Research Article

Cancer Epidemiology, Biomarkers & Prevention

Targetome Profiling, Pathway Analysis and Genetic Association Study Implicate miR-202 in Lymphomagenesis

Aaron E. Hoffman[1], Ran Liu[2,3], Alan Fu[3], Tongzhang Zheng[3], Frank Slack[4], and Yong Zhu[3]

Abstract

Background: miRNAs have been implicated in numerous tumorigenic pathways, and previous studies have associated miR-202 dysregulation with various cancer types, including follicular lymphoma.

Methods: The miR-202 targetome was identified by ribonucleoprotein immunoprecipitation-microarray (RIP-Chip), and functional interactions among identified targets were investigated using the Ingenuity Pathway Analysis tool. We also conducted a population-based genetic association study of a polymorphism within the miR-202 stem-loop sequence and risk of non-Hodgkin lymphoma. *In vitro* gain-of-function experiments were further conducted to elucidate the functional significance of the variant.

Results: A total of 141 potential members of the miR-202 targetome were identified by a transcriptome-wide RIP-Chip assay. Functional interactions among identified targets suggested that miR-202-regulated genes are involved in biologic pathways relevant for hematologic function and cancer. Consistent with this, a genetic association analysis using human blood samples revealed a significant association between a germline mutation (rs12355840) in the miR-202 precursor sequence and follicular lymphoma risk. An *in vitro* functional assay further showed that the variant allele resulted in diminished miR-202 levels, possibly by altering precursor-processing efficiency.

Conclusions: Taken together, our findings suggest that miR-202 is involved in follicular lymphomagenesis.

Impact: These findings implicate miR-202 as a potential tumor suppressor in follicular lymphoma and warrant the investigation of miR-202 as a novel biomarker of follicular lymphoma risk. *Cancer Epidemiol Biomarkers Prev; 22(3); 327–36. ©2013 AACR.*

Introduction

The 1993 discovery of *lin*-4 in *C. elegans* brought about a fundamental change in the understanding of how gene activity is regulated, helping to uncover an additional element of the central dogma of molecular biology. The 22-nucleotide *lin-4* transcript, with its post-transcriptional gene regulatory capacity, became the first identified member in what would later be the miRNA family of small nonprotein-coding RNAs (1). Since this initial discovery, more than 1,900 mature miRNAs (miRBase 18) have been experimentally verified in humans, regulating the activity of perhaps more than a third of all human genes (2, 3).

The role of miRNA in tumorigenesis has been extensively studied in recent years. For example, miRNAs are commonly located within fragile sites on chromosomes (4) and act as both oncogenes and tumor suppressors (5, 6). Further findings have shown that a global reduction in miRNA processing promotes carcinogenesis, and miRNA profile data have been successfully applied to classify tumors and predict prognosis in several cancer types (7–10). In addition, epidemiologic and experimental findings have implicated a role for miRNAs in virtually every human cancer. Previously, our group conducted a survey of miRNA-related single-nucleotide polymorphisms (SNP), followed by a genetic association study and functional analysis in breast cancer (11). We identified miR-196a-2 as a potential breast cancer-relevant miRNA, which is consistent with later findings implicating its tumorigenic relevance in other cancer types (12–14). Although a miR-202-associated SNP was identified in our survey, it was not associated with breast cancer risk in our population (11). However, prediction algorithms such as TargetScan (15) suggested that miR-202 may influence genes involved in hematologic function, such as *IL10* and *ABL2* (TargetScan release 5.2; June 2011). In addition, altered expression of miR-202 has been recently shown

Authors' Affiliations: [1]Department of Epidemiology, Tulane School of Public Health and Tropical Medicine and Tulane Cancer Center, New Orleans, Louisiana; [2]Key Laboratory of Environmental Medicine Engineering, School of Public Health, Southeast University, Nanjing, China; [3]Yale School of Public Health, Yale School of Medicine; and [4]Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

Corresponding Author: Yong Zhu, Yale School of Public Health, Yale School of Medicine, New Haven, CT 06520. Phone: 203-785-4844; Fax: 203-737-6023; E-mail: yong.zhu@yale.edu

doi: 10.1158/1055-9965.EPI-12-1131-T

in several cancer types such as breast (9), cervical (16), colorectal (17), and gastric tumors (18). Interestingly, a recent miRNA profiling experiment comparing follicular lymphoma tumor cells to follicular hyperplasia cells identified a 44-miRNA signature of follicular lymphoma, which included miR-202 (19).

Like many miRNAs, miR-202 is located within a chromosomal fragile site, in the subtelomeric region of chromosome 10, the deletion of which has been associated with endometrial (20) and brain tumors (21), as well as developmental issues in children, including hyperactivity (22) and speech delay (23). A recent study also found that miR-202 may negatively regulate the expression of the proto-oncogene, *MYCN*, resulting in the inhibition of neuroblastoma cell proliferation (24).

In this study, we experimentally identified the miR-202 targetome using a ribonucleoprotein immunoprecipitation-microarray (RIP-Chip)-based approach. Functional interactions between identified targets were then investigated for disease-relevance using the Ingenuity Pathway Analysis (IPA) tool. We also conducted a population-based genetic association analysis of a polymorphism within the miR-202 stem-loop sequence and risk of non-Hodgkin lymphoma (NHL). *In vitro* gain-of-function experiments were further conducted to elucidate the functional significance of the variant.

Materials and Methods

Cell culture and miRNA transfection

HeLa cells were cultured in (+)-glutamine RPMI-1640 supplemented with 10% heat-inactivated FBS and 1% penicillin/streptomycin (Invitrogen). Oligonucleotide transfection efficiencies were determined by fluorescence microscopy using the Cy3 DS Transfection Control (Integrated DNA Technologies, Inc), and cells were transfected with either a miR-202 mimic or a scrambled negative control (Qiagen) using the Lipofectamine RNAiMAX transfection reagent (Invitrogen), according to the manufacturer's protocol. Briefly, 120 pmol of miR-202 mimic or negative control was mixed with 20 μL of RNAiMAX reagent in 2 mL of OPTI-MEM (Invitrogen). The complex was then added into a 100 mm dish and incubated for 20 minutes at room temperature. Approximately 4 million cells were then seeded into the dish to a total volume of 10 mL and incubated for 12 hours at 37°C before harvesting.

RIP-Chip analysis for miR-202 target identification

Ribonucleoprotein (RNP) immunoprecipitation was conducted using the RIP-Assay kit for miRNA (MBL) according to the protocol described by the manufacturer. Briefly, anti-EIF2C2/Ago2 monoclonal antibody (Novus Biologicals, LLC) was incubated with Protein G plus agarose beads (Pierce) at 4°C overnight to prepare antibody-immobilized beads. 20 million cells were harvested and washed four times with ice-cold diethyl pyrocarbonate-treated PBS. The cell pellet was lysed with 500 μL of lysis buffer and the supernatant was incubated with Protein G agarose beads without antibody to reduce nonspecific adsorption. The cell lysate was then transferred into a tube containing antibody-immobilized Protein G agarose beads and incubated for 3 hours at 4°C. The complex was washed five times with wash buffer before total RNA isolation. Mouse IgG1 isotype control (Novus Biologicals, LLC) was used as a negative control for the immunoprecipitation procedure. To confirm proper functioning of the RIP assay, an aliquot of precleared cell lysate, postimmunoprecipitation beads, and the flow-through fraction of antibody-immobilized Protein G agarose beads–RNP complexes were used for analysis of Ago2 quantity by Western blot analysis. For use in the Western blot analysis, mouse anti-EIF2C2/Ago2 monoclonal antibody was diluted 1:500 and chicken anti-mouse immunoglobulin G (IgG) conjugated with horseradish peroxidase was diluted 1:2000. SuperSignal West Pico Chemiluminescent Substrate (ThermoScientific) was used for visualization.

Genome-wide expression levels were analyzed in total RNA samples from total cell lysate and antibody-immobilized Protein G agarose beads-RNP complexes from cells transfected with miR-202 mimic or negative control using the Agilent 44K 60-mer human whole-genome microarray. Signal hybridization and scans were conducted by MOGene, LC. The normalized signal intensities of probes from RNP-bead complexes (immunoprecipitation fraction) were divided by the signal intensities from total cell lysate (preimmunoprecipitation fraction) in miR-202 mimic-transfected cells. Transcripts were identified as members of the global miRNA targetome if they exhibited an enrichment fold change $\geq$2.0. From this list, immunoprecipitation to preimmunoprecipitation signal intensity ratios in miR-202 mimic-transfected cells were divided by immunoprecipitation to preimmunoprecipitation signal intensity ratios in negative control cells. Transcripts exhibiting normalized signal ratios of $\geq$1.5 were considered to be bound with miR-202 in the RNA-induced silencing complex (RISC), and therefore potential direct miR-202 targets. All microarray data were uploaded to the Gene Expression Omnibus database (accession # GSE42981).

Target screening

Enriched transcripts were screened for potential 3′untranslated region (UTR) miR-202 binding sites using four miRNA target prediction algorithms: miRanda (25), TargetScan/TargetScanS (26), PITA (27), and RNAhybrid (28). Prospective binding sites were predicted on the basis of the number of base pair matches to the 5′ miRNA seed region, the degree of compensatory 3′ nonseed matches, and the number and nature of mismatched pairs (miRanda, TargetScan/TargetScanS, and PITA), as well as thermodynamic stability (miRanda and RNAhybrid).

Luciferase reporter assay

The 3′UTRs of two randomly selected target genes that have been cloned into pLightSwitch_3UTR GoClone

vectors were purchased from Switchgear Genomics to confirm direct target binding. HeLa cells were seeded, in triplicate, onto a 96-well plate with a concentration of 10,000 cells/well in 100 μL of RPMI-1640 medium with 10% FBS and incubated for 24 hours. Cells were then cotransfected with a miR-202 mimic or a negative control and the GoClone vector (100 ng vector mixed with 1.5 pmol of miR-202 mimic or negative control in 10 μL of OPTI-MEM + 0.25 μL Lipofectamine 2000 transfection reagent (Invitrogen) in 10 μL of OPTI-MEM, followed by 20 minutes of incubation). A total of 20 μL of transfection mixture and 80 μL of antibiotics-free RPMI-1640 media were added into the 96-well plate and incubated at 37°C and 5% CO_2 for 24 hours. The LightSwitch Assay System (Switchgear Genomics) was used for luciferase signal detection. After a 30-minute incubation period, the signal from each well was read in a GLOMAX 96 Microplate Luminometer (Promega), and a ratio of luciferase signal intensities (negative control:miR-202) was calculated for each construct.

Network analysis

miR-202 targets identified by RIP-Chip were investigated for functional interrelatedness using the IPA software tool (Ingenuity Systems). This tool scans the set of input genes to identify networks using information in the Ingenuity Pathways Knowledge Base, a manually curated database of functional interactions previously identified in peer-reviewed publications (29). A Fisher's exact test based on the hypergeometric distribution was then done for each identified network to determine the likelihood of obtaining at least the same number of interrelated molecules by chance as actually exists in the input gene set.

NHL study population

The study population, which consists of female residents of Connecticut, has been described in detail elsewhere (30). Briefly, 455 cases were identified through Yale Cancer Center's Rapid Case Ascertainment between 1996 and 2000 as having incident, histologically confirmed NHL (ICD-O, M-9590-9642, 9690-9701, 9740-9750). A total of 527 population-based controls were identified either through random digit dialing (RDD; for those younger than 65 years) or through Health Care Financing Administration files (for controls 65 years or older). Controls were frequency matched to cases on age (±5 years) by occasionally adjusting the number of controls selected from each 5-year age strata. Participation rates were 72% for cases, 69% for RDD controls, and 47% for controls identified by health care financing records. Tumor specimens were classified according to the Revised European-American Lymphoma (REAL) system using the original pathology slides obtained for each patient. Further information on population characteristics, including demographic data and case pathology distributions, are presented in Supplementary Table S1. All participation was voluntary, and the study was approved by the Institutional Review Boards at Yale University, the Connecticut Department of Public Health, and the National Cancer Institute (Bethesda, MD). After obtaining written informed consent, those who agreed to participate were interviewed by trained study personnel either at the subject's home or at a mutually agreed upon location. All participants completed a questionnaire containing information on demographic characteristics, family history of cancer, past medical conditions and medication use, occupation, and lifestyle factors such as diet and smoking status. Following the interview, each participant provided a 10 mL peripheral blood sample from which genomic DNA was isolated.

Genotyping of SNP rs12355840 in the stem-loop of miR-202

The SNP (rs12355840) was identified using the miR-202 sequence and location data in the miRBase database (http://www.mirbase.org/) aligned with the SNP locations compiled in the HapMap database (HapMap genome browser release 22; http://hapmap.ncbi.nlm.nih.gov). Genotyping was conducted at Yale University's W.M. Keck Foundation Biotechnology Research Laboratory using the Sequenom MassARRAY multiplex genotyping platform (Sequenom, Inc.) according to the manufacturer's protocol. Duplicate samples from 100 study subjects and 40 replicate samples from each of two blood donors were interspersed throughout each batch, and the concordance rates for quality control samples was >99%. All genotyping scores, including quality control data, were rechecked by different laboratory personnel and the assay accuracy was confirmed. The secondary structure of miR-202 was predicted using the Vienna RNAfold prediction algorithm (31).

Statistical analysis of genetic association

All statistical analyses were conducted using the SAS statistical software, version 9.1 (SAS Institute), unless otherwise noted. For the case–control analyses, the allelic distribution of the SNP was tested in control subjects by goodness-of-fit χ^2-square for compliance with Hardy–Weinberg equilibrium, and no departure from equilibrium was detected ($P = 0.930$). OR and 95% confidence intervals (CI) were determined by unconditional multivariate logistic regression, including the following covariates: age, race, and family history of cancer in a first-degree relative.

Construction of the precursor miR-202 expression vector

Pre-miR-202-G and pre-miR-202-A constructs, consisting of the precursor miRNAs and a flanking region (299 bp on the left and 269 bp on the right), were synthesized using genomic DNA from individuals known to be homozygous for either the G or A allele of SNP rs12355840.

Each construct was cloned into a pMIRNA1 lentivector containing a GFP marker (System Biosciences). Vectors were then sequenced to verify the accuracy of the inserts and subsequently transfected into HeLa cells using the Lipofectamine 2000 transfection reagent (Invitrogen). Transfection efficiency was verified by fluorescence microscopy.

RNA isolation and miR-202 detection

Total RNA from precursor miR-202-G- and miR-202-A-transfected HeLa cells were isolated using the miRNeasy Mini Kit (QIAGEN) with on-column DNA digestion. Sequences for the forward and reverse primers used to detect miR-202 precursors were 5′-CCTCCCAGGCTCACGAGGCT-3′ and 5′-GGTGCAGGTGCACTGGTGCA-3′, respectively. Relative RNA abundance with respect to untransfected HeLa cells was assessed using the $2^{-\Delta\Delta C_t}$ method with RNA content normalized to the housekeeping gene *HPRT1*. To determine levels of mature miRNA, polyadenylated mature miRNA sequences were first generated and converted to cDNA using the NCode miRNA First-Strand cDNA Synthesis Kit (Invitrogen). The cDNA was then amplified using a custom miR-202-specific forward primer (5′-AGAGGTATAGGGCATGGGAA-3′) and a universal reverse primer targeting on the polyadenylated region of the miRNA. Mature miRNA levels relative to untransfected HeLa cells were assessed using the $2^{-\Delta\Delta C_t}$ method with normalization to miR-16. To more accurately quantify the effect of the A allele on mature miR-202 expression, mature miR-202 levels were additionally normalized to levels of precursor miR-202 using the $2^{-\Delta\Delta Ct}$ method, where $\Delta\Delta C_t = \left(C_{t_{pre-mir-202}} - C_{t_{HPRT1}}\right) - \left(C_{t_{miR-202}} - C_{t_{miR-16}}\right)$. All quantitative PCR reactions were conducted in triplicate on an ABI 7500 Fast Real-Time PCR instrument (Applied Biosystems) using the QuantiFast SYBR Green PCR Kit (QIAGEN).

Luciferase target confirmation in Farage cells

Confirmation of miR-202 targets in the Farage lymphoma cell line was conducted using the 2 luciferase-3′UTR constructs previously used for target confirmation in HeLa cells. Farage cells were transferred to a 96-well plate at a concentration of approximately 25,000 cells/well in 100 μL (+)-glutamine RPMI-1640 medium with 10% FBS. For each well, cells were treated with 100 ng of the GoClone vectors premixed with 0.6 μL of Lipofectamine LTX and 0.2 μL of PLUS Reagent (Invitrogen) stock solutions in OPTI-MEM followed by a 24-hour incubation period at 37ºC and 5% CO_2. After 24 hours, cells were treated in duplicate with 20 pmol of either the miScript Inhibitor Negative Control or the miScript miRNA Inhibitor Anti-hsa-miR-202 (QIAGEN), both premixed with 0.2 μL RNAiMAX transfection reagent (Invitrogen) in OPTI-MEM. Following treatment, cells were incubated for an additional 24 hours before evaluation in the luciferase assay, as described above, and an anti-miR-202 inhibitor: negative control luciferase signal intensity ratio was calculated for each gene.

Results

miR-202 targetome identification by RIP-Chip assay

Because miRNA function is mediated by argonaute (Ago) proteins in the RISC, coimmunoprecipitation of Ago with miRNA/mRNA complexes can be used to identify probable direct miRNA targets. First, we used an anti-Ago2 antibody to isolate global miRNA targets from HeLa cells (low expression of miR-202), which we identified using a genome-wide comparative hybridization microarray. Transcripts which were significantly enriched in the Ago2 antibody-treated fraction relative to total cell lysate ($\geq$2.0 $\times$ relative abundance) in cells treated with a miR-202 mimic were considered potential miRNA targets ($N = 857$). These transcripts were then compared in miR-202 mimic-treated cells and cells treated with a scrambled double-stranded negative control to enrich for direct miR-202 targets. As a previous quantitative reverse transcription (qRT)-PCR assay revealed a low endogenous level of the mature miR-202 species in HeLa cells (data not shown), the introduction of miR-202 mimics allowed us to investigate miR-202 target-binding with minimal loss of assay sensitivity resulting from background miR-202 expression. Transfection efficiencies and Ago2 Western blot analysis results showing significant Ago2 enrichment are presented in Fig. 1A and B, respectively. Transcripts which were enriched by 50% or greater in miR-202 mimic-treated cells relative to the negative control were considered potential direct miR-202 targets. In all, 141 unique transcripts from 216 probes were identified as potential members of the miR-202 targetome (Supplementary Table S2).

Validation of miR-202 target genes by luciferase reporter assay

To confirm that the RIP-Chip approach provides valid candidates for the miR-202 targetome, we selected, at random, *HAS2* and *FAM135A*, for further validation using a 3′UTR luciferase reporter assay. Figure 1C shows the luciferase signal results from cotransfections of 3′UTR constructs of each gene with the miR-202 mimic or negative control. The expression of both luciferase 3′UTR constructs were significantly reduced when cotransfected with the miR-202 mimic, with a greater reduction seen for *HAS2* (6.1-fold and 3.0-fold for *HAS2* and *FAM135A*, respectively). On the basis of a bioinformatics search of both genes using the TargetScan miRNA target prediction tool (Release 5.2), *HAS2* harbors an 8-mer site matched with the complementary sequence of the miR-202 seed region, whereas *FAM135A* carries a 7-mer site for the miR-202 seed sequence.

Interrogation of potential miR-202 targets using miRNA target prediction algorithms

137 of 141 potential targets were screened for prospective 3′UTR miR-202 binding sites. A total of 4 transcripts

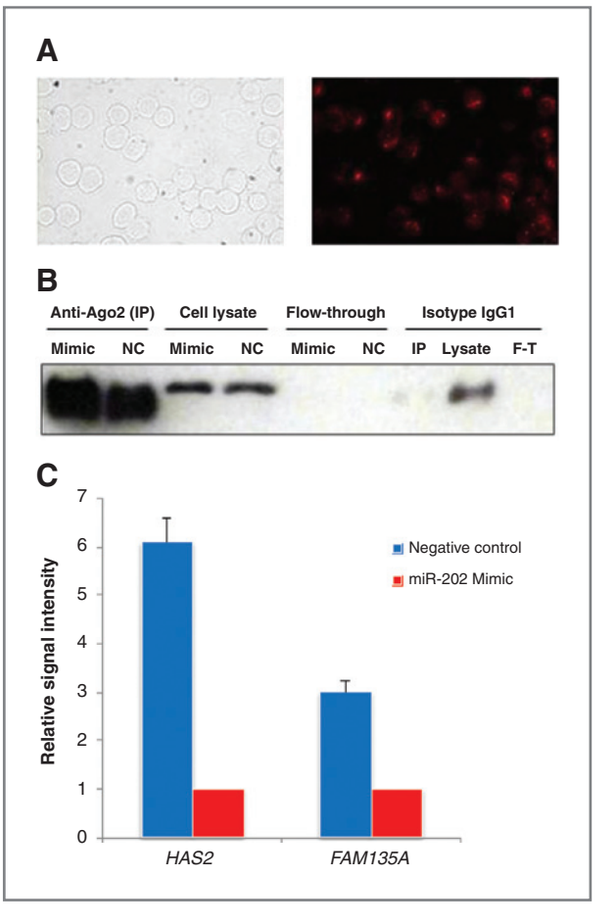

Figure 1. miR-202 Targetome identification using the RIP-Chip assay and validation with a luciferase reporter assay. A, transfection efficiencies were determined using a Cy3 DS transfection control. HeLa cells transfected with CY3 DS were screened using light microscopy (left panel) and fluorescence microscopy with a TRITC filter (right panel). The same transfection conditions were then used for miR-202 mimic and negative control (NC). B, an aliquot of postimmunoprecipitation beads, precleared cell lysate, and the flow-through fraction of the antibody-immobilized Protein G agarose beads-RNP complex from miR-202 mimic or negative control treated cells, and the equivalent from the mouse isotype IgG control were used in the analysis of Ago2 levels. C, a luciferase reporter assay for the 3′UTR of the RIP-Chip-identified genes *HAS2* and *FAM135A* was conducted to validate these genes as direct miR-202 targets. Transfection of miR-202 mimic results in decreased signal intensity values relative to negative control.

(BBIP1, LOC100389602, LOC100506305, and LOC387895) could not be recognized by one or more databases and were thus excluded from the analysis. 92 of 137 (69%) screened transcripts were identified as having miR-202-binding sites by 2 or more prediction algorithms based on the degree of base pairing complementarity to the seed region, the number of compensatory 3′ non-seed matches, and the thermodynamic stability of the predicted miR-202-mRNA hybrid. Figure 2 shows the distribution of positive target predictions among the 137 screened transcripts. The number of positive predictions per transcript is listed in Supplementary Table S3.

Lymphoma-related interaction network formed by identified miR-202 targets

The 141 identified transcripts comprising the miR-202 targetome were investigated for functional interrelatedness using the IPA software. The top interaction network identified using this method was designated by IPA as having relevance for "Hematological System Development and Function" ($P = 1.0E^{-46}$; Fig. 3) and the top disease associated with the transcripts was cancer ($P = 1.23E^{-4}$–$4.98E^{-2}$). Functional descriptors assigned by the Ingenuity Knowledge Database to the molecules in the top network included: "hematologic neoplasia," "differentiation of lymphocytes," "proliferation of hematopoietic cells," and "lymphomagenesis." Of the 35 genes in the top network, 22 (*B2M*, *CCL7*, *CRY1*, *ERCC8*, *FABP5*, *FUCA1*, *HAS2*, *HMGA2*, *HSPA14*, *IRF9*, *NKIRAS1*, *NLRP2*, *PFN2*, *RHEBL1*, *RIPK2*, *RPA2*, *RPS6KA5*, *SACS*, *SDC4*, *SLC5A5*, *SOCS1*, and *ZNF675*) were identified by our RIP-Chip assay as potential direct targets of miR-202. The full list of genes contained in the network, as well as notes on their relevance for lymphomagenesis, can be found in Supplementary Table S4.

Population-based genetic association

We further investigated miR-202's potential relevance in lymphomagenesis using human samples. A previous *in silico* analysis identified a SNP (rs12355840) located within the precursor region of miR-202 (11), which we genotyped in a population-based case–control study of NHL conducted in Connecticut ($N = 455$ cases; 527 controls). Selected participant demographics and case pathology information for this population are presented in Supplementary Table S1. After stratification by NHL subtype, an unconditional logistic regression analysis showed that the

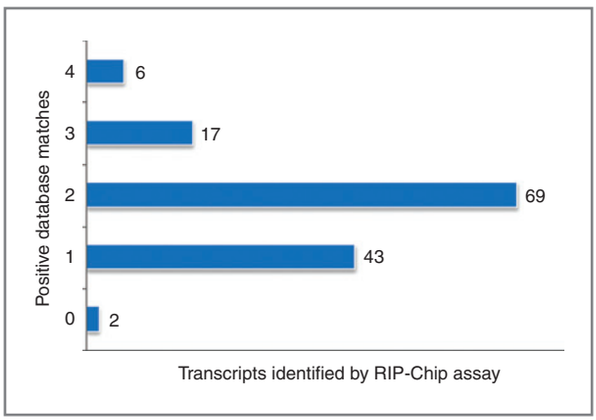

Figure 2. Distribution of positive target predictions among identified potential miR-202 targets. A total of 137 transcripts found to be enriched in the RIP-Chip assay were interrogated for 3′UTR miR-202 binding sites using the miRanda, TargetScan/TargetScanS, PITA, and RNAhybrid miRNA target prediction algorithms. A total of 92 of 137 (69%) screened transcripts were identified as having miR-202-binding sites by 2 or more prediction algorithms based on the degree of base pairing complementarity to the seed region, the number of compensatory 3′ nonseed matches, and the thermodynamic stability of the predicted miR-202-mRNA hybrid.

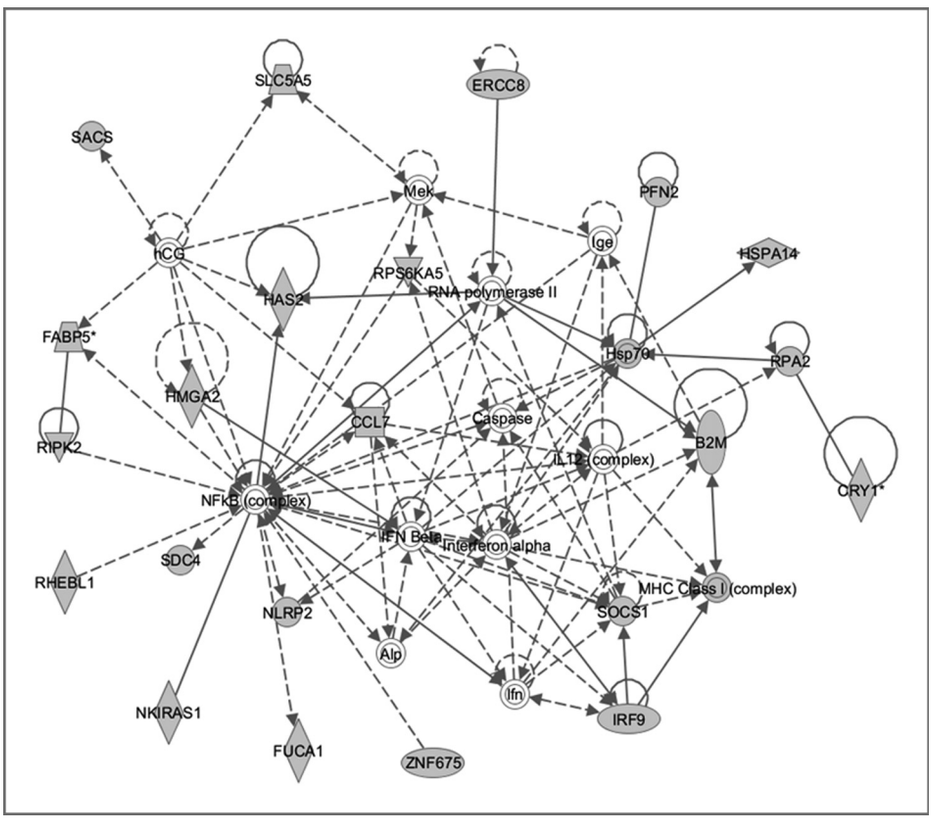

Figure 3. Top interaction network formed by identified miR-202 targets and functionally related molecules. The IPA software designated this network as having relevance for "Hematological System Development and Function" ($P = 1.0E^{-46}$). Functional descriptors assigned by the Ingenuity Knowledge Database to the molecules in this network included: "hematologic neoplasia," "differentiation of lymphocytes," "proliferation of hematopoietic cells," and "lymphomagenesis." Of the 35 genes in this network, 22 were identified by our RIP-Chip assay as potential direct targets of miR-202 (highlighted in gray).

variant alleles were strongly associated with follicular lymphoma risk ($P_{TREND} = 0.006$; Table 1). Specifically, individuals harboring a single variant allele at this locus had a significantly elevated risk of follicular lymphoma (OR = 1.77, 95% CI: 1.12–2.79; $P = 0.014$), and the magnitude of the association was even stronger among individuals with two copies of the variant allele, although this association did not reach statistical significance because of the small number of individuals with the homozygous variant genotype (OR = 2.62, 95% CI: 0.79–8.96; $P = 0.115$). Harboring one or both variant alleles (i.e., assuming a dominant model) was also associated with a significantly elevated risk of follicular lymphoma (OR = 1.83, 95% CI: 1.17–2.85; $P = 0.008$). The full genotyping results for all subtypes can be found in Supplemental Table S5.

Functional effect of SNP rs12355840 on miR-202 levels in vitro

As SNP rs12355840 is located within the stem-loop sequence of the miR-202 precursor (Fig. 4), it may influence miR-202 levels *in vivo* by interfering with the formation of the secondary structure and/or affecting the processing of the pre-miRNA to its mature form. To test this hypothesis, precursor miR-202 expression vectors with either the G or A allele were transfected into HeLa cells, and levels of both precursor and mature miR-202 were quantified. Relative to the negative control, both pre-miR-202-G- and pre-miR-202-A-transfected cells had approximately equal levels of the precursor miRNA (18.9- and 17.5-fold increase, respectively). In contrast, mature miR-202 levels were 6.0

Table 1. Association of rs12355840 in miR-202 with follicular lymphoma risk

	Genotype	Controls (%)	Cases (%)	Multivariate OR[a]
hsa-mir-202	AA	361 (70.4)	59 (57.8)	Ref.
rs12355840	GA	139 (27.1)	39 (38.2)	**1.77 (1.12–2.79)**
	GG	13 (2.5)	4 (3.9)	2.66 (0.79–8.96)
	GA+GG	152 (29.6)	43 (42.2)	**1.83 (1.17–2.85)**
	P_{trend}			**0.006**

[a]Adjusted for age, race, and family history of cancer. Statistically significant associations ($P < 0.05$) are shown in bold.

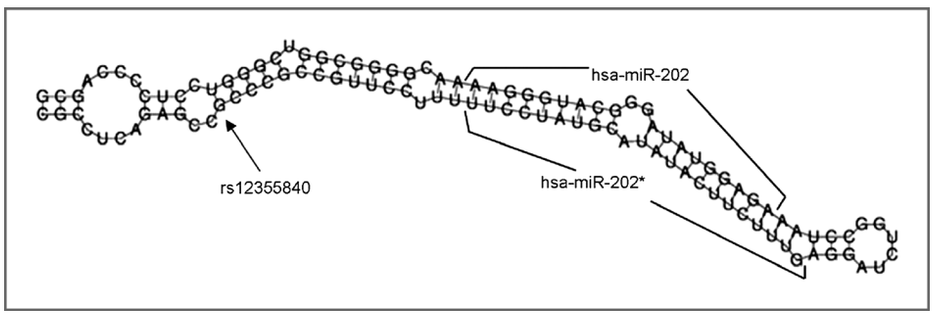

Figure 4. Schematic of secondary structure for miR-202 harboring a germline mutation rs12355840. The structure image was adapted from the Vienna RNA secondary structure prediction algorithm (http://www.tbi.univie.ac.at/RNA/).

times higher in pre-miR-202-A-transfected cells relative to the negative control but were approximately the same between pre-miR-202-G-transfected cells and the negative control (Fig. 5A). To further characterize the effect of the variant allele on pre-miR-202 processing, we conducted a second analysis in which mature miR-202 levels were normalized to precursor miR-202 levels. On the basis of equal quantities of precursor miRNA, pre-miR-202-A-transfected cells had 7.4 times more mature miRNA than pre-miR-202-G-transfected cells, suggesting that SNP rs12355840 influences processing of pre-miR-202 into the mature effector miRNA (Fig. 5B).

Luciferase miR-202 target confirmation in the Farage lymphoma cell line

Real-time RT-PCR revealed endogenous miR-202 levels to be 3.5- and 3.4-fold higher in Farage and Toledo B lymphoma cell lines, respectively, compared with levels in HeLa cells (data not shown). To validate our assumption that miR-202 targets are conserved between HeLa and B lymphocytes and investigate the endogenous functionality of miR-202 in lymphoma, we reaffirmed the target status of *HAS2* and *FAM135A* in the Farage lymphoma cell line using luciferase 3′UTR constructs and anti-miR-202 inhibitors. As expected, inhibiting miR-202 resulted in an increase in luciferase signal intensity for both *HAS2* (2.5-fold) and *FAM135A* (3.1-fold) relative to negative controls (Fig. 5C).

Discussion

Our RIP-Chip analysis of direct miR-202 targets identified several transcripts with potential relevance for lymphoma-related processes, and the gene interaction network most strongly associated with the potential targets contained several molecules with known involvement in lymphomagenesis. Examples include: suppressor of cytokine signaling (*SOCS1*), a tumor suppressor (32) and potential miR-202 target that is epigenetically silenced in some B-cell lymphomas (33), and *B2M*, a subunit required for MHC class I assembly that is frequently mutated in NHL (34, 35). Perhaps most interestingly, *DICER1* was identified as a potential miR-202 target. According to TargetScan, the *DICER1* 3′UTR contains three conserved miR-202 binding sites. This protein is essential for the biogenesis and proper function of all miRNAs, and thus has widespread biologic importance (36). Haploinsufficiency of *DICER1* has been linked to susceptibility to multiple nonhematopoietic tumors (37, 38). However, in B cells, deletion of *DICER1* results in suppression of lymphomagenesis *in vivo*, and inhibition of *DICER1* in existing B-cell lymphomas results in cell death via apoptosis, indicating that DICER1 is required for B-cell lymphoma development and survival (39).

Another important miR-202 target is the cell-cycle regulatory element and proto-oncogene *SKP2*. This protein regulates the G_1 to S-phase transition and encourages cell proliferation through the degradation of the cyclin-dependent kinase inhibitor p27 (40). SKP2 was shown to be causally involved in lymphomagenesis (41). In addition, inhibition of SKP2 in lymphoma cells leads to decreased growth and induces apoptosis, suggesting that SKP2 may be a potential target for therapeutic intervention in lymphoma patients (42). Indeed, bortezomib, which operates in part by degrading SKP2 (43), was recently approved by the U.S. Food and Drug Administration for the treatment of refractory multiple myeloma and mantle cell lymphoma (44).

The results of our network-based analysis were consistent with the finding of a significant association between a SNP within the miR-202 stem-loop sequence and lymphoma in our NHL study population. Altogether, these findings suggest that rs12355840 may represent a novel risk biomarker for follicular lymphoma, and implicate miR-202 in follicular lymphomagenesis. Interestingly, the risk allele for rs12355840 (G) in our population-based genetic association analysis was shown to be associated with diminished pre-miR-202-processing capacity, and thus reduced mature miR-202 levels. Because both target genes *DICER1* and *SKP2* are lymphoma related, individuals harboring the risk allele for rs12355840 may have increased levels of both of these genes, which, in both cases, would be consistent with increased lymphoma risk. Thus, miR-202 may operate as a tumor suppressor in follicular lymphoma. However, an important limitation of our study was the lack of RNA samples for the case–control population. Because only DNA was collected for study subjects, we are unable to correlate genotype to miRNA expression or processing directly, and are thus reliant on *in vitro* analyses.

So far, very few studies have investigated the role of miRNAs in lymphomagenesis, but there are a handful of

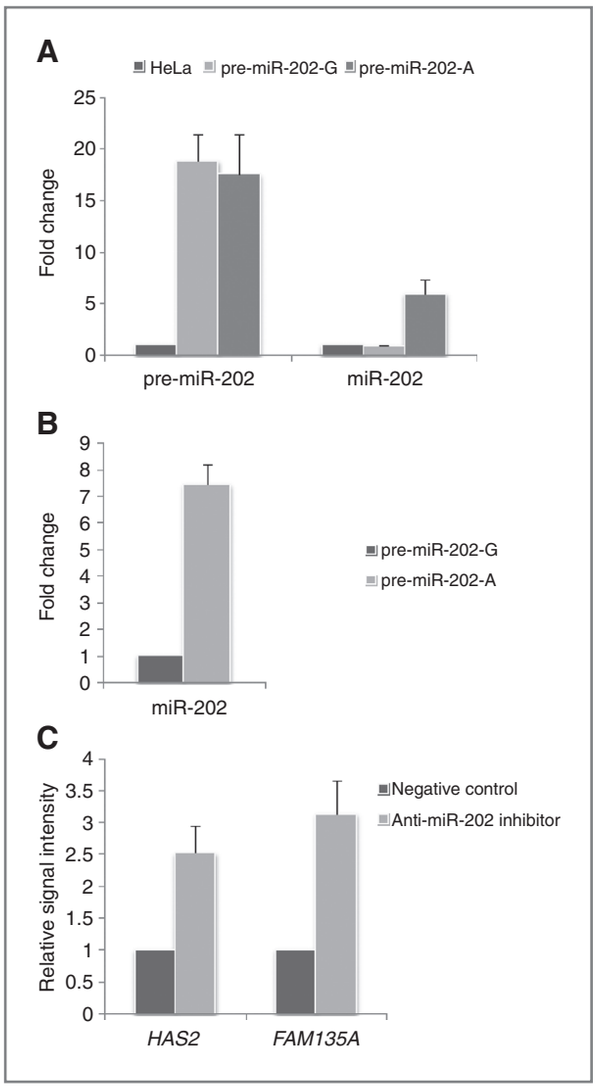

Figure 5. Functional assay to measure rs12355840's effect on pre-miR-202 processing efficiency and target validation in a lymphoma cell line. A, normalized pre-miR-202 levels were approximately the same between pre-miR-202-G- and pre-miR-202-A-transfected HeLa cells. In contrast, levels of mature miR-202 were several times higher in pre-miR-202-A-transfected cells compared with pre-miR-202-G transfected cells. B, quantification of miR-202 levels following normalization to pre-miR-202 levels. The 7.4-fold increase in mature miR-202 levels in pre-miR-202-A-transfected cells relative to pre-miR-202-G-transfected cells is attributable to molecular-processing events independent of initial precursor levels. These data indicate that the variant G allele may hinder the *in vitro* processing of precursor miR-202 transcripts into the mature form. C, Farage cells were transfected with vectors encoding luciferase fused to the 3′UTR of *HAS2* or *FAM135A* and either an anti-miR-202 or NC. Inhibition of miR-202 results in higher relative signal intensities for cells transfected with either 3′UTR construct relative to negative control.

previous observations with findings similar to the results of our association analysis. For example, the genomic region harboring miR-15a and miR-16 has been found to be commonly deleted in B-cell chronic lymphocytic leukemias (45), although a polycistronic miRNA cluster, miR-17–92, was found to be significantly amplified in B-cell lymphomas (46). In addition, miR-21 induction has been found to be sufficient to cause lymphoma in mouse models (6). Perhaps the most relevant analysis comes from a recent miRNA profiling experiment comparing follicular lymphoma tumor cells to follicular hyperplasia (19). This study describes a follicular lymphoma signature composed of three groups of miRNAs: (i) increased in follicular lymphoma; (ii) increased in most follicular lymphoma, but decreased in a subset of follicular lymphoma; and (iii) decreased in follicular lymphoma. miR-202 was one of only 11 miRNAs belonging to group 3, lending further support to our hypothesis that miR-202 may operate as a tumor suppressor in follicular lymphoma.

It should be noted that, apart from follicular lymphoma, our genetic association data were not significant for any other lymphoma subtype, a pattern which is consistent with the results from previous NHL genetic association analyses, including two recent genome-wide association studies (47, 48) and a large pooled case–control analysis (49), all of which noted their strongest associations in the follicular lymphoma subtype. This indication that follicular lymphoma may have a stronger genetic component than exists for other NHL subtypes is of particular interest, given a recent examination of etiologic heterogeneity among NHL subtypes, which identified no significant risk factors which were unique to follicular lymphoma alone (50).

As the RIP-Chip assay was conducted using HeLa cells, our analysis was limited to only those genes with detectable levels of endogenous expression in this cell model. As such, it is certain that a portion of the miR-202 targetome remains unelucidated. Moreover, HeLa cells may have a gene expression profile that differs from that which would be seen in lymphoma. Nonetheless, the fact that we were able to reaffirm *HAS2* and *FAM135A* as miR-202 targets in Farage lymphoma cell line lends credence to the supposition that miR-202 targets are conserved between HeLa cells and B lymphocytes. Also, as a subset of identified targets has previously been implicated in B-cell lymphomagenesis and hematopoiesis, their expression in B lymphocytes is almost certain at some stages of the lymphomagenic or hematopoietic process. Furthermore, it is evident that miR-202 has endogenous gene-silencing activity in B lymphocytes, as implicated by our finding that inhibition of miR-202 activity resulted in increased expression of miR-202 targets in the Farage lymphoma cell line. Therefore, it is likely that miR-202 modulates the activity of at least the lymphoma-relevant targets identified in HeLa cells, providing yet more evidence of miR-202's physiologic significance in lymphomagenesis.

In summary, we identified 141 potential members of the miR-202 targetome. Interrogation-enriched transcripts using miRNA target prediction algorithms revealed that a majority harbored predicted 3′UTR miR-202-binding sites. Potential targets were then investigated for disease-relevance using a network-based analytical

approach. We discovered that a large proportion of the molecular interactions between identified targets were relevant to lymphomagenesis and hematopoietic processes. We also investigated whether genetic variants within the miR-202 stem-loop could account for differences in susceptibility to lymphoma in a population of NHL patients. A miR-202 precursor SNP, rs12355840, was found to be significantly associated with follicular lymphoma. These findings need to be further confirmed in studies with larger sample sizes and expression levels of miR-202 and identified targets need to be examined in lymphoma clinical samples. A functional analysis of this polymorphism showed that the risk allele was also associated with reduced pre-miR-202-processing capacity, which is consistent with the findings of our network analysis. Taken together, these findings suggest a role for miR-202 in lymphomagenesis and the full extent of miR-202's role in lymphoma-related processes should be the focus of future investigations.

Disclosure of Potential Conflicts of Interest

F.J. Slack is a consultant/advisory board member of Mira Dx and Mirna Rx. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: T. Zheng, Y. Zhu
Development of methodology: A.E. Hoffman, R. Liu, T. Zheng
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A.E. Hoffman, T. Zheng
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A.E. Hoffman, A. Fu, T. Zheng
Writing, review, and/or revision of the manuscript: A.E. Hoffman, A. Fu, T. Zheng, F.J. Slack, Y. Zhu
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): T. Zheng
Study supervision: T. Zheng, F.J. Slack, Y. Zhu

Acknowledgments

The authors thank Irina Tikhonova at Yale University's W.M. Keck Foundation Biotechnology Research Laboratory for Sequenom genotyping analysis. The authors also thank Daniel Jacobs and Fengqin Shi for laboratory assistance.

Grant Support

This work was supported by the NIH grants CA154653 and CA122676.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 4, 2012; revised December 19, 2012; accepted January 8, 2013; published OnlineFirst January 18, 2013.

References

1. Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 1993;75:843–54.
2. Bader AG, Brown D, Stoudemire J, Lammers P. Developing therapeutic microRNAs for cancer. Gene Ther 2011;18:1121–6.
3. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer 2006;6:259–69.
4. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A 2004;101:2999–3004.
5. Trang P, Medina PP, Wiggins JF, Ruffino L, Kelnar K, Omotola M, et al. Regression of murine lung tumors by the let-7 microRNA. Oncogene 2010;29:1580–7.
6. Medina PP, Nolde M, Slack FJ. OncomiR addiction in an *in vivo* model of microRNA-21-induced pre-B-cell lymphoma. Nature 2010;467:86–90.
7. Kumar MS, Lu J, Mercer KL, Golub TR, Jacks T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. Nat Genet 2007;39:673–7.
8. Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci U S A 2006;103:2257–61.
9. Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, et al. MicroRNA gene expression deregulation in human breast cancer. Cancer Res 2005;65:7065–70.
10. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. Nature 2005;435:834–8.
11. Hoffman AE, Zheng T, Yi C, Leaderer D, Weidhaas J, Slack F, et al. microRNA miR-196a-2 and breast cancer: a genetic and epigenetic association study and functional analysis. Cancer Res 2009;69:5970–7.
12. Schimanski CC, Frerichs K, Rahman F, Berger M, Lang H, Galle PR, et al. High miR-196a levels promote the oncogenic phenotype of colorectal cancer cells. World J Gastroenterol 2009;15:2089–96.
13. Peng S, Kuang Z, Sheng C, Zhang Y, Xu H, Cheng Q. Association of microRNA-196a-2 gene polymorphism with gastric cancer risk in a Chinese population. Dig Dis Sci 2010;55:2288–93.
14. Tian T, Shu Y, Chen J, Hu Z, Xu L, Jin G, et al. A functional genetic variant in microRNA-196a2 is associated with increased susceptibility of lung cancer in Chinese. Cancer Epidemiol Biomarkers Prev 2009;18:1183–7.
15. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005;120:15–20.
16. Zhang Y, Dai Y, Huang Y, Ma L, Yin Y, Tang M, et al. Microarray profile of micro-ribonucleic acid in tumor tissue from cervical squamous cell carcinoma without human papillomavirus. J Obstet Gynaecol Res 2009;35:842–9.
17. Ng EK, Chong WW, Jin H, Lam EK, Shin VY, Yu J, et al. Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. Gut 2009;58:1375–81.
18. Jiang Z, Guo J, Xiao B, Miao Y, Huang R, Li D, et al. Increased expression of miR-421 in human gastric carcinoma and its clinical association. J Gastroenterol 2010;45:17–23.
19. Wang W, Corrigan-Cummins M, Hudson J, Maric I, Simakova O, Neelapu SS, et al. MicroRNA profiling of follicular lymphoma identifies microRNAs related to cell proliferation and tumor response. Haematologica 2012;97:586–94.
20. Peiffer SL, Herzog TJ, Tribune DJ, Mutch DG, Gersell DJ, Goodfellow PJ. Allelic loss of sequences from the long arm of chromosome 10 and replication errors in endometrial cancers. Cancer Res 1995;55:1922–6.
21. Lee SH, Davison JA, Vidal SM, Belouchi A. Cloning, expression and chromosomal location of NKX6B TO 10Q26, a region frequently deleted in brain tumors. Mamm Genome 2001;12:157–62.
22. Courtens W, Wuyts W, Rooms L, Pera SB, Wauters J. A subterminal deletion of the long arm of chromosome 10: a clinical report and review. Am J Med Genet A 2006;140:402–9.
23. Irving M, Hanson H, Turnpenny P, Brewer C, Ogilvie CM, Davies A, et al. Deletion of the distal long arm of chromosome 10; is there a characteristic phenotype? A report of 15 *de novo* and familial cases. Am J Med Genet A 2003;123A:153–63.
24. Buechner J, Tomte E, Haug BH, Henriksen JR, Lokke C, Flaegstad T, et al. Tumour-suppressor microRNAs let-7 and mir-101 target the proto-oncogene MYCN and inhibit cell proliferation in MYCN-amplified neuroblastoma. Br J Cancer 2011;105:296–303.

25. http://www.microrna.org.
26. http://www.targetscan.org.
27. http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html.
28. http://bibiserv.techfak.uni-bielefeld.de/rnahybrid.
29. http://www.ingenuity.com.
30. Zhang Y, Holford TR, Leaderer B, Boyle P, Zahm SH, Owens PH, et al. Blood transfusion and risk of non-Hodgkin's lymphoma in Connecticut women. Am J Epidemiol 2004;160:325–30.
31. Hofacker IL. Vienna RNA secondary structure server. Nucleic Acids Res 2003;31:3429–31.
32. Rottapel R, Ilangumaran S, Neale C, La Rose J, Ho JM, Nguyen MH, et al. The tumor suppressor activity of SOCS-1. Oncogene 2002;21: 4351–62.
33. Chim CS, Wong KY, Loong F, Srivastava G. SOCS1 and SHP1 hypermethylation in mantle cell lymphoma and follicular lymphoma: implications for epigenetic activation of the Jak/STAT pathway. Leukemia 2004;18:356–8.
34. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature 2011;476:298–303.
35. Challa-Malladi M, Lieu YK, Califano O, Holmes AB, Bhagat G, Murty VV, et al. Combined genetic inactivation of beta2-Microglobulin and CD58 reveals frequent escape from immune recognition in diffuse large B cell lymphoma. Cancer Cell 2011;20:728–40.
36. Park JE, Heo I, Tian Y, Simanshu DK, Chang H, Jee D, et al. Dicer recognizes the 5′ end of RNA for efficient and accurate processing. Nature 2011;475:201–5.
37. Slade I, Bacchelli C, Davies H, Murray A, Abbaszadeh F, Hanks S, et al. DICER1 syndrome: clarifying the diagnosis, clinical features and management implications of a pleiotropic tumour predisposition syndrome. J Med Genet 2011;48:273–8.
38. Kumar MS, Pester RE, Chen CY, Lane K, Chin C, Lu J, et al. Dicer1 functions as a haploinsufficient tumor suppressor. Genes Dev 2009; 23:2700–4.
39. Arrate MP, Vincent T, Odvody J, Kar R, Jones SN, Eischen CM. MicroRNA biogenesis is required for Myc-induced B-cell lymphoma development and survival. Cancer Res 2010;70:6083–92.
40. Carrano AC, Eytan E, Hershko A, Pagano M. SKP2 is required for ubiquitin-mediated degradation of the CDK inhibitor p27. Nat Cell Biol 1999;1:193–9.
41. Latres E, Chiarle R, Schulman BA, Pavletich NP, Pellicer A, Inghirami G, et al. Role of the F-box protein Skp2 in lymphomagenesis. Proc Natl Acad Sci U S A 2001;98:2515–20.
42. Uddin S, Hussain A, Ahmed M, Belgaumi A, Al-Dayel F, Ajarim D, et al. S-phase kinase protein 2 is an attractive therapeutic target in a subset of diffuse large B-cell lymphoma. J Pathol 2008;216: 483–94.
43. Uddin S, Ahmed M, Bavi P, El-Sayed R, Al-Sanea N, AbdulJabbar A, et al. Bortezomib (Velcade) induces p27Kip1 expression through S-phase kinase protein 2 degradation in colorectal cancer. Cancer Res 2008;68:3379–88.
44. Molineaux SM. Molecular pathways: targeting proteasomal protein degradation in cancer. Clin Cancer Res 2012;18:15–20.
45. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A 2002;99:15524–9.
46. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, et al. A microRNA polycistron as a potential human oncogene. Nature 2005;435:828–33.
47. Skibola CF, Bracci PM, Halperin E, Conde L, Craig DW, Agana L, et al. Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. Nat Genet 2009;41:873–5.
48. Conde L, Halperin E, Akers NK, Brown KM, Smedby KE, Rothman N, et al. Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. Nat Genet 2010;42:661–4.
49. Morton LM, Purdue MP, Zheng T, Wang SS, Armstrong B, Zhang Y, et al. Risk of non-Hodgkin lymphoma associated with germline variation in genes that regulate the cell cycle, apoptosis, and lymphocyte development. Cancer Epidemiol Biomarkers Prev 2009;18:1259–70.
50. Morton LM, Wang SS, Cozen W, Linet MS, Chatterjee N, Davis S, et al. Etiologic heterogeneity among non-Hodgkin lymphoma subtypes. Blood 2008;112:5150–60.