5-formylcytosine and 5-hydroxymethyluracil as surrogate markers of TET2 and SF3B1 mutations in myelodysplastic syndrome, respectively

TET2 is one of the most commonly mutated genes in adult myeloid malignancies, occurring in approximately 25% and approximately 50% of cases of myelodysplastic syndromes (MDS) and chronic myelomonocytic leukemia (CMML),[1-3] respectively. TET2 protein is a dioxygenase that converts 5-methylcytosine to 5-hydroxymethylcytosine, which it further oxidizes to 5-formylcytosine, and further still to 5-carboxycytosine.[4] TET2 has also been implicated in 5-hydroxymethyluracil generation.[5] These base modifications can be used to quantify the extent to which *TET2* mutations cause deficiencies in TET2 dioxygenase activity. We thus set out to characterize the linkage between various *TET2* mutations and 5-hydroxymethylcytosine, 5-formylcytosine, 5-carboxycytosine, and 5-hydroxymethyluracil, and linkage between these modifications and myeloid neoplastic disease. Measuring these modifications could help us form equivalence classes of disparate molecular abnormalities present in MDS and related disorders.

Accurate measurement of 5-formylcytosine, 5-carboxycytosine and 5-hydroxymethyluracil is challenging because levels of 5-formylcytosine and 5-carboxycytosine in mammalian genomes are 3-4 orders of magnitude lower than 5-hydroxymethylcytosine, which thus overshadows their detection. To address this challenge, we recently developed a rapid, highly-sensitive and specific isotope-dilution automated two-dimensional ultra-performance liquid chromatography tandem mass spectrometry method that is specifically tailored to analyze global levels of 5-methylcytosine, 5-hydroxymethylcytosine, 5-formylcytosine, 5-carboxycytosine, and 5-hydroxymethyluracil simultaneously.[6] Importantly, many previous studies used less-reliable semi-quantitative immunohistochemical methods to assess 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine; immunohistochemical methods are less reliable due to low levels of the modification in genomic DNA[7] and poor accessibility of DNA epigenetic modifications in cell nuclei under non-denaturing conditions. Our technique overcomes these challenges. It is highly sensitive and is often regarded as a gold standard.[8,9]

We hypothesized that there could be associations between the spectra of DNA epigenetic modifications 5-methylcytosine, 5-hydroxymethylcytosine, 5-formylcytosine, 5-carboxycytosine, and 5-hydroxymethyluracil and MDS- or CMML-associated mutations, and that characterizing them may further our understanding of these diseases. We analyzed a representative cohort of 17 patients with myeloid neoplasia: nine MDS, seven acute myeloid leukemia (AML), one CMML, and one aplastic anemia (AA). Results were compared with those obtained in five healthy controls. Samples were taken from patients with a broad range of different types of TET2 alterations and we detected a broad spectrum of DNA modifications.

Median levels of the most abundant substrate of TET2, 5-methylcytosine, were similar across mutation types (Figure 1A). Median levels of the most abundant product of TET2, 5-hydroxymethylcytosine, produced by TET2 acting on 5-methylcytosine, were approximately 36% lower in patients with *TET2* mutations *versus* non-mutated subjects (Figure 1B). Moreover, levels depended strongly on mutation position; e.g. mutations at D1384, which is involved in Fe(II) co-ordination in the TET2 catalytic domain, decreased 5-hydroxymethylcytosine levels almost 3-fold and simultaneously abolished all downstream modifications (*see below*) (Figure 1C-E). In contrast, missense mutation at R73, which is far from the TET2 catalytic domain, did not influence the measured base modifications (Figure 1B).

To our surprise, the best predictor/marker of TET2 mutations in patients with myeloid malignancies was 5-formylcytosine. Median levels of 5-formylcytosine were lower in patients with *TET2* mutations *versus* non-mutated subjects. In some samples, this modification was undetectable (with limit of detection 0.03 per 10^6 deoxynucleosides, dN). Median levels of the higher oxidation product 5-carboxycytosine were also lower in patients with TET2 mutations, albeit with marginal statistical significance (Figure 1D).

Levels of thymine in DNA are substantially higher than those of any other substrate, so although TET2 specific activity for thymine in DNA may not be that high, there may nevertheless be substantial amounts of 5-hydroxymethyluracil, the product of this reaction. Surprisingly, 5-hydroxymethyluracil levels were 7-fold higher in samples with splicing factor *SF3B1* mutations than in control samples (Figure 1F); no other differences were found for *SF3B1* mutations.

Receiver operating characteristic (ROC) curves for predicting TET2 mutated *versus* wild-type subjects show that a 5-formylcytosine threshold of 0.204 per 10^6dN yields a sensitivity of 100% and a specificity of 83% (*Online Supplementary Figure S1C*). Using 5-hydroxymethylcytosine and a threshold of 0.038 per 10^3dN yields a sensitivity of 67% and a specificity of 100% (*Online Supplementary Figure S1B*), and using 5-carboxycytosine and a threshold of 11.096 per 10^9dN yields a sensitivity of 60.0% and a specificity of 100% (*Online Supplementary Figure S1D*). Finally, using 5-methylcytosine and a threshold of 7.60 per 10^3dN yields a sensitivity of 53.3% and a specificity of 100% (*Online Supplementary Figure S1A*). ROC areas under the curves (AUC) capture these results in a manner that is independent of *ad hoc* choices of the thresholds. Results of this are 0.926, 0.838, 0.8, 0.733, and 0.838 (fractions of 1) for 5-formylcytosine, 5-hydroxymethylcytosine, 5-carboxycytosine, and 5-methylcytosine, respectively. For predictions of mutations in SF3B1, we found that a 5-hydroxymethyluracil threshold of 0.784 per 10^6dN yields 100% sensitivity and 87.5% specificity and AUC of 0.9 (*Online Supplementary Figure S2E*).

It has been established that TET's preferred substrate is 5-methylcytosine over 5-hydroxymethylcytosine[10] and thus that TET2 predominantly forms 5-hydroxymethylcytosine. However, we found that TET2 mutations are associated with decreases in 5-formylcytosine and 5-carboxycytosine, but have only a moderate influence on 5-hydroxymethylcytosine. Recently, Kohli's laboratory,[11] using an *in vitro* model, showed that certain types of *TET2* mutations (those in the active site scaffold Thr1372-Tyr1902) eliminate 5-formylcytosine and 5-carboxycytosine formation with no influence (or only a moderate one) on 5-hydroxymethylcytosine formation. Our results obtained in human marrow samples add to that finding, and demonstrate that several mutations located in the DSBH and Cys-rich domains, which form the catalytic site, have a similar effect on TET2 activity. In this work, all frameshift and non-sense *TET2* mutations, likely resulting in a truncated protein, yielded similar modified base spectra/profiles. These results suggest that *TET2* mutations alter protein structure-function in ways that weaken its activity, specifically with respect to

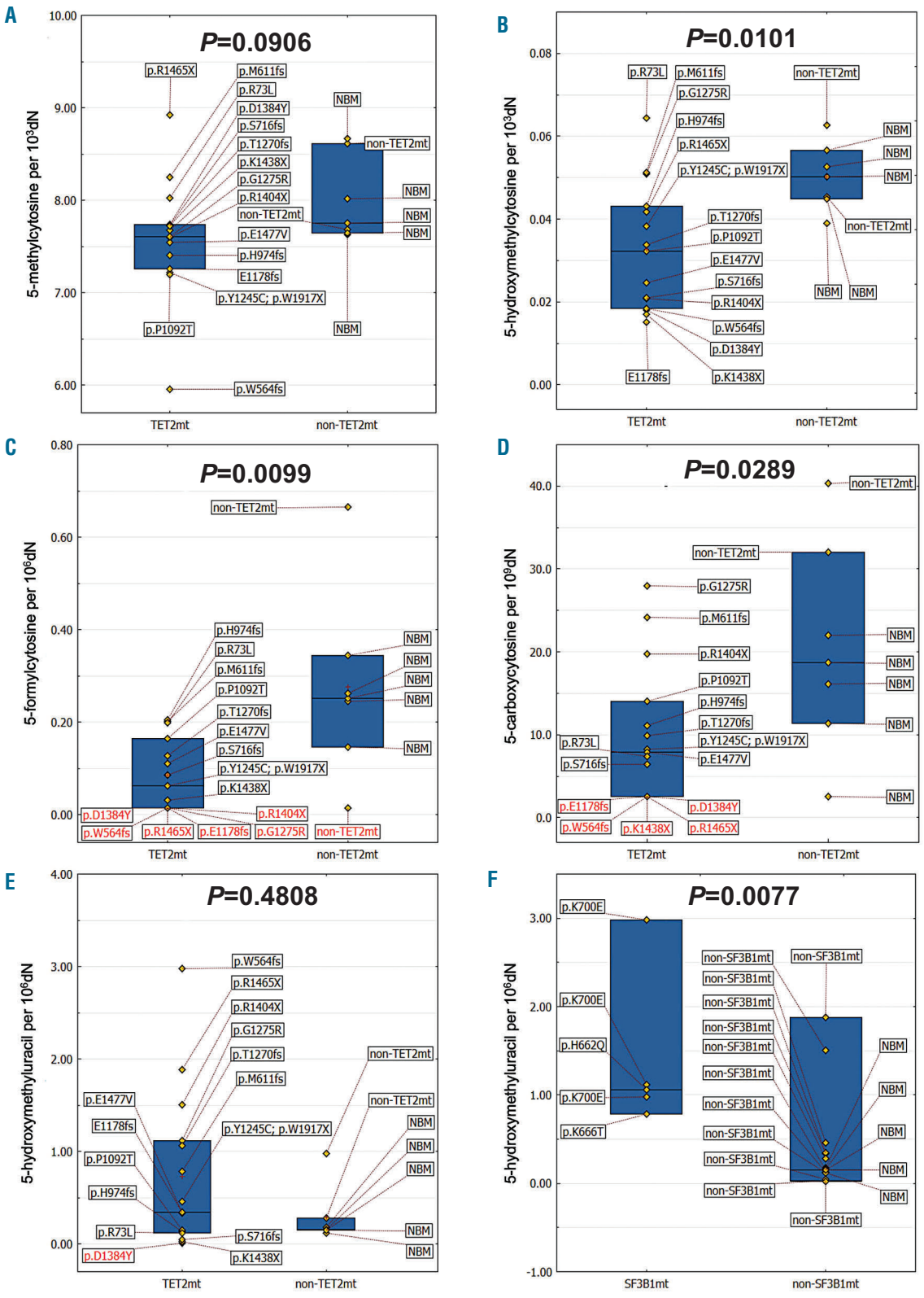

Figure 1. Levels of DNA markers of TET2 activity obtained using automated two-dimensional ultra-performance liquid chromatography tandem mass spectrometry with incorporation of stable isotope-labeled internal standards. (A) 5-methylcytosine. (B) 5-hydroxymethylcytosine. (C) 5-formylcytosine. (D) 5-carboxycytosine. (E) 5-hydroxymethyluracil. (F) 5-hydroxymethyluracil. Patients with *TET2* (A-E) or *SF3B1* (F) mutations (mt) *versus* controls (not mutated in *TET2* or *SF3B1* and normal bone marrow, NBM). *P*-values were obtained using two-sample *t*-tests. Medians and interquartile ranges are shown. Data below the limit of detection (LOD) were included in the statistical analyses as LOD/2 values. dN: deoxynucleosides.

higher order oxidation products. This is the first *in vivo* characterization of linkage between *TET2* mutations and TET2 enzyme activity gauged by its full spectrum of possible modified base products.

A recent observation of 5-formylcytosine enrichment at active enhancers involved in tissue development/differentiation[12] suggests that reduced levels of this base may be a characteristic feature of largely undifferentiated, malignant cells with *TET2* mutations. Others have also suggested that 5-formylcytosine levels in DNA can control gene expression.[13]

As mentioned, TET enzymes may also catalyze synthesis of 5-hydroxymethyluracil from thymine. According to Pfaffeneder *et al.*,[5] the level of 5-hydroxymethyluracil changes during the course of epigenetic reprogramming, following the same pattern as the other TET products 5-hydroxymethylcytosine, 5-carboxycytosine, and 5-formylcytosine. As TET enzymes are the most likely source of 5-hydroxymethyluracil, a question, unanswered here, remains: why do mutations in splicing factor gene *SF3B1* associate with 5-hydroxymethyluracil formation?

A partial explanation may be found in the study where it was demonstrated that disruption of *SF3B1* in the K562 cell line resulted in G2/M cell cycle arrest.[14] It is worth mentioning that we observed an approximately 19-35% enrichment of 5-hydroxymethyluracil in G2/M cell cycle phase (arrested by spindle poison KaryoMax) in FACS-sorted human hUES7 hESC cells (*Online Supplementary Figure S3*). Therefore, it is possible that high 5-hydroxymethyluracil levels, G2/M arrest, and *SF3B1* mutations may have something in common. Interestingly, our recently published data demonstrated that when *TET2* is the ancestral event, *SF3B1* is among the most common subsequent (secondary) hits.[15]

In summary, while previous studies of *TET2* mutations focused on 5-hydroxymethylcytosine content as a first product of the dioxygenase reaction, our results suggest analyses of 5-formylcytosine and 5-hydroxymethyluracil are also warranted. There is little consensus concerning a link between TET mutations and clinical outcome. Measurement of these alternately modified bases could improve our ability to cluster mutations in TET2 into subclasses that behave in a similar fashion in terms of downstream consequences. This, in turn, could improve our understanding of disease mechanisms and may prove valuable as a prognostic and diagnostic tool in myeloid malignancies.

Daniel Gackowski,[1] Maciej Gawronski,[1] Cassandra Kerr,[2] Tomas Radivoyevitch,[3] Ewelina Zarakowska,[1] Marta Starczak,[1] Abdulkadir Abakir,[4] Alexey Ruzov,[4] Jaroslaw P. Maciejewski[2] and Ryszard Olinski[1]

[1]Department of Clinical Biochemistry, Faculty of Pharmacy, Nicolaus Copernicus University in Torun, Ludwik Rydygier Collegium Medicum in Bydgoszcz, Poland; [2]Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH, USA; [3]Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA and [4]Wolfson Centre for Stem Cells, Tissue Engineering and Modelling (STEM), Division of Cancer and Stem Cells, School of Medicine, Centre for Biomolecular Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, UK

Funding: this work was supported in part by the Polish National Science Center (www.ncn.gov.pl) grants: UMO-2015/19/B/NZ5/02208 to DG and UMO-2017/27/B/NZ7/01487 to RO; by National Institutes of Health (Bethesda, MD, USA, NIH) grants, R01HL123904, R01HL132071, R35HL135795 to JPM; by Medical Research Council IMPACT DTP PhD Studentship (grant number MR/N013913/1) to AA and by Biotechnology and Biological Sciences Research Council (grant number BB/N005759/1) to AR.

Correspondence: RYSZARD OLINSKI - ryszardo@cm.umk.pl

JAROSLAW P. MACIEJEWSKI - maciejf@ccf.org

doi:10.3324/haematol.2019.224030

Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.

References

1. Jankowska AM, Szpurka H, Tiu RV, et al. Loss of heterozygosity 4q24 and TET2 mutations associated with myelodysplastic/myeloproliferative neoplasms. Blood. 2009;113(25):6403-6410.
2. Tefferi A, Pardanani A, Lim KH, et al. TET2 mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis. Leukemia. 2009;23(5):905-911.
3. Delhommeau F, Dupont S, Della Valle V, et al. Mutation in TET2 in myeloid cancers. N Engl J Med. 2009;360(22):2289-2301.
4. Tahiliani M, Koh KP, Shen Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324(5929):930-935.
5. Pfaffeneder T, Spada F, Wagner M, et al. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. Nat Chem Biol. 2014;10(7):574-581.
6. Gackowski D, Starczak M, Zarakowska E, et al. Accurate, Direct, and High-Throughput Analyses of a Broad Spectrum of Endogenously Generated DNA Base Modifications with Isotope-Dilution Two-Dimensional Ultraperformance Liquid Chromatography with Tandem Mass Spectrometry: Possible Clinical Implication. Anal Chem. 2016;88(24):12128-12136.
7. Yang H, Liu Y, Bai F, et al. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. Oncogene. 2013;32(5):663-669.
8. Branco MR, Ficz G, Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat Rev Genet. 2012;13(1):7-13.
9. Raiber EA, Hardisty R, van Delft P, Balasubramanian S. Mapping and elucidating the function of modified bases in DNA. Nat Rev Chem. 2017;1(9).
10. Hu L, Lu J, Cheng J, et al. Structural insight into substrate preference for TET-mediated oxidation. Nature. 2015;527(7576):118-122.
11. Liu MY, Torabifard H, Crawford DJ, et al. Mutations along a TET2 active site scaffold stall oxidation at 5-hydroxymethylcytosine. Nat Chem Biol. 2017;13(2):181-187.
12. Iurlaro M, McInroy GR, Burgess HE, et al. In vivo genome-wide profiling reveals a tissue-specific role for 5-formylcytosine. Genome Biol. 2016;17(1):141.
13. Raiber EA, Portella G, Martinez Cuesta S, et al. 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. Nat Chem. 2018;10(12):1258-1266.
14. Dolatshad H, Pellagatti A, Fernandez-Mercado M, et al. Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. Leukemia. 2015;29(5):1092-1103.
15. Hirsch CM, Nazha A, Kneen K, et al. Consequences of mutant TET2 on clonality and subclonal hierarchy. Leukemia. 2018;32(8):1751-1761.