Whole exome sequencing reveals activating *JAK1* and *STAT3* mutations in breast implant-associated anaplastic large cell lymphoma anaplastic large cell lymphoma

Piers Blombery,[1] Ella R. Thompson,[1,2] Kate Jones,[1] Gisela Mir Arnau,[1] Stephen Lade,[1,2] John F. Markham,[1] Jason Li,[1] Anand Deva,[3] Ricky W. Johnstone,[1] Amit Khot,[1] H. Miles Prince,[1] and David Westerman[1,2]

Peter MacCallum Cancer Centre, Melbourne; [2]University of Melbourne; and [3]Surgical Infection Research Group, Macquarie University, Sidney, Australia

Correspondence: piers.blombery@petermac.org

doi:10.3324/haematol.2016.146118

Supplementary data for *"Whole exome sequencing reveals activating JAK1 and STAT3 mutations in breast-implant associated anaplastic large cell lymphoma"*

Blombery P[1], Thompson E[1,2], Jones K[1], Mir Arnau G[1], Lade S[1,2], Markham JF[1], Li J[1], Deva A[3], Johnstone RW[1], Khot A[1], Prince HM[1], Westerman D[1,2].

[1] Peter MacCallum Cancer Centre, Melbourne, Australia

[2] University of Melbourne, Melbourne, Australia

[3] Surgical Infection Research Group, Macquarie University, Australia

Supplementary Methods

In-house bioinformatic pipeline was constructed using Seqliner v0.1a. Raw reads (fastq files) were quality checked using FastQC (0.11.2) and trimmed using cutadapt (1.7.1). Filtered reads were aligned to the Human reference genome (GRCh37/hg19) using BWA (0.7.12) and duplicate reads marked using Picard (1.119). Insertion/deletions (Indels) realignment and base quality recalibration were performed using GATK (3.2.2) prior to variant calling. Single nucleotide variants (SNVs) were identified using MuTect (2.7.1), Varscan (2.3.7) and GATK UG (3.2.2). GATK's Indel Genotyper (0.36) and Varscan (2.3.7) were used to call somatic indels. Copy number was estimated by taking the ratio at each interval with the corresponding median of four normal samples from the same batch and averaged into 1M-base regions for plotting. Variants in JAK1, JAK3 and STAT3 were confirmed by Sanger sequencing.

Supplementary Table 1 - Somatic variants detected in case 1 after filtering criteria (read depth >20x, excluding known polymorphic genes, coding variant/canonical splice site, present in tumour and not in germline, exclusion of artefacts based on manual inspection of aligned data)

Gene Name	Transcript	Nucleotide Change	Protein Change	Variant Allele Fraction (%)
AATK	NM_001080395.2	c.694C>T	p.R232C	42
AGBL2	NM_024783.3	c.1883C>T	p.T628I	46
AGK	NM_018238.3	c.677A>G	p.Y226C	57
ANKAR	NM_144708.3	c.4057G>A	p.G1353R	19
ASAH2	NM_019893.2	c.337G>T	p.G113*	45
ASS1	NM_000050.4	c.1069C>A	p.Q357K	36
FKBPL	NM_022110.3	c.922G>A	p.D308N	17
CAND1	NM_018448.4	c.548A>C	p.Q183P	16
CCDC40	NM_001243342.1	c.910G>A	p.E304K	46
CTTNBP2	NM_033427.2	c.530A>G	p.E177G	45
DCDC5	NM_020869.3	c.2433C>G	p.H811Q	56
TSSK2	NM_053006.4	c.418G>A	p.E140K	48
DIAPH3	NM_001258370.1	c.2047A>G	p.S683G	46
EPB41L3	NM_012307.3	c.2012C>G	p.A671G	41
EPHA6	NM_001080448.2	c.618C>A	p.N206K	37
ESCO2	NM_001017420.2	c.1418C>T	p.P473L	30
FRAS1	NM_001166133.1	c.1278G>C	p.L426F	33
GRAMD3	NM_023927.2	c.1141A>T	p.I381F	43
GSTA4	NM_001512.3	c.662G>C	p.R221T	31
ITGA1	NM_181501.1	c.1525C>G	p.L509V	31
KCNAB1	NM_172160.2	c.130G>C	p.D44H	26
KDR	NM_002253.2	c.400G>A	p.G134R	43
LMTK3	NM_001080434.1	c.4459G>A	p.V1487M	26
MAP1B	NM_005909.3	c.4612G>A	p.V1538I	18
MDGA2	NM_182830.4	c.1009G>T	p.A337S	31
MED12	NM_005120.2	c.4394G>C	p.R1465P	15
MRO	NM_031939.3	c.194G>A	p.R65H	60
MYOD1	NM_002478.4	c.133G>C	p.D45H	42
NCKAP1	NM_205842.2	c.436T>C	p.Y146H	14
OXR1	NM_181354.4	c.1189G>T	p.E397*	15
PDZRN4	NM_001164595.1	c.802G>T	p.D268Y	23
PKP3	NM_007183.3	c.2268C>G	p.D756E	13
RALGDS	NM_006266.3	c.764A>T	p.E255V	46
RBM41	NM_018301.3	c.392G>A	p.R131H	32
RPF1	NM_025065.6	c.364G>C	p.E122Q	15
RREB1	NM_001003698.3	c.2914T>C	p.C972R	45
SETD1A	NM_014712.2	c.1640A>C	p.N547T	28
SGPL1	NM_003901.3	c.433G>A	p.A145T	57
SLC10A3	NM_019848.4	c.80G>C	p.S27T	63
SPATA5	NM_145207.2	c.1058_1067delinsTCCATGAT	p.Asp353Valfs*7	39
STAT3	NM_139276.2	c.1840A>C	p.S614R	59*
TMEM176A	NM_018487.2	c.116C>T	p.A39V	29
TPSG1	NM_012467.3	c.86C>T	p.P29L	33
TRIP12	NM_004238.2	c.1588-1G>T	-	29
TTBK2	NM_173500.3	c.3685A>C	p.T1229P	37
VCAN	NM_004385.4	c.7796C>A	p.T2599K	53
ZBTB11	NM_014415.3	c.3035C>G	p.S1012C	27
ZFAT	NM_020863.3	c.16G>A	p.A6T	50
ZFR2	NM_015174.1	c.284A>T	p.Q95L	14
ZNF177	NM_003451.2	c.2T>G	p.M1R	33
ZNF385A	NM_015481.2	c.451C>G	p.Q151E	31

***** 22 variant reads out of 37 total reads (59.5%). Quantification by conventional Sanger sequencing indicated variant allele fraction of <50% (approximately 40-50%).

Supplementary Table 2 - Somatic variants detected in case 2 after filtering criteria (read depth >20x, excluding known polymorphic genes, coding variant/canonical splice site, present in tumour and not in germline, exclusion of artefacts based on manual inspection of aligned data)

Gene Name	Transcript	Nucleotide Change	Protein Change	Variant Allele Fraction (%)
CFAP47	NM_001304548.1	c.4565C>T	p.S1522F	9
COL4A3	NM_000091.4	c.3779G>T	p.G1260V	26
CPSF1	NM_013291.2	c.1519G>A	p.G507R	14
CTSK	NM_000396.3	c.955G>A	p.G319S	20
CYLD	NM_015247.2	c.2273G>A	p.R758Q	14
DOCK5	NM_024940.6	c.1412C>T	p.S471F	13
GTPBP3	NM_032620.3	c.667G>A	p.D223N	21
JAK1	NM_002227.2	c.3290_3291delinsTT	p.G1097V	20
KIAA1109	NM_015312.3	c.3472G>T	p.A1158S	10
KLHL34	NM_153270.2	c.913G>A	p.A305T	10
KLHL38	NM_001081675.2	c.1332C>G	p.N444K	14
MYT1	NM_004535.2	c.344C>T	p.A115V	13
OPLAH	NM_017570.4	c.3277G>A	p.A1093T	16
OR8K3	NM_001005202.1	c.412C>T	p.R138*	13
PLXNA4	NM_181775.3	c.649G>A	p.V217I	16
RSPH4A	NM_001010892.2	c.833T>A	p.L278H	12
SLITRK3	NM_014926.2	c.1973delT	p.L658fs*135	13
SMC2	NM_006444.2	c.2575G>A	p.A859T	12
SOX6	NM_033326.3	c.1670C>T	p.S557L	14
SPEG	NM_005876.4	c.4846C>T	p.R1616W	10
SVEP1	NM_153366.3	c.8677A>C	p.T2893P	39
SYN3	NM_003490.3	c.145G>T	p.A49S	15
TRHDE	NM_013381.2	c.3044T>C	p.F1015S	12
ZHX3	NM_015035.3	c.616G>A	p.V206I	10