Clonal independence of *JAK2* and *CALR* or *MPL* mutations in comutated myeloproliferative neoplasms demonstrated by single cell DNA sequencing

Ella R. Thompson,[1,2] Tamia Nguyen,[1*] Yamuna Kankanige,[1,2] Paul Yeh,[1,2,3,4] Michael Ingbritsen,[1] Michelle McBean,[1] Timothy Semple,[3] Gisela Mir Arnau,[2,3] Kate Burbury,[4] Nora Lee,[4] Amit Khot,[4] David Westerman[1,2,4] and Piers Blombery[1,2,4]*

[1]Pathology Department, Peter MacCallum Cancer Center; [2]Sir Peter MacCallum Department of Oncology, University of Melbourne; [3]Research Division, Peter MacCallum Cancer Center and [4]Clinical Haematology, Peter MacCallum Cancer Centre/Royal Melbourne Hospital, Melbourne, Victoria, Australia

**ERT and TN contributed equally as co-first authors.*

Correspondence: ELLA R. THOMPSON - ella.thompson@petermac.org.

doi:10.3324/haematol.2020.260448

Supplementary methods

Patient samples. Peripheral blood and bone marrow aspirate samples were referred to the Molecular Haematology Laboratory at Peter MacCallum Cancer Centre for molecular profiling. Two patients harboring mutations in *JAK2* and *CALR* genes (first case) and *JAK2* and *MPL* genes (second case) were identified and their peripheral blood collected for single cell analysis. Mononuclear cells from the first patient were isolated using Ficoll-Paque media following the manufacturer's instruction and peripheral blood from the second patient was lysed with a non-fixative ammonium chloride based lysing solution to collect all nucleated cells. Harvested cells were stored in liquid nitrogen until use. The study was conducted under local institutional ethics approval (03/90).

Single cell library preparation. Single cell library preparation using the Tapestri platform according to the manufacturer's instructions (Mission Bio). Briefly, cryopreserved cells were thawed and quantified using a Countess Automated Cell Counter (Invitrogen). Based on the quantification, cells were normalised to a concentration of 4,500 cells/µL using a density gradient medium which prevented cell sedimentation. Next, they were loaded into a microfluidics cartridge where, following a pre-programmed protocol on a Tapestri instrument, cells were encapsulated in oil droplets, together with lysis buffer containing proteases. Droplets were incubated for 60 min at 50^{o}C to allow for cell lysis and cellular protein digestion. Protease enzymes were then heat deactivated. Droplets containing cell lysate were barcoded with a unique barcode and DNA was amplified with 70 custom primer pairs targeting 17 haematological genes of interest including *JAK2*, *CALR*, *MPL* and *TP53* (but not *SRSF2*). The PCR conditions were as follows: 6 min at 98^{o}C (4^{o}C/s ramp rate); 10 cycles of 30 s at 95^{o}C, 10 s at 72^{o}C, 3 min at 61^{o}C and 20 s at 72^{o}C (1^{o}C/s ramp rate); 10 cycles of 30 s at 95^{o}C, 10 s at 72^{o}C, 3 min at 48^{o}C and 20 s at 72^{o}C (1^{o}C/s ramp rate); and a final extension at 72^{o}C for 2 min then hold at 4^{o}C (4^{o}C/s ramp rate).

Single cell library sequencing. Following amplification, droplets were demulsified to harvest the library which was cleaned up enzymatically to remove excess primers and primer dimers. The cleaned up products were purified using 0.72x SPRI beads (Beckman Coulter) before being dual-indexed with i5/i7 index adapters (Illumina). The indexed library was purified with 0.69x SPRI beads (Beckman Coulter) and run on the 4200 TapeStation system (Agilent) to assess quantity and quality. Sequencing was performed on a NextSeq instrument using 300-cycle Mid Output Kit (Illumina) and 25% PhiX to generate 150 base pair paired-end reads.

Single cell data variant analysis. ***Pipeline*** **–** FASTQ files were analysed using the cloud-based Tapestri bioinformatics pipeline to perform adapter trimming, barcode correction, cell identification, read alignment to human hg19 genome, and variant calling using a GATK-based algorithm. To ensure complete cells, only barcodes with (1) at least 10 x the number of amplicons (70) and (2) at least 0.2X mean amplicon coverage across all amplicons were included in the variant analysis. The genotypes and the cell matrix were consolidated into a loom file for subsequent clonal analysis using Tapestri Insights v2.2 software (Mission Bio). ***Data filtering*** – Data from the loom file were subjected to a series of filters to ensure only cells containing high quality genotype calls were included in the clone analysis. Specifically, for every cell, a genotype was retained only if it had a quality score of 30 or more, a read depth of 10 reads or above and for non-reference genotype calls only, an alternate allele frequency of no less than 20%. Then, any variant genotyped in less than 2% of all cells were removed from the dataset and cells with retained genotypes for less than 50% of the called variants were also removed. Lastly, variants present in less than 1% of the cells were excluded. ***Clone analysis*** – Cells were classified into clonal populations according to their known mutations. Cells missing genotypes for these mutations were excluded from the analysis. Further, minor clones accounting for less than 1% of the cell population were also

excluded from analysis. Finally, clones were visually assessed for evidence of cell doublets (observed as increased read depth per variant) or allele dropout (observed as decreased read depth and decreased quality per variant) and these were also excluded. Figures were generated in R studio using filtered data exported from Tapestri Insights v2.2.

Single cell data amplicon-based copy number analysis. Copy number analysis was performed in R studio using a modified version of the tapestri cnv R package and the loom file generated by the Tapestri pipeline. Filtered variants from the variant analysis were utilised to identify the normal diploid clone which was considered to be the reference clone to calculate copy numbers of individual cells per amplicon. Mean copy number of all the amplicons within a gene was considered to be the estimated copy number of that gene. Supplementary Figure 1 shows the copy number estimates of individual *MCL1* amplicons and the gene for Patient 1.

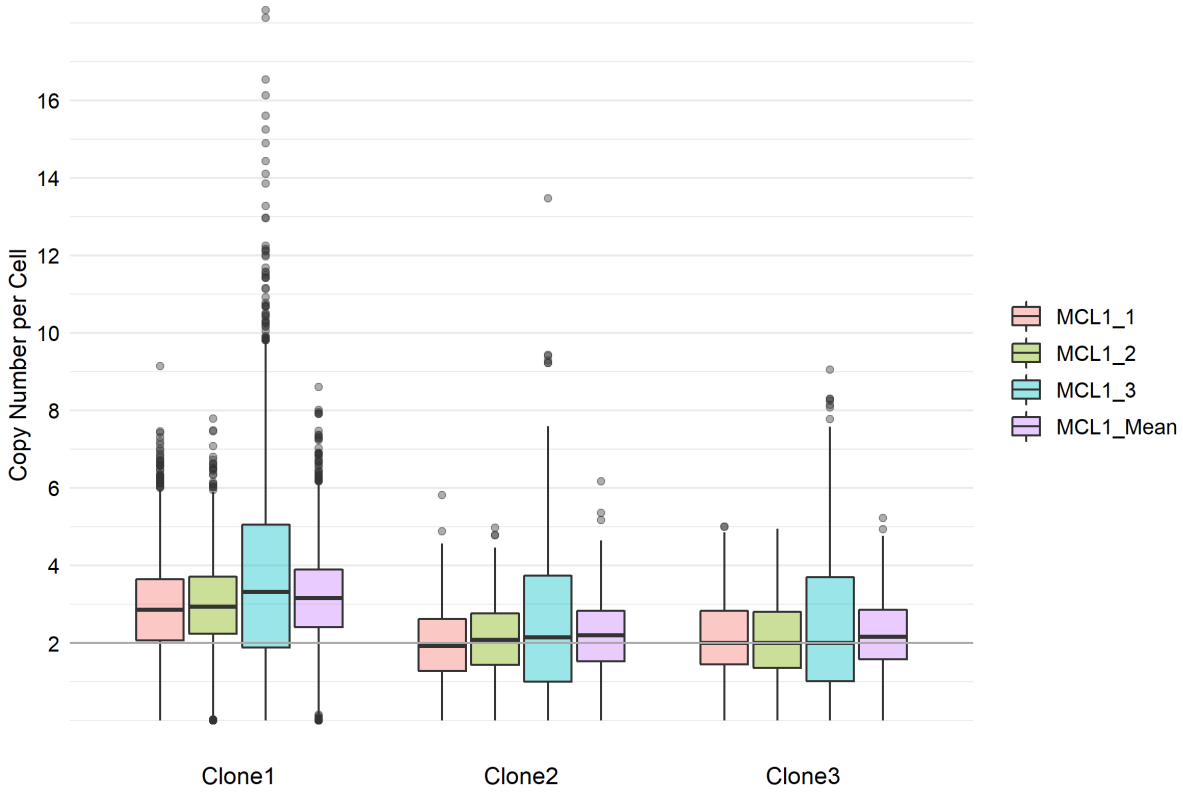
Supplementary Table 1. Single cell analysis of patients 1 and 2

	Patient 1	**Patient 2**
Sequenced cells	10,564	8,181
Sequenced cells post-filtering	9,485	7,187
Variants post-filtering	15	19
Cells genotyped, n (% of post-filtered cells)		
JAK2 Val617Phe	9,067 (96%)	6,559 (91%)
CALR Lys385Asnfs*47	4,965 (52%)	-
MPL Trp515Ser	-	3,149 (44%)
TP53 His178Pro	8,914 (94%)	-
All variants	4,547 (48%)	2,894 (40%)
Cells missing one or more genotypes	*4,938 (52%)*	*4,293 (60%)*
Cells mutated, n (% of genotyped cells)		
JAK2 Val617Phe	5,837 (64%)	1,378 (21%)
CALR Lys385Asnfs*47	1,314 (26%)	-
MPL Trp515Ser	-	713 (23%)
TP53 His178Pro	2,593 (29%)	-
Aggregate single cell VAF (by read count)		
JAK2 Val617Phe	62%	11%
CALR Lys385Asnfs*47	16%	-
MPL Trp515Ser	-	13%
TP53 His178Pro	16%	-
Clones identified	19	7
Clones excluded with <1% total cells	*13*	*2*
Clones excluded as cell doublets	*2*	*0*
Clones excluded as allele dropout	*1*	*2*
Major clones identified	3	3

Supplementary Table 2. Clone analysis of patients 1 and 2

Major clone	Identity	Cells	Clone size (% of major clones)	Cell fraction (% of post-filtered cells)
Patient 1				
Clone 1	*JAK2*^{hom*};*CALR*^{wt};*TP53*^{wt}	2,621	72.48%	27.63%
Clone 2	*JAK2*^{wt};*CALR*^{het};*TP53*^{het}	508	14.05%	5.36%
Clone 3	*JAK2*^{wt};*CALR*^{wt};*TP53*^{wt}	487	13.47%	5.13%
Patient 2$^{\#}$				
Clone 1	*JAK2*^{wt};*MPL*^{wt}	1,591	62.17%	22.14%
Clone 2	*JAK2*^{het};*MPL*^{wt}	539	21.06%	7.50%
Clone 3	*JAK2*^{wt};*MPL*^{het}	429	16.76%	5.97%

Hom – homozygous, het – heterozygous, wt - wildtype
*Trisomy 1q21.3 also observed in clone 1, absent in clones 2 and 3.
#*SRSF2*^{het} mutation presumed present in all clones



Supplementary Figure 1.