MicroRNA-mediated alteration of TET2 interaction network in myeloproliferative neoplasms

Feng-Feng Wang, Lawrence WC Chan*, SP Yip, Benjamin YM Yung
Department of Health Technology and Informatics
Hong Kong Polytechnic University
Hong Kong, China
* wing.chi.chan@inet.polyu.edu.hk

Abstract—**Myeloproliferative neoplasms (MPNs) constitute a type of proliferative and dysplastic myeloid tumors, which are frequently found in the elderly people. Although some kinds of gene mutations in MPNs have been studied, the biological mechanisms behind this disease are still not very clear. Researchers have found that MPN patients with TET2 mutations have low level of 5hmC. However, they do not give reasons why there is also low level of 5hmC in patients with wild-type TET2. The aim of this study is to investigate into the role of TET2 and its interacting proteins in MPN under the repression by microRNAs. MicroRNAs are short, endogenous, non-coding RNA molecules, which regulate the target genes expression. We hypothesize that microRNAs lead to low level of 5hmC by down-regulating the expression of TET2 and other proteins interacting with TET2 in MPN patients with wild-type TET2, which is similar to the function of TET2 mutations. Bioinformatics tools were performed in this study. There were 11 databases considered, only 3 of which predicted microRNAs binding to TET2. Moreover, 10 proteins were found to be associated with TET2 according to *STRING* database and their targeting miRNA predictions was compared with that of TET2. The hypothesis can be supported by he predicted simultaneous repression of DNMT-1 and TET2 by *miR-152*.**

Keywords-MPN; TET2; miRNAs; protein interaction

I. INTRODUCTION

Myeloproliferative neoplasms (MPNs) are a kind of diseases of the bone marrow in which many different types of blood cells are generated. The North American Association of Central Cancer Registries (NAACCR) reported that among 82% of US population the average incidence rate of MPN was 2.1 per 100,000 per year in 2001-2003[1]. Moreover, because of the imperfect diagnosis, there may be some of MPN cases that cannot be detected or reported to this report [1]. Ten-Eleven Transcription 2 (TET2) located on 4q24 can catalyze the conversion from 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) in DNA [2, 3]. TET2 mutations are found in MPN patients, which lead to low level of 5hmC. So far, a host of TET2 mutations have been found, such as nonsense mutations, out-of-frame insertions, deletions, and splice site mutations [4, 5, 6]. It has been reported that TET2 mutations are associated with myeloid malignancies damaging catalytic activity, converting from 5mC to 5hmC [7]. The samples from MPN patients with TET2 mutations show low level of 5hmC, while patients with wild type TET2 reveal both low level and high level of 5hmC compared with healthy controls [7].

MicroRNAs (miRNAs) are short, endogenous, non-coding RNA fragments of 21-25 nucleotides. It can regulate the expression of target genes by binding to complementary sequences in 3'-UTR site of mRNA [8], which will ultimately lead to the down-regulation of protein expression [9]. It has been found that miRNAs play an important role in human tumors with alteration of target oncogenes or tumor suppressor genes [10]. As the validated data about miRNA targets are not complete at moment, various computational prediction databases are used to analyze the profile of miRNA targets [11].

The interaction among different proteins plays an important role in cellular function, which describes the protein interaction relationship in cells. In our study, we hypothesize that miRNAs can directly regulate TET2 transcript expression by binding to 3'-UTR of mRNA, or indirectly regulate TET2 by down-regulating the expression of other proteins interacting with TET2 in patients with wild type TET2 and low level of 5hmC. In this study, we applied bioinformatics tools to search for miRNAs regulating TET2 and proteins interacting with TET2.

II. MATERIALS AND METHODS

A. Databases searching for miRNAs

In our study, we considered 11 databases, including 8 target prediction databases and 3 validated databases supported by experiments or literatures (Table I). These databases are divided into two generations [12]. The representative examples of the first generation method are *TargetScan* and *DIANA-microT*. There are 3 basic criterions exiting in the first generation method: a. complementarity

between miRNA and mRNA in seed regions; b. free energy in folding the miRNA-mRNA duplex; c, conservation among different species on mature miRNA, the binding sites and the duplex [12]. The next generation method bases on machine learning approaches to rate interaction between miRNA and mRNA. One example of the next generation is *miRanda-miSVR*. It utilizes support vector regression (SVR) considering a number of features, such as accessibility of complementary site and conservation [13]. Apart from the prediction databases, there are 3 manually curated databases: *TarBase*, *miRWalk-validated* and *miR2Disease*. All the data from these 3 databases are derived from experiments or literatures. The systematic search is aimed to find out the algorithmically predicted or experimentally confirmed miRNAs in human. Most of the considered resources facilitate online database search using Entrez Gene symbol, for example TET2, but one resource, *PITA*, requires the entry of 3'-UTR sequence of gene transcript. Usually, a gene has several transcripts, but only one can fit the requirement of our study. Therefore, we should firstly choose the right one. In our study, we chose transcript variant 1 of TET2, since it is the longest variant and has the complete structure of catalytic domain. The 3'-UTR sequence can be extracted from the *NCBI* database and input to *PITA*. Homo sapiens should be chosen if a species field is provided by the graphical user interface (GUI) of the database search. At last, we obtained a set of candidate miRNAs.

B. The proteins interacting with TET2

In order to explore the proteins interacting with TET2, we preformed *STRING* database searching. The *STRING* database is an updated search tool for seeking interacting genes or proteins, which provides both validated and predicted interaction information [14]. This database searches by using the name of protein (e. g. TET2). After that, we performed miRNA searching for all these proteins from *STRING* database based on the same set of candidate miRNAs regulating TET2.

C. Statistical Analysis

The significance of enrichment is usually tested by using the hypergeometric distribution [15]. This statistical method is a discrete probability distribution that characterizes the number of successes in a sequence of n randomly selects from a finite population of N items without replacement [15]. The hypergeometric probability is referred to the probability that an n-trial hypergeometric experiment leads to precisely x successes [15]. There are some notations in hypergeometric distribution: N is defined as the number of items in the population; k is defined as the number of items in the population that stands for successes; n is defined as the number of items in the sample; x is defined as the number of items in the sample that stands for successes; ${}_kC_x$ is defined as the number of mixture of k items, taken x at a time; h(x;N,n,k) represents hypergeometric distribution [15]. The formula is:

$$h(x;N,n,k) = \frac{{}_kC_x \cdot {}_{N-k}C_{n-x}}{{}_NC_n} \qquad (1)$$

TABLE I. DATABASES FOR MIRNA TARGETS IDENTIFICATION

Databases	Methods	Sources
TargetScan	1st generation	http://www.targetscan.org/vert_50/
MicroCosm-Targets	1st generation	http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/#
miRWalk-predicted	1st generation	http://www.umm.uni-heidelberg.de/apps/zmf/mirwalk/index.html
DIANA-microT	1st generation	http://diana.cslab.ece.ntua.gr/microT/
miRanda-mirSVR	2nd generation	http://www.microrna.org/
miRDB	2nd generation	http://mirdb.org/miRDB/
PicTar	2nd generation	http://pictar.mdc-berlin.de/
PITA	2nd generation	http://genie.weizmann.ac.il/pubs/mir07/
Tarbase	Validated version	http://diana.cslab.ece.ntua.gr/tarbase/
miR2Disease	Validated version	http://mlg.hit.edu.cn:8080/miR2Disease/
miRWalk-validated	Validated version	http://www.umm.uni-heidelberg.de/apps/zmf/mirwalk/index.html

This statistical method was used to analyze the significance of enrichment on the overlapping of miRNAs targeting TET2 from different databases, in order to see if the common miRNAs included in any two databases are obtained by chance.

III. RESULTS

A. The candidate miRNAs regulating TET2

In our study, only 3 databases revealed the potential miRNAs binding to TET2 transcript variant 1. Most importantly, all the 3 databases are prediction databases: *MicroCosm-Targets* and *miRWalk-predicted* belong to the first generation method; while *miRanda-miSVR* is in the next generation group. There were 100 candidate miRNAs predicted in these 3 databases. Some candidate miRNAs were mutual in both two databases: *hsa-miR-23b* and *hsa-miR-202* were included in *MicroCosm-Targets* and *miRanda-miSVR*; there were 16 miRNAs included in *miRWalk-predicted* and *miRanda-miSVR*, such as *hsa-miR-29a and hsa-miR-29b*. Most importantly, the enrichment of these 16 miRNAs was significant ($p<0.05$) after the analysis of the hypergeometric distribution. There were no candidate miRNAs included among all the 3 databases (Table II & Fig. 1). Why did *MicroCosm-Targets* have only 2 common candidate miRNAs with *miRanda-miSVR*; while *miRanda-miSVR* and *miRWalk-predicted* have 16 common miRNAs? The algorithm and score

system in these 2 databases may account for it. The miRanda Algorithm is used in *MicroCosm-Targets*, which is a weighted scoring system and rewards complementarity at the 5' end of the miRNA. In this algorithm strict complementarity is performed at seed region, and regions between miRNA and mRNA where more than one base is not to a target site in this algorithm will be ruled out [16]. However, *miRWalk-predicted* uses miRWalk Algorithm, which is based on a computational method to recognize the longest complementary between miRNA and mRNA [17]. The discipline in *MicroCosm-Targets* is more strict than in *miRWalk-predicted*, which leads to the different results. Furthermore, the score system based on their algorithm is another factor. The score in *MicroCosm-Targets* stands for weighted complementarity in binding sites generating from dynamic programming alignment [16]; while P-value is used in *miRWalk-predicted* for the score scoring system, which is a probability distribution of random matches of miRNA 5' end sequence in mRNA sequence. It is calculated by making use of poisson distribution [17]. In a word, different databases have different results based on different conditions.

B. MiRNA-regulated proteins interaction with TET2

Here, 10 proteins were found to have relationship with TET2 (KIAA1546 is one gene symbol of TET2) from *STRING* database (Fig. 2). There was one protein (ASXL3) without any candidate miRNA binding to TET2. Another one (PDC) was not found Homo sapiens related information from *NCBI* database. As a consequence, there were 8 proteins (TET3, CXXC5, ASXL2, CXXC4, ASXL1, CXXC6, DNMT1 and TBC1D4) studied in this research. Among these 8 proteins one candidate miRNA interacted with several proteins. On the contrary, one protein also interacted with several candidate miRNAs. In all these candidate miRNAs, there were only 2 miRNAs (*miR-152 and miR-29b*) validated from literature and experiment. It has been reported that *miR-152* can down-regulate DNMT-1 expression in some human cancers, such as liver cancer and cholangiocarcinomas [18, 19]. Methyltransferase-1 (DNMT-1) is greatly abundant in mammalian cells, which plays a vital role in the process of DNA methylation [20, 21]. The repression of DNMT-1 can result in low level of methylation. In our research, we studied why some MPN patients have wild-type TET2 but low level of 5hmC. From database researching, we knew that *miR-152* was also one of the candidate miRNAs regulating TET2. We infer that over-expression of *miR-152* in MPN patients may lead to down-regulation of TET2 and DNMT-1, and then lead to decreased 5hmC next to reduction of the substrate, 5mC (Fig. 3). However, the exact relationship between TET2 and DNMT-1 is not very clear at the moment. Although information on other proteins interacting with TET2 is not validated for now, we did some prediction by using bioinformatics tools. In this study, we hypothesize that miRNAs down-regulate the expression of these proteins interacting with TET2 (e.g. DNMT-1), and indirectly affect the function of TET2, leading to the low level of 5hmC.

TABLE II. CANDIDATE MIRNAS INCLUDED IN EITHER 2 DATABASES

Databases	Common miRNAs
MicroCosm-Targets miRanda-miSVR	hsa-miR-23b, hsa-miR-202
miRWalk-predicted miRanda-miSVR	hsa-miR-101, hsa-miR-144, hsa-miR-17, hsa-miR-148b, hsa-miR-152, hsa-miR-20b, hsa-miR-26a, hsa-miR-26b, hsa-miR-29a, hsa-miR-29b, hsa-miR-29c, hsa-miR-520e, hsa-miR-520a-3p, hsa-miR-520b, hsa-miR-520c-3p, hsa-miR-520d-3p

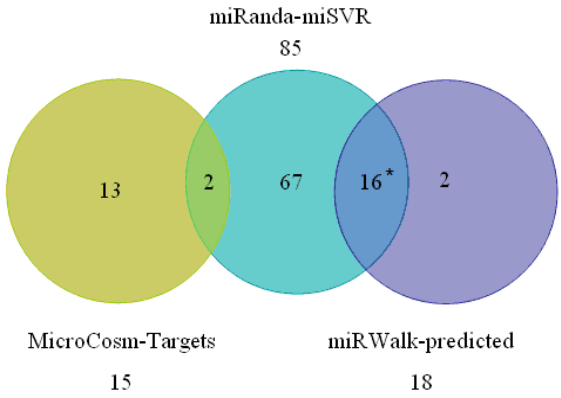

Figure 1. Prediction of miRNAs targeting TET2 from databases (* indicates $p<0.05$ for hypergeometric statistics).

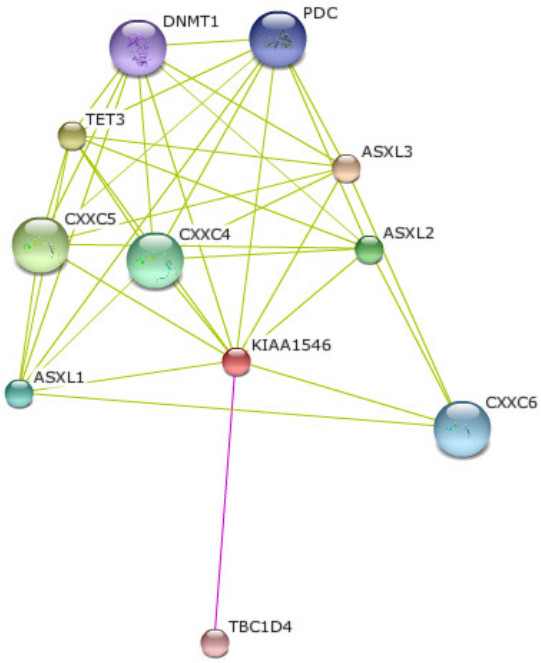

Figure 2. miRNA-regulated proteins interaction with TET2 (KIAA1546) generated by *STRING* database.

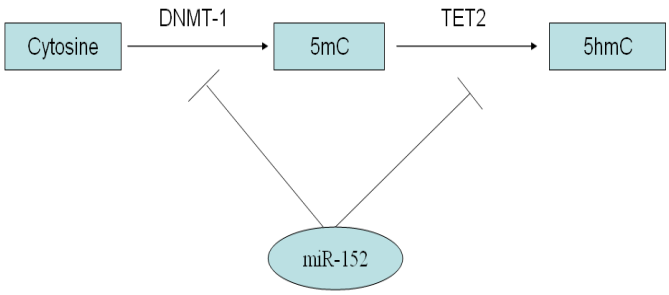

Figure 3. Schematic representation of the relationship among miR-152, DNMT-1 and TET2.

IV. DISCUSSION

The role of miRNA on repression of mRNA is becoming increasingly hot topic, but remains poorly understood in more detailed. With the development of science and technology, bioinformatics tools are becoming increasingly important and useful in prediction aspect, owning to the imperfect validated data at the moment.

As for miRNAs, different databases are based on different conditions to predict targets. Usually, the results from different databases are somehow different. Besides, we also can find the validated information on miRNAs from the databases, which is derived from experiments or literatures. In our study, we got a set of miRNAs derived from 11 databases searching, and then enriched the data to form a profile of candidate miRNAs. Actually, there were only 3 databases that had the potential miRNAs regulating TET2 in this study. Furthermore, we performed protein interaction searching and generated 8 proteins interacting with TET2 by using *STRING* database. However, in this database, we only can know that there is a relationship between these proteins and TET2. The exact relationship, such as upregulation and downregulation, is not very clear. It was found that some of the candidate miRNAs of TET2 also regulated the 8 proteins, which demonstrated that miRNAs can directly and indirectly affect the function of TET2. The interaction between proteins and miRNAs can form a very complete network, which ultimately affect the normal function of TET2. Maybe these miRNAs are over-expressed in MPN patients with wild-type TET2 and low level of 5hmC compared with healthy controls. Among these 8 proteins, DNMT-1 is the most important one related to methylation. The process of adding methyl group to cytosine is an important epigenetic modification [17], which is catalyzed by DNA methyltransferases, such as DNMT-1. We just hypothesize that the conversion from cytosine to 5mC are damaged owning to the down-regulation of DNMT-1 by miRNAs. As a result, lead to low level of 5hmC. From the database searching and literature review, we found that *miR-152* represses the expression of DNMT-1 [22]. Most importantly, *miR-152* also down-regulates the expression of TET2 from database prediction. All theses two aspects result in low level of 5hmC. However, the exact relationship between TET2 and DNMT-1 is not very clear at the moment.

In sum, our results showed that there were miRNAs regulating TET2 and proteins interacting with TET2 by performing bioinformatics tools. However, all the data on miRNAs regulating TET2 were predicted by different databases and no validated data available now. Moreover, among these 8 proteins interacting with TET2, only DNMT-1 had the validated miRNA (*miR-152*) and no validated information on other proteins. Besides, the exact relationship between these 8 proteins and TET2 is not very clear. As a result, there are still quite a few problems await solution in the future.

Actually, the processes happening in cells are complex. Every process is a network involving various factors. In similar manner, the network of miRNAs regulating targets also comprises too many members. Only one miRNA can not function well. Moreover, competitions exit in different miRNAs but with same binding sites. As a consequence, it is a long way to study the function of miRNAs involved in a certain process in human body clearly.

REFERENCES

[1] D.E. Rollison, et al, Epidemiology of myelodysplastic syndromes and chronic myeloproliferative disorders in the United States, 2001–2004, using data from the NAACCR and SEER programs. *Blood* **112,** 45–52 (2008).

[2] M. Tahiliani, et al, Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–935 (2009).

[3] L. M. Iyer, M. Tahiliani, A. Rao, and L. Aravind, Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8,** 1698–1710 (2009).

[4] A. Tefferi, et al, TET2 mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis. *Leukemia* **23,** 905–911 (2009).

[5] O. Kosmider, et al, TET2 mutation is an independent favorable prognostic factor in myelodysplastic syndromes (MDS). *Blood* **114,** 3285–3291 (2009).

[6] C. Saint-Martin, et al, Analysis of the teneleven translocation 2 (TET2) gene in familial myeloproliferative neoplasms. *Blood* **114,** 1628–1632 (2009).

[7] K. Myunggon, et al, Impaired hydroxylation of 5-methylcytosine in myeloid Cancers with mutant TET2. *Nature* **468,** 839-843 (2010).

[8] M.S. Kumar, J. Lu, and K.L. Mercer, T.R. Golub, and T. Jacks, Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat Genet.* **39,** 673-677 (2007).

[9] M. Eder, and M. Scherr, MicroRNA and lung cancer. *N Engl J Med.* **352,** 2446-2448 (2005).

[10] P.P. Medina, and F.J. Slack, MicroRNAs and cancer: an overview. *Cell Cycle* **7,** 2485-2492 (2008).

[11] C.Y. Chen, S.T. Chen, C.S. Fuh, H.F. Juan, and H.C. Huang, Coregulation of transcription factors and microRNAs in human transcriptional regulatory network. *BMC Bioinformatics* **12,** S41-S51 (2011).

[12] L. Li, J. Xu, D. Yang, X. Tan, and H. Wang, Computational approaches for microRNA studies: a review. *Mamm Genome* **21,** 1-12 (2010).

[13] D. Betel, A. Koppal, P. Agius, C. Sander, and C. Leslie, Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* **11,** R90-R103 (2010).

[14] D. Szklarczyk, et al, The STRING database in 2011: functional interaction network of proteins globally integrated and scored. *Nucleic Acids Res.* **39,** D561-D568 (2011).

[15] S.B. Plaisier, R. Taschereau, J.A. Wong, and T.G. Graeber, Rank-rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Res.* **38,** e169 (2010).

[16] S. Griffiths-Jones, H.K. Saini, S. Van Dongen, and A.J. Enright, MiRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36,** D154-D158 (2008).

[17] P. Shahi, et al, Argonaute--a database for gene regulation by mammalian microRNAs. *Nucleic Acids Res.* **34,** D115-D118 (2006).

[18] J.F. Huang, Y. Wang, Y. Guo, and S. Sun, Down-Regulated MicroRNA-152 Induces Aberrant DNA Methylation in Hepatitis B Virus–Related Hepatocellular Carcinoma by Targeting DNA Methyltransferase 1. *Hepatology* **52,** 61-70 (2010).

[19] C. Braconi, N. Huang, and T. Patel, MicroRNA-Dependent Regulation of DNA Methyltransferase-1 and Tumor Suppressor Gene Expression by Interleukin-6 in Human Malignant Cholangiocytes. *Hepatology* **51,** 881-890 (2010).

[20] D.R. Hodge, et al, Interleukin-6 regulation of the human DNA methyltransferase (HDNMT) gene in human erythroleukemia cells. *J Biol Chem.* **276,** 39508–39511 (2001).

[21] F. Meng, H. Wehbe-Janek, R. Henson, H. Smith, and T. Patel, Epigenetic regulation of microRNA-370 by interleukin-6 in malignant human cholangiocytes. *Oncogene* **27,**378-386 (2008).

[22] A. Bird, DNA methylation patterns and epigenetic memory. *Genes Dev.* **16,** 6-21 (2002).