Zhang, et al.

Supplemental materials

Longitudinal study of two patients with cyclic thrombocytopenia, *STAT3* **and** *MPL* **mutations**

Supplemental Methods

Sample collection and clinical assays

Peripheral blood samples were drawn in EDTA tubes, and complete blood counts were monitored. Peripheral blood mononuclear cells (PBMCs) were obtained by Ficoll-Paque density gradient centrifugation.

T- and B-cell receptor clonality assays provide next-generation sequencing (NGS)-based determination of clonal rearrangements and the frequency distribution of T cell receptor beta locus (*TRB*), immunoglobulin heavy locus (*IGH*, from the conserved framework regions 1, 2, and 3 [FR1, FR2, FR3]), and immunoglobulin kappa locus (*IGK*) (Invivoscibe). The results were interpreted as previously described.¹

Stanford Actionable Mutation Panel for Hematopoietic and Lymphoid Malignancies (Heme-STAMP) panel is a CLIA-validated, Stanford-developed NGS assay targeting germline and somatic variants in 164 genes associated with hematolymphoid neoplasms.

Plasma TPO assay

Platelet poor plasma was collected for each blood sample point, and aliquots were stored at - 80°C freezer before the assay. Plasma TPO levels were measured using Quantikine ELISA for Human Thrombopoietin Immunoassay (R&D Systems).

Cell lines and bioassays for MPL functions

Ba/F3 cells (DSMZ), a murine interleukin (IL)-3 dependent pro-B cell line, was used as an *in vitro* model system for assessing biofunctions and downstream signaling of MPL mutation. Wild type (WT) human *MPL* cDNA clone (OriGene) was inserted into pCMV-Neo vector (OriGene). *MPL* c.1210G>A mutant was constructed using QuickChange lightning site-directed mutagenesis kit (Agilent) and confirmed by Sanger sequencing. These constructs were transfected into Ba/F3 cells, and G418-resistant stable clones were obtained for functional analysis. For growth assay, IL-3 (10 ng/mL) or TPO (10 ng/mL, or as indicated) was added to the media. Cell numbers were counted by hemocytometer, and metabolic levels were monitored by CellTiter-Glo Luminescent cell viability Assay (Promega). For TPO-uptake assay, 5x10⁶ cells were incubated for 0, 30, or 60 min after the addition of equal amounts of TPO. After incubation, TPO concentrations in cell culture supernatant were measured by Quantikine ELISA assay. For the MPL internalization assay, cells were washed and incubated with TPO for the indicated

time. Cell surface MPL was measured by flow cytometry analysis. Protein expression of MPL and phosphorylation of STAT3, STAT5, and ERK1/2 were measured using Western blot.

Blood transcriptome and data analysis

3SEQ (3'-end sequencing for expression quantification) analyses were performed on blood RNA samples. 3SEQ is a type of RNA-seq that focuses on quantitative analysis of transcriptome by generating a directional sequencing library targeting 3'UTRs and flanking regions upstream of poly-A tail, ensuring that one read is produced and measured per transcript. 3SEQ libraries were constructed based on the previously published method^{2,3} with modifications. Briefly, mRNAs were enriched by poly-A selection using Dynabeads mRNA purification kit (Thermo Fisher Scientific, Waltham, MA) and heat-fragmented to 100-200 nucleotides. First-strand cDNAs were synthesized using Superscript III reverse transcriptase with Rd2SP-oligodT primer, followed by second-strand cDNA synthesis. Adenine was added to the 3'-end of the doublestranded cDNA and then ligated to the P5-R1SP adapter. The ligated product was amplified by PCR for 15 cycles using primers P5-Rd1SP and P7-index-Rd2SP. Six Illumina indexes, each with six bases, were introduced in the P7 primers. Sequence of the primers used in 3SEQ library construction is provided in Supplemental Table 3. Qualities of the libraries were examined using Agilent DNA 1000 kit on Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Libraries were quantified by Qubit 2.0 fluorometer using Qubit dsDNA BR assay kit (Thermo Fisher Scientific, Waltham, MA). Five or six samples with different indexes were pooled together for next generation sequencing on HiSeq 2000 system for patient 1 (CT1) and on NextSeq (Illumina Inc., San Diego, CA) for patient 2 (CT2). Index associated 36 bp sequences from the P5 primer end, providing the 5'-end sequences of the polyA-containing mRNA fragments, were generated.

3SEQ data were filtered and mapped to human transcriptome hg19, and read counts for each gene were generated as previously described⁴. For each patient, Significance Analysis of Microarrays-Seq (SAMseq) algorithm was used to obtain genes that are quantitatively associated with platelet count, both aligned to the platelet count measurements and precede or succeed of the platelet count measurements for 1, 2, or 3 time points (which are 3-4 days, 7 days, or 10-11 days respectively). This generated total of 7 quantitative SAMseq analyses for each patient. Significant genes (q<0.05) were compiled and duplicate genes were removed to obtain a list of genes that correlated with platelet count measurements, or correlated with platelet count patterns yet precede/succeed of platelet count measurements. Gene list was further confined to genes that displayed substantial fluctuations within each platelet count cycle

3

with coefficient of variation (CV) > 0.2 for each platelet count cycle. CV was calculated based on the gene expression values, normalized in transcripts per million (TPM), for each platelet count cycle. Specifically, for CT1, platelet count cycle 1 was designated as from CT1-1 to CT1-12, and cycle 2 from CT1-13 to CT1-24; for CT2, cycle 1 was from CT2-1 to CT2-7, and cycle 2 from CT2-8 to CT2-15. Next, genes that have a mean TPM > 0.25 across all sample points of each patient were selected. Collectively, this blood transcriptome analysis yielded a group of genes that correlated with platelet count patterns, showed substantial fluctuations over each platelet count cycle, and had reliable overall expression levels. This yielded 667 and 590 genes that were quantitatively correlated with platelet count in CT1 and CT2, respectively. Of those, 236 genes were shared by both patients.

The 236 genes were parsed into several groups (supplemental Table 1) based on the unsupervised gene clustering of their longitudinal expression profiles over all sample points in each patient. Of the 151 genes that showed platelet-specific pattern in both patients, 34 exclusive platelet-specific genes were obtained using following criteria: have over 50-fold inductions in CT1 and 10-fold inductions in CT2 over the sapling period; have a maximum expression level > 5 TPM in both patients.

Cluster 3.0 (https://www.encodeproject.org/software/cluster/) was used for hierarchical clustering: expression data (in TPM) of selected genes in all sample points of each patient were adjusted by "log transform data", and "center genes-median", then clustered using "complete linkage" method. The clustered heatmaps were visualized using Java TreeView (Version 1.1.6r4, [https://sourceforge.net/projects/jtreeview/files/\)](https://sourceforge.net/projects/jtreeview/files/) with yellow and blue color indicates high and low expression respectively. And gray means not detected.

Supplemental Figure 1. Neutrophil count fluctuations in CT1. (A) Neutrophil count (orange) and platelet count (blue) measured in CT1. Neutrophil count showed peaks prior to platelet count peaks in both cycles. (B) Periodogram of platelet count (upper) and neutrophil count (lower). Platelet count showed a significant oscillation with a period of 39 days, and neutrophil count showed a cyclical pattern with the same period. (C) Neutrophil count (orange) and the immature neutrophil gene expression (light blue) pattern, which showed that the immature neutrophil gene expression is preceding the neutrophil count in CT1.

Supplemental Figure 2. Characterization of the novel c.1210G>A *MPL* **mutation identified in CT1.** (A) The MPL mutation resulted in a pG404R amino acid substitution. (B) The alignment of MPL amino acid sequence around the G404 position in various species, the query sequence is WT human MPL. The alignment shows that the glycine residue G404 (boxed and red arrow pointed) is conserved across species. (C) The mutant MPL required about 15 times higher TPO concentration to initiate cell growth. Equal numbers of Ba/F3 cells expressing WT or c.1210G>A *MPL* were seeded in wells containing TPO at different concentrations ranging from 19.5 to 10,000 pg/mL, or in wells containing IL-3 (10 ng/mL) as a positive control. The cell growth after 3 days of incubation was measured by cellular metabolic activity, and the TPO-stimulated growth was plotted as a relative level to the IL-3 stimulated growth. The minimum TPO concentration that is required to initiate cell growth through mutant MPL was 1250 pg/mL (red arrow), which is 15 times higher than that for WT MPL (78 pg/mL, blue arrow). The double-headed green arrow span the range of plasma TPO levels measured in the patient. (D) The MPL-mutant Ba/F3 cells had impaired internalization of membrane MPL. Normally, the extracellular TPO binds MPL on the plasma membrane, and the complex is subsequently internalized, which removes TPO from extracellular environment and reduces the number of cell surface MPL. After incubation with TPO for 15 min, the WT cells had about 40% membrane MPL remaining compared with the baseline level as measured by flow cytometry analysis for the median fluorescence intensity of membrane MPL. MPL mutant cells did not show MPL internalization until after at least 45 min of TPO incubation. The delays of internalization of mutated MPL indicates impairment in TPO uptake.

Supplemental Figure 3. Large granular lymphocyte cells observed in patients. These

images show the large granular lymphocyte cells observed in CT1 in peripheral blood**.** Wright's-Giemsa staining was performed on freshly prepared peripheral blood smears. Photos were taken using Olympus DP22 camera under 100x oil objective, and were white balanced and cropped in Adobe Photoshop.

Supplemental Figure 4. Clonality of B cells in CT1 and CT2. (A) Histograms show the frequencies of the top 10 clones of B-cell receptor genes in a sample at platelet count descending phase. *IGH*-FR3: immunoglobulin heavy locus from the conserved framework region FR3, *IGK*: immunoglobulin kappa locus, CT1 had clonal *IGH* and *IGK*. CT2 had polyclonal *IGH* and *IGK*. (B) The top clones of the B-cell receptors in 6 sequential samples in CT1 were identical. Their clonal frequencies were plotted and linked to show their stable longitudinal patterns over a platelet count cycle.

Supplemental Table 1. The 236 genes that were quantitatively correlated with platelet count in both patients and their expression patterns.

TPM, transcripts per million; Plt, platelet; Neut, neutrophil; Ery, erythrocyte.

Supplemental Table 2. Gene Ontology biological process enrichment test of the 236 genes that were quantitatively correlated with platelet count in both patients.

Analysis summary:

Analysis Type: PANTHER Overrepresentation Test (Released 20200728)

Annotation Version and Release Date: GO Ontology database DOI: 10.5281/zenodo.4437524 Released 2021-01-01

Analyzed List: CT1_CT2_shared_plt_236.csv (Homo sapiens)

Reference List: Homo sapiens (all genes in database)

Test Type: Fisher's Exact

Correction: Calculate False Discovery Rate

Biological processes that have a false discovery rate < 0.05 and fold enrichment >5 were listed in the table.

Supplemental Table 3. Primers used for 3SEQ library construction.

a Corresponding index sequence in the primer was underlined.

Supplemental References

- 1. Ho CC, Tung JK, Zehnder JL, Zhang BM. Validation of a Next-Generation Sequencing-Based T-Cell Receptor Gamma Gene Rearrangement Diagnostic Assay: Transitioning from Capillary Electrophoresis to Next-Generation Sequencing. *J Mol Diagn*. 2021;23(7):805-815.
- 2. Beck AH, Weng Z, Witten DM, et al. 3'-end sequencing for expression quantification (3SEQ) from archival tumor samples. *PLoS One*. 2010;5(1):e8768.
- 3. Zhang H, Zhang BM, Guo X, et al. Blood transcriptome and clonal T-cell correlates of response and non-response to eltrombopag therapy in a cohort of patients with chronic immune thrombocytopenia. *Haematologica*. 2020;105(3):e129-132.
- 4. Guo X, Zhu SX, Brunner AL, van de Rijn M, West RB. Next generation sequencingbased expression profiling identifies signatures from benign stromal proliferations that define stromal components of breast cancer. *Breast Cancer Res*. 2013;15(6):R117.