In silico profiling of miRNAs and their target polymorphisms in leukemia associated genes

C. George Priya Doss*, B. Rajith, T. Samuel Jeba Seelan, R. Santhosh Kumar

Medical Biotechnology Division, School of Biosciences and Technology, VIT University, Vellore, 632014, Tamil Nadu, India

Received 3 January 2013; accepted 22 January 2013
Available online 20 February 2013

Abstract  Single Nucleotide Polymorphisms (SNPs) within microRNA (miRNA) encoding regions of the genome are a large potential source for biologically relevant variation. SNPs along with miRNA act as a powerful tool to study the biology of a disease and also have the potential in monitoring disease prognosis and diagnosis. Therefore, evaluating the functional role of target mRNA will be a major challenge of future studies in the field of cancer biomarker research in leukemia. To assess, whether miRNA target SNPs are implicated in leukemia associated genes, we conducted an in silico approach along with the availability of publicly available web based tools for miRNA prediction and comprehensive genomic databases of SNPs. In this in-depth report, we attempted to use two computational approaches: prediction of miRNA in leukemia associated genes, and identifying the functional role of mRNAs targeted by miRNA. Our results from this study suggest that the application of in silico algorithms miRdSNP, PupaSuite and UTRScan analyses might provide an alternative approach to select target untranslated region (UTR) SNPs and understand the effect of SNPs on the functional attributes or molecular phenotype of a protein.

1. Introduction

Leukemia is a type of cancer that can affect the bone marrow, blood cells, lymph nodes and other parts of the lymphatic system. This disease probably results from acquired mutations to the DNA of a single lymph or blood forming stem cell. The abnormal cells multiply and survive without the usual controls that are in place for healthy cells. The accumulation of these cells in the marrow, blood and/or lymphatic tissue interferes with production and functioning of red cells, white cells and platelets. The disease process can lead to severe anemia, bleeding, an impaired ability to fight infection, or death [1]. miRNAs are short, endogenous, non-coding RNA fragments of 21–25 nucleotides that post-transcriptionally repress protein translation or degrade the coding message by binding with imperfect complementarity to the 3' UTRs of target messenger RNAs (mRNAs) [2]. Recent reports from GWAS have shown that mutations in these regions are associated with several neurological disorders, muscular hypertrophy, gastric mucosal atrophy, cardiovascular disease and Type II diabetes [3]. Accumulation of the large amount of biological data suggests that a

* Corresponding author. Tel.: +91 9500040439.
E-mail addresses: georgecp77@yahoo.co.in (C. George Priya Doss), rajith.b@gmail.com (B. Rajith), jebaseelan.samuel@gmail.com (T. Samuel Jeba Seelan), rsanthosh003@gmail.com (R. Santhosh Kumar).
Peer review under responsibility of Ain Shams University.
single miRNA could bind to hundreds of mRNA targets, and these targets could be implicated in the regulation of almost every biological process [4]. Loss or amplification of miRNA genes has been reported in a variety of cancers such as lung, breast, and colorectal, and altered patterns of miRNA expression may affect cell cycle and survival programs [5]. The importance of miRNAs in gene regulation and their potential significance in both cancer biology and gene evolution suggests that mutations in miRNA target sites might well also be important in the etiology of human associated disease. Because small variation in the quantity of miRNAs may have an effect on thousands of target mRNAs and result in diverse functional consequences, the most common genetic variation such as SNPs, in mRNA sequences may also be functional and therefore, may represent ideal candidate biomarkers for cancer prognosis [6]. When SNPs occur in 3' UTR, they may interfere with mRNA stability and translation by altering polyadenylation, protein: mRNA, and miRNA:mRNA regulatory interactions. An increasing number of programs are available on the World Wide Web to predict miRNA-target mRNA sites; however, it is advisable to use multiple algorithms to confirm the predictions. The role of SNPs in the leukemia associated genes by GWAS has been studied extensively. There are queries on which SNPs within miRNAs that target associated genes can regulate leukemia. However, there is no such report to evaluate the role of SNPs in miRNA::mRNA gene regulation and its effect on leukemia associated genes. These motivate us to predict the impact of SNPs on candidate miRNAs and investigate the availability of these candidate miRNA regulators in leukemia by in silico approach. To answer this question, in the absence of other experimental investigations, we have presented an integrated computational method for the first time which allows the use of miRdSNP [7] PupaSuite [8] and UTRScan [9] as a pipeline for the prediction of miRNAs and their targets, and also we evaluated the functional role of targeted mRNAs. The proposed computational approach could scale up in two ways. As a first step, we used miRdSNP to scan the whole genome to find possible SNPs located at experimentally proved miRNAs. In the second step, we applied PupaSuite and UTRScan to evaluate the functional role of target mRNA SNPs. This analysis will definitely hold a promise in future for studying the prognosis and diagnosis of leukemia, as well as pharmacogenomics, molecular epidemiology and for individualized medicine.

2. Materials and methods

2.1. Gene and SNP dataset considered for our evaluation

A list of 167 leukemia associated genes was retrieved from the Internet resource, Leukemia Gene Database (LeGenD), available at (http://www.bioinformatics.org/legend/leuk_db.htm). Ten genes namely CCND1, EVI1, NUP214, PDGFRB, PER1, PICALM, PBX1, PTPN11, RPL22 and TAL2 were involved in two different forms of leukemia. The Atlas of Genetics and Cytogenetics in Oncology and Haematology was used to validate the involvement of gene in mutation; where this gene is implicated in diseases, with the prognosis and data on oncogenesis; a selected bibliography with hyperlinks to MEDLINE abstracts. The annotated sequence is required to determine the functional class (e.g., intron, exon, exonic UTR) of the SNP so that appropriate in silico tools can be selected for analyses. Functional class information regarding the SNPs was obtained from NCBI. In this work as a first step, we used miRdSNP which incorporates a set of computational tools like miRTarBase, TarBase, miR2desease and miRecords to identify experimentally validated miRNAs and their target miRNAs in leukemia associated genes.

3. Analyzing miRNAs and their target sites

In our study, experimentally confirmed miRNA target genes in leukemia were considered. Systematic analysis of miRNAs targeting leukemia associated genes was performed based on three curated target prediction database; miRdSNP [7]. This provides a comprehensive data source for exploring the effect of SNPs on miRNA binding in relation to human diseases. Gene names under Homo sapiens were chosen as search criteria provided in graphical user interface of the database search. miRdSNP incorporates data from four curated databases TarBase, miRTarBase, miRecords and miR2disease which collect experimentally confirmed miRNA target interactions in order to address low prediction specificity.

4. PupaSuite

PupaSuite [8] are now synchronized to deliver annotations for both noncoding and coding SNPs, as well as annotations for the Swiss-Prot set of human disease mutations. In this approach, we used the input option as list of genes and specified the type of gene identifiers by selecting either Ensembl or an external database (which includes GenBank, Swissprot/TremBL and other gene IDs supported by Ensembl). PupaSuite finds all the SNPs mapping in locations that might cause a loss of functionality in the genes. PapasView retrieves SNPs that could affect conserved regions that the cellular machinery uses for the correct processing of genes (intron/exon boundaries or exonic splicing enhancers) and miRNAs. It uses either sequence-based or structural bioinformatics tools additional methods for predicting SNPs in TFBSs and splice sites.

5. Scanning of UTR SNPs in UTR site

5' and 3' untranslated regions (UTR) of eukaryotic mRNAs are involved in many post transcriptional regulatory pathways that control mRNA localization, stability and translation efficiency [9]. We used the program UTRScan for this analysis. UTRScan looks for UTR functional elements by searching through user-submitted query sequences for the patterns defined in the UTRsite collection. UTRsite is a collection of functional sequence patterns located in 5' or 3' UTR sequences. Briefly, two or three sequences of each UTR SNP that have a different nucleotide at an SNP position are analyzed by UTRScan, which looks for UTR functional elements by searching through user-submitted sequence data for the patterns defined in the UTRsite and UTR databases. If different sequences for each UTR SNP are found to have different functional patterns, this UTR SNP is predicted to have functional significance. The internet resources for UTR analysis were UTRdb and UTRsite. UTRdb contains experimentally proven biological activity of functional patterns of UTR se-
sequence from eukaryotic mRNAs. The UTRsite has the data collected from UTRdb and also is continuously enriched with new functional patterns.

6. Results

6.1. miRNA and their target mRNA

By comparing the results of all the three miRNA target databases, out of 167 only 23 (14%) genes namely ABL1, CBFB, CDX2, CHIC2, EP300, HOXC13, KIT, LASP1, NCOA2, PI, CALM, RUNXI, ARNT, CCND1, CDK6, JAK2, LMO2, MLL, NOTCH1, BCL2, BCL6, CCND2, CAV1 and ETS1 displayed miRNAs and their targets (Table 1). These genes interacted with several hundreds of mRNA targets via perfect or imperfect base-pairing primarily in the 3' UTR. We identified experimentally verified 56 miRNAs associated with 2962 target mRNAs that can potentially create, destroy, or modify miRNA binding sites by our computational approach as depicted in Supplementary Table 1[10–43].

7. Polymorphism in miRNA target sites

To validate the computational predictions and the biological relevance of target SNPs, we retrieved the SNP related information from dbSNP for all the 23 genes predicted by miRTarBase, TarBase and miRecords. As a next step, we submitted the SNP information in PupaSuite and UTRScan to identify the functional significance of the targeted mRNAs. By PupaSuite, 505 SNPs in mRNAs were predicted to disrupt exon splicing enhancers and 57 were predicted to disrupt exon splicing silencer. Notably, 35 SNPs were predicted to disrupt both exon splicing enhancer and exon splicing silencer. Next, putative UTR functional elements potentially affected by cancer-associated SNPs were searched for using UTRScan. As shown in Supplementary Table 1, a total of six regulatory elements out of the 31 included in the UTRsite database were located near or at cancer-associated SNP sites. Based on the UTRScan analysis, 35 SNPs showed a functional pattern change of 15-LOX-DICE, 287 SNPs showed a functional pattern change of IRES, 8 SNPs showed a functional pattern change of GYBox, 62 SNPs showed a functional pattern change of K-Box, 22 SNPs showed a functional pattern change of Brd-Box and 5 SNPs showed a functional pattern change of TOP, respectively.

8. Discussion

Founding members of miRNAs were discovered by genetic screening approaches, experimental approaches were limited by their low efficiency, time consuming, and high cost. As a consequence, several web-based or non web-based computer software programs are publicly available for predicting miRNAs and their targets have been devised in order to predict targets for follow up experimental validation. Recent work by Machová et al. [44] has identified miRNAs associated with CML pathogenesis by in silico approach based on integrated miRNA expression profiling. Similarly, Chan et al. [45] performed an in silico study applying genomic sequence analysis for the prediction of miRNAs that target EGFR in lung cancer. Even though many computational methods for the identification of miRNA may have their own limitations, there is no other option now other than to use computational methods for miRNA predictions. The next step in miRNA research is to identify and experimentally validate their mRNA targets. Since direct experimental methods for discovering miRNA targets are lacking, a large number of target prediction algorithms

<table>
<thead>
<tr>
<th>Table 1</th>
<th>List of leukemia associated genes taken for our analysis.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute myeloid leukemia</td>
<td>ABL2, AF10, AFI5Q14, ARHGEF12, CBFA2T2, CBFA2T3, CBFB, CBL, CDX2, CEBPA, CHIC2, CREBBP, DDX10, DEK, ELF4, EP300, ERG, EVI1, FAC1, FLT3, FBNP1, FUS, GAST, GMPS, GRAP, HEAB, HLXB9, HOX9, HOX13, HOX13, HOXD11, IRF1, KIT, LASP1, LCX, LCP1, MDSI, MLI, MLI, MLI, MLI, MYH1, MYH1, MYST4, NCOA2, NPM1, NSD1, NU1P98, NUP214, PDGFRB, PERL1, PICALM, PMX1, PNUTL1, PRDM16, PSIP2, PTPN11, RGS2, RPL22, RPN1, RUNX1, RUNXBP2, SEPT6, SET, SHHSP1, TOP1, TRIP11</td>
</tr>
<tr>
<td>Acute myeloblastic leukemia</td>
<td>ARNT</td>
</tr>
<tr>
<td>Acute lymphoid leukemia</td>
<td>AF1Q, AF3p21, AF5q31, BCL9, BCL11B, BCR, CCND1, CDK6, ELL, EPS15, FBXW7, FCGR2B, FLT3, FOXO3A, GPHN, HLF, JAK2, LAF4, LCK, LMO1, LMO2, LYL1, MLL, MLL1, MLL2, MLL3, MLL4, MLL6, MLL7, MLL7, MLLT1, MLLT2, MLLT3, MLLT4, MLLT6, MLLT7, NOTCH1, NU1P24, PICALM, RANBP17, RAPID5S1, SHGGLI, SIL, TAF15, TAL2, TCF3, TCL6, TPTT, TLL, TXL1, TLX3, ZNF385, ZNF11A</td>
</tr>
<tr>
<td>Chronic myeloid leukemia</td>
<td>ABL1, AXL, BCR, D10S170, EVI1, HIP1, HOX11, MSI2, PDGFRB, PER1, RABSEP, RPL22</td>
</tr>
<tr>
<td>Chronic lymphoid leukemia</td>
<td>BCL2, BCL3, BCL5, BCL6, BTG1, CCND1, CCND2, DLEU1, DLEU2, FSTL3, IGH@, TCL1A</td>
</tr>
<tr>
<td>B-cell acute lymphocytic leukemia</td>
<td>PBX1, STL</td>
</tr>
<tr>
<td>Juvenile myelomonocytic leukemia</td>
<td>HCMOGT1, PTPN11</td>
</tr>
<tr>
<td>T-cell acute lymphoblastic leukemia</td>
<td>OLG2, TAL2, TR@, TRB@, HOX11, BAX</td>
</tr>
<tr>
<td>T-cell prolymphocytic leukemia</td>
<td>ATM</td>
</tr>
<tr>
<td>Murine leukemia</td>
<td>EVI2A, EVI2B</td>
</tr>
<tr>
<td>Myeloid leukemia</td>
<td>CDC25, CLC</td>
</tr>
<tr>
<td>Pre B-cell leukemia</td>
<td>PBX1, PBX2, PBX3</td>
</tr>
<tr>
<td>T-cell leukemia</td>
<td>TCL1B, MTC1P1, LDB1</td>
</tr>
<tr>
<td>Human monocytic leukemia</td>
<td>ETS1</td>
</tr>
<tr>
<td>T-cell leukemia</td>
<td>CA1</td>
</tr>
<tr>
<td>Mixed linkage leukemia</td>
<td>ML3</td>
</tr>
<tr>
<td>Acute promyelocytic leukemia</td>
<td>NUMAI, PML, RARA, THRA, TIF1, ZNF145</td>
</tr>
</tbody>
</table>
have been developed. In this study, publicly available SNP data have been analyzed in context with miRNAs and their target sites. This allows the use of miRdSNP, PupaSuite and UTRScan as an integrated pipeline for the prediction of miRNAs and their targets, and also evaluated the functional role of targeted miRNAs. By our analysis we found experimentally verified 56 miRNAs associated with 2962 target miRNAs. 562 SNPs were found to be functionally significant by PupaSuite and 419 SNPs by UTRScan. In silico methods provide a useful tool for an initial approach to any mutation suspected of causing aberrant RNA processing. These mutations can result either in complete skipping of the exon, retention of the intron or in the introduction of a new splice site within an exon or intron. Nonsense and missense mutations can disrupt exonic splicing enhancers (ESEs) and cause the splicing machinery to skip the mutant exon, with dramatic effects on the structure of the gene product [46]. ESEs are common in alternative and constitutive exons, where they act as binding sites for Ser/Arg-rich proteins (SR proteins), a family of conserved splicing factors that participate in multiple steps of the splicing pathway [47]. Recent studies show that regulatory mutations could make a significant contribution to genetic variation, including disease susceptibility leading to the identification of mutations in regulatory variants (rSNPs) affecting transcript levels in cis [48]. The most common interpretation of such cis effects is that the corresponding variants are modulating the activity of regulatory elements, including promoters and enhancers. Recent studies showed that mutations in cis splicing regulating sequences will lead to pathogenicity with increased cancer-prone potential [49]. Several studies have related the occurrence of SNPs in human 3’ UTRs that may modulate the expression of computationally predicted miRNA target sites [50]. Our results from this study suggest that the application of computational algorithms, PupaSuite and UTRScan analyses might provide an alternative approach to select target SNPs by understanding the effect of SNPs on the functional attributes or molecular phenotype of a protein. In the absence of experimental evidence, the potential functional consequences of a SNP can be predicted using various bioinformatics tools. However, these in silico tools are often not familiar to epidemiologists. There is a need for a simple guide for non-bioinformatics researchers to obtain this information, particularly when genomic data are continually updated and revised as new discoveries are made. The computational protocol proposed in this study is based on integrating relevant biomedical information sources to provide a systematic analysis of complex disease miRNA identification and target prediction.

Author’s contributions

RB, SJST, and SKR were involved in design, acquisition of data, analysis and interpretation of the data. CGPD and RB were involved in interpretation of the data and drafting the manuscript.

CGPD supervised the entire study and was involved in design, acquisition of data, analysis and Interpretation of the data and drafting the manuscript.

Acknowledgments

The authors take this opportunity to thank the management of VIT University for providing the facilities and encouragement to carry out this work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ejmhg.2013.01.004.

References

In silico profiling of miRNAs and their target polymorphisms in leukemia associated genes


