Bioinformatics Prioritization of SNPs Perturbing MicroRNA Regulationof Hematological malignancies-implicated Genes

Hamid Ghaedi, Milad Bastami, Davood Zare-Abdollahi, Behnam Alipoor, Abolfazl Movafagh, Reza Mirfakhraie, Mir Davood Omrani, Andrea Masotti

PII: S0888-7543(15)30038-0
DOI: doi: 10.1016/j.ygeno.2015.10.004
Reference: YGENO 8776

To appear in: *Genomics*

Received date: 4 August 2015
Revised date: 19 October 2015
Accepted date: 27 October 2015

Please cite this article as: Hamid Ghaedi, Milad Bastami, Davood Zare-Abdollahi, Behnam Alipoor, Abolfazl Movafagh, Reza Mirfakhraie, Mir Davood Omrani, Andrea Masotti, Bioinformatics Prioritization of SNPs Perturbing MicroRNA Regulationof Hematological malignancies-implicated Genes, *Genomics* (2015), doi: 10.1016/j.ygeno.2015.10.004

Bioinformatics Prioritization of SNPs Perturbing MicroRNA Regulation of Hematological malignancies-implicated Genes

Hamid Ghaedi[1], Milad Bastami[1], Davood Zare-Abdollahi[1], Behnam Alipoor[2], Abolfazl Movafagh[1], Reza Mirfakhraie[1], Mir Davood Omrani[1,§], Andrea Masotti[3,§]

1- Medical Genetics Department, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran
2- Clinical Biochemistry Department, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran
3- Bambino Gesù Children's Hospital-IRCCS, Gene Expression - Microarrays Laboratory, V.le San Paolo 15, 00146, Rome, ITALY

§ These two are the corresponding authors

Corresponding authors mailing addresses:

Mir Davood Omrani

E-mail: davood_omrani@yahoo.co.uk

Medical Genetics Department, Faculty of Medicine, Shahid Beheshti University

Andrea Masotti

E-mail: andrea.masotti@opbg.net

Bambino Gesù Children's Hospital-IRCCS, Gene Expression - Microarrays Laboratory, Polo di Ricerca - V.le San Paolo 15, 00146 Rome – ITALY, Tel: +39 06 6859 2650, Fax: +39 06 233234321

Abstract

The contribution of microRNAs (miRNAs) to cancer has been extensively investigated and it became obvious that a strict regulation of miRNA-mRNA regulatory network is crucial for safeguarding cell health. Apart from the direct impact of miRNAs dysregulation in cancer pathogenesis, genetic variations in miRNAs are likely to disrupt miRNA–target interaction. Indeed, many evidences suggested that SNPs within miRNA regulome are associated with the development of different hematological malignancies. However, a full catalog of SNPs within miRNAs target sites of genes relevant to hematopoiesis and hematological malignancies is still lacking. Accordingly, we aimed to systematically identify and characterize such SNPs and provide a prioritized list of most potentially disrupting SNPs. Although in the present study we did not address the functional significance of these potential disturbing variants, we believe that our compiled results will be valuable for researchers interested in determining the role of target-SNPs in the development of hematological malignancies.

Key words: hematological malignancy, leukemia, microRNA, single nucleotide polymorphism

1. Introduction

MicroRNAs (miRNAs) are a class of small noncoding RNAs that regulate nearly all cellular processes altered during tumorigenesis. The contribution of these regulators to cancer has been therefore extensively investigated. MiRNAs modulate gene expression by binding most often to 3' UTR target sites of mRNAs and interfere with target translation or promote its cleavage and degradation (1). In light of these premises, an efficient and strict regulation of miRNA-mRNA regulatory network is critical for safeguarding cell health. Over the last few years, many papers have pointed to the miRNA-mRNA regulatory network alterations and their fundamental contribution during tumorigenesis (2-5).

MiRNAs could function as oncogenes or tumor suppressors. Oncogenic miRNAs (e.g. miR-17-92 cluster, miR-21, miR-155 and miR-372/373) are amplified or overexpressed in cancers. On the other hand, in cancer tissues tumor suppressor miRNAs (e.g. miR-15a and miR-16-1, miR-34 family, let-7 family and miR-29) exhibit a reduced expression profile or may be deleted from the genome (3-5). Moreover, about half of miRNAs are located at fragile sites and genomic regions frequently amplified or deleted in human cancers (6, 7). Regarding hematopoiesis, recent findings indicated that a functional miRNA-mRNA regulatory network is crucial for hematopoietic stem cells (HSCs) maintenance and for preserving HSCs potency to produce all the cells of the hematopoietic system throughout the organism lifetime. For example, miR-223 regulates granulopoiesis (8), miR-221 and miR-222 negatively control erythropoiesis (9) while miR-146, miR-150, and miR-181 exert critical roles in promotion of B-lymphocyte development (10). In this framework and given the fundamental roles of miRNAs in regulating normal hematopoiesis, it is not surprising that miRNA dysregulation may contribute to the development of hematologic malignancies (HMs) (11). Moreover miRNA expression profiles could be used to determine leukemias prognosis and to discriminate cancer types including AML (acute myeloid leukemia) and ALL (acute lymphoblastic leukemia) (12, 13).

It is not surprising that genetic and epigenetic aberrations might lead to miRNA misexpression (14-17). The most frequent genetic variation in the human genome are SNPs (single nucleotide polymorphisms), occurring every one in 300 bp throughout the human genome and not necessarily with equal spacing. SNPs have been

extensively studied for defining candidate disease-causing gene regions, and establishing functional relationships between genotypic and phenotypic differences (18, 19). Until few years ago, there was a huge interest in nonsynonymous SNPs because they shift the codons and often change the protein structure and function. The majority of SNPs in the genome are not nonsynonymous SNPs and they occur in non-coding regions including intronic, intergenic regions and untranslated regions (UTRs). These SNPs could affect individuals susceptibility to complex diseases through modulating gene expression quantitatively (19, 20).

Unlike nonsynonymous SNPs, elucidation of non-coding SNPs functions is not straightforward because gene regulatory elements could not be accurately defined in a complex gene regulation process. However, since miRNA-mRNA interaction depends on the thermodynamics of RNA duplex, sequence variations such as SNPs at miRNA-binding sites may affect gene expression and mRNA translation. The rationale for this assumption is based on the principle of miRNA targeting process which initiates with crucial perfect base paring step between target mRNA and the 'seed region' (a sequence of 7 nt from position 2 to 8 at 5'-terminus of miRNAs) (21).

Accordingly, a SNP in the miRNA-binding site of a target mRNA (hereinafter referred to as 'target-SNP') is likely to disrupt miRNA–target interaction, therefore leading to gene expression/translation dysregulation. In this regard, the effect of such SNPs on miRNA-mRNA networks is predictable. Target-SNPs could deregulate gene expression/translation in hematopoietic tissue and might contribute to the pathogenesis of hematological malignancies (20).

The seminal study in the research field of target-SNPs association with cancer risk, was the 2008 Landi *et al.* paper (22). Thereafter, researchers have found that allele-specific miRNA-mediated regulation of gene expression may play a pivotal role in different aspects of HMs including individual's predisposition, prognosis and outcome. For example, it has been reported that rs3660 within *KRT81* is a target-SNP (for miR-17, miR-20a, miR-20b, miR-93, miR-106a, miR-106b, and miR-519d) and associates with longer overall survival in patients with multiple myeloma (23, 24). Moreover, homozygous polymorphic genotypes of rs1573613 (in *ETV6*) were associated with increased ALL risk, while carriers of the variant allele of rs9479 (in *PML*) were at

lower risk of ALL and AML (25). Homozygous polymorphic genotypes of rs187729 (in *ARHGAP26*) and rs10514611 (in *IRF8*) conferred increased risk of CML (chronic myelogenous leukemia). Variant alleles of rs187729, rs1573613 and rs9479 were shown to affect binding of miR-18a-3p, miR-34c-5p/miR-449b-5p and miR-510-5p, respectively (25). Recently, target-SNPs that occur in genes that are related to specific cancers have been identified, catalogued and proposed as candidate disease susceptibility variants. However, a comprehensive study that integrates multiple computational evidences to effectively catalog target-SNPs and predict their potential consequences in genes relevant to HMs is still lacking.

In this study, we systematically identified target-SNPs in genes relevant to hematopoiesis and HMs and employed various computational tools to predict their functional consequences.

2. Results

2.1. SNPs residing in miRNA-mRNA binding sites

All SNPs from dbSNP build 138 have been mapped to the 606,409 experimentally verified miRNA target sites. We obtained a final list of 392,753 SNPs residing in miRNA-mRNA binding sites, which has been considered for further analyses.

2.2. Polymorphic miRNA–mRNA binding sites in genes of HM pathways

Following our workflow, we obtained the full list of polymorphic miRNA-mRNA binding sites in all of the genes present in the KEGG pathways related to HMs and we found 373 target-SNPs potentially affecting 816 unique miRNA-mRNA interaction (interaction of 253 different miRNAs with 46 genes) (see Supplementary Material - Table S1).

2.3. GWAS variants and polymorphic miRNA target sites

Currently, the NHGRI GWAS Catalog (26) contains more than 30 published association studies and 152 associated variants related to HMs. We screened GWAS to find genetic variants located in experimentally

identified miRNA target sites. We found that one HM-associated GWAS variant (rs7097 associated with diffuse large B-cell lymphoma, p-value $<7\times10^{-6}$) and three proxy variants lay in miRNA binding sites (Table 1). Moreover, rs7097 The complete list of variants has been reported in Supplementary Table S1.

Table 1. HMs associated SNPs in GWAS located in miRNA target sites.

SNP	GWAS variant	LD (r2)	Population	Target gene	microRNA
rs7097	rs7097a	1	ASN	POLR1D	miR-425-3p, miR-544a,miR-507
rs8243	rs2041840	0.99	EUR	PRKD3	hsa-miR-329-3phsa-miR-495-3p
rs2160395	rs2041840	0.98	EUR	PRKD3	hsa-miR-145-5p,hsa-miR-27b-3p, hsa-miR-27a-3p
rs1317082	rs10936599	1	EUR	MYNN	hsa-miR-135b-5p,hsa-miR-135a-5p
rs78176302	rs948562	0.95	EUR	ZFP91-CNTF	hsa-miR-204-5p
rs442220	rs2191566	0.9	EUR	ZNF155	hsa-miR-708-5p,hsa-miR-28-5p

2.4. Evolutionary conservation of target sites

We computed PhastCons scores for target sites, which represent an evolutionary conservation measure based on 17 vertebrate species multiple alignment. The results showed that 168 out of 462 target sites (~36%) were highly conserved (PhastCons score > 0.7), while 179 out of 462 (~39%) were poorly conserved (PhastCons score <0.1) (Fig. 1A). Consistent with previous studies, this may represent that a considerable portion of miRNA binding sites in HMs-related genes are species specific.

2.5. Impact of HMs attributed target-SNPs on miRNA-mRNA hybrid stability

Target-SNPs can affect the hybridization of miRNA-target mRNA through creation or disruption of base pair(s). To predict the effect of these SNPs on miRNA-target mRNA hybrid stability, we computed $\Delta\Delta G_{hybrid}$ for each interaction affected by target-SNP, where $\Delta\Delta G_{hybrid}$ represents change in the stability of hybridization

between miRNA and mRNA. A positive value of ΔΔG represents a decrease in hybrid stability due to the target-SNP, whereas a negative value indicates an increase in hybrid stability. The stability of miRNA-target hybrid is decreased in 86 out of 740 interactions from 1 up to 6.6 kcal/mol and increased in 347 out of 740 interactions from 1 up to 7.2 kcal/mol for effect of the SNPs (Fig. 1B).

2.6. Impact of HM target-SNPs on local RNA structure

SNPs can also affect miRNA binding by altering local secondary structure, which is not quantified by $\Delta\Delta G_{hybrid}$. Therefore, we employed RNAsnp to measure the effect of HM target-SNPs on RNA structure. A total of 68 target-SNPs were predicted as structure disruptive variants by d_{max}, (*P* value threshold of 0.2) (Fig. 1C).

2.7. Impact of HM target-SNPs on structural accessibility of target sites

Previous studies have shown that a change in structural accessibility of target site can alter target recognition (27). Therefore, we examined whether the structural accessibility of target sites can be modified by HM target-SNPs. The SNP-induced change in target site accessibility was measured by ΔPu. The structural accessibility of 290 out of 740 target sites were decreased from 1 up to 30.6% and increased in 321 out of 740 target sites from 1 up to 43% by the corresponding target-SNPs, which indicates that many target-SNPs can affect the accessibility of target sites and hereby modulate the expression of target genes (Fig. 1D).

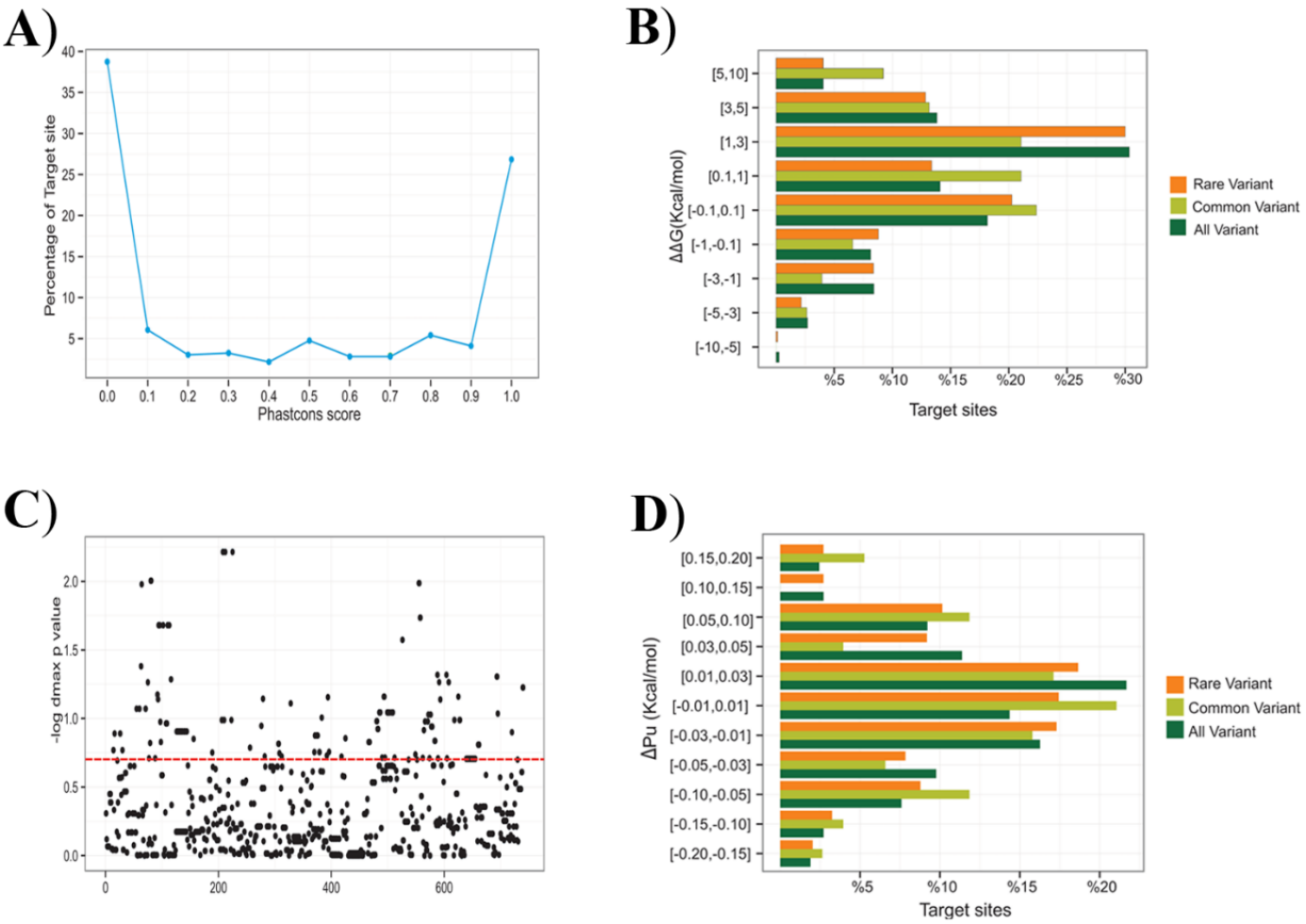

Figure1. (A) PhastCons scores for target sites. **(B)** The bar plot of $\Delta\Delta G_{hybrid}$ for target sites regarding to alternative SNPs alleles. **(C)** d_{max} p value distribution. **(D)** The bar plot of ΔPu for HM target-sites, shows distribution of target site accessibility changes (in Kcal/mol).

2.8 Prioritized target-SNPs in gene relevant to HMs

The results of the abovementioned computational measures were combined to prepare a catalogue of prioritized target-SNPs in genes relevant to HMs (Table 2). The complete list of target-SNPs used for prioritization has been reported in Supplementary Table S2.

Table 2. Prioritized target-SNPs relevant to HMs. [a] Obtained from RegulomDB and HaploReg database. [b] TF (transcription factor) Motifs affected by SNP. [c,d] GERP and Siphy Cons are measures of evolutionary conservation and they measure conservation pressure on a single base pair instead of a stretch of base pairs, as calculated by PhastCons algorithm.

SNP	ΔG	d_{max}p value	phastCons score	dPu	TF Motifs[a,b]	GERP[a,c]	SiPhy Cons[a,d]	Interaction
rs1051062	4.4	0.11	0	-4.71	AhR::Arnt	+	-	CCND1::hsa-miR-186-5p
rs112973508	2.3	0.13	0.4	-2.67	AP-1,HNF1, Lhx8,Pax-2	-	-	AFF1::hsa-miR-302c-3p, AFF1::hsa-miR-520d-3p
rs139771367	3.4	0.01	1	-1.07	-	+	+	CBFB::hsa-miR-30e-5p, CBFB::hsa-miR-30b-5p
rs146967889	4.3	0.09	0	8.57	HNF4	-	-	BCL2::hsa-let-7e-5p, BCL2::hsa-miR-4770, BCL2::hsa-miR-202-3p
rs187904389	2	0.07	1	8.33	Ets,Nrf-2,PLZF, TATA	+	-	CCND1::hsa-miR-142-3p
rs188589315	6	0.02	0	4.42	EBF,Ets,Foxa ,GR,TCF12,VDR	-	-	PAX5::hsa-miR-193a-3p, PAX5::hsa-miR-370-3p
rs200416354	2.2	0.15	0.5	3.72	Arid3a,Esx1, Hoxa3,Hoxd8	+	+	CBFB::hsa-miR-218-5p
rs200993614	1.5	0.14	0	9.20	DBP,Foxa,Foxd3, Nanog,Pou2f2, Smad	-	-	MS4A1::hsa-miR-379-5p
rs367727337	1.8	0.18	1	3.81	TEF	NA	NA	MECOM::hsa-miR-382-5p, MECOM::hsa-miR-208b-3p
rs442220	5	0.06	0	20.24	Maf,SIX5,SP1, Znf143	-	-	ZNF155::hsa-miR-28-5p, ZNF155::hsa-miR-708-5p
rs61764366	2	0.15	0	1.17	AIRE,EBF	+	-	KRAS::hsa-miR-525-3p, KRAS::hsa-miR-524-3p
rs696	1.1	0.07	0	-4.55	STAT	-	-	NFKBIA::hsa-miR-208b-3p, NFKBIA::hsa-miR-208a-3p
rs743582	1.5	0.11	0	1.64	Hsf	-	-	PML::hsa-miR-206, PML::hsa-miR-1
rs80330323	3.4	0.01	0.3	9.76	Foxc1	-	-	FAS::hsa-miR-34a-5p, FAS::hsa-miR-452-5p, FAS::hsa-miR-34c-5p

rs180805186	2.4	0.20	1	7.51	TATA	+	+	RUNX1T1::hsa-miR-107, RUNX1T1::hsa-miR-103a-3p, RUNX1T1::hsa-miR-503-5p, RUNX1T1::hsa-miR-153-3p, RUNX1T1::hsa-miR-15a-5p ,RUNX1T1::hsa-miR-144-3p, RUNX1T1::hsa-miR-16-5p, RUNX1T1::hsa-miR-29b-3p, RUNX1T1::hsa-miR-29c-3p, RUNX1T1::hsa-miR-15b-5p, RUNX1T1::hsa-miR-410-3p, RUNX1T1::hsa-miR-424-5p, RUNX1T1::hsa-miR-376c-3p, RUNX1T1::hsa-miR-195-5p, RUNX1T1::hsa-miR-497-5p, RUNX1T1::hsa-miR-29a-3p, RUNX1T1::hsa-miR-338-3p

3. Discussion

Deciphering molecular events and determinants underlying HMs requires demanding and challenging researches as HMs still remain a class of diseases with a complex pathogenesis. Results pointed to cytogenetic abnormalities and other gene mutations as common events usually occurring in HMs. These abnormalities often are integral to the pathogenesis and the subsequent evolution of a hematologic malignancy. In spite of this longstanding endeavor to define the genetic lesions that drive HMs and parallel targeted therapy improvement, overall survival of HMs (apart from acute promyelocytic leukemia and chronic myeloid leukemia) is generally poor. Therefore, studies aimed at finding new molecular marker to refine prognostic and predictive models, improve the accuracy of disease diagnosis and identify novel therapeutic strategies are welcomed.

Recently, many studies have focused on the analysis of cancer-related miRNA regulome variations and strongly suggested that target-SNPs may play an important role in modulation of individuals cancer

predisposition (28). Therefore, these studies emphasized the importance of bioinformatics analysis as preliminary steps to identify the most relevant functional SNPs that potentially confer susceptibility to cancer, and cost-effectively facilitate the prioritization of variants for further experimental validation. In the present study, we exploited a bioinformatics approach to systematically identify candidate target-SNPs in HMs. We further used several computational tools to predict their functional consequences and characterize the most relevant candidate functional variants. Given the intricate network of relationships between miRNAs and their mRNA targets, target-SNPs could be considered as mediators and provide new potential diagnostic and prognostic marker in initial risk stratification and treatment decision of HMs. Of course, the functional significance of these variants remains to be fully elucidated with functional studies (that are beyond the scope of the present study) that are needed to confirm the relevance or association of these identified candidate target-SNPs in HMs.

It is generally acknowledged that common algorithms for miRNA target predictions suffer from high false positive and false negative rates (27) that, in turn, may lead to an inaccurate selection of candidate variants. In order to reduce this limitation, we focused our analysis on a general miRNA targetome dataset in which interactions were experimentally supported/validated by CLIP and CLASH experiments. Generally, there is a tradeoff between sensitivity and specificity as a consequence of the selection of predictive or experimentally validated target sites to study. We decided to choose only experimentally validated target sites, making the study more specific but decreasing sensitivity. The inclusion of CLASH based interactions in our miRNA targetome dataset enabled us to analyze both canonical and non-canonical binding sites. This dataset includes high-confidence 3'UTR, 5'UTR and CDS (coding region) interactions. We combined the results of different measures to catalogue and prioritized target-SNPs in HMs relevant genes (Table 2). Interestingly, from the listed SNPs, rs1051062 was used in a panel of genetic variants for assessing the risk of developing lung cancer (29), whereas rs696 has been reported to be associated with increased risk and bad prognosis in colorectal cancers (30) and involved also in the progression of cervical carcinomas (31). The bioinformatics criteria we

applied to prioritized SNPs are maybe too stringent as some target-SNPs (e.g. rs1573613, rs9479, rs187729 and rs10514611) previously reported to be associated with leukemias (2) were not prioritized.

Overall, our data provides a comprehensive catalog of HMs target-SNPs, their predicted functional consequences and important biological processes putatively influenced by these variations. We believe that the results of our study will be valuable for researchers interested in determining the role of target-SNPs in HMs development. Along with roles in HMs pathogenesis, target-SNPs may emerge as new potent markers for efficient diagnosis and prognosis and can be considered as an attractive target for innovative treatment of HMs patients.

4. Materials and Methods

The general bioinformatics workflow employed in this study is depicted in **Figure 2**. We mapped all SNPs from dbSNP (build 138) to an experimentally supported miRNA-mRNA binding sites to obtain a dataset of target-SNPs. Further, we intersected target-SNPs dataset with the datasets of genes related to HMs pathways (including pre-B cell ALL, pre-T cell ALL, AML, CML, CLL (chronic lymphocytic leukemia), MM (multiple myeloma), Hodgkin lymphoma, Burkitt lymphoma and lymphoplasmacytic lymphoma), GWAS-SNPs associated with HM conditions (including different type of leukemias and lymphomas) and proxy SNPs of GWAS-SNP. Then, we tried to predict functional consequences of target-SNPs by performing different analyses as follows; (i) evolutionary conservation analysis of target sites, (ii) miRNA-target gen hybrid stability analysis, (iii) local RNA structure analysis, (iv) target site accessibility analysis and (v) annotation with regulatory elements.

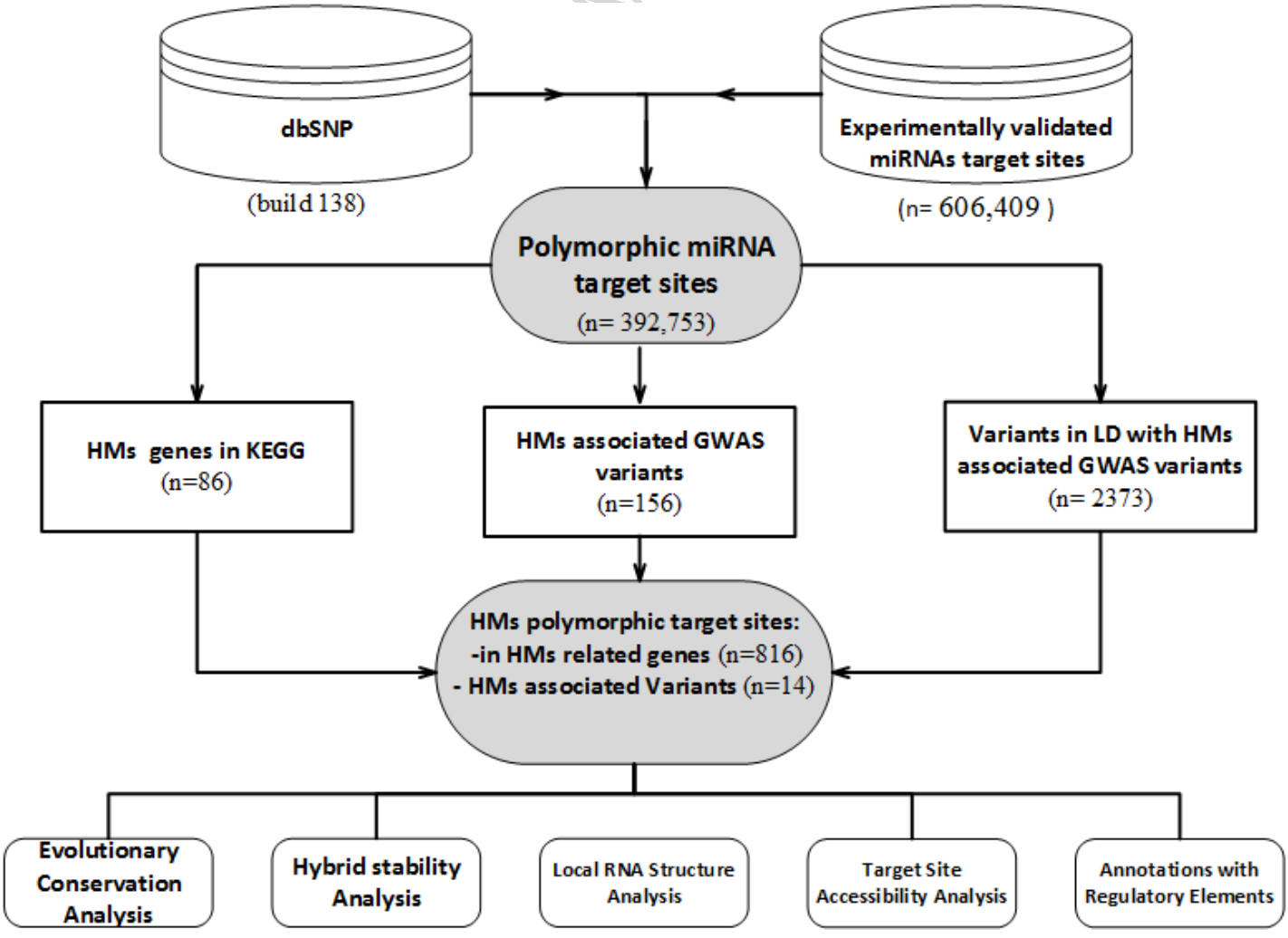

Figure 2. Schematic representation of study workflow.

4.1. Datasets

4.1.1. SNPs. Human SNP annotations were downloaded from dbSNP build 138 ftp site (ftp://ftp.ncbi.nih.gov/snp/organisms/human_9606/) (32). For each SNP, we extracted the following information; ID, genomic coordinates, type of polymorphism (microsatellites were discarded from the dataset), global minor allele frequency, functional consequences and validation status.

4.1.2. Experimentally supported miRNA target sites. It is quite well acknowledged that the commonest algorithms for miRNA target sites prediction suffer from a high number of false positives (27). Moreover, the lack of experimentally validated data on miRNA target sites is a further limitation. For these reasons and to focus the analysis on functionally verified targets sites, we exploited data from sequencing-based technologies (HITS-CLIP, PAR-CLIP and CLASH) for high throughput target identification. We compiled a dataset of experimentally-supported miRNA target sites, which includes CLIP-Seq supported interactions from starBase version 2 (33) and CLASH verified interactions extracted from PolymiRTS database (34).

4.1.3. HMs-related genes. We interrogated the Kyoto Encyclopedia of Genes and Genomes (KEGG) (35) and extracted all the gene information to identify HMs-implicated genes. We downloaded the gene lists and the information related to pre-B cell ALL (KEGG id: H00001), pre-T cell ALL (KEGG id: H00002), AML (KEGG id: H00003), CML (KEGG id: H00004), CLL (KEGG id: H00005), MM (KEGG id: H00010), Hodgkin lymphoma (KEGG id: H00007), Burkitt lymphoma (KEGG id: H00008) and lymphoplasmacytic lymphoma (KEGG id: H00011) directly from the KEGG web site.

4.1.4. GWAS-SNPs associated with HMs. We mined the NHGRI GWAS catalog (accessed April 2015) to obtained all SNPs associated (*p-value* $< 1.0 \times 10^{-5}$) with HMs disease/trait. To defined proxy SNPs for GWAS-SNPs, we compared subject population in each GWAS with 1000 Genomes Project super-populations. A SNP in strong linkage disequilibrium (defined as $r^2 \geq 0.8$) with any GWAS-SNP was considered as proxy SNP. We mined the Haploreg version 2 database (http://www.broadinstitute.org/mammals/haploreg/haploreg_v2.vcf.gz) to obtain LD scores (36).

4.2. Analysis

4.2.1. Extraction of polymorphic miRNA target sites in HMs. We compiled the polymorphic miRNA target sites dataset (target-SNPs dataset) by intersecting SNPs with the experimentally validated miRNA target site dataset, using specific tools available on the Galaxy web portal for large-scale interactive data analysis (37). To find which genes in the KEGG pathways of different HM conditions has polymorphic miRNA binding sites; we intersected the list of these genes with the target-SNPs dataset. To define which

GWAS variants and variants in LD is a target-SNP, we intersected the correspondent lists with Galaxy tools (37).

4.2.2. Conservation score of target sites. We used *Get Genomic Score* tool implemented in Galaxy (38) to calculate phastCons conservation score for each target sites. The average phastCons score for a target site was considered as its conservation score. The more conserved target site, the more severe consequences would be expected if SNP occur therein.

4.2.3. Impact of Target-SNPs on hybrid stability. We used the RNAhybrid v2.1.2 software (39) to compute $\Delta G_{hybrid.}$ This value represents the free energy of hybridization and was used as a measure of hybrid stability. To assess the effects of target-SNPs on the stability of miRNA-mRNA hybrids, we defined the $\Delta\Delta G_{hybrid}$. $\Delta\Delta G_{hybrid}$ score considered as the differences between ΔG_{hybrid} for alternative (ALT) allele and wild-type (WT) of a given SNP ($\Delta\Delta G_{hybrid} = \Delta G_{hybrid}$ (ALT) – ΔG_{hybrid} (WT)).

4.2.4. Impact of target-SNPs on local RNA secondary structure. We used the RNAsnp program (version 1.1) (40) to measure the effect of target-SNPs on local RNA secondary structure. The RNAsnp generates the base-pairing probability matrix for the WT and ALT alleles using RNA sequence. Then, the program computes two statistics; (1) Euclidian distance (d) and (2) Pearson correlation coefficient (r) for all sequence intervals (≥50 base) that forms internal base pairs to assess structural difference between WT and ALT allele and reports the interval with the maximum base pairing distance (d_{max}) or minimum correlation coefficient (r_{min}) along with the corresponding empirical *P* values (40). A SNP considered as structure disruptive target-SNPs whenever d_{max} p-value was <0.2 (significance threshold is P<0.2 as defined by RNAsnp).

4.2.5. Impact of target-SNPs on target sites accessibility. The accessibility of target sites in both WT and ALT form were defined as the average probability that each nucleotide in the corresponding sequence is unpaired (P_u). P_u values for nucleotides in target sites were computed with RNAplfold program (41) using -u option and a window size (W) of 80 and a maximum pairing distance (L) of 40, as used in prior studies (42). We considered the SNP-induced changes in structural accessibility as ΔPu (So $\Delta P\text{u} = P_{u\ (ALT)} - P_{u\ (WT)}$)

4.2.6. Regulatory annotations of target-SNPs. RegulomeDB (Version 1.1) (43) and Haploreg V2 (36) were employed to annotate target-SNPs with known and predicted regulatory elements of the genome including regions of DNAase hypersensitivity, binding sites and motifs of transcription factors, chromatin state and expression quantitative trait loci.

Acknowledgements

The Italian Ministry of Health is gratefully acknowledged for financial support to A.M. (Ricerca Corrente 2015).

References:

1. Saadatian Z, Masotti A, Fam ZNS, Alipoor B, Bastami M, Ghaedi H. Single-Nucleotide Polymorphisms Within MicroRNAs Sequences and Their 3'UTR Target Sites May Regulate Gene Expression in Gastrointestinal Tract Cancers. Iranian Red Crescent Medical Journal. 2014;16(7).

2. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis. 2010;31(1):27-36. Epub 2009/09/16.

3. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12(12):861-74. Epub 2011/11/19.

4. Lujambio A, Lowe SW. The microcosmos of cancer. Nature. 2012;482(7385):347-55. Epub 2012/02/18.

5. Kong YW, Ferland-McCollough D, Jackson TJ, Bushell M. microRNAs in cancer management. Lancet Oncol. 2012;13(6):e249-58. Epub 2012/06/02.

6. Fatica A, Fazi F. MicroRNA-regulated pathways in hematological malignancies: how to avoid cells playing out of tune. International journal of molecular sciences. 2013;14(10):20930-53.

7. Yendamuri S, Calin G. The role of microRNA in human leukemia: a review. Leukemia. 2009;23(7):1257-63.

8. Fazi F, Rosa A, Fatica A, Gelmetti V, De Marchis ML, Nervi C, et al. A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBPalpha regulates human granulopoiesis. Cell. 2005;123(5):819-31. Epub 2005/12/06.

9. Felli N, Fontana L, Pelosi E, Botta R, Bonci D, Facchiano F, et al. MicroRNAs 221 and 222 inhibit normal erythropoiesis and erythroleukemic cell growth via kit receptor down-modulation. Proc Natl Acad Sci U S A. 2005;102(50):18081-6. Epub 2005/12/07.

10. Khalaj M, Tavakkoli M, Stranahan AW, Park CY. Pathogenic microRNA's in myeloid malignancies. Front Genet. 2014;5:361. Epub 2014/12/06.

11. Zhao H, Wang D, Du W, Gu D, Yang R. MicroRNA and leukemia: tiny molecule, great function. Critical reviews in oncology/hematology. 2010;74(3):149-55.

12. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature. 2005;435(7043):834-8. Epub 2005/06/10.

13. Li Z, Lu J, Sun M, Mi S, Zhang H, Luo RT, et al. Distinct microRNA expression profiles in acute myeloid leukemia with common translocations. Proc Natl Acad Sci U S A. 2008;105(40):15535-40. Epub 2008/10/04.

14. Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE, et al. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. New England Journal of Medicine. 2005;353(17):1793-801.

15. Marcucci G, Mrózek K, Radmacher MD, Garzon R, Bloomfield CD. The prognostic and functional role of microRNAs in acute myeloid leukemia. Blood. 2011;117(4):1121-9.

16. San José-Enériz E, Román-Gómez J, Jiménez-Velasco A, Garate L, Martin V, Cordeu L, et al. MicroRNA expression profiling in Imatinib-resistant Chronic Myeloid Leukemia patients without clinically significant ABL1-mutations. Mol Cancer. 2009;8(1):69-72.

17. Melo SA, Esteller M. Dysregulation of microRNAs in cancer: playing with fire. FEBS letters. 2011;585(13):2087-99.

18. Engle L, Simpson C, Landers J. Using high-throughput SNP technologies to study cancer. Oncogene. 2006;25(11):1594-601.

19. Bernig T, Chanock SJ. Challenges of SNP genotyping and genetic variation: its future role in diagnosis and treatment of cancer. 2006.

20. Yu Z, Li Z, Jolicoeur N, Zhang L, Fortin Y, Wang E, et al. Aberrant allele frequencies of the SNPs located in microRNA target sites are potentially associated with human cancers. Nucleic acids research. 2007;35(13):4535-41.

21. Brodersen P, Voinnet O. Revisiting the principles of microRNA target recognition and mode of action. Nature reviews Molecular cell biology. 2009;10(2):141-8.

22. Landi D, Gemignani F, Naccarati A, Pardini B, Vodicka P, Vodickova L, et al. Polymorphisms within micro-RNA-binding sites and risk of sporadic colorectal cancer. Carcinogenesis. 2008;29(3):579-84.

23. Dzikiewicz-Krawczyk A. MicroRNA-binding site polymorphisms in hematological malignancies. Journal of Hematology & Oncology. 2014;7(1):1-6.

24. de Larrea CF, Navarro A, Tejero R, Tovar N, Díaz T, Cibeira MT, et al. Impact of MiRSNPs on survival and progression in patients with multiple myeloma undergoing autologous stem cell transplantation. Clinical Cancer Research. 2012;18(13):3697-704.

25. Dzikiewicz-Krawczyk A, Macieja A, Ma E, Januszkiewicz-Lewandowska D, Mosor M, Fichna M, et al. Polymorphisms in microRNA target sites modulate risk of lymphoblastic and myeloid leukemias and affect microRNA binding. Journal of Hematology & Oncology. 2014;7(1):43.

26. Hindorff LA, Junkins HA, Mehta J, Manolio T. A catalog of published genome-wide association studies. National Human Genome Research Institute. 2010.

27. Liu C, Rennie WA, Carmack CS, Kanoria S, Cheng J, Lu J, et al. Effects of genetic variations on microRNA: target interactions. Nucleic Acids Res. 2014;42(15):9543-52.

28. Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the implications for cancer research. Nature reviews Cancer. 2010;10(6):389-402. Epub 2010/05/25.

29. Young RP. Methods for the assesssment of risk of developing lung cancer using analysis of genetic polymorphisms. Google Patents; 2006.

30. Gao J, Pfeifer D, He L-J, Qiao F, Zhang Z, Arbman G, et al. Association of NFKBIA polymorphism with colorectal cancer risk and prognosis in Swedish and Chinese populations. Scandinavian journal of gastroenterology. 2007;42(3):345-50.

31. Pallavi S, Anoop K, Showket H, Alo N, Mausumi B. NFKB1/NFKBIa polymorphisms are associated with the progression of cervical carcinoma in HPV-infected postmenopausal women from rural area. Tumor Biology. 2015:1-12.

32. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic acids research. 2001;29(1):308-11. Epub 2000/01/11.

33. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic acids research. 2014;42(Database issue):D92-7. Epub 2013/12/04.

34. Bhattacharya A, Ziebarth JD, Cui Y. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic acids research. 2014;42(Database issue):D86-91. Epub 2013/10/29.

35. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014;42(Database issue):7.

36. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012;40(Database issue):7.

37. Blankenberg D, Johnson JE, Taylor J, Nekrutenko A. Wrangling Galaxy's reference data. Bioinformatics. 2014;30(13):1917-9.

38. Goecks J, Nekrutenko A, Taylor J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome biology. 2010;11(8):R86. Epub 2010/08/27.

39. Kruger J, Rehmsmeier M. RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res. 2006;34(Web Server issue):W451-4.

40. Sabarinathan R, Tafer H, Seemann SE, Hofacker IL, Stadler PF, Gorodkin J. RNAsnp: efficient detection of local RNA secondary structure changes induced by SNPs. Human mutation. 2013;34(4):546-56. Epub 2013/01/15.

41. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. Algorithms for molecular biology : AMB. 2011;6:26. Epub 2011/11/26.

42. Marin RM, Voellmy F, von Erlach T, Vanicek J. Analysis of the accessibility of CLIP bound sites reveals that nucleation of the miRNA:mRNA pairing occurs preferentially at the 3'-end of the seed match. RNA (New York, NY). 2012;18(10):1760-70. Epub 2012/08/24.

43. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 2012;22(9):1790-7.

Highlights

- Genetic variations in miRNAs are likely to disrupt miRNA–target interaction and are associated with the development of different hematological malignancies.

- Bioinformatics pipelines are useful to prioritize relevant SNPs in miRNA-targets associated with hematological malignancies.

- Prioritized SNPs in miRNAs and their targets can represent useful molecular determinants to investigate in order to unravel their crucial role in hematological malignancies.