygous SNPs, 2 in the case of heterozygous SNPs).

The mean level of background noise in our assay was thus determined as 0.31% (range 0.0-0.8%) across all SNP loci in all samples, and was consistently low both between the SNPs (mean 0.31%, range 0.1-0.8%), and between the samples (mean 0.33%, range 0.25-0.55). Interestingly, the background level at heterozygous *loci* was lower than that at homozygous *loci* (0.2% and 0.4% respectively).

Taken together, we therefore defined the sensitivity of the panel at 1-3% depending on the locus examined and the variant caller software.

JAK2^{V617F} pyrosequencing

JAK2^{V617F} (c.1849G>T) mutation was analysed using primers as previously described.[3] In brief, DNA was amplified in 25µl reactions, containing 2x Qiagen Multiplex PCR Master Mix (Qiagen), 5x Q Solution (Qiagen) and 5mM each of reverse and biotinylated forward primers. Cycling conditions consisted of an initial denaturation step of 97^{o}C for 15 minutes followed by 35 cycles of 30 seconds at 97^{o}C, 90 seconds at 62 ^{o}C and 2 minutes at 72 ^{o}C. The resulting biotinylated PCR product was subjected to pyrosequencing using a Pyromark Q24 System (Qiagen). Pyromark Q24 allele quantification (AQ) software was used to quantify the level (if any) of *JAK2*^{V617F} variant present in each sample.

FLT3-ITD ARMS-PCR

FLT3-ITD mutations were analysed using primers as previously described,[4] modified with WellRED fluorescent dyes.[4,5] In brief, DNA was amplified in 25µl reactions, containing 2x Qiagen Multiplex PCR Master Mix (Qiagen), 5x Q Solution (Qiagen) and 5mM each of forward and reverse primers. Cycling conditions consisted of an initial denaturation step of 95^{o}C for 15 minutes followed by 35 cycles of 30 seconds at 95^{o}C, 1 minute at 56^{o}C and 2 minutes at 72^{o}C, with a final extension step of 10 minutes at 72^{o}C. The resulting PCR product was diluted 1:10. 2µl of diluted PCR product was mixed with 40µl Sample Loading Solution

(Beckman Coulter) and 0.5µl GenomeLab DNA Size Standard 600 (Beckman Coulter) and subjected to capillary electrophoresis on a CEQ8000 Genetic Analysis System (Beckman Coulter). Data analysis was performed using CEQ analysis software version 9.0.25.

NPM1 fragment analysis

Validation of the *NPM1* mutation was performed by fragment analysis, using primers as previously described.[6] DNA was amplified in 25µl reactions containing 2x Qiagen Master Mix (Qiagen), 10pmol of forward and reverse primers and sterile water up to the final 25µl volume. Cycling conditions consisted of an initial denaturation step of 95^{o}C for 15 minutes followed by 40 cycles of 30 seconds at 92^{o}C, 30 seconds at 58^{o}C and 20 seconds at 72^{o}C, with a final extension step of 10 minutes at 72^{o}C. The resulting PCR product was diluted 1:10. 2µl of diluted PCR product was mixed with 40µl Sample Loading Solution (Beckman Coulter) and 0.5µl GenomeLab DNA Size Standard 600 (Beckman Coulter) and subjected to capillary electrophoresis on a CEQ8000 Genetic Analysis System (Beckman Coulter). Data analysis was performed using CEQ analysis software version 9.0.25.

Sanger Sequencing

Mutations discovered in the validation cohort in *TET2, RUNX1, SF3B1* and *FLT3* were confirmed by Sanger sequencing. DNA was amplified in 25µl reactions containing 2x Qiagen Master Mix (Qiagen) and 5mM of forward and reverse primers. 5x Q Solution (Qiagen) was used where indicated (Table S2). Cycling conditions for all targets consisted of an initial denaturation step of 97^{o}C for 15 minutes followed by 35 cycles of 30 seconds at 92^{o}C, 30 seconds at 55^{o}C (*RUNX1* and *FLT3*) or 60^{o}C (*TET2* and *SF3B1*) and 20 seconds at 72^{o}C, with a final extension step of 10 minutes at 72^{o}C. The PCR products were purified using MicroClean (Cambio) and 1µl of purified PCR product was used for sequencing with the Big Dye terminator v3.1 chemistry (Applied Biosystems) with either the forward or reverse primer. After ethanol/EDTA precipitation, the samples underwent electrophoresis on an ABI 3130 Genetic Analyzer (Applied Biosystems).

Genome-wide DNA-methylation

The DNA methylation profiles of 14 cases were analysed using Illumina HumanMethylation 27 BeadChip (Illumina, Inc., San Diego, CA, USA). Those 14 cases included 11 5q- syndrome, 1 del(5q) RA with additional cytogenetic aberrations and 2 advanced del(5q) cases. To ensure karyotypic homogeneity, only the DNA methylation profiles of the 11 5q- syndrome cases was further analysed based on the mutational status of the genes involved in epigenetic regulation included in our TSCA. Within these 11 5q- syndrome cases 1 had a *DNMT3A* mutation, 2 had an *ASXL1* mutation, and 1 had concomitant *ASXL1* and *TET2* mutations.

Data analysis was carried out using R/Bioconductor. Before selection of differentially methylated probes a filtering process based on the mean β-values for each gene mutated under study (*DNMT3A, ASXL1, ASXL1* and *TET2, ASXL1* or *TET2*) was performed to focus the analysis on genes with large differences in their methylation status. Briefly, the obtained mean value was categorized in three states: unmethylated state (mean value < 0.3), partially methylated state (mean value > 0.3-<0.7) and methylated state (mean value > 0.7). We assigned a value of 0, 1 or 2 to each probe in function of its methylation state and calculated the difference between states for each comparison. All probes with differential methylated state equal to 0 were filtered out. Finally, fold-change of mean β-values was used to find out the probes that showed significant differential methylation patterns. Probes were selected as significant using a logFC cut off of 1.5.

In order to investigate the potential effect on DNA methylation of mutations in genes involved in the epigenetic regulation of the cell, the following comparisons were run:

- 2 *ASXL1*-mut cases *versus* 7 cases with no epigenetic gene mutations. Number of differentially methylated genes (DMG): 422.
- 1 *DNMT3A*-mut case *versus* 7 cases with no epigenetic gene mutations. Number of DMG:144.
- 1 *ASXL1* & *TET2* mutant cases *versus* 7 cases with no epigenetic gene mutations. Number of DMG:156.
- 3 *ASXL1*-mut cases *versus* 7 cases with no epigenetic gene mutations. Number of DMG: 205.

The lists of DMG were used to generate supervised clusters on all 11 5q- syndrome cases. None of the analyses managed to cluster the samples based on their mutations in epigenetic genes. Based on these results, we cannot attribute any specific DNA methylation profile to the mutations detected in genes involved in the epigenetic regulation of the cell.

SNP mapping assay and data analysis

The SNP mapping assay was performed according to the protocol supplied by the manufacturer (Affymetrix, Santa Clara, CA, USA). Briefly, 250 ng DNA were digested with Hind III, ligated to the adaptor, and amplified by polymerase chain reaction (PCR) using a single primer. PCR products were purified with the DNA amplification clean-up kit (Clontech) and the amplicons were quantified. The 40 µg of purified amplicons were fragmented, end-labeled and hybridized to a Genechip Mapping 50K Hind III array at 48°C for 16–18 hours in a Hybridization Oven 640 (Affymetrix). After washing and staining in a Fluidics Station 450 (Affymetrix), the arrays were scanned with a GeneChip Scanner 3000 (Affymetrix).

Cell intensity calculations and scaling were performed using GeneChip Operating Software (GCOS). Data were analyzed using GeneChip Genotyping Analysis Software Version 4.0 (Affymetrix) and CNAG software version 2.0. Quality control was performed within the Genotyping software after scaling the signal intensities of all arrays to a target of 100%. DNA copy number was analyzed with both the chromosome copy number tool (CNAT) version 3.0 and CNAG version 2.0. CNAT compares obtained SNP hybridization signal intensities with SNP intensity distributions of a reference set from more than 100 healthy individuals of different ethnicity. For analysis with CNAG we used a pool of 45 healthy controls as a reference set.[7]

References

1. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res 2011;21(6):936-9.

2. Rimmer A, Mathieson I, Lunter G, McVean G. Platypus: An Integrated Variant Caller (www.well.ox.ac.uk/platypus). 2012.

3. Jones AV, Kreil S, Zoi K, Waghorn K, Curtis C, Zhang L, et al. Widespread occurrence of the JAK2 V617F mutation in chronic myeloproliferative disorders. Blood 2005;106(6):2162-8.

4. Murphy KM, Levis M, Hafez MJ, Geiger T, Cooper LC, Smith BD, et al. Detection of FLT3 internal tandem duplication and D835 mutations by a multiplex polymerase chain reaction and capillary electrophoresis assay. J Mol Diagn 2003;5(2):96-102.

5. Kiyoi H, Naoe T, Yokota S, Nakao M, Minami S, Kuriyama K, et al. Internal tandem duplication of FLT3 associated with leukocytosis in acute promyelocytic leukemia. Leukemia Study Group of the Ministry of Health and Welfare (Kohseisho). Leukemia 1997;11(9):1447-52.

6. Scholl S, Mugge LO, Landt O, Loncarevic IF, Kunert C, Clement JH, et al. Rapid screening and sensitive detection of NPM1 (nucleophosmin) exon 12 mutations in acute myeloid leukaemia. Leuk Res 2007;31(9):1205-11.

7. Wang L, Fidler C, Nadig N, Giagounidis A, Della Porta MG, Malcovati L, et al. Genome-wide analysis of copy number changes and loss of heterozygosity in myelodysplastic syndrome with del(5q) using high-density single nucleotide polymorphism arrays. Haematologica 2008;93(7):994-1000.

Table S1. Summary of *JAK2*^{V617F} variant allele frequencies (VAF).

	RT-PCR			*MiSeq*			
Sample ID	**JAK2 WT Copy Number**	**V617F Copy Number**	**VAF**	**Total Depth**	**Reference Depth**	**Variant Depth**	**VAF**
JAK2_A	60035	1865	0.03	8205	7564	628	0.08
JAK2_B	51617	5929	0.10	7924	6540	1366	0.17
JAK2_C	58408	9834	0.14	7883	5859	2015	0.26
JAK2_D	52331	7917	0.13	7828	6342	1472	0.19
JAK2_E	59013	852	0.01	7411	7219	177	0.02
JAK2_F	50564	10411	0.17	8139	6390	1719	0.21
JAK2_G	36490	11804	0.24	7637	5290	2333	0.31

Table S2. Sanger sequencing primers and PCR conditions.

Target	***Forward Primer***	***Reverse Primer***	***PCR Conditions***	***Reference***
TET2	AGACTTATGTATCTTTCATCTAGCTCTGG	ACTCTCTTCCTTTCAACCAAAGATT	60°C	Gelsi-Boyer *et al.*
RUNX1	GCTGTTTGCAGGGTCCTAA	CCTGTCCTCCCACCACCCTC	5x Q Solution, 55°C	
SF3B1	CTGCAGTTTGGCYGAATAGTTG	AAAATTCTGTTAGAACCATGAAACA	60°C	Papaemmanuil *et al.*
FLT3	CCGCCAGGAACGTGCTTG	GCAGCCTCACATTGCCCC	5x Q Solution, 55°C	Nakao *et al.*

Table S3. Detailed description of non-synonymous variants with a COSMIC ID or not reported in dbSNP.

sample ID	*Diagnostic*	*Gene*	*Genome coordinates*	*DNA change*	*Protein change*	*Q-score*	*Variant call ratio [% (variant/total)]*	*COSMIC ID*	*dbSNP ID*	*Polyphen2 (score, sensitivity, specificity)*
MDS16	RA (5q-syndrome)	*RUNX1*	chr21:36259324	A>AG	L29S	99	31.9 (23/72)	COSM24756	rs111527738	Probably damaging (0.999 0.14, 0.99)
MDS15	RA (5q-syndrome)	*SF3B1*	chr2:198266834	T>TC	K700E	99	11.4 (170/1496)	COSM84677	NA	Probably damaging (1.000, 0.00, 1.00)
MDS07	RA (5q-syndrome)	*DNMT3A*	chr2:25457242	C>CA	R882L	75	7.8 (92/1176)	NA	NA	Probably damaging (0.982, 0.75, 0.96)
MDS08	RA (5q-syndrome)	*ASXL1*	chr20:31022449	insG	G646WfsX12	99	44.7 (174/389)	COSM34210	NA	Truncated protein
MDS08	RA (5q-syndrome)	*WT1*	chr11:32413565	C>CT	R462Q	99	49.0 (174/355)	COSM21408	NA	Probably damaging (1.000, 0.00,1.00)
MDS14	RA (5q-syndrome)	*TET2*	chr4:106193748	C>CT	R1404X	99	45.1 (309/685)	COSM42037	NA	Truncated protein
MDS12	RA (5q-syndrome)	*ASXL1*	chr20:31022449	insG	G646WfsX12	99	10.5 (37/351)	COSM34210	NA	Truncated protein
MDS12	RA (5q-syndrome)	*SF3B1*	chr2:198266834	T>TC	K700E	99	40.0 (620/1549)	COSM84677	NA	Probably damaging (1.000, 0.00, 1.00)
MDS12	RA (5q-syndrome)	*TET2*	chr4:106164896	insA	fs (Y1255X)	99	5.3 (41/771)	COSM110747	NA	Truncated protein
MDS06	RA (5q-syndrome)	*TET2*	chr4:106197552	C>CT	P1962L	99	50.3 (303/602)	COSM41894	NA	Probably damaging (0.974, 0.76, 0.96)
MDS11	RA (5q-syndrome)	*ASXL1*	chr20:31022902	G>GA	W796X	99	35.8 (144/402)	COSM53207	NA	Truncated protein
MDS10	RA (5q-syndrome)	*TP53*	chr17:7578413	C>CG	V173L	99	41.1 (109/265)	COSM43559	NA	Probably damaging (0.979, 0.76, 0.96)
MDS29	RA (5q-syndrome)	*JAK2*	Chr9:5073770	G>GT	V617F	99	7	COSM12600	rs77375493	Probably damaging (0.996, 0.55, 0.98)
MDS34	RA (5q-syndrome)	*JAK2*	Chr9:5073770	G>GT	V617F	99	28	COSM12600	rs77375493	Probably damaging (0.996, 0.55, 0.98)
MDS29	Del(5q) RA with additional cytogenetic abnormalities	*DNMT3A*	chr2:25457176	G>GA	P904L	99	44.0 (198/450)	COSM52989	rs149095705	Probably damaging (0.995, 0.68, 0.97)
MDS28	Del(5q) RA with additional cytogenetic abnormalities	*U2AF1*	chr21:44514777	T>TC	Q157R	99	38.3 (242/632)	COSM144989	NA	Probably damaging (0.997, 0.41, 0.98)
MDS30	Del(5q) RA with additional cytogenetic	*CBL*	chr11:119149332	C>CT	A447V	99	43.6 (99/227)	NA	NA	Possibly damaging (0.717, 0.86, 0.92)

	abnormalities									
MDS26	Del(5q) RA with additional cytogenetic abnormalities	*TP53*	chr17:7577553	A>AG	M243T	99	28.0 (327/1166)	COSM43726	NA	Probably damaging (1.000, 0.00, 1.00)
MDS37	Advanced del(5q) MDS (RAEB)	*TP53*	chr17:7577120	C>CT	R273H	99	82.2 (620/754)	COSM10660	rs28934576	Possibly damaging (0.831, 0.84, 0.93)
MDS42	Advanced del(5q) MDS (CMML)	*ASXL1*	chr20:31023821	G>GT	E1102D	99	45.4 (366/806)	COSM36205	rs139115934	Possibly damaging (0.779, 0.85, 0.93)
MDS42	Advanced del(5q) MDS (CMML)	*CBL*	chr11:119149004	G>GT	W408C	99	96.0 (267/278)	COSM34072	NA	Probably damaging (0.996, 0.55, 0.98)
MDS36	Advanced del(5q) MDS (RAEB)	*ASXL1*	chr20:31024704	G>GA	G1397S	99	49.9 (875/1755)	COSM133033	rs146464648	Possibly damaging (0.792, 0.85, 0.93)
MDS33	Advanced del(5q) MDS (RAEB)	*TET2*	chr4:106196850	insCATG	E1728Dfs*13	99	17.0 (121/713)	COSM211745	NA	Truncated protein
MDS43	Advanced del(5q) MDS (RAEB)	*TET2*	chr4:106164880	G>GT	E1250X	99	27.3 (313/1145)	NA	NA	Truncated protein
MDS43	Advanced del(5q) MDS (RAEB)	*ASXL1*	chr20:31022449	insG	G646WfsX12	99	42.8 (470/1097)	COSM34210	NA	Truncated protein
MDS39	Advanced del(5q) MDS (RAEB)	*TP53*	chr17:7578190	T>TC	Y220C	99	38.7 (48/124)	COSM99719	rs121912666	Probably damaging (1.000, 0.00, 1.00)
MDS39	Advanced del(5q) MDS (RAEB)	*TP53*	chr17:7578275	G>GA	Q192X	99	49.3 (99/201)	COSM117949	NA	Truncated protein
MDS38	Advanced del(5q) MDS (RAEB)	*TP53*	chr17:7577538	C>CA	R248L	99	44.1 (1168/2648)	COSM6549	rs11540652	Probably damaging (1.000, 0.00, 1.00)
MDS38	Advanced del(5q) MDS (RAEB)	*TP53*	chr17:7577568	C>CT	C238Y	99	37.5 (998/2664)	COSM11059	NA	Probably damaging (1.000, 0.00, 1.00)

Table S4. Detailed description of synonymous variants with a COSMIC ID.

sample ID	*Diagnostic*	*Gene*	*Genome coordinates*	*DNA change*	*Protein change*	*Q-score*	*Variant call ratio [% (variant/total)]*	*COSMIC ID*	*dbSNP ID*
MDS04	RA (5q- syndrome)	*IDH1*	chr2:209113192	G>GA	G105G	99	49.2 (445/904)	COSM253316	rs11554137
MDS13	RA (5q- syndrome)	*IDH1*	chr2:209113192	G>GA	G105G	99	49.3 (465/943)	COSM253316	rs11554137
MDS01	RA (5q- syndrome)	*IDH1*	chr2:209113192	G>GA	G105G	99	49.3 (421/854)	COSM253316	rs11554137
MDS24	Del(5q) RA with additional cytogenetic abnormalities	*IDH1*	chr2:209113192	G>GA	G105G	99	49.6 (483/973)	COSM253316	rs11554137
MDS30	Del(5q) RA with additional cytogenetic abnormalities	*IDH1*	chr2:209113192	G>GA	G105G	99	49.4 (356/721)	COSM253316	rs11554137
MDS27	Del(5q) RA with additional cytogenetic abnormalities	*FLT3*	chr13:28608459	T>TC	L561L	99	53.6 (149/278)	COSM19740	rs34374211
MDS43	Advanced del(5q) MDS (RAEB)	*FLT3*	chr13:28608459	T>TC	L561L	99	52.1 (173/332)	COSM19740	rs34374211
MDS42	Advanced del(5q) MDS (CMML)	*KIT*	chr4:55599268	C>CT	I798I	99	55.1 (162/294)	COSM1307	rs55789615
MDS26	Del(5q) RA with additional cytogenetic abnormalities	*KIT*	chr4:55599268	C>CT	I798I	99	45.5 (150/330)	COSM1307	rs55789615
MDS02	RA (5q- syndrome)	*KIT*	chr4:55599268	C>CT	I798I	99	50.0 (166/332)	COSM1307	rs55789615
MDS05	RA (5q- syndrome)	*PDGFRA*	chr4:55152040	C>CT	V824V	99	55.6 (280/504)	COSM22413	rs2228230
MDS08	RA (5q- syndrome)	*PDGFRA*	chr4:55152040	C>CT	V824V	99	53.8 (271/504)	COSM22413	rs2228230
MDS09	RA (5q- syndrome)	*PDGFRA*	chr4:55152040	C>CT	V824V	99	49.0 (251/512)	COSM22413	rs2228230
MDS11	RA (5q- syndrome)	*PDGFRA*	chr4:55152040	C>CT	V824V	99	48.0 (210/437)	COSM22413	rs2228230
MDS14	RA (5q- syndrome)	*PDGFRA*	chr4:55152040	C>CT	V824V	99	47.5 (308/648)	COSM22413	rs2228230
MDS16	RA (5q- syndrome)	*PDGFRA*	chr4:55152040	C>CT	V824V	99	52.2 (251/481)	COSM22413	rs2228230
MDS01	RA (5q- syndrome)	*PDGFRA*	chr4:55152040	C>CT	V824V	99	53.0 (231/436)	COSM22413	rs2228230
MDS27	Del(5q) RA with additional cytogenetic abnormalities	*PDGFRA*	chr4:55152040	C>CT	V824V	99	51.4 (360/701)	COSM22413	rs2228230
MDS35	Advanced del(5q) MDS (RAEB)	*PDGFRA*	chr4:55152040	C>CT	V824V	99	45.6 (312/684)	COSM22413	rs2228230
MDS42	Advanced del(5q) MDS (CMML)	*PDGFRA*	chr4:55152040	C>CT	V824V	99	50.8 (332/654)	COSM22413	rs2228230
MDS32	Advanced del(5q) MDS (RAEB)	*TP53*	chr17:7578210	T>TC	R213R	99	44.9 (137/305)	COSM249885	rs1800372

Table S5. Genomic array results for 33 del(5q) cases analysed, including 18 5q- Syndrome cases. Brackets show several metrics of the detected alterations: coordinates mapping the alteration (start-end); SNPs within it (start-end); length (bp); SNPs contained (number); copy number. It is noted if any of the 25 genes analysed in this study was encompassed in that region. UPD: uniparental dysomy. NA: not available.

Sample ID	***Diagnosis***	***Age/Sex***	***Karyotype***	***Deletions***	***UPD***	***Gains***
MDS02	RA, 5q- Syndrome	NA/F	NA	2p23.3 (25119937-26655307; 4992-5005; 1535370; 14; 1.17; 0.67) **DNMT3A** 5q22.1-q33.2 (110998762-154437028; 20461-21565; 43438266; 1105; 1.38; 0.18)	13q14.11-q14.13 (42627520-44793601; 43906-43966; 2166081; 61; 1.99; 0.25)	Whole Chr8 (272252-146052174; 29522-33067; 145779922; 3546; 2.28; 0.34)
MDS03	RA, 5q- Syndrome	48/M	46,XY,del(5)(q13:q33)	5q14.3-q34 (87824784-167184563; 19911-21920; 79359779; 2010; 1.40; 0.18)	12q15 (67170328-69219954; 42039-42099; 2049626; 61; 2.10; 0.36)	6q24.1 (139269078-139910603; 25404-25432; 641525; 29; 2.30; 0.38) 10p14-p13 (11264836-13110988; 35711-35761; 1846152; 51; 2.20; 0.31) 12q24.22-q24.31 (125358706; 43078-43146; 9511465; 69; 2.18; 0.34) 16q22.3-q23.1 (72348989-73935434; 50009-50037; 1586445; 29; 2.25; 0.30) 17q23.2-q23.3 (53845988-58047700; 51071-51112; 4201712; 42; 2.24; 0.33)
MDS04	RA, 5q- Syndrome	88/F	46,XX,del(5)(q13:q33)	5q14.3-q34 (86862506-166939254; 19896-21913; 80076748; 2018; 1.61; 0.20)	4q13.1-q13.2 (65578799-67713359; 14910-14985; 2134560; 76; 2.00; 0.23) 4q26-q27 (117680512-121421102; 16225-16298; 3740590; 74; 2.02; 0.27) 4q31.21-q31.23 (146661914-149093249; 16855-16925; 2431335; 71; 1.99; 0.24)	ChrX. **ATRX, ZRSR2**
MDS05	RA, 5q- Syndrome	60/F	46,XX,del(5)(q13:q33)	5q14.2-q33.3 (81866327-156969197; 19783-21650; 75102870; 1868; 1.41; 0.34)	1p31.2-p31.1 (68657746-70841176; 975-1035; 2183430; 61; 1.93; 0.35) 5q11.1-q11.2 (50213917-52264199;	

					18981-19036; 2050282; 56; 2.34; 2.02)
MDS06	RA, 5q-Syndrome	68/F	46,XX,del(5)(q14-15:q33)	5q14.3-q33.3 (87875023-156072147; 199912-21634; 68197124; 1723; 1.82; 0.25)	6q14.3-q15 (85411634-88478204; 24018-24102; 3066570; 85; 1.97; 0.28)
MDS07	RA, 5q-Syndrome	NA/F	46,XX,del(5)(q13:q33)	5q14.3-5q34 (89303345-163980289;19955-21829; 74676944; 1875; 1.54; 0.26)	4q21.21-21.22 (80195990-82802565; 15257-15352; 2606575; 96; 1.98; 0.43) 13q21.2-q21.31 (58895681-61586211; 44290_44371; length; 2690530; 82; 1.96; 0.30)
MDS08	RA, 5q-Syndrome	NA/M	NA	5q14.3-q33.3 (89917534-158840805; 19974-21705; 68923271; 1732; 1.47; 0.31)	6q13-q14.1 (75035581-77537444; 23758-23818; 2501863; 61; 1.99; 0.25)
MDS09	RA, 5q-Syndrome	84/F	46,XX,del(5)(q13:q33)	5q14.3-q33.2 (85182021-154953129; 19863-21600; 69771108; 1738; 1.44; 0.21)	
MDS10	RA, 5q-Syndrome	76/F	46,XX,del(5)(q13:q33)	5q12.3-q13.1 (65535157-67677808; 19411-19484; 2142651; 74; 1.41; 0.14) 5q14.3-q15 (83343740-95777305; 19825-20080; 12433565; 256; 1.40; 0.17) 5q21.1-q34 (97786027-163782378; 20142-21815; 65996351; 1674; 1.39; 0.18)	4q12-q13.1 (58804522-61855787; 14735-14810; 3051265; 76; 2.07; 0.33) 13q21.31-q21.32 (62298667-65258156; 44390-44460; 2959489; 71; 1.97; 0.26)
MDS11	RA, 5q-Syndrome	81/F	46,XX,del(5)(q13:q33)	5q21.1-q34 (98822612-164720069; 20174-21840; 65897457; 1667; 1.52; 0.22)	2q23.3-q24.1 (152916479-155038402; 7616-7694; 2121923; 79; 2.00; 0.30) 7p15.2-p15.1 (25388631-28776965; 26914-27025; 3388334; 112; 1.96; 0.27)
MDS12	RA, 5q-Syndrome	77/F	46,XX,del(5)(q22:q35)	5q14.3-q33.1 (86463622-151297473; 19893-21476; 64833851; 1584; 1.48; 0.18)	6q13-q14.1 (74620278-77463618; 23753-23813; 2843340; 61; 2.00; 0.24)
MDS13	RA, 5q-Syndrome	64/F	46,XX,del(5)(q33:q34)	5q32-q34 (148469763-167102662; 21427-21918; 18632899; 492; 1.46;; 0.28)	1q31.1 (185202380-187449830; 3205-3265; 22474450; 61; 1.95; 0.30)
MDS14	RA, 5q-Syndrome	66/F	46,XX,del(5)(q31:q33)[8] /46,XX[31]	5q31.3-q33.3 (142271912-156074292; 21232-21637; 13802380; 406; 1.42; 0.23)	3p24.1-p23 (29906836-34218933; 10495-10560; 4312097; 66; 1.96; 0.24)

					6q22.33-q23.1 (128419190-130577705; 25114-25209; 2158515; 96; 1.92; 0.23)	
MDS15	RA, 5q-Syndrome	72/F	46,XX,del(5)(q13:q33)[6]/46,XX[4]	5q14.3-q33.3 (84641203-158921205; 19851-21707; 74280002; 1817; 1.69; 0.20)	4q26-q27 (120601325-122789184;16270-16340; 2187859; 71; 1.93; 0.29) 10q21.2-q21.3 (61596872-64250581; 36741-36796; 2653709; 56; 2.04; 0.33)	
MDS16	RA, 5q-Syndrome	74/F	NA	5q21.1-q33.3 (102215241-156099317; 20248-21641; 53884076; 1394; 1.34; 0.23)	1p32.3-33 (48755449-54062185; 462-534; 5306736; 73; 2.09; 0.36) 4q21.3-22.1 888367780-90815461; 15505-15585; 2447681; 81; 1.97; 0.43)	
MDS18	RA, 5q-Syndrome	66/F	46,XX,del(5)(q14-15:q33)	5q21.3-q34 (106062626-163950485; 20323-21828; 57887789; 1506; 1.35; 0.31)	10q23.1 (83732605-85765313; 37101-37181; 2032708; 81; 2.22; 0.54)	
MDS20	RA, 5q-Syndrome	24/F	46,XX,del(5)(q31:q33)	5q31.3q.33.3 (141347991-158590590; NA; 17242599; 491;; 1.37; 0.17)	3q25.1-q25.2(151639927-154140946; 12710-12780; 2501019; 71; 2.00; 0.27) 7q31.33 (123603987-125743270; 28927-28987; 2139283; 61; 2.04; 0.33) 17q23.2-q24.1 (53562730-61202528; 51069-51137;7639798; 69; 2.02; 0.26)	
MDS21	RA, 5q-Syndrome	70/M	46,XY,del(5)(q13:q33)	5q14.3-5q33.3 (19918-21643; 87898589-156124093; 68225504; 1726; 1.65; 0.30)	5q11.2-q12.1 (57355239-62126018; 19149-19330; 4770779; 182; 1.94; 0.39) 11q14.3-q21 (91659665-94441966; 39765-39825; 2782301; 61; 1.95; 0.38)	Multiple small gains
MDS23	RA	77/F	46,XX,del(5)(q13:q33),del(11)(q22)[3]/46,XX[2]	5q14.3-q34 (82810660-163854743; 19815-21825; 81044083; 2011; 1.78; 0.20) 11q22.3-q25 (106376702-134173875; 40174-40623; 27797173; 450; 1.77; 0.19) (Only CN loss) **CBL**	19p12-q12 (21633219-33888946; 53124-53185; 12255727; 62; 1.95; 0.27)	
MDS24	RA	72/M	46,Y,der(X)t(X;12)(p22;q21),del(5)(q14-15;q33-	5q15-q33 (91919548-158401872; 20009-21692; 66482324;1684; 1.60;	6q24.1-q24.2 (141617059-143973425; 25458-25518;2356366;	

			34),der(12),del(12)(p11q13)[7]/46,XY[3]	0.16) 6q23.2-q23.3 (135131247-138523284; 25324-25394; 3392037; 71; 1.68; 0.15) (Only CN loss) 12p11.23-p13.31 (9809369-27307682; 40761-41192; 17498313; 432; 1.63; 0.13) (Only CN loss) **ETV6** 12q21.33-q22 (89389338-94118553; 42583-42677; 4729215; 95; 1.69; 0.19) (Only CN loss)	61; 1.96; 0.21)	
MDS25	RA	78/F	46,XX,del(5)(q14:q34),t(1,3)(p33:p14)[21]/46,XX[4]	5q14.3-q33.2 (86226079-154919227; 19885-21598; 68693148; 1714; 1.73; 0.33)	12q21.2-q21.31 (78226836-81842401; 42349-42425; 3615565; 77; 2.05; 0.44)	6p21.2-p22.1 (length: 11429636; 131; 2.20; 0.36) 22q13.1-q13.31 (length: 8390013; 88; 2.21; 0.42)
MDS26	RA	85/F	46,XX,del(5)(13:q33),+8	5q14.3-q33.3 (86607880-157924393; 19894-21673; 71316513; 1780; 1.60; 0.19)	13q21.1 (55328914-58382524; 44214-44273;3053610;60; 2.02; 0.24)	Some small CN changes Whole Chr8 (228574-143783463; 29518-33065; 143554889; 3548; 2.25; 0.29)
MDS29	RA	78/F	46,XX,del(5)(q13:q33)[18]/46,XX,del(5)(q13:q33),-7[1]	5q14.3-q33.3 (90357044-158432337; 19981-21696; 68075293; 1716; 1.64; 0.19)	9q21.13 (71990110-74653742; 34293-34368; 2663632; 76; 1.98; 0.25)	
MDS31	RA	73/F	46,XX,del(5)(q13:q31)[18]/48,XX,del(5)(q13:q31),idic(21)(q22),+2mar[2]/46,XX[1]	5q21.3-q34 (104537088-167772186; 20302-21937; 63235098; 1636; 1.47; 0.16)	8q21.11 (75717988-78220362; 31423-31498; 2502374; 76; 1.95; 0.20)	
MDS32	RAEB	52/F	46,XX,del(5)(q13:q33)	5q14.3-33.2 (87217489-153708130; 19903-21559; 66490641; 1657; 1.47; 0.21)	13q13.1-14.11 (31506479-39779549; 43528-43816; 8273070; 289; 1.98; 0.27)) 14q12 (24397576-26870017; 45947-46026; 2472441; 80; 2.00; 0.23) 21q11.1-q22.3 (10000969-46844296; 54389-55266; 36843327; 878 whole Chr; 1.99; 0.29) **RUNX1, U2AF1**	
MDS35	RAEB	58/M	92,XXYY,del(5)(q14:q33)	5q21.3-q35.3 (107008082-180607628; 20365-22122; 73599546; 1758; 1.51; 0.21) **NPM1**	13q21.33-q22.1 (70401435-73818187; 44605-44740; 3416752; 136; 2.02; 0.26) 16p11.1-q24.3 (34953675-88143266; 49577-50361; 53189591; 785; 1.99; 0.27)	13q31.2-q34 (88103652-113215972; 45121-45851; 25112320; 731; 2.45; 0.38)

MDS36	RAEB	NA/F	46,XX,del(5)(q13:q33),del(11)(q23)	5q23.1-q33.2 (116859235-155177249; 20656-21605; 38318014; 950; 1.45; 0.20)	6q23.3 (135385192-138089687; 25328-25388; 2704495; 61;2.02; 0.29) 9q12.1 (56241373-59006184; 30908-30988; 2764811; 81; 2.01; 0.25)	15q13.1-q13.2 (26237007-28085050; 47841-47866; 1848043; 26; 2.12; 0.35)
MDS37	RAEB	58/M	43-45,XY,del(5)(q31),der(7)t(7;12)(q22;q1?3),-12,-13,-19,?del(20)(q1?3)[cp4]	5q21.1-q35.3 (98243608-180607628; 20163-22122; 82364020; 1960; 1.42; 0.20) **NPM1** 7q11.22-q36.3 (69377470-158624663; 27729-29517; 89247193; 1789; 1.43; 0.20) **EZH2** 12p12.1-p13.2 (10219902-22122693; 40783-41038; 11902791; 256; 1.46; 0.22) **ETV6** 13q14.11q14.2 (40036389-46217079; 43817-44001; 6180690; 185; 1.38; 0.18) 13q14.2-q21.1 (47592092-55466888; 44039-44218; 7874796; 180; 1.43; 0.17) 15q12-q13.2 (24374497-28800086; 47820-47867; 4425589; 48; 1.46; 0.27) 17p11.2-p13.3(450509-19519465; 50362-50609; 19068956; 248; 1.39; 0.18) **TP53** 20q11.21-q13.13 (29933631-48271268; 53958-54172; 18337637; 215; 1.37; 0.22) **ASXL1**	5q15 (92105855-95229134; 20011-20072; 3123279; 62; 2.12; 0.44)	6p22.3 (22218694-23666140; 22688-22735; 1447446; 48) 9q21.31-q21.32 (80782354-82945497; 34556-34600; 2163143; 45)
MDS39	RAEB	82/F	46,XX,del(5)(q13:q33),t(6;12)(q13;p12)[2]/45,XX,-7,-22/46,XX,del(5)(q13:q33),t(6;12)(q13;p12),+mar[15]/46,XX[3]	5q14.2-q34 (81511479-161557314; 19775-21753; 80045835; 1979; 1.53; 0.44) 7p22.3-p11.2 (250149-56479844; 26082-27645; 56229695; 1564; 1.58; 0.45) 7q21.3-q36.3 (94919442-158624663;	2p22.2-p22.1 (38319370-41333317; 5245-5334; 3013947; 90; 2.29; 0.70)	Multiple gain of copy number

				28305-29517; 63705221; 1213; 1.57; 0,47) **EZH2**		
MDS40	RAEB	54/F	46,XX,del(5)(q14:q34)	5q14.3-q34 (87425364-161996101; 19904-21761; 74570737; 1858; 1.47; 0.17)	6q13-q15 (72449224-89814330; 23699-24130; 17365106; 432; 1.99; 0.30) **SRSF2** 9p21.3-p22.2 (18466830-21763347; 33640-33730; 3296517; 91; 1.98; 0.32) 9p21.1-p21.2 (26316539-30212869; 33860-34045; 3896330; 186; 2.00; 0.24) 12q24.13-q24.21 (111751790-115292641; 43011-43072; 3540851; 62; 2.02; 0.32)	
MDS41	RAEB	56/M	46, XY,del(5)(q14:q34)[2];47 ,XY,del(5)(q14:q34),+21[20]	5q22.3-q34 (20585-21899; 115079389-166312668; 51233279; 1315; 1.57; 0.39)	5q14.3 (86007785-88632975; 19881-19936; 2625190; 56; 1.88; 0.40)	Multiple gains of copy number
MDS42	CMML	45/M	46,XY,del(5)(q13:q33),d el(13)(q12:q22)	5q14.3-q33.3 (85143956-159665477; 19861-21719; 74521521; 1859; 1.45; 0.24) 13q13.2-q21.31 (33495418-61943642; 43601-44385; 28448224; 785; 1.44; 0.21	8q21.11 (75982355-78287079; 31427-31506; 2304724; 80; 2.06; 0.32) 11q22.1-q25 (97063972-134173875; 39904-40623; 37109903; 720; 2.00; 0.31) **CBL**	4q21.21 (79349747-81034492; 15226-15290; 1684745; 65; 2.28; 0.44)
MDS43	RAEB	79/F	46,XX,del(5)(q15:q33)	5q21.1-q33.2 (101389190-154492074; 20226-21568; 53102884; 1343; 1.50; 0.21)	4q13.1 (60577199-62928065; 14761-14841; 2350866; 81; 1.89; 0.24)	

Table S6. List of genes affected by cytogenetic loss.

	5q- Syndrome (n=18)	*RA del(5q) with additional karyotypic abnormalities (n=6)*	*Advanced del(5q) cases (n=9)*
EZH2			2
NPM1			2
TP53			1
ETV6		1	1
ASXL1			1
CBL		1	
DNMT3A	1		

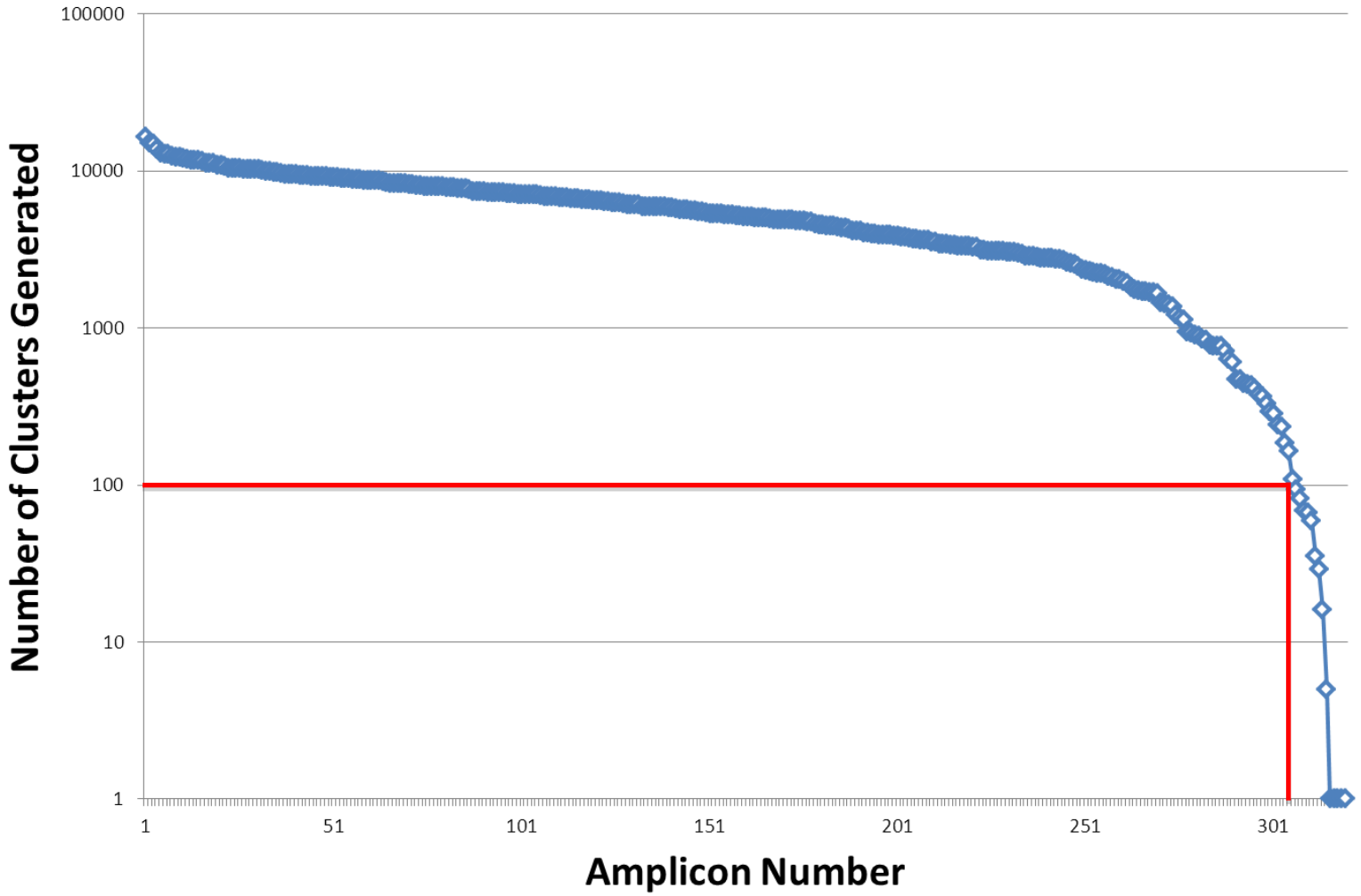

Figure S1. Number of clusters generated per amplicon in the panel during the MDS del(5q) cohort MiSeq sequencing run. A total of 96% (308/322) of all amplicons generated at least 100 clusters during sequencing (average 5,362 clusters/amplicon).

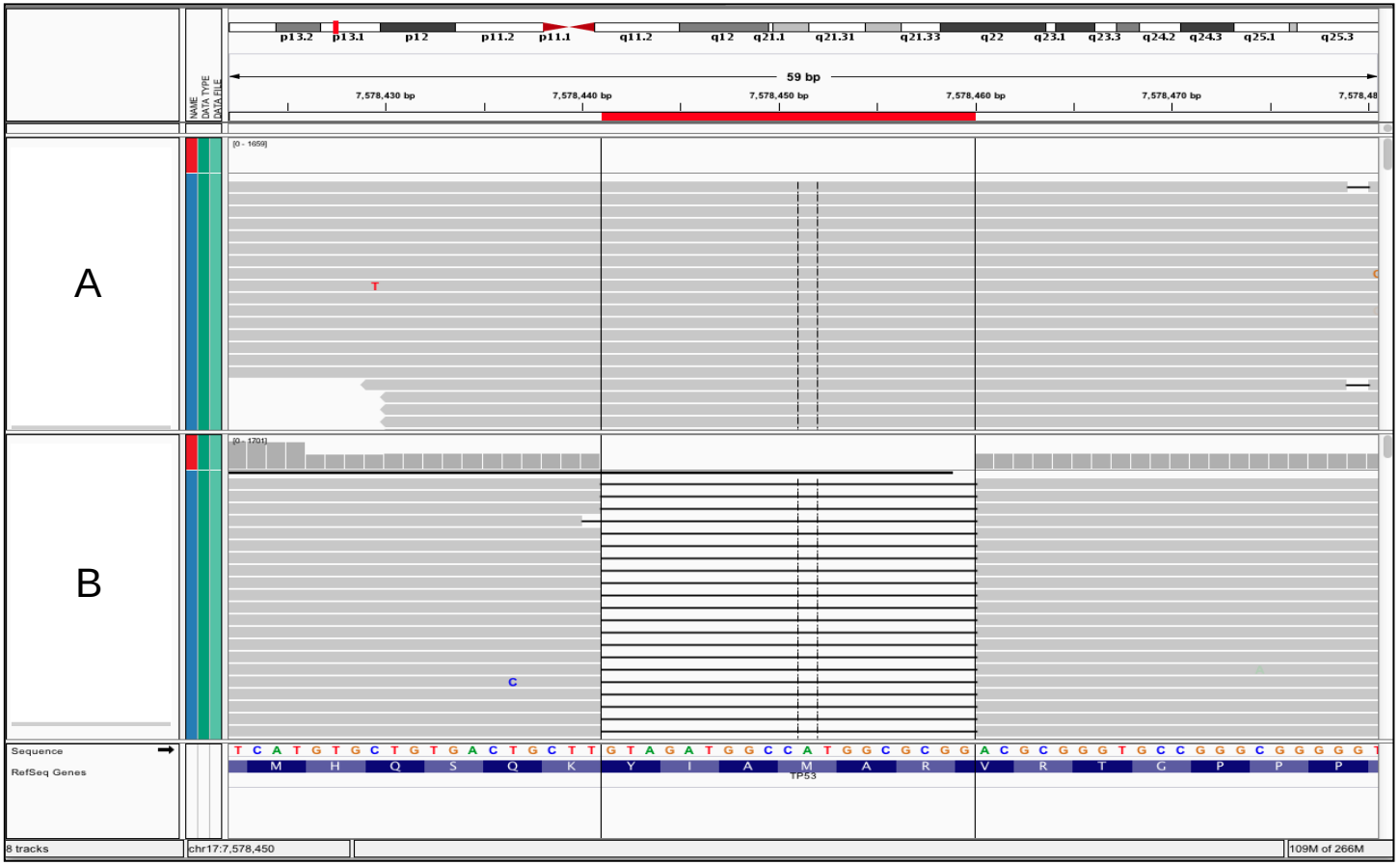

Figure S2. Comparison of read alignments covering the 19bp *TP53* deletion in sample TEST009. The initial read alignment and variant calling (BaseSpace, **A**) failed to align any reads containing deletions to the reference genome, resulting in a much lower read depth across this locus (~30x). By comparison, re-analysis of the same data using the Stampy and Platypus pipeline (**B**) resulted in a greater number of aligned reads, giving a higher read depth (>700x) and successfully identified the deletion.

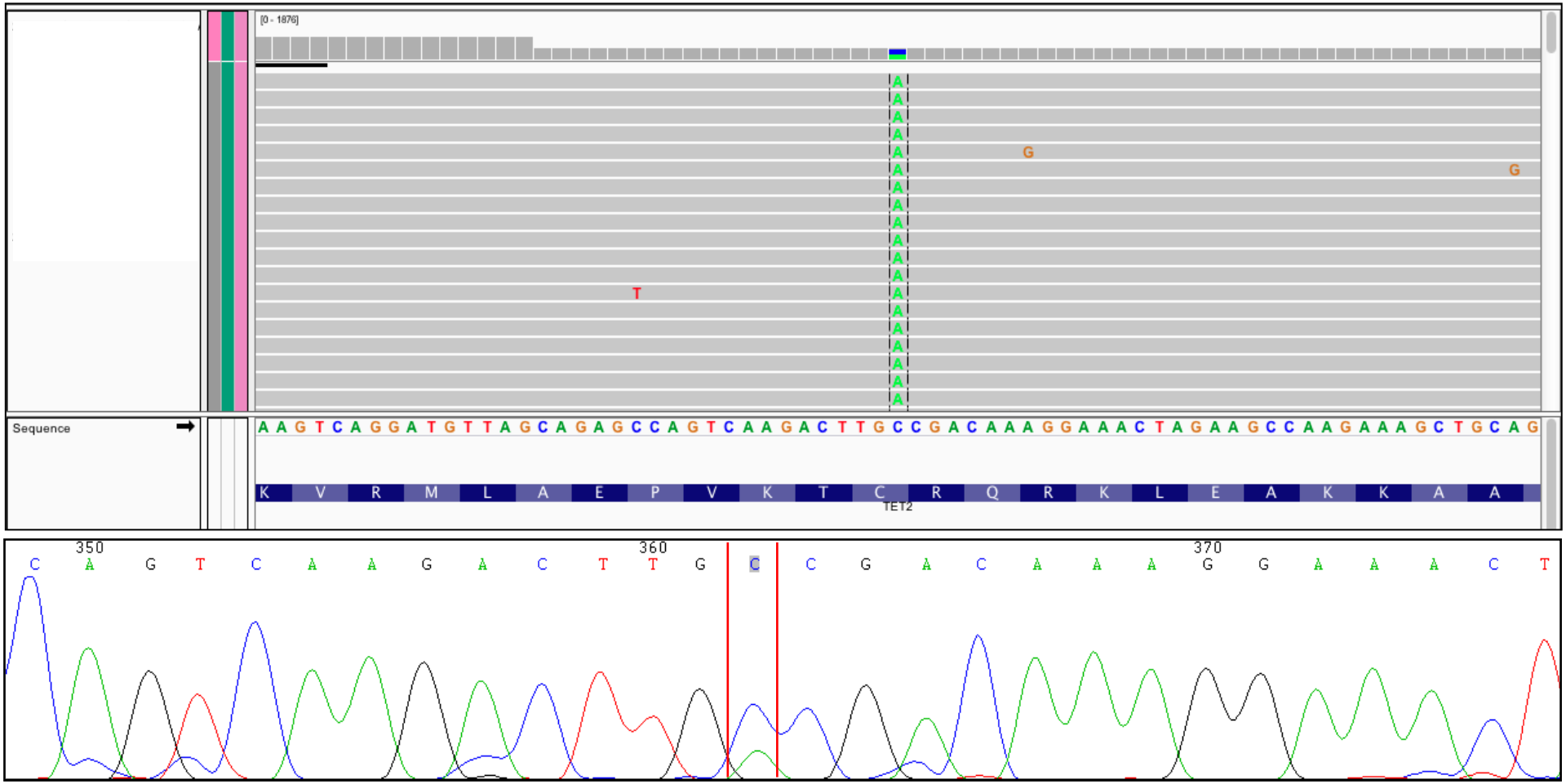

Figure S3. Comparison of the *TET2* C1464X mutation in sample TEST001 by Sanger and next-generation sequencing. The C1464X variant was detected and called in the MiSeq data (top) at a frequency of 47% (200/423 reads). The variant can be seen in the Sanger sequencing trace (bottom), but was not identified by the Mutation Surveyor software due to the relatively high background noise in the data.

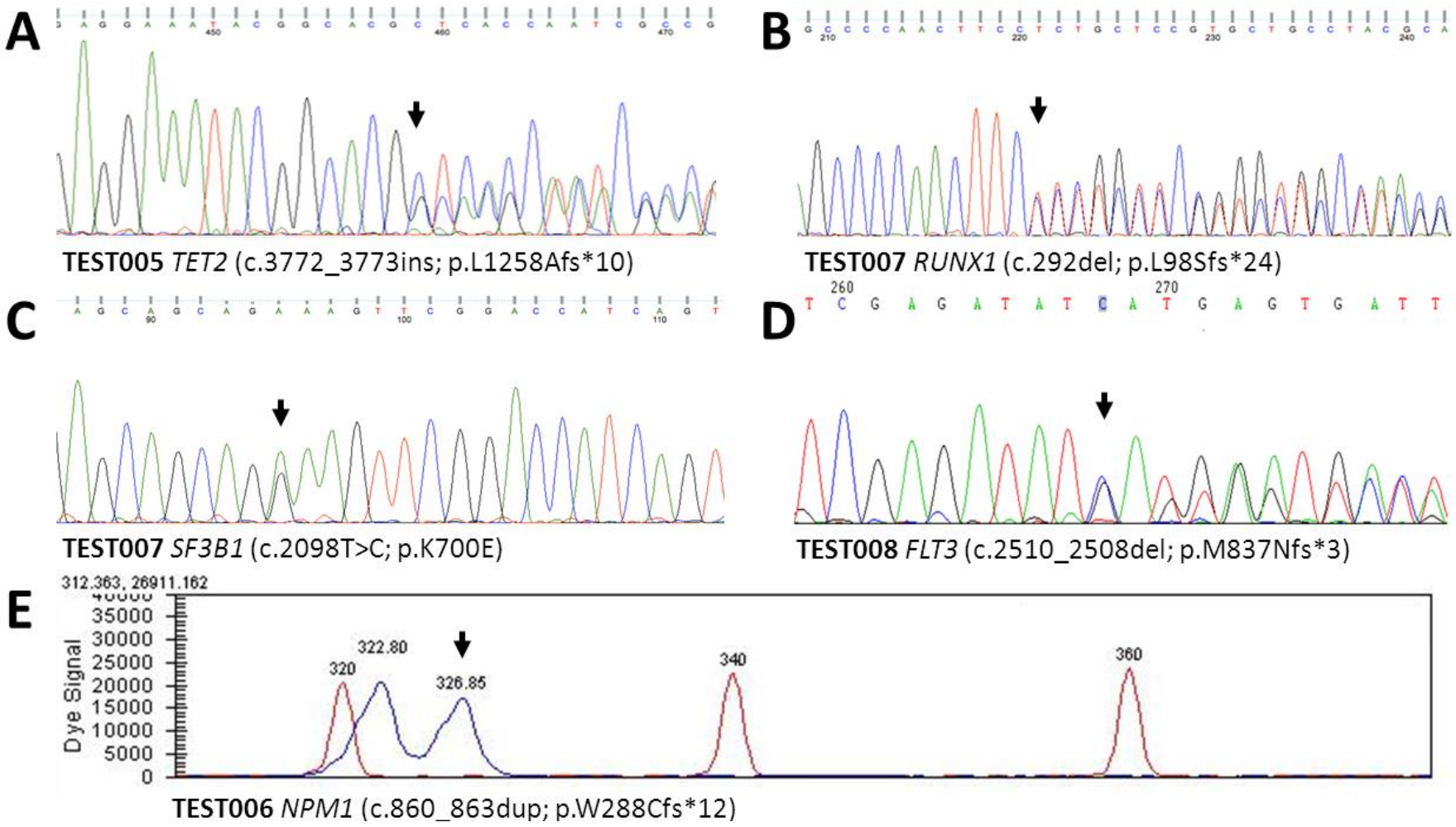

Figure S4. **Validation of new mutations found by MiSeq in the validation cohort in addition to *TET2* C1464X.** The remaining new mutations were confirmed by Sanger sequencing (**A-D**) or fragment analysis (**E**).

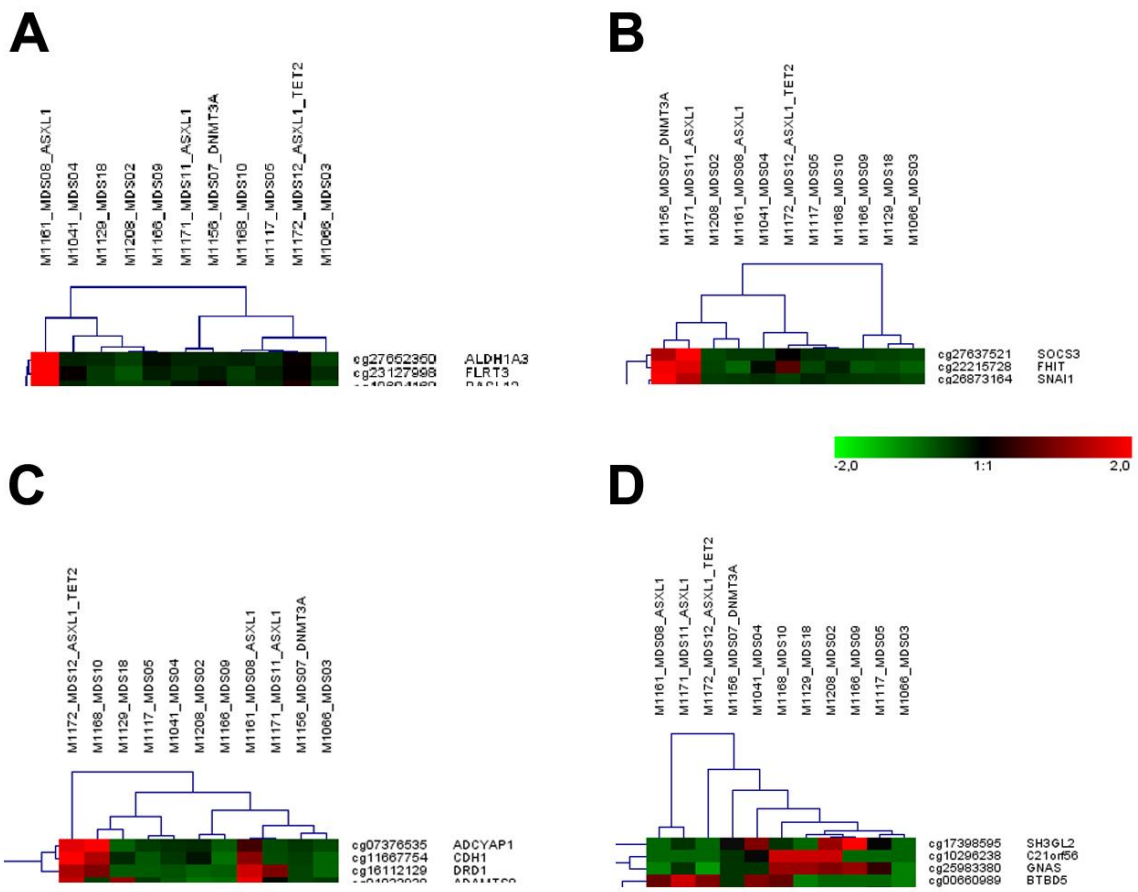

Figure S5. Supervised clustering using methylation data from 11 5q- syndrome cases. All pictures have been cropped to show the hierarchical clustering at the top. (**A**) Clustering using 422 differentially methylated genes between 2 *ASXL1*-mut cases and 7 cases with no epigenetic gene mutations. (**B**) Clustering using 144 differentially methylated genes between 1 *DNMT3A*-mut case and 7 cases with no epigenetic gene mutations. (**C**) Clustering using 156 differentially methylated genes between 1 *ASXL1* & *TET2*-mut case and 7 cases with no epigenetic gene mutations. (**D**) Clustering using 205 differentially methylated genes between 3 *ASXL1*-mut cases and 7 cases with no epigenetic gene mutations.