Distinct molecular abnormalities underlie unique clinical features of essential thrombocythemia in children

Supplementary methods

Patients and diagnostic criteria

The study was approved by the hospital-based ethics committees according to the guidelines of the Helsinki Declaration. Informed consent was obtained from the patients' legal guardians. Major bleeding events were defined in agreement with the international recommendations.[1] Major thrombotic events were defined as ischemic stroke, cerebral transient ischemic attacks, acute myocardial infarction, or thrombosis in vital visceral organs or peripheral vessels.[2] The diagnosis of post-ET myelofibrosis was made according to the criteria suggested by the International Working Group for Myelofibrosis Research and Treatment.[3]

Investigation of JAK2, MPL, CALR mutations and the JAK2 46/1 haplotype

Genomic DNA was extracted from bone marrow cells or granulocytes isolated from peripheral blood samples using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA) according to the manufacturer's instructions. Germline DNA was extracted from nails or hair follicles using the QIAamp DNA Investigator Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. JAK2 V617F mutation was detected by a quantitative real-time polymerase chain reaction (PCR) assay, using 20ng genomic DNA as described previously.[4] Briefly, PCR amplification was performed on an ABI StepOne Detection System (Applied Biosystems, Foster City, CA, USA) according to the following cycling conditions: 10 minutes at 95℃ followed by 40 cycles of 15 seconds at 95℃ and 60 seconds at 60℃. Primers for JAK2 V617F mutation (forward primer

5’-AAGCTTTCTCACAAGCATTTGGTTT-3’ and reverse primer 5’-AGAAAGGCATTAGAAAGCCTGTAGTT-3’) were used together with TaqMan probes (Applied Biosystems, Foster City, CA, USA) specific for either wild type (VIC-5’-TCTCCACAGACACATAC-3’-MGB) or mutant JAK2 allele (FAM-5’-TCCACAGAAACATAC-3’-MGB). DNA from a healthy control homozygous for wild type JAK2 and from a polycythemia vera patient harboring 100% mutant allele were mixed in various proportions to generate a standard curve for $JAK2_{V617F}/JAK2_{total}$ against ΔCt ($Ct_{JAK2\ V617F}$ - $Ct_{JAK2\ WT}$). All samples were measured in triplicate, and the mean ΔCt was used to calculate $JAK2_{V617F}/JAK2_{total}$. Genotypes of MPL exon 10, CALR exon 9, and the JAK2 46/1 haplotype (rs12340895) were assessed by Sanger sequencing. Primers used for amplificating CALR exon 9, MPL exon 10 and JAK2 46/1 haplotype were as follows: forward primer 5’-AAAGCAAGGGCTATCGGGTAT-3’ and reverse primer 5’-GGGGACATCTTCCTCCTCATC-3’ (CALR); forward primer 5’-CCGAAGTCTGACCCTTTTTG-3’ and reverse primer 5’-ACAGAGCGAACCAAGAATGC-3’ (MPL)[5]; forward primer 5’-TTTAAGAAACTGCCACAGCTACC-3’ and reverse primer 5’-CCTGGATTCCTTCATGGTCC-3’ (JAK2 46/1 haplotype).

Selection of candidate genes and target enrichment kit design

DNA samples from granulocytes were available for next-generation sequencing in 25 patients. The list of the 55 candidate genes is available in Supplementary Table 1. Candidate genes were selected based on the data in the Catalogue of Somatic Mutations in Cancer (COSMIC) (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/) and in the previous literature.[5-13] The coordinates of the sequence data were based on the National Center for Biotechnology Information (NCBI) build 37 (UCSC hg19).[14] We designed a customized Agilent SureSelect target capture kit

(https://earray.chem.agilent.com/suredesign/index.htm) to enrich for candidate genes. Sureselect baits were designed to capture the 2-kb upstream region of the transcription start site, all exons, and 20 bp of each exon-intron junction, with a total size of ~0.35 Mb.

Library preparation, target enrichment, sequencing and data analysis

Sample preparation was performed according to the manufacturer's instructions (SureSelect XT Custom 1kb-499kb library, Cat. No. 5190–4806, SureSelect Library prep kit; Agilent Technologies, Inc.). DNA paired-end 200bp short-insert libraries were constructed and captured libraries were then loaded on the Illumina Hiseq2000 platform. The bioinformatics analysis began from the sequencing data (raw data) which generated from the Illumina pipeline. The reads with too many N bases (> 10%) or low base quality (> 50% bases with base quality < 5) were discarded. Alignment of the sequence data was performed using the Burrows-Wheeler Aligner,[15] to hg19.[14] SAMTools[16] was used to remove duplicated reads. SOAPsnp was used to detect single-nucleotide polymorphisms (SNP).[17] Small Insertion/Deletions were detected by SAMTools.[16] We used ANNOVAR[18] to annotate the confident variant results. SNP in the 1000 Genomes Project (http://www.1000genomes.org) was excluded. Mutations that were marked as polymorphisms or non-cancer specific variations in the SM2PH (http://decrypthon.igbmc.fr/sm2ph/cgi-bin/home), SNPeffect (http://snpeffect.switchlab.org/) or CanProVar 2.0 database (http://lifecenter.sgst.cn/CanProVar/index.html) were also excluded.

Assessment of the possible effects of somatic mutations

According to their effect, somatic mutations were classified into three groups: oncogenic mutations, possible oncogenic mutations, and mutations of unknown effect.[19] Mutations were considered to be oncogenic in case of a truncating variant or a missense variant described as pathogenic in previous

literature. Novel variants that were not present in the dbSNP137 database (http://www.ncbi.nlm.nih.gov/SNP) were called possible oncogenic mutations if they cluster (± 3aa) with known oncogenic variants in the COSMIC database.[19] Conserved residues in different mammals usually play an important role for maintaining the normal structure and function of the protein. So variants in conserved residues were also considered as possible oncogenic mutations. Protein sequences, gene functions, and pathway analysis were based on the NCBI resources (http://www.ncbi.nlm.nih.gov/protein/) and the Gene Ontology website (http://www.geneontology.org/). The three-dimensional structure was analyzed using the Cn3D software (http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml).

PCR amplification conditions

PCR amplification conditions were 95 ℃ for 5 min; 94 ℃ for 30 s, 56 ℃ to 65 ℃ for 30 s, 72 ℃ for 30 s for 35 cycles; and 72 ℃ for 10 min. All PCR products were confirmed by 2% agarose gel electrophoresis, purified using QIAquick Spin Kit (Qiagen, SantaClara, CA, USA) and sequenced using ABI PRISM 3730 ×l DNA Analyzer (Applied Biosystems, Foster City, CA, USA).

Statistical analysis

Data were analyzed using the software SPSS 16.0. Categorical variables were compared by χ^2 test or Fisher exact test. The Mann–Whitney *U*-test or *t*-test was used for quantitative values. A two-tailed $P \leq 0.05$ was considered statistically significant.

References

1. Schulman S, Kearon C. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. *J Thromb Haemost* 2005; **3:** 692-694.

2. Barbui T, Finazzi G, Carobbio A, Thiele J, Passamonti F, Rumi E *et al.* Development and validation of an International Prognostic Score of thrombosis in World Health Organization-essential thrombocythemia (IPSET-thrombosis). *Blood* 2012; **120:** 5128-5133.

3. Barosi G, Mesa RA, Thiele J, Cervantes F, Campbell PJ, Verstovsek S *et al.* Proposed criteria for the diagnosis of post-polycythemia vera and post-essential thrombocythemia myelofibrosis: a consensus statement from the International Working Group for Myelofibrosis Research and Treatment. *Leukemia* 2008; **22:** 437-438.

4. Carobbio A, Antonioli E, Guglielmelli P, Vannucchi AM, Delaini F, Guerini V *et al.* Leukocytosis and risk stratification assessment in essential thrombocythemia. *J Clin Oncol* 2008; **26:** 2732-2736.

5. Brecqueville M, Rey J, Bertucci F, Coppin E, Finetti P, Carbuccia N *et al.* Mutation analysis of ASXL1, CBL, DNMT3A, IDH1, IDH2, JAK2, MPL, NF1, SF3B1, SUZ12, and TET2 in myeloproliferative neoplasms. *Genes Chromosomes Cancer* 2012; **51:** 743-755.

6. Reuther GW. Recurring mutations in myeloproliferative neoplasms alter epigenetic regulation of gene expression. *Am J Cancer Res* 2011; **1:** 752-762.

7. Hou Y, Song L, Zhu P, Zhang B, Tao Y, Xu X *et al.* Single-cell exome sequencing and

monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* 2012; **148:** 873-885.

8. Greenberg PL. Molecular and genetic features of myelodysplastic syndromes. *Int J Lab Hematol* 2012; **34:** 215-222.

9. Tefferi A. Novel mutations and their functional and clinical relevance in myeloproliferative neoplasms: JAK2, MPL, TET2, ASXL1, CBL, IDH and IKZF1. *Leukemia* 2010; **24:** 1128-1138.

10. Larsson CA, Cote G, Quintas-Cardama A. The changing mutational landscape of acute myeloid leukemia and myelodysplastic syndrome. *Mol Cancer Res* 2013; **11:** 815-827.

11. Thol F, Kade S, Schlarmann C, Loffeld P, Morgan M, Krauter J *et al.* Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood* 2012; **119:** 3578-3584.

12. Harada H, Harada Y, Niimi H, Kyo T, Kimura A, Inaba T. High incidence of somatic mutations in the AML1/RUNX1 gene in myelodysplastic syndrome and low blast percentage myeloid leukemia with myelodysplasia. *Blood* 2004; **103:** 2316-2324.

13. Bejar R, Stevenson K, Abdel-Wahab O, Galili N, Nilsson B, Garcia-Manero G *et al.* Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med* 2011; **364:** 2496-2506.

14. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J *et al.* Initial sequencing and analysis of the human genome. *Nature* 2001; **409:** 860-921.

15. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform.

Bioinformatics 2009; **25:** 1754-1760.

16. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25:** 2078-2079.

17. Li R, Li Y, Fang X, Yang H, Wang J, Kristiansen K. SNP detection for massively parallel whole-genome resequencing. *Genome Res* 2009; **19:** 1124-1132.

18. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; **38:** e164.

19. 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:** 3616-3627.

20. Zhong Y, Wu J, Ma R, Cao H, Wang Z, Ding J *et al.* Association of Janus kinase 2 (JAK2) polymorphisms with acute leukemia susceptibility. *Int J Lab Hematol* 2012; **34:** 248-253.

21. Gelsi-Boyer V, Brecqueville M, Devillier R, Murati A, Mozziconacci MJ, Birnbaum D. Mutations in ASXL1 are associated with poor prognosis across the spectrum of malignant myeloid diseases. *J Hematol Oncol* 2012; **5:** 12.

22. Wu SJ, Tang JL, Lin CT, Kuo YY, Li LY, Tseng MH *et al.* Clinical implications of U2AF1 mutation in patients with myelodysplastic syndrome and its stability during disease progression. *Am J Hematol* 2013; **88:** E277-282.

23. Przychodzen B, Jerez A, Guinta K, Sekeres MA, Padgett R, Maciejewski JP *et al.* Patterns of missplicing due to somatic U2AF1 mutations in myeloid neoplasms. *Blood* 2013; **122:**

999-1006.

24. Klco JM, Spencer DH, Miller CA, Griffith M, Lamprecht TL, O'Laughlin M *et al.* Functional heterogeneity of genetically defined subclones in acute myeloid leukemia. *Cancer Cell* 2014; **25:** 379-392.

25. De Filippi P, Zecca M, Lisini D, Rosti V, Cagioni C, Carlo-Stella C *et al.* Germ-line mutation of the NRAS gene may be responsible for the development of juvenile myelomonocytic leukaemia. *Br J Haematol* 2009; **147:** 706-709.

26. Murphy SJ, Hart SN, Lima JF, Kipp BR, Klebig M, Winters JL *et al.* Genetic alterations associated with progression from pancreatic intraepithelial neoplasia to invasive pancreatic tumor. *Gastroenterology* 2013; **145:** 1098-1109.

Supplementary Table 1. List of the 55 candidate genes in high-throughout targeted capture sequencing

No.	Genes	NCBI：ID	Position	No.	Genes	NCBI：ID	Position
1	ASXL1	171023	20q11.1	29	PDGFRA	5156	4q12
2	ATRX	546	Xq21.1	30	PHF6	84295	Xq26.2
3	BCOR	54880	Xp11.14	31	PTEN	5728	10q23.3
4	CBL	867	11q23.3	32	PTPN11	5781	12q24.1
5	CDKN2A	1029	9p21	33	RUNX1	861	21q22.3
6	CEBPA	1050	19q13.1	34	SF3B1	23451	2q33.1
7	CSF1R	1436	5q2	35	SH2B3	10019	12q24.12
8	DNMT3A	1788	2p23	36	SOCS1	8651	16p13.13
9	ETV6	2120	12p13.2	37	SRSF2	6427	17q25.1
10	EZH2	2146	7q35-q36	38	TET2	54790	4q24
11	FLT3	2322	13q12	39	TP53	7157	17p13.1
12	GATA1	2623	Xp11.23	40	U2AF1	7307	21q22.3
13	GATA2	2624	3q21.3	41	WT1	7490	1p13
14	GNAS	2778	20q13.3	42	ZRSR2	8233	Xp22.1
15	HRAS	3265	11p15.5	43	THPO	7066	3q27
16	IDH1	3417	2q33.3	44	EED	8726	11q14.2-q22.3
17	IDH2	3418	15q26.1	45	JARID2	3720	6p24-p23
18	IKZF1	10320	7p13	46	RBBP4	5928	1p35.1
19	JAK2	3717	9p24	47	SUZ12	23512	17q11.2
20	JAK3	3718	19p13.1	48	STAT5B	6777	17q11.2
21	KIT	3815	4q12	49	STAT3	6774	17q21.31
22	KRAS	3845	12p12.1	50	SOCS2	8835	12q
23	MECOM	2122	3q26	51	SOCS3	9021	17q25.3
24	KMT2A	4297	11q23	52	CBFB	865	16q22.1
25	MPL	4352	1p34.2	53	PPP2R4	5524	9q34
26	NF1	4763	17q11.2	54	IER3	8870	6p21.3
27	NPM1	4869	5q35	55	SPARC	6678	5q31.3-q32
28	NRAS	4893	1p13.2				

Abbreviations: NCBI, National Center of Biotechnology Information

Supplementary Table 2. Primers for validating mutations by Sanger sequencing

Mutations	Forward (5'-3')	Reverse (5'-3')	Length (bp)	T (℃)
JAK2 I354T	ATTGGTATGTAGGGCTGATGC	TGGGCCATGACAGTTGCT	799	60
JAK2 G127D	CCTTTCAGTATGCTGTAGGTGACT	ACCTGAGCAAAGAGGTAAGACA	211	60
ASXL1 M1096L	CACGGATACAGCCTCTGACTT	ACTTGGGACTGCCTTGCA	634	57
ASXL1 R693X	GGTCAGCCCACTTACCAGAT	TCCTTGCTCCTCATCATCACT	654	57
ASXL1 R1068X	CACGGATACAGCCTCTGACTT	ACTTGGGACTGCCTTGCA	634	57
ASXL1 R1273H	GCAGTTCTCTTCCTTTAGTT	CCTGTAACATTGCTCTGAT	198	58
SH2B3 P126Q	CGCGACTACCGGGACACA	AGGCTCCAGGGCAGGAACT	302	56
U2AF1 S34F	ACAGCAAGGAAGAGGAGGTG	AAACAAGGAGTGGTGGTCTCA	448	65
FLT3 R372G	CTCAATACGTTTCCCTATGAA	TGTATCCGTTATCAAGACCC	229	60
NRAS G13D	CAGACAGTCTCGCTACTATGGC	AGGGAGAATGGGAAGGAGTT	766	57
MLL G73E	GGCGCAGAGCTGGTTAGG	GCAGGTTGGTGCCGATG	488	65
GNAS R186H	TCCGTTGAGCCTGACCTTG	AGAGCGTGAGCAGCGACC	591	57
RUNX1 P76L	AACTTCCTCTGCTCCGTGCTG	TGCTGCTAGTGTAGTAGGGTGA	997	60
CALR L367fs*46	AAAGCAAGGGCTATCGGGTAT	GGGGACATCTTCCTCCTCATC	407	57
WT1 P136delinsPP	TGTGCCCTGCCTGTGAGC	TGGACAGTGAAGGCGCTCAG	215	58
ZRSR2 G438delinsGSR	AGAAATGGGGAATCCGAGAG	CCATCTGCGTTCATAGCCTTC	361	57
STAT5B K71delinsKK	GGATGTGTTGAGGCTGTATTTA	AGCTGTGTGGCATAGTGC	227	60
SH2B3 R140H	CGCGACTACCGGGACACA	AGGCTCCAGGGCAGGAACT	302	56
U2AF1 G217S	TTCCACAAGAAATAAGTTACACCA	GAAGAAGACAGAAGATGCAAGG	567	57
FLT3 G282E	CGGGAAAAGGAAGAGGGT	TGCTGTGCCAGAAATGAACT	937	57
IDH1 I189V	AGAATCATAGGGATAGGGAGATACA	ATTTGGTTGTGGTGGGTGA	359	60
JAK3 R175Q	ATTGGTTTGGGCTGGAGAAG	GGGTCATAGGAACACCCTGA	575	58
GATA1 H71P	GGGGTGGAAGGATTTCTGTG	GCAAAATCAGTCTAGGGTCAGC	764	57
CSF1R R748W	CCCCATTTGCTTGATAACA	GCCAGGTTCCTACTCACATT	212	58

Abbreviations: T, annealing temperature

Supplementary Table 3. Clinical and laboratory characteristics of 63 children diagnosed with essential thrombocythemia

	Total (n = 63)	JAK2 V617F + (n = 14)	JAK2 V617F – CALR- (n = 49)	*P*
Clinical characteristics at diagnosis				
M/F	38/25	8/6	30/19	0.783
Age, median (range), y	11 (3–14)	10 (5–14)	12 (3–14)	0.153
Symptoms, n (%)	39 (61.9)	8 (57.1)	31 (63.3)	0.677
Organomegaly, n (%)	33 (52.4)	7 (50.0)	26 (53.1)	0.840
Hematologic features				
WBCs, median (range), $\times 10^9$/L	10.6 (4.3–24.1)	11.8 (8.1–18.7)	10.0 (4.3–24.1)	0.126
HGB, median (range), g/L	127 (110–167)	135 (110–167)	126 (110–163）	0.118
PLTs, median (range), $\times 10^9$/L	1224 (513–4378)	1109 (714–3958)	1320 (513–4378)	0.381
Complications and Clinical outcomes				
Major thrombosis, n (%)	3 (4.8)	2 (14.3)	1 (2.0)	0.121
Microvascular disturbances, n (%)	30 (47.6)	5 (35.7)	25 (51.0)	0.312
Major bleeding events, n (%)	0	0	0	
Minor bleeding events, n (%)	9 (14.3%)	4 (28.6%)	5 (10.2)	0.194
Transformation to myelofibrosis	2 (3.2)	1 (7.1)	1 (2.0)	0.398
Secondary malignancies, n (%)	0	0	0	
Follow-up, n (%)				
Median (range), months	48 (12–302)	60 (12–302)	47 (12–156)	0.153

Abbreviations: F, female; HGB, hemoglobin; M, male; PLT, platelet; WBC, white blood cell.

Supplementary Table 4. Clinical and hematologic features of 25 children with essential thrombocythemia involved in targeted high-throughout sequencing

	Total pediatric ET (n = 25)	Mutated (n = 14)	Non-mutated (n = 11)	*P*
Clinical characteristics at diagnosis				
M/F	13/12	7/7	6/5	1.000
Age, median (range), y	10 (3-14)	11.5 (3-14）	9 (4-13)	0.144
Symptoms, n (%)	16 (64.0)	8 (57.1)	8 (72.7)	0.677
Organomegaly, n (%)	9 (36.0)	5 (35.7)	4 (36.4)	1.000
Hematologic features				
WBCs, median (range), $\times 10^9$/L	11.2 (6.8-23.4)	13.0 (7.2-23.4)	9.6 (6.8-15.3)	**0.025**
HB, median (range), g/L	125 (110-167)	128 (110-167)	124 (113-149)	0.584
PLTs, median (range), $\times 10^9$/L	1399 (513-4378)	1362 (714-3958)	1399 (513-4378)	0.622
Complications				
Major thrombosis	1 (4.0)	1 (7.1)	0	
Major bleeding events	0	0	0	
Treatment				
No treatment	5 (20.0)	3 (21.4)	2 (18.2)	1.000
Antiplatelet therapy	14 (56.0)	9 (64.3)	5 (45.5)	0.435
Cytoreductive drugs	15 (60.0)	7 (50.0)	8 (72.7)	0.414
Follow-up, n (%)				
Median, months (range)	46 (12-180)	52 (18-180)	43 (12-166)	

Supplementary Table 5. Total single nucleotide variants identified by targeted high-throughout sequencing

Gene	CHR	POS	rsID	1kg Frequency	*Cosmic	Exonic Function	TransID	c.	p.	No. of Samples
HRAS	chr11	534242	rs12628	0.19	1	synonymous	NM_001130442	c.T81C	p.H27H	10
JAK2	chr9	5022042	rs150159583	.	0	nonsynonymous	NM_004972	c.A55T	p.I19L	1
JAK2	chr9	5044432	rs56118985	0.01	0	nonsynonymous	NM_004972	c.G380A	p.G127D	1
JAK2	chr9	5050706	rs2230722	0.26	0	synonymous	NM_004972	c.C489T	p.H163H	12
JAK2	chr9	5064887	.	.	0	nonsynonymous	NM_004972	c.T1061C	p.I354T	1
JAK2	chr9	5073770	rs77375493	.	1	nonsynonymous	NM_004972	c.G1849T	p.V617F	6
JAK2	chr9	5081780	rs2230724	0.41	0	synonymous	NM_004972	c.G2490A	p.L830L	15
JAK2	chr9	5126407	rs140219534	0.01	0	synonymous	NM_004972	c.T3252C	p.N1084N	2
TP53	chr17	7578456	rs139200646	.	1	synonymous	NM_001126115	c.C78T	p.R26R	1
TP53	chr17	7578546	.	.	1	synonymous	NM_000546	c.T384G	p.P128P	1
TP53	chr17	7579472	rs1042522	0.61	0	nonsynonymous	NM_000546	c.C215G	p.P72R	19
ETV6	chr12	11992168	rs11611479	0.19	0	synonymous	NM_001987	c.G258A	p.T86T	12
JARID2	chr6	15496662	rs742099	0.28	0	synonymous	NM_004973	c.C1206T	p.P402P	9
JARID2	chr6	15513482	rs2235258	0.66	1	synonymous	NM_004973	c.G3279A	p.L1093L	19
JARID2	chr6	15513542	rs200784767	0.0017	0	synonymous	NM_004973	c.C3339T	p.H1113H	1
ZRSR2	chrX	15821890	rs188867560	0.0047	0	nonsynonymous	NM_005089	c.G283A	p.A95T	2
ZRSR2	chrX	15838366	rs2301724	0.61	0	synonymous	NM_005089	c.C864T	p.N288N	19
JAK3	chr19	17953878	.	.	0	nonsynonymous	NM_000215	c.G524A	p.R175Q	1
KRAS	chr12	25368462	rs4362222	1	0	synonymous	NM_033360	c.G483A	p.R161R	24
DNMT3A	chr2	25536827	rs41284843	0.08	0	synonymous	NM_022552	c.C27T	p.P9P	6
FLT3	chr13	28599041	.	.	0	synonymous	NM_004119	c.G2247A	p.S749S	1
FLT3	chr13	28608459	rs34374211	0.16	1	synonymous	NM_004119	c.A1683G	p.L561L	6
FLT3	chr13	28622503	.	.	0	nonsynonymous	NM_004119	c.A1114G	p.R372G	1
FLT3	chr13	28622544	rs34172843	0.03	0	nonsynonymous	NM_004119	c.A1073T	p.D358V	2
FLT3	chr13	28623809	.	.	0	nonsynonymous	NM_004119	c.G845A	p.G282E	1
FLT3	chr13	28624294	rs1933437	0.75	0	nonsynonymous	NM_004119	c.C680T	p.T227M	24
FLT3	chr13	28636084	rs7338903	1	0	synonymous	NM_004119	c.C288T	p.D96D	25
FLT3	chr13	28674628	rs12872889	0.31	0	nonsynonymous	NM_004119	c.A20G	p.D7G	10
NF1	chr17	29490284	rs146691765	0.01	0	synonymous	NM_000267	c.C369G	p.T123T	1
NF1	chr17	29508775	rs1801052	0.59	0	synonymous	NM_000267	c.G702A	p.L234L	22
NF1	chr17	29552200	rs146051850	0.01	0	nonsynonymous	NM_000267	c.A1933G	p.M645V	1
NF1	chr17	29553485	rs2285892	0.51	0	synonymous	NM_000267	c.G2034A	p.P678P	21
NF1	chr17	29677247	rs201287021	0.0035	0	synonymous	NM_000267	c.A7305G	p.K2435K	2
NF1	chr17	29686006	rs78610439	.	0	nonsynonymous	NM_000267	c.G8070T	p.L2690F	2
SUZ12	chr17	30264415	.	.	0	synonymous	NM_015355	c.C150T	p.G50G	1
ASXL1	chr20	31019164	.	.	0	synonymous	NM_015338	c.G759A	p.E253E	1
ASXL1	chr20	31022469	rs3746609	0.08	0	nonsynonymous	NM_015338	c.G1954A	p.G652S	5

ASXL1	chr20	31022592	.	.	1	stopgain	NM_015338	c.C2077T	p.R693X	1
ASXL1	chr20	31022959	rs6058694	1	0	nonsynonymous	NM_015338	c.T2444C	p.L815P	25
ASXL1	chr20	31023717	.	.	1	stopgain	NM_015338	c.C3202T	p.R1068X	1
ASXL1	chr20	31023801	.	.	0	nonsynonymous	NM_015338	c.A3286C	p.M1096L	1
ASXL1	chr20	31024001	.	.	0	synonymous	NM_015338	c.T3486G	p.V1162V	1
ASXL1	chr20	31024274	rs4911231	0.74	0	synonymous	NM_015338	c.T3759C	p.S1253S	25
ASXL1	chr20	31024333	.	.	1	nonsynonymous	NM_015338	c.G3818A	p.R1273H	1
WT1	chr11	32417945	rs16754	0.71	1	synonymous	NM_000378	c.A1056G	p.R352R	20
WT1	chr11	32456562	rs1799925	0.7	0	synonymous	NM_000378	c.C330T	p.P110P	22
WT1	chr11	32456694	rs2234582	0.06	0	synonymous	NM_000378	c.G198T	p.P66P	1
RBBP4	chr1	33133968	rs2762904	1	0	synonymous	NM_001135255	c.T450C	p.V150V	25
RUNX1	chr21	36259183	.	.	0	nonsynonymous	NM_001001890	c.C227T	p.P76L	1
BCOR	chrX	39911528	rs200163930	.	0	nonsynonymous	NM_001123384	c.T4946G	p.V1649G	8
BCOR	chrX	39913225	.	.	0	synonymous	NM_001123384	c.C4734T	p.D1578D	1
BCOR	chrX	39931908	rs3810693	0.03	0	synonymous	NM_001123383	c.G2691C	p.S897S	1
BCOR	chrX	39932907	rs6520618	0.69	0	synonymous	NM_001123383	c.A1692G	p.A564A	20
BCOR	chrX	39933339	rs5917933	0.97	0	synonymous	NM_001123383	c.T1260C	p.D420D	25
BCOR	chrX	39933378	.	.	0	synonymous	NM_001123383	c.C1221T	p.P407P	1
STAT5B	chr17	40362471	rs75130700	0.01	0	synonymous	NM_012448	c.C1725T	p.D575D	1
STAT5B	chr17	40371364	.	.	0	nonsynonymous	NM_012448	c.C799G	p.P267A	1
STAT5A	chr17	40459446	.	.	0	synonymous	NM_003152	c.C1707T	p.T569T	1
STAT5A	chr17	40459464	.	.	0	synonymous	NM_003152	c.C1725T	p.D575D	1
STAT5A	chr17	40459737	rs1135669	0.35	0	synonymous	NM_003152	c.C1902T	p.D634D	7
STAT3	chr17	40475056	rs117691970	0.03	0	synonymous	NM_003150	c.C1854T	p.G618G	2
STAT3	chr17	40489599	.	.	0	synonymous	NM_003150	c.C651T	p.I217I	1
MPL;MPL	chr1	43805240	rs16830693	0.02	0	synonymous	NM_005373	c.A690G	p.E230E	3
U2AF1	chr21	44513286	.	.	0	nonsynonymous	NM_001025203	c.G649A	p.G217S	1
U2AF1	chr21	44524456	.	.	1	nonsynonymous	NM_001025203	c.C101T	p.S34F	1
GATA1	chrX	48649690	rs139614533	0.01	1	synonymous	NM_002049	c.G174A	p.A58A	1
GATA1	chrX	48649728	.	.	0	nonsynonymous	NM_002049	c.A212C	p.H71P	1
IKZF1	chr7	50367292	rs61732861	.	0	synonymous	NM_001220765	c.C99T	p.I33I	1
IKZF1	chr7	50467767	rs61731355	0.05	1	synonymous	NM_001220772	c.C192A	p.P64P	3
IKZF1	chr7	50467941	rs61731356	0.15	0	synonymous	NM_001220772	c.C366T	p.N122N	3
PDGFRA	chr4	55130078	rs2229307	0.21	0	synonymous	NM_006206	c.T612C	p.N204N	6
PDGFRA	chr4	55133726	rs4358459	0.21	0	synonymous	NM_006206	c.T939G	p.G313G	6
PDGFRA	chr4	55139771	rs35597368	0.17	0	nonsynonymous	NM_006206	c.T1432C	p.S478P	5
PDGFRA	chr4	55141055	rs1873778	1	1	synonymous	NM_006206	c.A1701G	p.P567P	25
PDGFRA	chr4	55143577	rs10028020	0.18	0	synonymous	NM_006206	c.G1809A	p.A603A	5
PDGFRA	chr4	55152040	rs2228230	0.17	1	synonymous	NM_006206	c.C2472T	p.V824V	6
PDGFRA	chr4	55161391	rs7685117	1	0	synonymous	NM_006206	c.T3222C	p.D1074D	25
KIT	chr4	55561861	rs201872586	0.0017	0	nonsynonymous	NM_000222	c.C251T	p.T84M	1

KIT	chr4	55593464	rs3822214	0.05	1	nonsynonymous	NM_000222	c.A1621C	p.M541L	1
KIT	chr4	55593481	rs55986963	0.04	1	synonymous	NM_000222	c.A1638G	p.K546K	1
KIT	chr4	55602765	rs3733542	0.05	1	synonymous	NM_000222	c.G2586C	p.L862L	1
GNAS	chr20	57478807	rs7121	0.64	0	synonymous	NM_001077489	c.C348T	p.I116I	21
GNAS	chr20	57484421	rs121913495	.	1	nonsynonymous	NM_001077489	c.G557A	p.R186H	1
CBFB	chr16	67116169	rs78053453	0.02	0	synonymous	NM_001755	c.C453T	p.R151R	2
SOCS3	chr17	76354832	.	.	0	synonymous	NM_003955	c.C345T	p.P115P	1
ATRX	chrX	76937963	rs3088074	0.49	0	nonsynonymous	NM_138270	c.C2671G	p.Q891E	12
ATRX	chrX	76938208	rs45624939	0.03	0	nonsynonymous	NM_138270	c.T2426C	p.F809S	3
EED	chr11	85968623	rs974144	0.33	1	synonymous	NM_003797	c.C619T	p.L207L	16
IDH2	chr15	90628537	rs11540478	0.17	0	synonymous	NM_002168	c.C1050T	p.T350T	10
IDH2	chr15	90628549	rs190078206	0.01	0	synonymous	NM_002168	c.C1038T	p.A346A	1
IDH2	chr15	90630372	rs16943901	0.06	0	synonymous	NM_002168	c.A939G	p.G313G	2
SOCS2	chr12	93968598	.	.	0	synonymous	NM_003877	c.C240T	p.Y80Y	1
TET2	chr4	106155185	rs12498609	0.23	0	nonsynonymous	NM_001127208	c.C86G	p.P29R	9
TET2	chr4	106155751	rs6843141	0.07	0	nonsynonymous	NM_001127208	c.G652A	p.V218M	1
TET2	chr4	106158216	rs3796927	0.07	0	synonymous	NM_001127208	c.G3117A	p.S1039S	1
TET2	chr4	106190862	rs3733609	0.07	0	synonymous	NM_001127208	c.T4140C	p.H1380H	1
TET2	chr4	106196829	rs34402524	0.04	0	nonsynonymous	NM_001127208	c.T5162G	p.L1721W	3
TET2	chr4	106196951	rs2454206	0.19	0	nonsynonymous	NM_001127208	c.A5284G	p.I1762V	6
SH2B3	chr12	111856326	.	.	0	nonsynonymous	NM_005475	c.C377A	p.P126Q	1
SH2B3	chr12	111856368	.	.	0	nonsynonymous	NM_005475	c.G419A	p.R140H	1
SH2B3	chr12	111856673	rs78894077	0.08	0	nonsynonymous	NM_005475	c.C724T	p.P242S	7
SH2B3	chr12	111884608	rs3184504	0.99	0	nonsynonymous	NM_005475	c.T784C	p.W262R	25
SH2B3	chr12	111885984	rs140649197	0.01	0	nonsynonymous	NM_005475	c.G1606A	p.A536T	1
SH2B3	chr12	111886081	rs199803113	0.0017	0	nonsynonymous	NM_005475	c.T1703C	p.I568T	1
NRAS	chr1	115258744	rs121434596	.	1	nonsynonymous	NM_002524	c.G38A	p.G13D	1
MLL	chr11	118307385	rs9332747	.	0	nonsynonymous	NM_001197104	c.C158T	p.A53V	1
MLL	chr11	118307445	.	.	0	nonsynonymous	NM_001197104	c.G218A	p.G73E	1
MLL	chr11	118343848	rs9332773	0.04	0	synonymous	NM_001197104	c.C1974T	p.P658P	3
MLL	chr11	118355642	rs9332801	0.04	0	synonymous	NM_001197104	c.A4284C	p.I1428I	1
MLL	chr11	118365477	.	.	0	synonymous	NM_001197104	c.A5358C	p.S1786S	1
MLL	chr11	118368665	rs7107305	0.03	0	synonymous	NM_001197104	c.A5679G	p.L1893L	1
MLL	chr11	118373183	rs201483860	0.0017	0	synonymous	NM_001197104	c.C6576T	p.S2192S	1
MLL	chr11	118373677	rs9332838	0.01	0	nonsynonymous	NM_001197104	c.C7070G	p.P2357R	2
MLL	chr11	118373861	rs2071702	0.03	0	synonymous	NM_001197104	c.C7254T	p.N2418N	1
MLL	chr11	118375998	rs150804738	0.02	0	nonsynonymous	NM_001197104	c.G9391A	p.G3131S	2
MLL	chr11	118382707	rs184820566	0.0017	0	nonsynonymous	NM_001197104	c.T11113A	p.S3705T	1
MLL	chr11	118390741	rs2276058	0.02	0	synonymous	NM_001197104	c.T11391C	p.N3797N	1
CBL	chr11	119168130	rs143840974	.	0	synonymous	NM_005188	c.G2190C	p.T730T	1
CBL	chr11	119168175	.	.	0	synonymous	NM_005188	c.C2235T	p.T745T	1

GATA2	chr3	128204693	rs78245253	0.05	0	nonsynonymous	NM_001145662	c.C748G	p.P250A	3
GATA2	chr3	128204951	rs2335052	0.37	1	nonsynonymous	NM_001145662	c.G490A	p.A164T	15
GATA2	chr3	128205860	rs1573858	0.66	0	synonymous	NM_001145662	c.C15G	p.P5P	25
EZH2	chr7	148511087	.	.	0	synonymous	NM_001203249	c.C1647T	p.S549S	1
EZH2	chr7	148525904	rs2302427	0.15	0	nonsynonymous	NM_152998	c.G436C	p.D146H	6
CSF1R	chr5	149436927	.	.	0	nonsynonymous	NM_005211	c.C2242T	p.R748W	1
CSF1R	chr5	149450132	rs10079250	0.37	0	nonsynonymous	NM_005211	c.A1085G	p.H362R	15
CSF1R	chr5	149456893	rs3829986	0.03	1	nonsynonymous	NM_005211	c.G835A	p.V279M	4
CSF1R	chr5	149457678	rs2228422	0.19	1	synonymous	NM_005211	c.C726T	p.T242T	10
CSF1R	chr5	149460343	rs17652007	0.24	1	synonymous	NM_005211	c.C294T	p.H98H	10
CSF1R	chr5	149460459	rs55865465	0.03	0	synonymous	NM_005211	c.C178T	p.L60L	1
CSF1R	chr5	149460553	rs216123	0.1	0	synonymous	NM_005211	c.T84C	p.P28P	6
SPARC	chr5	151054227	rs2304052	0.1	0	synonymous	NM_003118	c.A66G	p.E22E	3
IDH1	chr2	209108284	rs62193615	.	0	nonsynonymous	NM_005896	c.A565G	p.I189V	1
IDH1	chr2	209108317	rs34218846	0.0035	1	nonsynonymous	NM_005896	c.G532A	p.V178I	1
IDH1	chr2	209113192	rs11554137	0.0035	1	synonymous	NM_005896	c.C315T	p.G105G	1

Abbreviations: 1kg: 1000 Genomes Project; c.: nucleotide; CHR: chromosome; p.: amino acid; POS: position;

*1 represents mutations present in COSMIC database; 0 represents mutations absent in COSMIC database.

Supplementary Table 6. Total single nucleotide variants and indels after filtration and confirmed by Sanger sequencing

Gene	CHR	POS	Exonic Function	c.	p.	Sample ID	Somatic or Germline	#Previously reported	*Cosmic	Annotation
JAK2	chr9	5073770	nonsynonymous	c.G1849T	p.V617F	E1	Somatic		1	Oncogenic
JAK2	chr9	5073770	nonsynonymous	c.G1849T	p.V617F	E2	Somatic		1	Oncogenic
JAK2	chr9	5073770	nonsynonymous	c.G1849T	p.V617F	E3	Somatic		1	Oncogenic
JAK2	chr9	5073770	nonsynonymous	c.G1849T	p.V617F	E5	Somatic		1	Oncogenic
JAK2	chr9	5073770	nonsynonymous	c.G1849T	p.V617F	E7	Somatic		1	Oncogenic
JAK2	chr9	5073770	nonsynonymous	c.G1849T	p.V617F	E8	Somatic		1	Oncogenic
JAK2	chr9	5064887	nonsynonymous	c.T1061C	p.I354T	E8	Somatic		0	Possible oncogenic
JAK2	chr9	5044432	nonsynonymous	c.G380A	p.G127D	E3	Somatic	Acute leukemia susceptibility[20]	0	Possible oncogenic
ASXL1	chr20	31023717	stopgain	c.C3202T	p.R1068X	E3	Somatic	Malignant myeloid diseases[21]	1	Oncogenic
ASXL1	chr20	31023801	nonsynonymous	c.A3286C	p.M1096L	E4	Somatic		0	Unknown
ASXL1	chr20	31022592	stopgain	c.C2077T	p.R693X	E1	Somatic	Malignant myeloid diseases[21]	1	Oncogenic
ASXL1	chr20	31024333	nonsynonymous	c.G3818A	p.R1273H	E24	Somatic	Endometrial carcinoma (Cosmic)	1	Oncogenic
SH2B3	chr12	111856326	nonsynonymous	c.C377A	p.P126Q	E11	Somatic		0	Possible oncogenic
U2AF1	chr21	44524456	nonsynonymous	c.C101T	p.S34F	E3	Somatic	Malignant myeloid diseases[22,23]	1	Oncogenic
FLT3	chr13	28622503	nonsynonymous	c.A1114G	p.R372G	E21	Somatic		0	Possible oncogenic
NRAS	chr1	115258744	nonsynonymous	c.G38A	p.G13D	E1	Somatic	Malignant myeloid diseases[24,25]	1	Oncogenic
MLL	chr11	118307445	nonsynonymous	c.G218A	p.G73E	E11	Somatic		0	Possible oncogenic
GNAS	chr20	57484421	nonsynonymous	c.G557A	p.R186H	E18	Somatic	Pancreatic carcinoma[26]	1	Oncogenic
RUNX1	chr21	36259183	nonsynonymous	c.C227T	p.P76L	E18	Somatic		0	Possible oncogenic
WT1	chr11	32456484	nonframeshift	c.408_409insGCC	p.P136delinsPP	E15	Somatic		0	Possible oncogenic
ZRSR2	chrX	15841230	nonframeshift	c.1314_1315insAGCCGG	p.G438delinsGSR	E13	Somatic		0	Unknown
STAT5B	chr17	40379618	nonframeshift	c.213_214insCTT	p.K71delinsKK	E19	Somatic		0	Possible oncogenic
‡CALR	Chr19	13054572	frameshift	c.1099_1150del52	p.L367fs*46	E4	Somatic			Oncogenic
SH2B3	chr12	111856368	nonsynonymous	c.G419A	p.R140H	E4	Germline		0	NA
U2AF1	chr21	44513286	nonsynonymous	c.G649A	p.G217S	E16	Germline		0	NA
FLT3	chr13	28623809	nonsynonymous	c.G845A	p.G282E	E14	Germline		0	NA
IDH1	chr2	209108284	nonsynonymous	c.A565G	p.I189V	E1	Germline		0	NA
JAK3	chr19	17953878	nonsynonymous	c.G524A	p.R175Q	E10	Germline		0	NA
GATA1	chrX	48649728	nonsynonymous	c.A212C	p.H71P	E12	Germline		0	NA

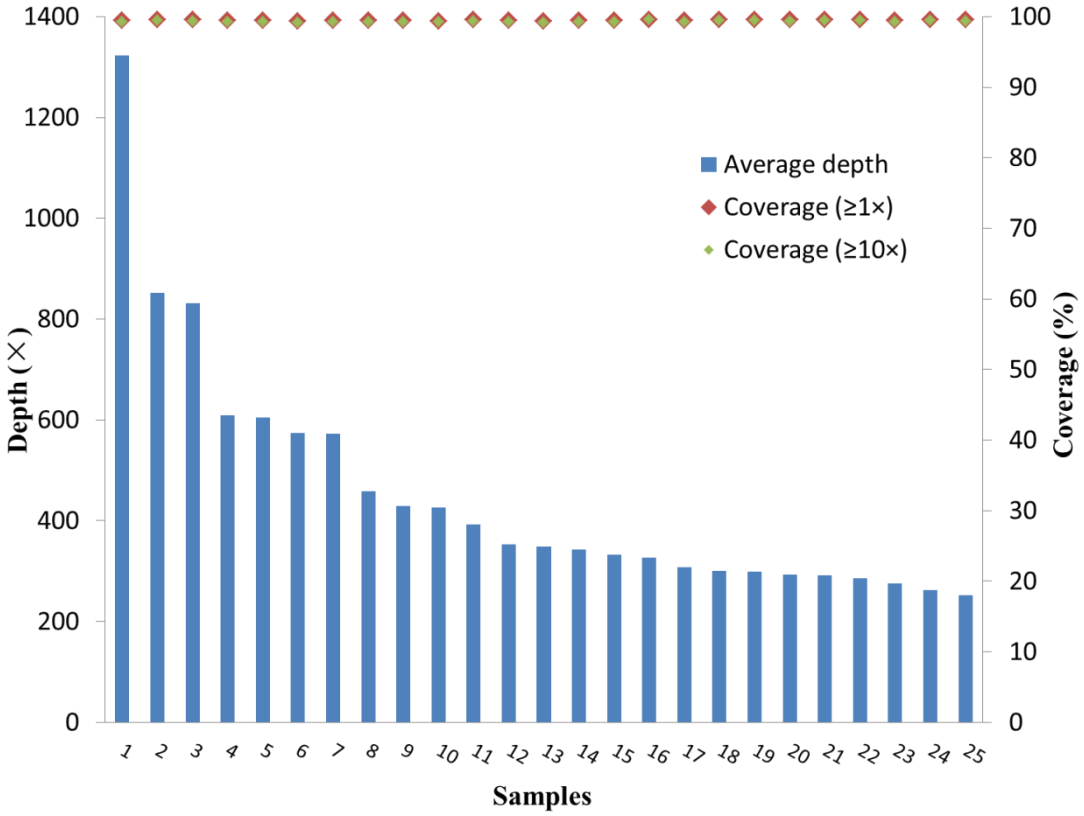
CSF1R	chr5	149436927	nonsynonymous	c.C2242T	p.R748W	E19	Germline	0	NA

Abbreviations: c: nucleotide; CHR: chromosome; NA: not applicable; p: amino acid; POS: position.

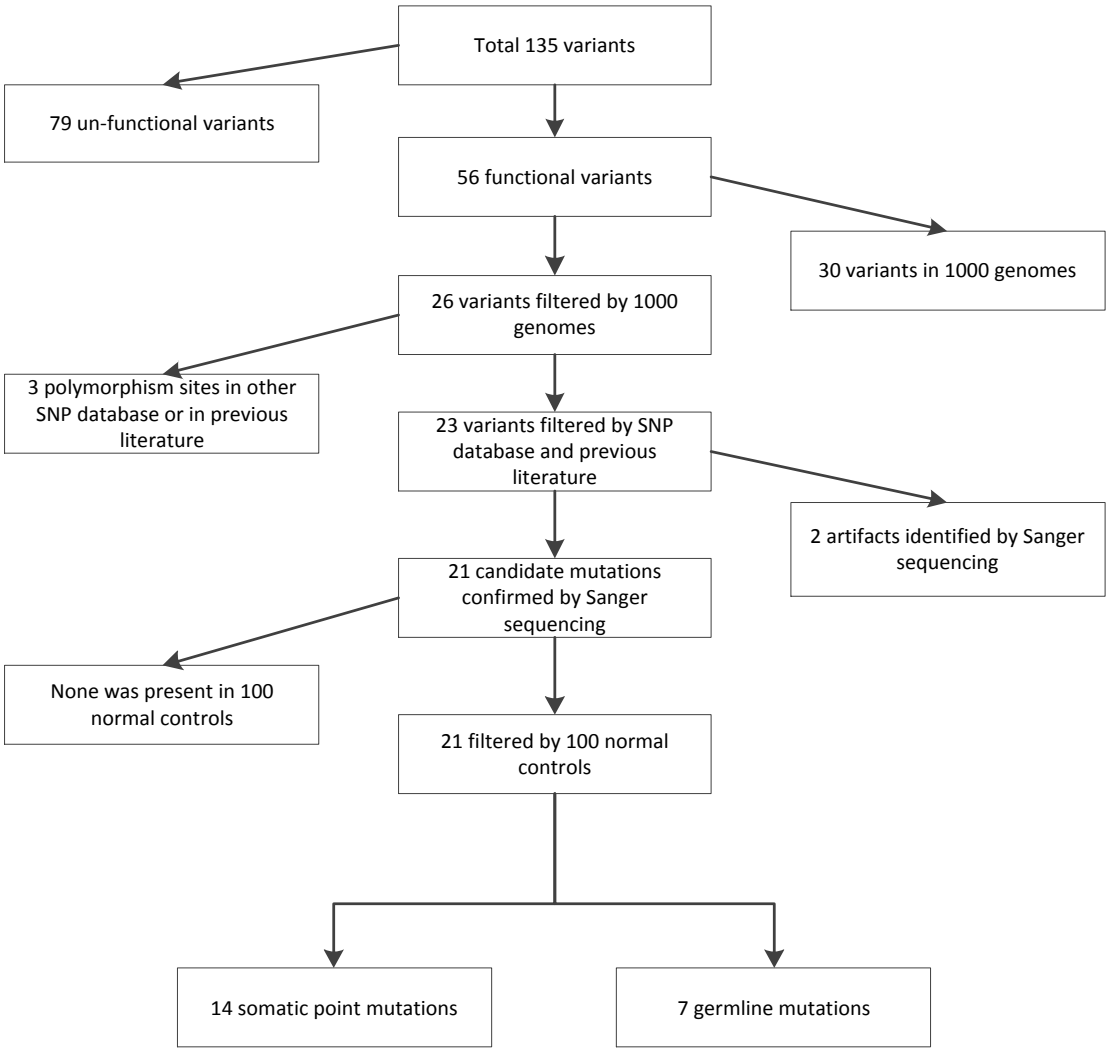
[#] Only mutations other than JAK2 V617F and CALR mutations are shown.

*1 represents mutations present in COSMIC database; 0 represents mutations absent in COSMIC database.

[‡]CALR mutation was detected by Sanger sequencing.



Supplementary Figure 1. Average depth and coverage of target region



Supplementary Figure 2. Single nucleotide variant filtration pipeline in targeted high-throughout sequencing analysis