UNIVERSITÉ DE GENÈVE

Archive ouverte UNIGE
https://archive-ouverte.unige.ch

Thèse 2012

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

Computational prediction of microRNA targets: thermodynamic, probabilistic and evolutionary models parameterized by genome-scale experimental data

Vejnar, Charles

How to cite

VEJNAR, Charles. Computational prediction of microRNA targets: thermodynamic, probabilistic and evolutionary models parameterized by genome-scale experimental data. Doctoral Thesis, 2012. doi: 10.13097/archive-ouverte/unige:23938

This publication URL: https://archive-ouverte.unige.ch/unige:23938
Publication DOI: 10.13097/archive-ouverte/unige:23938

© This document is protected by copyright. Please refer to copyright holder(s) for terms of use.

UNIVERSITÉ DE GENÈVE

Département de Médecine génétique et Développement

FACULTÉ DE MÉDECINE
Professeur Evgeny M. Zdobnov

Département d'Informatique

FACULTÉ DES SCIENCES
Professeur Ron D. Appel

Computational prediction of microRNA targets: thermodynamic, probabilistic and evolutionary models parameterized by genome-scale experimental data

THÈSE

présentée à la Faculté des Sciences de l'Université de Genève pour obtenir le grade de Docteur ès Sciences, mention Bioinformatique

par

Charles E. Vejnar

de

Versailles (France)

Thèse N°4475

Genève, 2012

Charles E. Vejnar: *Computational prediction of microRNA targets: thermodynamic, probabilistic and evolutionary models parameterized by genome-scale experimental data*, © 2012

SUPERVISOR
Professor Evgeny M. Zdobnov

LOCATION
Genève, Faculté de Médecine

TIME FRAME
2007-2012

VERSION
25 October 11:29

UNIVERSITÉ DE GENÈVE

FACULTÉ DES SCIENCES

Doctorat ès sciences
Mention bioinformatique

Thèse de *Monsieur Charles VEJNAR*

intitulée :

" Computational Prediction of MicroRNA Targets : Thermodynamic, Probabilistic and Evolutionary Models Parameterized by Genome-scale Experimental Data "

La Faculté des sciences, sur le préavis de Messieurs E. ZDOBNOV, professeur associé et directeur de thèse (Faculté de médecine, Département de médecine génétique et développement), R. APPEL, professeur ordinaire et codirecteur de thèse (Département d'informatique), I. XENARIOS, professeur (Center for Integrative Genomics, University of Lausanne, Switzerland) et H. KAESMANN, professeur (Center for Integrative Genomics, University of Lausanne, Switzerland), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 15 octobre 2012

Thèse - 4475 -

Le Doyen, Jean-Marc TRISCONE

N.B.- La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "Informations relatives aux thèses de doctorat à l'Université de Genève".

Remerciements

Je remercie en premier lieu le Pr Evgeny Zdobnov de m'avoir permis d'effectuer mon travail de doctorat dans son laboratoire, le Computational Evolutionary Genomics Group (CEGG). Nos échanges d'idées, sa connaissance et son inspiration m'ont permis de devenir chercheur. Je le remercie de m'avoir accordé sa confiance et une grande liberté.

Je remercie aussi mon rapporteur à la Faculté des Sciences, le Pr Ron D. Appel, ainsi que les autres membres de mon jury de thèse, le Pr Henrik Kaessmann et le Pr Ioannis Xenarios.

Le jeune doctorant que j'étais en arrivant à Genève en 2007 a pu compter sur la bienveillance et la compétence du Dr Stefan Wyder, post-doctorant du CEGG. Merci beaucoup de m'avoir encadré et soutenu.

Un travail de doctorat est un exercice solitaire mais c'est aussi faire partie d'un groupe, le CEGG. Je remercie tous ses membres pour leur bonne humeur et leurs compétences grâce auxquelles j'ai travaillé dans une ambiance très motivante. En particulier je remercie le Dr Daniel Gerlach avec qui j'ai partagé mes premiers pas de doctorant ainsi que le Dr Thomas Junier qui m'a fait découvrir les helvétismes avec humour. Merci beaucoup aux Dr Robert Waterhouse et Dr Thomas Petty pour toutes ces heures à corriger mes écrits en langue anglaise avec beaucoup de sérieux. Pour toute l'entraide dans le travail et toutes nos activités extra-académiques, je remercie le Dr. Fredrik Tegenfeldt, le Dr. Evgenia Kriventseva, Isabelle Cosandier, Jia Li, Matthias Blum, Adrian Cesar, Ismaël Padioleau et Aline Dousse.

Les collaborations sont un aspect particulièrement agréable, enthousiasmant et productif de la recherche et en particulier lorsqu'on collabore avec des chercheurs talentueux. Je tiens à remercier tout d'abord le Pr David Gatfield pour ma première collaboration qui fut très enrichissante et qui j'espère pourra un jour se poursuivre. Merci aussi au Dr Marilena Papaioannou, au Pr Serge Nef, au Dr Isabelle Dunand-Sauthier et au Pr Walter Reith pour tous ces échanges qui furent pour moi motivants et enrichissants.

Je tiens aussi à remercier l'Institut Suisse de Bioinformatique, en particulier pour l'organisation du *PhD Training Network* qui est un espace d'échanges sympathiques, utile pour ma thèse et pour mon avenir.

Enfin, je remercie chaleureusement ma famille et mes amis. Merci à mes parents, à mes grands-parents et à mon frère Gabriel pour leur infaillible soutien et de n'avoir jamais douté de ma réussite depuis le début.

CONTENTS

1 ABSTRACT

In animals, the expression of genes is a regulated process shaping cellular gene expression profiles, at the origin of tissue identity. While transcription factors regulate the expression of genes at the transcription level, microRNAs (miRNAs) induce a post-transcriptional repression of protein-coding genes. Embedded in the RNA-Induced Silencing Complex (RISC), miRNAs act as a recognition element for driving the RISC to repress targeted mRNAs. Repression is mainly achieved by degrading targeted mRNAs, but also by inhibiting translation. Although identification of miRNA targets remains challenging, partial base pairing between the miRNA 5′ region, the so-called "seed", and the targeted mRNA in its 3′-untranslated region (UTR) triggers repression of mRNA expression. As the human genome encodes over 1000 miRNA genes, searching for miRNA seeds identifies many potential targets. For example, simple seed-match searches suggest that miR-122 has about 13′300 potential target sites in 3′-UTRs of about 7′600 human genes.

Prioritization of targets for any miRNA functional analysis is therefore of critical importance. This necessitates the ranking of potential miRNA targets bearing a seed-match, not only predicting in a binary manner if an mRNA is a target or not. A biologically meaningful ranking criterion is the miRNA-mediated repression strength that can be experimentally measured as the effect on mRNA or protein levels. I employed a collection features to computationally predict the miRNA repression strength from additional information beyond the seed-match, and thereby rank putative miRNA-mRNA interactions in a biologically relevant manner.

I developed an open source software library, miRmap, which for the first time comprehensively covers thermodynamic, evolutionary, probabilistic, and sequence-based approaches. Accessibility of mRNAs to miRNA binding and stability of miRNA-mRNA duplexes were estimated with RNA-folding algorithms. The significance of the target site's evolutionary conservation was assessed non-empirically with the Siepel, Pollard, and Haussler test, which evaluates the significance of negative selection. The statistical significance of seed occurrence(s) in 3′-UTR sequences was calculated with an approximate, but also an exact probability distribution. In total, eleven features are implemented in miRmap, three of which are novel.

The predictive power of miRmap features was evaluated in an unbiased way using high throughput experimental data from immunopurification, transcriptomics, proteomics and polysome fractionation experiments, covering recognition, mRNA stability and translational miRNA-mediated repression aspects. Overall, target site accessibility is the most predictive feature. My linear model combining all features almost doubles the predictive power of the renowned TargetScan tool. It increases the proportion of variance explained from 7.5% to 13% of miRNA over-expression effects measured at the transcriptome level. Prediction features were also tested with experimental data, obtained in collaboration with Swiss research groups, on tissue samples instead of cell line cultures: I investigated the role of miR-122 in the regulation of the hepatocyte transcriptome, of miR-155 in dendritic cell maturation and function, and the effects on transcriptome of knocking-out all miRNAs in mouse testis.

Available as an open source Python library, miRmap establishes a solid foundation for the future development of approaches to miRNA target prediction, facilitating meaningful comparisons between existing and new features, and providing the community with direct access to state-of-the-art analytical tools.

RÉSUMÉ 2

Chez les animaux, l'expression des gènes est un processus régulé qui aboutit aux profils d'expression cellulaire des gènes, eux-mêmes à l'origine de l'identité tissulaire. Alors que les facteurs de transcription régulent l'expression des gènes au niveau de leur transcription, les microARN (miARN) induisent une répression post-transcriptionnelle des gènes codants pour des protéines. Chargé dans le RNA-Induced Silencing Complex (RISC), les miARN servent d'élément de reconnaissance pour diriger le RISC vers les ARNm ciblés. La répression s'effectue principalement par la dégradation de l'ARNm ciblé mais aussi par l'inhibition de la traduction. Bien que l'identification des cibles des miARN reste difficile, un appariement partiel entre la région 5′ du miARN, dite «région d'ancrage», et la région 3′ non-traduites (RNT) de l'ARNm ciblé provoque la répression de l'expression de l'ARNm. Le génome de l'Homme codant pour environ 1′000 gènes de miARN, la recherche des régions d'ancrage de miARN permet l'identification d'un grand nombre de cibles potentielles. Par exemple, une simple recherche de région d'ancrage suggère que le miR-122 a environ 13′300 cibles potentielles dans les 3′-RNT d'environ 7′600 gènes.

Dans le cadre d'une analyse fonctionnelle, le choix des cibles à inclure dans l'étude est ainsi primordial. Ce choix nécessite un classement des cibles potentielles des miARN, et non une simple prédiction binaire qui distingue les ARNm ciblés de ceux qui ne le sont pas. Un critère de classement, qui fait sens d'un point de vue biologique, est l'intensité de la répression associée au miARN. La mesure de cette intensité peut être l'effet sur les quantités d'ARNm ou de protéines et a donc un sens biologique. J'ai utilisé une série de critères pour prédire *in silico* l'intensité de la répression associée au miARN, à partir d'autres informations que la seule présence d'une région d'ancrage.

J'ai développé une librairie, nommée miRmap et dont le code est librement accessible, qui pour la première fois couvre les approches thermodynamique, de l'évolution, probabiliste et fondée uniquement sur la séquence. L'accessibilité des ARNm à l'appariement avec un miARN et la stabilité des paires miARN-mRNA sont estimées en utilisant des algorithmes de repliement d'ARN. La significativité de la conservation des sites cibles est évaluée par le test non-empirique de Siepel, Pollard et Haussler, qui mesure la significativité de la sélection négative. La significativité statistique de la présence des régions d'ancrage dans les séquences des 3′-RNT est calculée grâce à une approximation ainsi qu'avec une solution exacte de la distribution de probabilité. Au total, onze critères sont implémentés au sein de miRmap, dont trois critères originaux.

Les performances des critères de prédictions de miRmap ont été évaluées de façon non-biaisée en utilisant des données d'expériences à haut-débits d'immunopurification, de transcriptomique, de protéomique et de fractionnement de polysomes, qui couvrent les différents aspects de répression par les miARN tels que la reconnaissance des cibles, la stabilité de l'ARNm et la traduction. Généralement, l'accessibilité du site cible est le meilleur critère de prédiction. Mon modèle linéaire qui combine l'ensemble des critères de prédiction double presque les performances de l'outil de référence TargetScan. Il accroît le pourcentage de variance expliquée de 7.5% à 13% des effets d'une sur-expression de miARN mesurés au niveau du transcriptome. Les critères de prédiction ont aussi été évalués avec des données expérimentales produites à partir d'échantillons de tissue, obtenues en collaboration avec des laboratoires de recherche suisses, au lieu de cultures cellulaires: j'ai examiné le rôle de

miR-122 dans la régulation du transcriptome des hépatocytes, de miR-155 dans la maturation et le fonctionnement des cellules dendritiques, ainsi que les effets sur le transcriptome d'un knock-out de tous les miARN dans le testicule de souris.

Disponible sous forme de logiciel libre écrit en Python, miRmap établit de solides fondations pour les développements futurs des méthodes de prédiction des cibles de miARN, en facilitant les comparaisons entre les outils existants et nouveaux et en donnant accès à la communauté à un outil d'analyse de pointe.

3 INTRODUCTION

3.1 Biology and evolution of miRNAs

In the central dogma of molecular biology (Crick [1]), the most important class of RNA is messenger RNA (mRNA). In the past, mostly non-coding RNAs (ncRNAs) associated with translation, such as transfer RNA (tRNA) and ribosomal RNA (rRNA), were described. Recently, new classes of ncRNAs have been discovered like small interfering RNA (siRNA) and microRNA (miRNA). About 20 years ago, studies in the worm *Caenorhabditis elegans* led to the discovery of the first miRNA, lin-4 (Lee et al. [2]). Most of the characteristics of lin-4 were later extended to the numerous miRNAs that were discovered. The lin-4 gene is intronic, as a substantial part of today's known miRNAs. The lin-4 small regulatory RNA was found to be responsible for the decrease of lin-14 protein level. It acts post-transcriptionally through RNA-RNA interaction with the lin-14 mRNA by partial complementary pairing to seven possible sites in the 3' untranslated region of lin-14. The pairing between the 5' end of the miRNA and the mRNA was later described to be a general requirement for target recognition: mutations in this region, called the "seed", disrupt miRNA-mediated repression (Brennecke et al. [3]).

While the majority of miRNA genes are found in intergenic regions, about 30% of miRNA genes are located on the sense strand of coding-gene introns (Rodriguez et al. [4]). Although these miRNA genes can have their own regulatory elements, most of them are transcribed together with the host gene. The transcription of miRNA genes into long primary miRNA (pri-miRNA) is generally performed by RNA polymerase II (**Figure 1**). Once transcribed, these pri-miRNAs adopt a stem-loop structure that is recognized by an RNAse III-like enzyme, the Drosha protein. Drosha releases a precursor miRNA (pre-miRNA) of 60 to 70 nucleotides. Additionally, intron splicing can directly generate a pre-miRNA from a so-called mirtron, in a Drosha independent pathway, when the intron already adopts the pre-miRNA structure. The pre-miRNAs are exported into the cytoplasm by exportin-5, where they are processed in miRNA/miRNA* duplexes by the cytoplasmic Dicer enzyme. Only a single-stranded mature miRNA is kept and bound to an argonaute (Ago) protein to form the core of the RNA-Induced Silencing Complex (RISC) (Bartel [5]), while the passenger strand, named miRNA*, is discarded. Vertebrates have four argonaute proteins. The mammalian Ago2 and the *Drosophila* Ago1 have also a slicer activity (Liu et al. [6]).

The miRNA reference database, miRBase (Kozomara and Griffiths-Jones [7]), currently contains approximately 1500 miRNA genes for human, 750 for mouse, 240 for *Drosophila melanogaster*, and 220 for *Caenorhabditis elegans*. The miRNA genes are also found outside the bilaterian clade. Eight putative miRNAs were indeed found in the sponge *Amphimedon queenslandica* at the root of Metazoa, but miRNAs are lost in *Trichoplax adhaerens* (Grimson et al. [8]). The miRNA repertoire expanded during the metazoan evolution and expanded even further for vertebrates (Grimson et al. [8] and Kozomara and Griffiths-Jones [7]). MiRNA genes are absent from *Saccharomyces cerevisiae* and present in plants with 300 miRNA genes in *Arabidopsis thaliana*, yet with a different biogenesis. They are often organized in clusters, next to each other: up to 40% of miRNAs in animals are clustered. Most of these clusters (80% to 95%) are composed of miRNA genes encoding unrelated mature miRNAs, implying a different mode of evolution than amplification by duplication (Axtell et al. [9]) as this mode

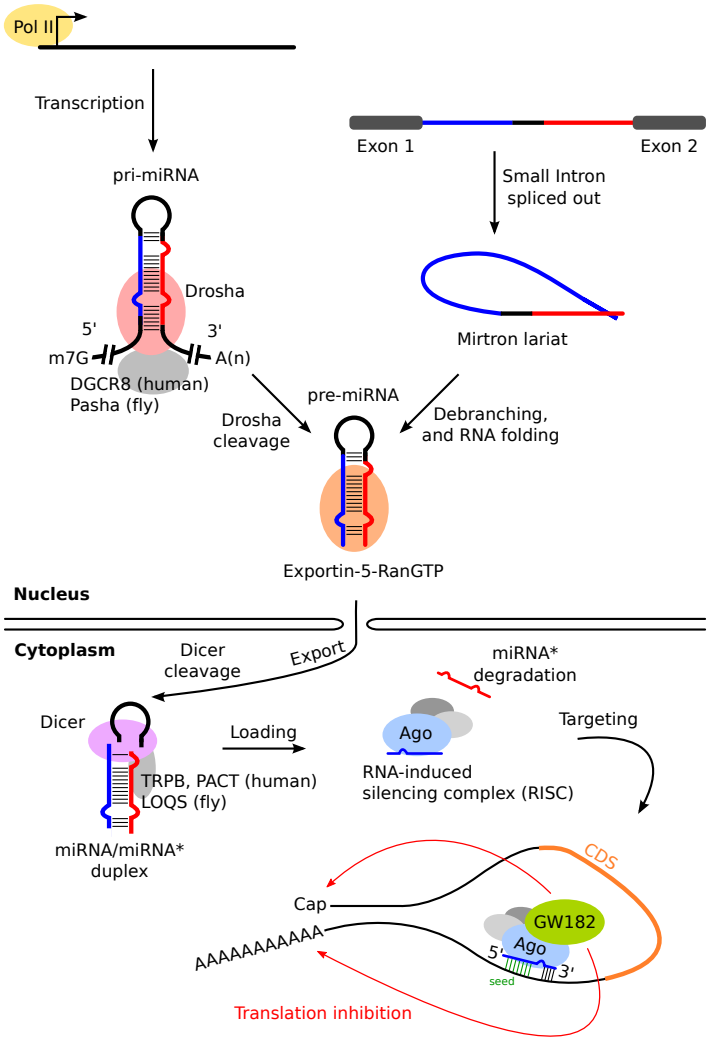


Figure 1 *Biogenesis pathway of miRNAs.* miRNA genes are first transcribed, then recognized and processed by the Drosha enzyme, before being exported from the nucleus to the cytoplasm. An alternative pathway through direct intronic sequence export, called "mirton", is also depicted. The RNA duplex is processed by the Dicer enzyme, and a single RNA strand is loaded into the RNA-Induced Silencing Complex (RISC). The mature RISC binds to 3'-UTRs, and inhibits mRNA translation mainly through mRNA destabilization. The key factor in this process is GW182. (Adapted from Daniel Gerlach, personal communication)

of evolution implies the presence of identical miRNA clusters. Due to the low complementarity of miRNA targeting in animals, it it conceivable that new miRNAs appear through *de novo* emergence from the transcription of the many existing hairpins in the genome. Indeed, the human genome contains 11 million hairpins (Bentwich et al. [10]) that, once transcribed, can potentially give rise to a functional mature miRNA. As nascent miRNAs can diminish the quantity of many mRNAs, they are expected to be expressed at low levels, as they could otherwise have very deleterious effects. They would then either disappear or their expression level would increase due to the advantageous nature of the new regulation. This theory is supported by the fact that (i) among highly expressed miRNAs, all are deeply conserved (for example to the level of vertebrate for human), and that (ii) nascent miRNAs have low

levels of expression (Berezikov et al. [11]). Despite the abundance of hairpins, new regulation would rarely arise from newly transcribed hairpins, as most of them would not be recognized by Drosha and processed efficiently (Berezikov et al. [11]). Yet natural selection can also act to select hairpins better processed by Drosha to increase the efficiency and the level of the beneficial miRNA repression. Even careful efforts to distinguish the functional pri-miRNA recognized by Drosha are challenged by experimental work (Chiang et al. [12] and Berezikov et al. [11]), calling into question existing miRNA annotations.

The miRNA machinery composed of Drosha and Dicer and their co-factors is present in all known animal species (Reviewed in Muljo et al. [13]). The existence of the two processing steps performed by these enzymes can be traced back to the last eukaryotic common ancestor, from which they evolved in different directions: plants have different miRNA biogenesis and targeting rules whereas *Saccharomyces cerevisiae* has lost the RNAi (RNA interference) machinery (Muljo et al. [13]). Interestingly, this loss is advantageous to yeast, as it avoids the processing and consequent destruction of the beneficial Killer virus (Drinnenberg et al. [14]) by the RNAi machinery. This virus encodes a protein toxin that kills nearby cells while conferring immunity to cells making the toxin: cells with a functional RNAi machinery are not immune to the toxin, as the viral dsRNA (double-stranded RNA) is destroyed by RNAi.

In a comparison involving nematodes, insects, and vertebrates, only five miRNA-target relationships were found to be conserved across all the lineages, implying that the miRNA regulatory network might be lineage specific. Over 250 were shared by at least one pair of lineages (Chen and Rajewsky [15]). On the contrary, in the mammalian lineage, 85% of the target sites have conserved positions (Vejnar and Zdobnov [16]), implying a large conservation of miRNA-target relationships. By tuning the expression pattern of genes, miRNAs have a role in the specification of structures in animals, and are involved in the development (Bartel [5]). Conservation of the miRNA-mediated regulatory network is therefore expected.

Figure 2 *Seed definition.* Target recognition by the RISC is mainly driven by base-pairing of the 5′ part of the miRNA with the 3′-UTR part of mRNAs. This pairing region from nucleotide 2 to 7 or 8 of the miRNA is called the "seed".

In animals miRNAs recognize their targeted mRNAs through imperfect complementarity. However, perfect complementarity between the 5′ region of the miRNA (positions 2 to 7 or 8) and the target, called the “seed” (**Figure 2**), is present in most of the described miRNA-target relationships. The pairing between a seed and target, however, is not always sufficient for a functional interaction, and in a few cases such pairing is not required (Didiano and Hobert [17]) or non-canonical pairing with G:U wobbles or mismatches may be acceptable (Brennecke et al. [3]). Furthermore other types of binding have been described, such as centered pairing sites (Shin et al. [18]). These sites require a long 11-nt pairing in the center of the miRNA and can trigger *in vitro* mRNA cleavage in some conditions (in elevated Mg^{2+}). In humans they are two orders of magnitude less frequent than seed-based sites. The key contribution to the stability of the binding is the seed pairing together with a set of arginines, which has also been shown at the molecular level with molecular dynamics (Wang

et al. [19]) on a ternary complex including the crystal structures of a *Thermus thermophilus* argonaute complex (Wang et al. [20]) and two RNAs. Overall, most of the miRNA target sites are found in the 3'-UTR of mRNAs. However, some functional sites were found at the end of the coding sequence (CDS), but their efficacy is limited because they compete with ribosomes (Gu et al. [21]) to bind to CDSs. Yet CDSs are "ribosome-free" once the translation is inhibited by miRNAs, explaining how a few RISCs can bind to the CDS or the 5'-UTR.

Known miRNAs act post-transcriptionally on gene expression by degrading targeted mRNAs and/or inhibiting translation. There is evidence of deadenylation and decapping of miRNA-targeted mRNAs that impair mRNA stability and initiation of translation (Eulalio et al. [22] and Filipowicz et al. [23]). The key effector protein, bound to Ago, is the GW182 protein implicated in mRNA deadenylation (Reviewed in Huntzinger and Izaurralde [24]). Repressed mRNAs and RISCs also accumulate in processing bodies (P-bodies), where they are sequestered and stay untranslated (Pillai et al. [25] and Filipowicz et al. [23]). The major effect is on the stability of the mRNA, with estimates of about 75% (Hendrickson et al. [26]) to 84% (Guo et al. [27]) of miRNA repression attributable to decreased mRNA levels. These measurements were obtained with polysome fractionation, also named ribosome profiling, that measures translation rate. Coupled with microarray or RNA-Seq to measure mRNA abundance, the relative contribution of miRNA repression levels can be estimated. The effect on mRNA translation, studied on a large scale with about 5000 protein levels measured with pSILAC (pulsed Stable Isotope Labeling by Amino acids in cell Culture), is indeed relatively mild (Baek et al. [28] and Selbach et al. [29]). The two effects are correlated without evidence for the existence of large classes of miRNAs triggering specifically repression by degradation or by translation inhibition (Baek et al. [28], Selbach et al. [29] and Hendrickson et al. [26]).

3.2 Computational prediction of miRNA targets

Rules predicting miRNA repression can be inferred from the biological knowledge of the participants, enzymes and co-enzymes, their properties and structures, their mechanisms of action, in addition to *in vivo* miRNA machinery measurements. For instance, the mandatory RNA-RNA interaction has been studied by exhaustive mutation experiments in *Drosophila melanogaster* (Brennecke et al. [3]) and with an endogenous neuronal miRNA in *Caenorhabditis elegans* (Didiano et al. [30]), where every possible interacting nucleotide in the miRNA and the mRNA has been probed. Most new miRNA-mediated regulations are tested with reporter constructs, and collected in databases like TarBase (Vergoulis et al. [31]) and miRTarBase (Hsu et al. [32]). However, the rules need to be tested and/or parameterized with studies that have statistically relevant sizes, and that probe the different aspects of miRNA repression, notably mRNA destabilization and translational repression.

First miRNA target prediction tools were fully rule-based without any parameterization with experimental data that were used only for benchmarking, while more recent tools use high-throughput data to fit simple linear models (Grimson et al. [33] and Vejnar and Zdobnov [16]), or more complex models like Support Vector Machine (SVM) (Saito and Sætrom [34]). However, all miRNA target prediction tools largely share a common set of prediction features derived from the knowledge of the miRNA repression pathway. SVM is a supervised machine-learning method used to solve classification problems by finding an optimal hyperplane separating different populations of points. The hyperplane is defined by the "support vectors", or points that best separate the different populations.

As mentioned above, mRNA repression by a miRNA implies the pairing of both RNAs. Rules describing the characteristics of this pairing have been described, the most used and predictive being the presence of a seed-match in the mRNA. If the pairing in the seed is not

canonical, *i.e.* with G:U wobble(s) or mismatch(s), pairing with the 3′ part of the miRNA (the 5′ part being the seed) can compensate the imperfect pairing and maintain the same repression level (Brennecke et al. [3]). Features counting canonical pairing have been derived from this observation. Furthermore, the overall pairing can be determined by computing the folding of the miRNA-mRNA duplex and calculating its Gibbs free energy, or ΔG. In the various miRNA prediction tools, different ΔGs are computed, where some consider the duplex energy, and some consider the internal structures of the mRNA and miRNA themselves *etc*.

The energies of many small RNAs were measured to developed a thermodynamic model for RNA folding (Mathews et al. [35]) used to obtain the 2D structure of RNAs with a dynamic programming algorithm (Nussinov and Jacobson [36]). The structure with the lowest predicted energy, or the most stable structure, has the Minimum Free Energy (MFE). *In vivo*, a population of RNAs adopts different sub-optimal structures than the MFE structure. Sub-optimal structures (Zuker [37]) and their contributions to the ensemble of structures weighted by their Boltzmann probabilities (McCaskill [38]) can be computed. This approach allows more realistic energy computation of the mRNA-miRNA duplex. As mRNAs are long molecules with a high number of possible nucleotide pairings, they can fold locally differently, with local structures affecting miRNA binding, but still globally have an energy close to the MFE. *Ab initio* methods, those which rely solely on a thermodynamic model, were shown to have an accuracy of 73% for known canonical base pairs in sequences smaller than 700 nt (Mathews [39]).

As described above, the miRNA machinery, the miRNA genes and the miRNA targets are conserved among species. Depending on natural selection and the conferred advantage of the miRNA repression, targets can either be conserved in multiple species or disappear. This degree of conservation can be used as a proxy for the functional relevance of each miRNA target. From the simplest search in a multiple species sequence alignment to more complex methods, a wide range of methods measures the target conservation (see below). Since conservation can also be observed by chance, these methods try to assess the significance of the observed target conservation. The main information used for this purpose is that, within 3′-UTRs, only certain sequence regions have regulatory or structural roles. These regions can therefore be considered as islands of natural selection in a sea of mostly neutrally evolving sequence; about 5% of the human 3′-UTR bases are constrained (Lindblad-Toh et al. [40]). This distinction can be exploited to propose statistical test for measuring the significance of conservation. It can also be used within a probabilistic framework to distinguish the background sequence composition from the target site composition.

The first prediction step of most software tools is to search for seed-matches in the 3′-UTR of genes. This definition has a major impact on the sensitivity of the tools (Ellwanger et al. [41]), as the number of seed-matches can be orders of magnitude different with different seed lengths. The widely accepted definition is a perfect match with nucleotides 2 to 8 of the miRNA for restrictive sites and 2 to 7 to capture the majority of potential target sites. Experimental methods can identify Ago-miRNA-mRNA complexes using an *in vivo* cross-linking protocol with subsequent high-throughput sequencing. In the miRNA-mRNA binding map based on Ago HITS-CLIP (HIgh-Throughput Sequencing of RNA isolated by CrossLinking ImmunoPrecipitation) (Chi et al. [42]) and PAR-CLIP (PhotoActivatable-Ribonucleoside-enhanced CrossLinking and ImmunoPrecipitation) (Hafner et al. [43]), 67% of the sites were 6-mer seed-matches, but about 55% of them were conserved whereas about 70% of 7-mer seed-matches were conserved in 17 vertebrates (Ellwanger et al. [41]). Some approaches consider GU wobbles or even mismatches in the miRNA-mRNA pairing, while others add the presence of an A in the first position. For higher confidence in the predictions, canonical pairing and long seeds should be used. Otherwise, shorter seeds and the presence of GU

wobble and mismatches decrease the prediction sensitivity. In the second step, the potential target sites defined only by the presence of the seed-match are selected based on different criteria. In the following paragraphs, the major miRNA prediction tools and their criteria are described.

3.2.1 Existing miRNA target prediction tools: Description

RNAhybrid (Rehmsmeier et al. [44]) relies on the MFE of the miRNA-mRNA duplex to predict miRNA targets without relying on the seed, whose presence is optional. It also evaluates the statistical significance of the computed energy by first normalizing the energy with the miRNA and mRNA lengths, and second by estimating the probability distribution stated to be an Extreme Value Distribution (EVD). The parameters of the EVD were estimated for each miRNA.

TargetScan is based on a sequence approach, combining sequence features into a context score (Grimson et al. [33]). These features are the position in the 3'-UTR, the nucleotide composition surrounding the seed and the pairing in the 3' part of the miRNA. A conservation score was later added (Friedman et al. [45]), derived from the evolutionary feature described by Stark *et al.* (Stark et al. [46]). Predictions available at targetscan.org are based on the context scores and conserved targets are obtained by thresholding the evolutionary feature. The threshold is determined on an empirical probability distribution, a refined version of Stark et al. [46]. Recently, two novel features were added measuring the target site abundance and the seed-pairing stability (Garcia et al. [47]). They improve the miRNA target ranking by trying to integrate the kinetic effect of the target site abundance for the first feature, and to better score the low energy miRNA-mRNA seed duplex for the second.

The first version of miRanda (John et al. [48]) was a two-step tool. In the first step, miRNA and mRNA sequences are locally aligned using a scoring matrix including weights for GU wobbles. The score of the miRNA first half, containing the seed, was multiplied by a scaling factor of 2.0. In the second step, the MFE of the identified duplexes is computed, on which a threshold was applied to predict potential target sites. Predictions were published on the microrna.org web site (Betel et al. [49]). The miRanda pipeline was later extended by mirSVR (Betel et al. [50]) that learns the feature weights using the Support Vector Regression (SVR) algorithm. The features include the TargetScan context score features and an mRNA accessibility measure computed with a thermodynamic model. Overall, mirSVR marginally improves the performance of TargetScan 4 (Grimson et al. [33]) by about 5%. The source code of miRanda is available, but not for mirSVR.

Thermodynamic evaluation of the mRNA accessibility was first introduced in PITA (Kertesz et al. [51]). The free energy of the mRNA constrained to maintain the target site single-stranded is subtracted from the free energy of the same unconstrained mRNA to obtain the accessibility of the target site. The PITA score is obtained by adding this accessibility energy to the MFE of the miRNA-mRNA duplex.

PicTar (Krek et al. [52]) is based on a Hidden Markov Model (HMM) trained on a set of 3'-UTRs targeted by coexpressed miRNAs. Stable, with low MFE, and conserved seed-matches are considered as anchors in the training set of 3'-UTRs. The other nucleotides considered as the background are modeled as a Markov chain of order 0, where the anchors are the two hidden states of the HMM. The model is trained using the Baum-Welch algorithm and used to compute the likelihood of potential target sites.

With the Teiresias algorithm (Rigoutsos and Floratos [53]), Rna22 (Miranda et al. [54]) determines representative 4-mer patterns in a set of miRNA sequences. These patterns are then

searched in 3′-UTRs sequences to determine "target islands". A target island is predicted to be a target site if the MFE with a given miRNA is below a defined threshold.

ElMMo (Gaidatzis et al. [55]) uses a Bayesian framework to model target site evolution and assess their conservation by evaluating the probability for each seed-match to be conserved more than the background. The conservation pattern of each seed-match is computed as the presence/absence in the 3′-UTRs sequence pairwise alignments of the considered species. These frequency patterns are compared with the frequency pattern of random patterns of the seed length. Finally, for each putative target site, the posterior probability that the site is functional given its conservation pattern is computed.

Target sites are, from a probabilistic point of view, rare events in the 3′-UTR sequence. The binomial distribution is, therefore, an appropriate probability distribution for computing the statistical significance of target site presence. PACMIT (Marín and Vanícek [56]) uses this method and adds a seed-match accessibility criteria. It computes the statistical significance of accessible target sites.

The DIANA-microT algorithm was first presented in 2004 (Kiriakidou et al. [57]) with refinement published at a later time (Maragkakis et al. [58] and Maragkakis et al. [59]). A minimal length of 7 nucleotides is required by DIANA-microT to select potential target sites. Shorter seeds are kept if their MFE is below a defined threshold. A conservation score is attributed to every potential target site if found at the same position in a multiple species sequence alignment. The final score is the ratio between this conservation score and the score of the randomized miRNA. Multiple target site scores on a 3′-UTR are summed with higher weights for longer sites (Maragkakis et al. [58]). In the latest DIANA-microT version (version 4), mock miRNAs are replaced by high throughput proteomics experimental data (Selbach et al. [29]).

The tools described above have many aspects in common. They all have a limited number of features, are mostly *de novo* predictors, and rely on only experimental data for a statistical evaluation of the significance of their predictions (RNAhybrid for example) or for benchmarking (PITA or PACMIT for example). Another approach is to include as many as possible criteria that can together potentially distinguish a functional from a non-functional target site, and to use a machine learning algorithm to weigh and combine the features. The most commonly used algorithm for this purpose is the SVM. Even if SVMs are an efficient machine learning tool, in the case of miRNA target prediction, they are not the most appropriate. Training an SVM requires both a positive (miRNA targets) and a negative (miRNA non-targets) datasets. Given the definition of functional miRNA targets is not completely determined, defining a negative dataset for training remains difficult. Moreover this framework ignores the fact that miRNA repression strength is continuous, ranging from strong to weak effects, making the distinction between target versus non-target a matter of an arbitrary cut-off choice. Contrary to machine learning methods, the other methods that use a negative dataset are less dependent on it, as the negative dataset is used for example to determine the statistical significance of a previously computed score. However, many tools use this approach, two of the most promising are described below: the Saito tool (Saito and Sætrom [34]) and SVMicrO (Liu et al. [60]).

Both methods first predict individual target sites with a large number of features, all of which are already described and published. The Saito tool has 24 features covering the seed type and length, position in the 3′-UTR, mRNA accessibility and conservation. SVMicrO covers every pairing position in the miRNA, seed type and conservation, position in the 3′-UTR and mRNA accessibility with 113 features. Most of the SVM-based prediction tools only have this first target site level SVM. Interestingly, both tools also have a UTR level SVM that provide the means to combine each target site score into a global score for the whole

mRNA. The Saito tool UTR level SVM has 17 features, SVMicrO has 30. Both have features with site numbers, one for each seed type, and 3'-UTR length. SVMicrO has features with target site densities. The tools are different by how they include each target site score produced by the first SVM. SVMicrO only adds features with counts of sites in a top category whereas the Saito tool represents the score distribution in a 16-cell vector. The boundaries of each cell are appropriately chosen to increase the performance prediction; they are closer for higher scores.

Most of the source code of the miRNA target prediction tools is not available. In a recent review (Saito and Saetrom [61]), over the 30 tools examined by the authors, only 7 have software available. Of the described tools here, only the source code of miRanda, PITA, RNAhybrid, SVMicrO, and TargetScan (with some delay after manuscript publication) are available.

3.2.2 Existing miRNA target prediction tools: Performance

The unavailability of the source code for most tools makes performance comparison difficult, as any comparison will rely on pre-computed predictions (for example in Rajewsky [62]). Up to 30% of the differences among the prediction tools could be due to the mapping step mandatory to any comparison (Nikolaus Rajewski, personal communication), as the predictions are based on different miRNA and mRNA annotations. However, to limit such a bias, comparisons using only recent target prediction sets were performed. They cover the different aspects of miRNA repression: the recognition step with IP pull-down (**Figure 3**), the mRNA destabilization with transcriptomics (**Figure 4**), and the translation inhibition with proteomics experiments (**Figure 5**). Unsurprisingly, each tool is the best according to its authors (Diana-microT is slightly better than PicTar in Selbach et al. [29], but has less sensitivity with number of predicted targets divided by two compared to the other tools in the same comparison). According to these studies, the best performing tool overall is TargetScan, followed by PITA, PicTar and Diana-microT. SVM based tools seem to be efficient but independent benchmarking is not available. Moreover, their performance compared to non machine-learning methods is close, making it difficult to justify their added complexity.

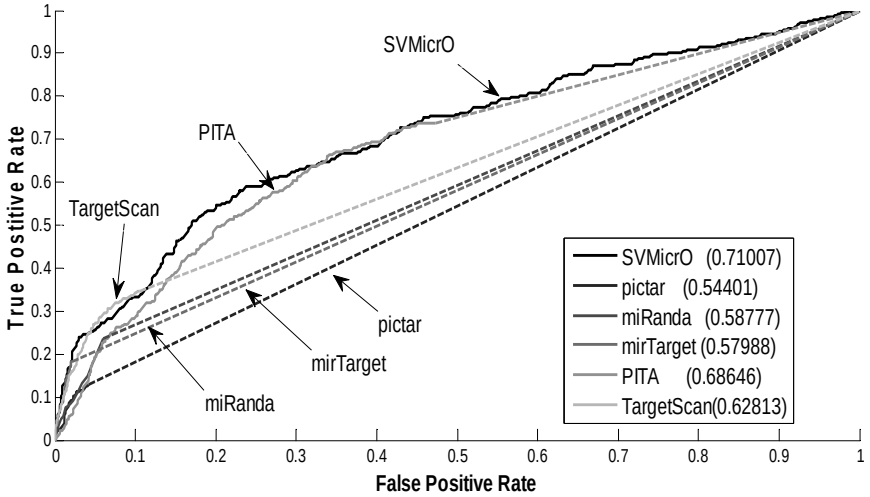


Figure 3 *Target predictions tested on an IP pull-down experiment.* ROC curves (Liu et al. [60]) based on 388 high confidence positive targets of miR-124 determined by IP pull down experiment (Hendrickson et al. [63]).

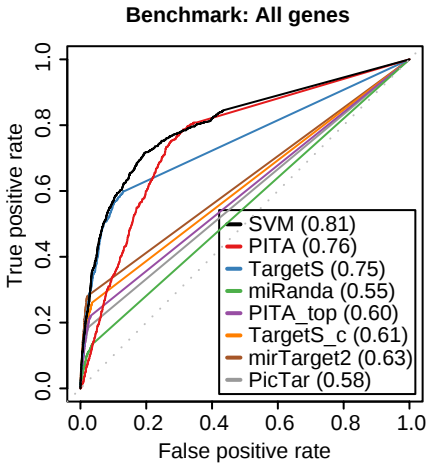


Figure 4 *Target predictions tested on a transcriptomics experiment.* ROC curves (Saito and Sætrom [34]) based on mRNA measurements from Linsley et al. [64]. The AUC (Area Under the Curve) is shown in parentheses.

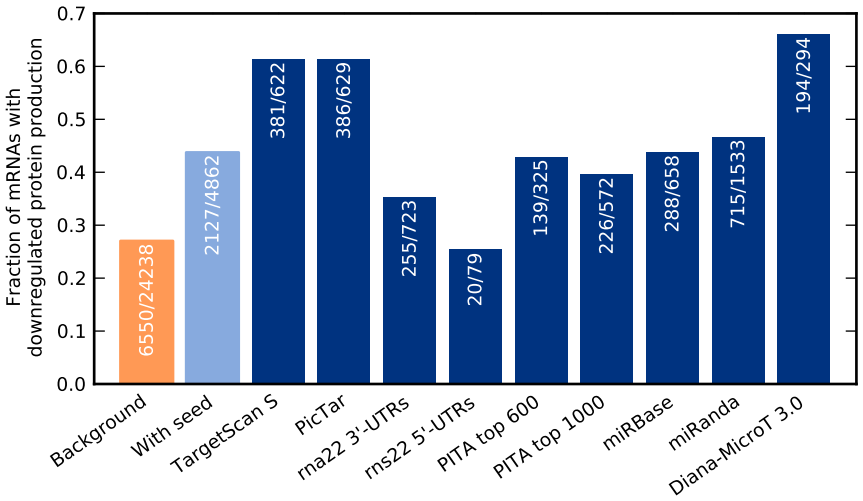


Figure 5 *Target predictions tested on a proteomics experiment.* The fraction of computationally predicted target mRNAs with reduced protein production (log2-fold change < -0.1) is calculated for the five miRNAs of the study (Selbach et al. [29]).

3.3 Thesis contributions

While the biological knowledge of the miRNA pathway remains incomplete, a substantial amount of information has accumulated over the years. Multiple approaches to predict miRNA targets can be developed based on this knowledge. Moreover, high-throughput experimental data are publicly available. Indeed, the effects of overexpressing or knocking-out a specific or all miRNAs has been measured at the genome-wide scale, opening the possibility to model these data with a bioinformatics approach. In other words, there is both a multiplicity of approaches to tackle the target prediction problem and of experimental datasets to parameterize and test these approaches. Most current studies choose to focus on a single aspect of the miRNA repression, the target recognition for example (Wen et al. [65]), and

test a limited set of appropriate prediction features with a limited appropriate experimental dataset.

The goal of this thesis work is to computationally predict the targets of miRNAs by developing a comprehensive set of predictive features and testing them on a complete range of miRNA repression assays. I developed the **miRmap** Python library covering thermodynamic, probabilistic, evolutionary, and sequence-based features. I evaluated their individual predictive power, and measured their intercorrelations on immunopurification, transcriptomics, proteomics, and polysome fractionation experimental data.

During the development of miRmap, I collaborated with three research groups to guide the choice of prediction features, and had direct access to non-cell-line-based experiments to test my predictions in biologically relevant situations. In these studies, a commonly used experimental scheme is the measurement of the changes, at the level of the transcriptome for example, in response to overexpression or depletion of one or all miRNAs. I developed an approach to statistically test the correlation between such perturbation of the miRNAome and the effect on mRNA transcriptome linked by my computationally predicted potent miRNA-mRNA interactions (**Figure 6**).

I first participated in the investigation of the role of the Dicer enzyme (miRNA processing, previously described in **Section 3.1**), at the transcriptome and proteome level in the mouse testis. The study including my analysis was published in the two following articles:

- Sertoli cell Dicer is essential for spermatogenesis in mice.
  Papaioannou MD, Pitetti JL, Ro S, Park C, Aubry F, Schaad O, **Vejnar CE**, Kühne F, Descombes P, Zdobnov EM, McManus MT, Guillou F, Harfe BD, Yan W, Jégou B, Nef S.
  Developmental biology. February 2009

- Loss of Dicer in Sertoli cells has a major impact on the testicular proteome of mice.
  Papaioannou MD, Lagarrigue M, **Vejnar CE**, Rolland AD, Kühne F, Aubry F, Schaad O, Fort A, Descombes P, Neerman-Arbez M, Guillou F, Zdobnov EM, Pineau C, Nef S.
  Molecular and cellular proteomics. April 2011

Some miRNAs are abundantly expressed in specific tissues (Lagos-Quintana et al. [66]), making their impact on the transcriptome *a priori* more measurable, and therefore an appropriate *in situ* test case for a miRNA target prediction tool. In our case, we studied two miRNAs transiently expressed in the liver for miR-122, and in dendritic cells for miR-155. I participated in the analysis of the experimental data, performing all miRNA target predictions. In particular, I presented statistical tests and controls to distinguish the effects of the miRNA regulation in the experiments. The study including my analysis was published in the two following articles:

- Integration of microRNA miR-122 in hepatic circadian gene expression.
  Gatfield D, Le Martelot G[*], **Vejnar CE**[*], Gerlach D, Schaad O, Fleury-Olela F, Ruskeepää AL, Oresic M, Esau CC, Zdobnov EM, Schibler U.
  Genes and development. June 2009
  [*]These authors contributed equally to this work.

- Silencing of c-Fos expression by microRNA-155 is critical for dendritic cell maturation and function.
  Dunand-Sauthier I, Santiago-Raber ML, Capponi L, **Vejnar CE**, Schaad O, Irla M,

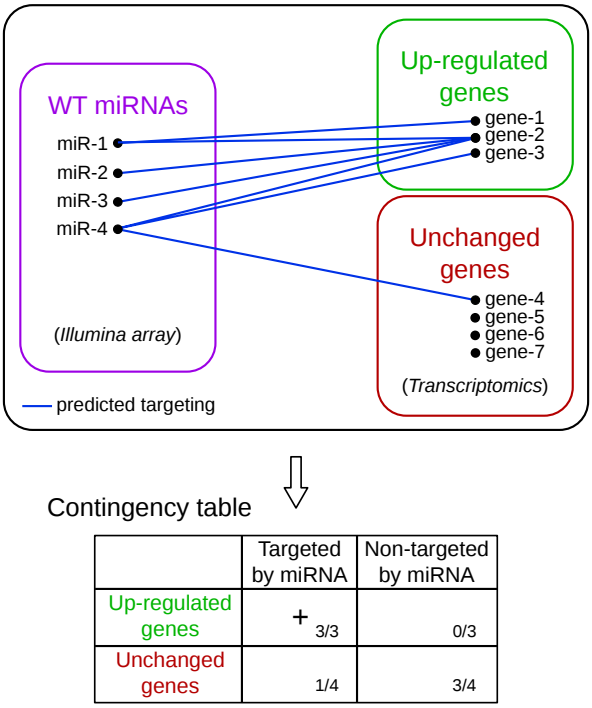


Figure 6 *miRNA target site enrichment analysis.* For all miRNAs (left), targets were predicted on the identified up-regulated or unchanged mRNAs (right) (each predicted relation is represented as a blue line). A contingency table was deduced from the targeting graph; proportions of the case depicted in this example are indicated in lower right corner of the cells. In the case of knock-down and knock-out described here, up-regulated mRNAs/proteins are expected to be targeted by miRNAs, which is marked by the "+" sign in the contingency table. This enrichment is tested on the contingency table with a one-sided Fisher test.

Seguín-Estévez Q, Descombes P, Zdobnov EM, Acha-Orbea H, Reith W. Blood. April 2011

In the miRmap library, I implemented published prediction features as well as three novel methods including (i) a more accurate way to compute the binding energy between the miRNA and the mRNA based on the ensemble free energy instead of the minimum free energy, (ii) an exact method to compute the probability that the seed-match is an over-represented motif in the 3′-UTR and (iii) a non-empirical statistical test to assess the significance of the target site evolutionary conservation.

Finally, I proposed a novel model for miRNA target predictions based on the linear combination of the eleven features implemented in my library. This model predicts the repression strength of each individual miRNA-mRNA relationship established with the seed pairing. The overall predictive power of my model appears to almost double that of the most renowned TargetScan software, and outperform PITA and PACMIT that are single and double feature tools respectively. I presented the miRmap library in the following article:

- miRmap: Comprehensive prediction of microRNA target repression strength.
  Vejnar CE, Zdobnov EM.
  In submission

In addition, I was involved in studies related to miRNAs but unrelated to miRNA target prediction that I include in this thesis as appendices. These studies were published in the two following articles:

- miROrtho: computational survey of microRNA genes.
  Gerlach D, Kriventseva EV, Rahman N, **Vejnar CE**, Zdobnov EM.
  Nucleic Acids Research. January 2009

- Identification of cis- and trans-regulatory variation modulating microRNA expression levels in human fibroblasts.
  Borel C, Deutsch S, Letourneau A, Migliavacca E, Montgomery SB, Dimas AS, **Vejnar CE**, Attar H, Gagnebin M, Gehrig C, Falconnet E, Dupré Y, Dermitzakis ET, Antonarakis SE.
  Genome Research. January 2011

3.4 Thesis outline

Following this introduction is a results part composed of the manuscripts described in the Thesis contributions. Each manuscript is preceded by an overview of the study and my analysis performed in the frame of that study. The last manuscript presenting miRmap closes the results part. Finally, the discussion part places the developments and results in perspective.

RESULTS 4

4.1 Impact of Dicer loss in Sertoli cells on the testicular transcriptome and proteome of mice

In the seminiferous tubules of the testis, the Sertoli cells (SCs) provide structural and nutritional support to the germ cells during adulthood, and play a critical role in the formation of the testis during embryonic development (Brennan and Capel [67]). Papaioannou et al. [68] investigated the implication of miRNA regulation in mouse SCs by knocking-out (KO) the RNase III Dicer, essential to miRNA processing (see introduction). The deletion of Dicer was done specifically in SCs with a Cre-LoxP recombination system, as the complete loss of Dicer is lethal at embryonic stage. The absence of Dicer in SCs with a full penetrance, starting at embryonic day E13.5, caused a complete loss of mature miRNA and had severe consequences on testis development leading to infertility. At post-natal day P5, massive apoptosis in the mutant testes was observed. At P60, the mutant testis had several defects including vacuolization, presence of Sertoli-cell-only (SCO) tubes, spermatogenic arrest and internal disorganization.

	% targeted genes			Fisher test p-value	
Fraction / Test	**Up**	**Unchanged**	**Down**	**Up vs Unchanged balanced**	**Down vs Unchanged balanced**
Genes	122	15236	135		
All seeds	98.3	89.2	88.9	10^{-4}	0.62
ΔG duplex	97.2	87.4	85.7	10^{-6}	0.61
Seed conservation	86.4	70.8	65.0	10^{-4}	0.98
TargetScan context score	98.3	89.1	88.9	10^{-4}	0.56
Above 3 filters	76.7	54.8	41.5	10^{-7}	1

Table 1 *Target site enrichment with different prediction filters for Dicer KO at the mRNA level.* The first three columns indicate the percentage of targeted genes in each gene fraction (up-regulated, unchanged and down-regulated). The last two columns provide the one-sided Fisher test p-values comparing the up- and down-regulated fractions with the unchanged fraction. Different features of miRmap were used to refine miRNA target predictions.

To investigate the consequences of Dicer loss at the transcriptome level, mRNA profiling was first performed on the wild-type and Dicer KO animals at P0 (new born animal) and P5. At P0, a large proportion of transcripts were expressed, as 67% of the about 45000 Affymetrix probe sets showed detectable levels of hybridization. With a stringent threshold of 2.0 on fold-changes between control and mutant testes, 77 probe sets were up-regulated and 68 were down-regulated at P0. To investigate the role of miRNA regulation, miRNA profiling was

performed on wild-type SCs, where a total of 248 miRNAs were cloned. As reported in the manuscript, comparisons of miRNA target distributions between up- and down-regulated mRNA fractions were not statistically significant.

The direct effects of miRNA regulation are expected to be detected on the mRNAs of up-regulated fraction as the ablation of Dicer removes all miRNA repression. Indirect effects can lead to up-regulation but in normal conditions only indirect effects can lead to down-regulation (Vasudevan et al. [69]). Based on this observation, I compared the percentage of targeted genes between the up-regulated and unchanged fractions, with a less stringent threshold of 1.5 on fold-changes. Considered globally, 98.3% and 89.2% of up-regulated and unchanged genes respectively have at least one seed-match for one of the miRNA expressed in wild-type testes, and are significantly different ($p=10^{-4}$ with Fisher's exact test). When I refined the potential target sites to those with lower energies, a conserved seed-match, a high TargetScan context-score (computed with miRmap with the 3 features of TargetScan), significant proportions of targeted genes were observed in the up-regulated fraction (**Table 1**). Interestingly, a very stringent potential target site refinement combining the three previous filters increased the ratio of targeted genes between up-regulated and unchanged fractions from 1.1 to 1.4. On the contrary, as a control of the previous tests, no significant differences were observed between the down-regulated and unchanged fractions.

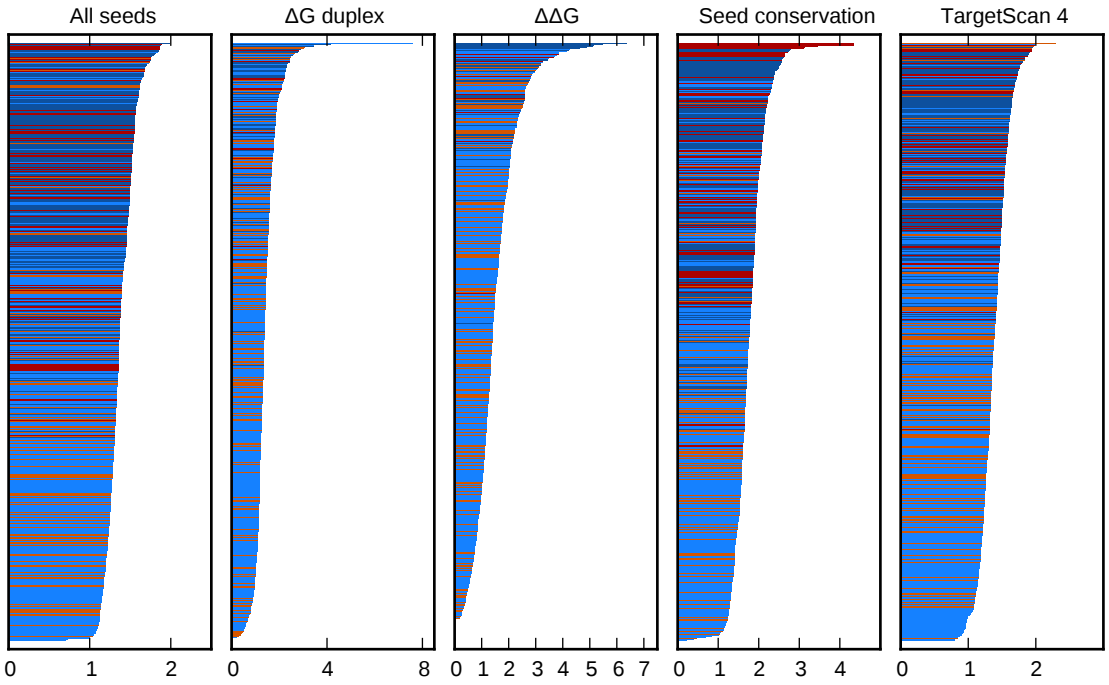


Figure 1 *Target site enrichment for miRNA target sites in the up-regulated mRNA fraction.* For each miRNA annotated in the mouse genome, ratios of targeted mRNA in up-regulated and unchanged fractions are represented. Significant ratios were determined with one-sided Fisher test with Dunn-Šidák multiple test correction at 10%. The color code is explained on **Table 2**.

Expression / Test	Non-significant	Significant
Expressed	Orange	Red
Non-expressed	Light blue	Dark blue

Table 2 *Color code.* Color code used in "all miRNAs enrichment" analyses.

The ratios of targeted genes between up-regulated and unchanged fractions were computed above for all miRNAs at the same time. To refine the analysis, target site enrichments were computed separately for each miRNA, either expressed in wild-type SCs or not, in order to observe differences between expressed miRNAs and non-expressed miRNAs used as controls (**Figure 1**). The individual effects of specific sets of miRNAs were not distinguishable.

To further investigate miRNA regulation at the translational level, proteome measurements were performed on 20 mutant and control P0 testes. Quantitative mass spectrometry allowed the detection of 130 proteins with a mutant to control ratio, of which 50 were up-regulated, 77 unchanged, and 3 down-regulated with a 1.3 fold-change threshold. With the same type of analysis, proportions of targeted genes were compared between up-regulated and unchanged fractions. Enrichments were statistically significant only when considering conserved seed-matches (**Table 3**).

	% targeted genes		Fisher test p-value
Fraction / Test	**Up**	**Unchanged**	**Up vs Unchanged**
Proteins	59	96	**balanced**
All seeds	93.2	90.6	0.40
ΔG duplex	77.8	65.2	0.072
Seed conservation	74.6	77.1	0.71
TargetScan context score	78.0	59.4	0.013

Table 3 *Target site enrichment with different prediction filters for Dicer KO at the protein level.* See **Table 1** for the column description.

Developmental Biology 326 (2009) 250–259

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/developmentalbiology

Sertoli cell Dicer is essential for spermatogenesis in mice

Marilena D. Papaioannou [a], Jean-Luc Pitetti [a], Seungil Ro [c], Chanjae Park [c], Florence Aubry [d], Olivier Schaad [b], Charles E. Vejnar [a], Francoise Kühne [a], Patrick Descombes [b], Evgeny M. Zdobnov [a], Michael T. McManus [e], Florian Guillou [f], Brian D. Harfe [g], Wei Yan [c], Bernard Jégou [d], Serge Nef [a,*]

[a] Department of Genetic Medicine and Development, University of Geneva Medical School, 1, rue Michel Servet, 1211 Geneva 4, Switzerland
[b] Genomics Platform, National Center of Competence in Research 'Frontiers in Genetics', University of Geneva, 1211 Geneva 4, Switzerland
[c] Department of Physiology and Cell Biology, University of Nevada School of Medicine, Reno, NV 89557, USA
[d] Inserm, U625, Université Rennes I, IFR140, GERHM, F-35042, Rennes, France
[e] Department of Microbiology and Immunology Diabetes Center, UCSF, CA 94143, USA
[f] Unité PRC, UMR 6175 INRA-CNRS-Université de Tours-Haras Nationaux, 37380 Nouzilly, France
[g] Department of Molecular Genetics and Microbiology, University of Florida College of Medicine, Gainesville FL 32610, USA

ARTICLE INFO

Article history:
Received for publication 6 September 2008
Accepted 17 November 2008
Available online 28 November 2008

Keywords:
Dicer
MicroRNAs
Sertoli cells
Germ cells
Testis
Spermatogenesis

ABSTRACT

Spermatogenesis requires intact, fully competent Sertoli cells. Here, we investigate the functions of Dicer, an RNaseIII endonuclease required for microRNA and small interfering RNA biogenesis, in mouse Sertoli cell function. We show that selective ablation of *Dicer* in Sertoli cells leads to infertility due to complete absence of spermatozoa and progressive testicular degeneration. The first morphological alterations appear already at postnatal day 5 and correlate with a severe impairment of the prepubertal spermatogenic wave, due to defective Sertoli cell maturation and incapacity to properly support meiosis and spermiogenesis. Importantly, we find several key genes known to be essential for Sertoli cell function to be significantly down-regulated in neonatal testes lacking *Dicer* in Sertoli cells. Overall, our results reveal novel essential roles played by the Dicer-dependent pathway in mammalian reproductive function, and thus pave the way for new insights into human infertility.

© 2008 Elsevier Inc. All rights reserved.

Introduction

Spermatogenesis refers to the development of mature haploid spermatozoa from diploid spermatogonial germ cells and ensures continuous gamete production throughout the adult life of males. Sertoli cells (SCs), one of the somatic constituents of the testis, have long been known to play an essential role in spermatogenesis. They extend from the base to the apex of the seminiferous epithelium, and are in direct physical association with all types of germ cells. During embryonic development, SCs play a critical role in the formation of the testis (for review see Brennan and Capel, 2004), whereas during adulthood they are entirely committed to sustaining spermatogenesis. Adult SCs provide germ cells with structural and nutritional support, assist their movement, produce seminiferous fluid and support spermiation (reviewed in Jegou, 1992). Importantly, one SC can support only a finite number of germ cells, therefore the ultimate adult testis size and eventual sperm production is directly linked to the total SC number (Orth et al., 1988). The latter is already established by around P15 in mice, when, after extensive proliferative activity, SCs cease dividing and switch from a fetal, 'immature' to an adult, 'mature' state. This maturation is characterized by radical morphological and functional changes, the most characteristic being the formation of the blood–testis barrier (BTB) at the level of adjacent SC tight junctions and the commitment of SCs to sustain germ cell progression through meiosis and differentiation into spermatozoa (reviewed in Mruk and Cheng, 2004; Sharpe et al., 2003). Thus, the adult spermatogenic outcome is not only dependent on the SC number, but also on their functional integrity.

Regulation of spermatogenesis at the post-transcriptional level, particularly during spermiogenesis, was earlier shown to be of crucial importance (reviewed in Braun, 1998). Recently, a novel mechanism of post-transcriptional regulation mediated by microRNAs (miRNAs) has emerged (for review see Pillai et al., 2007). MicroRNAs are endogenous, small (19–25 nucleotides), non-coding RNAs that act as negative post-transcriptional regulators of gene expression and control diverse aspects of development in several species. In animals, the majority of miRNAs regulate their target mRNAs by inhibiting their translation; however some may regulate their targets by inducing their degradation (Lim et al., 2005). Dicer is an RNaseIII endonuclease essential for miRNA processing; its deletion, which leads to a complete loss of mature miRNAs, is lethal at E7.5 in mice. Importantly, Dicer acts as a 'transcriptional

* Corresponding author. Fax: +41 22 379 5260.
E-mail address: Serge.Nef@medecine.unige.ch (S. Nef).

0012-1606/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2008.11.011

regulator' itself, since it plays a role in the structural maintenance and silencing of centromeres in murine ES cells (Kanellopoulou et al., 2005).

The functional relevance of Dicer and miRNAs in spermatogenesis is only starting to be unraveled. Several miRNAs are specifically expressed or enriched in the testis (Ro et al., 2007a; Yan et al., 2007) and all essential members of the RNA interference (RNAi) machinery (Drosha, Dicer, Ago2) are expressed in SCs, meiotic and postmeiotic germ cells (Gonzalez-Gonzalez et al., 2008; Kotaja et al., 2006). In fact Dicer was recently reported to be required for primordial germ cell development and spermatogenesis (Hayashi et al., 2008). The purpose of our study was to investigate the role of Dicer and miRNAs in SC function, and thereby their involvement in spermatogenesis. We found that male mice in which *Dicer* was deleted specifically in SCs were infertile, due to defective SC function preceded by down-regulation of SC-specific genes known to be essential for spermatogenesis. Our results demonstrate for the first time the crucial importance of Dicer—and thereby miRNAs in SC function, and thereby unravel the existence of post-transcriptional control in the supporting cell lineage of the testis.

Materials and methods

Animals

Dcr^{flox} (Dcr^{fx}) and Mis-Cre (Amh-Cre) mice were kindly provided by B. Harfe and F. Guillou respectively, and were genotyped as described (Harfe et al., 2005; Lecureuil et al., 2002). To achieve selective inactivation of *Dcr* in Sertoli cells, we mated transgenic *MisCre* female mice expressing Cre recombinase under the control of the *Mis* gene promoter with male mice carrying two floxed *Dcr* alleles in order to generate 50% $Dcr^{fx/wt}$;*MisCre* and 50% $Dcr^{fx/wt}$ mice. These animals were then intercrossed to produce $Dcr^{fx/fx}$;*MisCre* as well as control littermates, namely $Dcr^{fx/fx}$ and $Dcr^{fx/wt}$;*MisCre* mice. The genetic background of these mice is a mixed C57BL/6J and SV129. Protocols for the use of animals were approved by the Commission d'Ethique de l'Expérimentation Animale of the University of Geneva Medical School and the Geneva Veterinarian Office.

Fertility tests and sperm analysis

$Dcr^{fx/fx}$;*MisCre* males ($n=8$) and control littermates ($Dcr^{fx/wt}$;*MisCre*, $n=5$ and $Dcr^{fx/fx}$, $n=5$) were each bred with two 6-week-old wild type C57BL/6J female mice for 6 months. The number of litters and pups/litter were systematically recorded. Epididymal sperm count was performed with sperm extracted from the caudal epididymis and ductus deferens of adult (P60) male mice and was analyzed for its concentration as previously described (Guerif et al., 2002).

Histology and immunohistochemistry

Tissues were fixed overnight either in 4% paraformaldehyde (PFA) or in Bouin's fixative and embedded in paraffin. Five-µm sections were stained with haematoxylin and eosin (H&E) or processed for immunohistochemistry (IHC). For IHC analysis, PFA-fixed sections were incubated overnight at 4 °C with the following antibodies: anti-GATA4 (sc-9053, Santa Cruz Biotechnology, 1:50), anti-ZO1 (#61-7300, Zymed, 1:250), anti-β galactosidase (ab9361, Abcam, 1:500), anti-MVH (1:1000, gift from Toshiaki Noce) and anti-3β-HSD (1:500, gift from Ian Mason). For fluorescent staining, Alexa-conjugated secondary antibodies (Invitrogen) were used for signal revelation, whereas for stable staining, signals were revealed with DAB (Sigma). All images were obtained with a Zeiss Axioscop microscope and processed using the AxioVision software. For X-gal coloration tissues were fixed in 4% PFA for 2 h, immersed in PBS1×/sucrose 25% and then stained with X-gal (1 M MgCl2, 10% sodium deoxycholate, 10% NP40, X-gal 20 mg/ml, 200 mM potassium ferricyanide, 200 mM potassium ferrocyanide) overnight at 37°.

Proliferation and apoptosis assays

Fifteen regions from 3 different animals per genotype were randomly selected to count proliferating or TUNEL (TdT-mediated X-dUTP nicked labeling)-positive cells. The proliferation assay was performed with PFA-fixed sections double stained with anti-GATA4 (sc-9053, Santa Cruz Biotechnology, 1:50) and anti-Ki67 (BD, 1:100) overnight at 4 °C. Values were expressed as the percentage of proliferating SCs over the total number of SCs counted in a given region. Apoptotic assays were performed both by means of TdT-mediated X-dUTP nicked labeling (TUNEL) reaction using the *In Situ* Cell death kit (Roche) and double IHC using anti-cleaved caspase3 (1:200, Cell Signaling, #9661L) and anti-GCNA1 (1:50, gift from G. Enders) so as to reveal the identity of TUNEL-positive cells. The percentage of apoptotic, TUNEL positive cells within seminiferous tubules was expressed as the average number of apoptotic cells within 20 seminiferous tubes.

Microarray analysis

Total RNAs from 3 control ($Dcr^{fx/fx}$) and 3 mutant ($Dcr^{fx/fx}$;*MisCre*) P0 and P5 pairs of testes were extracted individually using the RNeasy Micro kit (Qiagen) according to the manufacturer's protocol, and their quality was assessed using Agilent Biosizing Total RNA NanoChips. To minimize biological variability, mutant and control pups originated from the same litters. Briefly, for each of the 12 independent samples, 1 µg of total RNA was reverse transcribed and amplified using the MessageAmpTM II-Biotin Enhanced Single Round aRNA Amplification Kit (#1791, Ambion). For each probe, 20 µg of the amplified biotinylated cRNA was fragmented and hybridized to Mouse Genome 430 2.0 Arrays (Affymetrix, High Wycombe, UK) as described (Cederroth et al., 2007). All microarray data are available through ArrayExpress (http://www.ebi.ac.uk/arrayexpress/, accession #E-TABM-426).

Classification and functional analysis of genes

Classification of differentially expressed genes was performed using the Ingenuity Pathways Knowledge Base (Ingenuity Systems, www.Ingenuity.com), based on their involvement in diverse biological processes. In short, a data set containing the Affymetrix gene identifiers, their corresponding expression values and *p*-values, was used to map to the corresponding gene object in the Ingenuity Pathways Knowledge Base. A fold-change cutoff of at least 2 was set between control and mutant testes to further filter genes whose expression was significantly altered. These genes, called "focus genes", were used as the starting point for generating biological networks. To start building networks, the program queries the Ingenuity Pathways Knowledge Base for interactions between focus genes and all other gene objects stored in the knowledge base, and generates a set of networks. IPA then computes a score for each network based on how well it fits to the set of focus genes. The score is derived from a *p*-value and indicates the likelihood of the Focus Genes in a network being found together due by chance. Scores of 2 or higher represent a 99% confidence level. Biological functions are then calculated and assigned to each network.

MicroRNA expression profiling

Purification of P6 SCs, small RNA isolation and cloning were performed as described (Ro et al., 2007a). MicroRNA profiling on purified spermatogenic populations was performed using quantitative PCR as described (Ro et al., 2006). Oligonucleotides used for qRT-PCR are listed on Supplementary Table 4.

MicroRNA target recognition analysis

The prediction model employed for the identification of potential miRNA–target interactions is similar to that of Kertesz et al., (2007) and relies on: (i) the initial identification of seeds for miRNAs, followed by (ii) the evaluation of the free energy gain (ΔΔG) resulting from the formation of the miRNA–target duplex, which takes into consideration the competing internal mRNA structures, and requires at least 10 nucleotides upstream and 15 downstream of the target site to be unfolded. For seed identification, we used standard parameters, requiring seed length to be 6–8 bases from position 2 of the miRNA, and not allowing mismatches except a single G:U wobble in 7-mers and 2 G:U in 8-mers. The stringency cut-off we used yields over 80% specificity as estimated from published luciferase assays.

Real-time quantitative PCR

Total RNAs from 6 control ($Dcr^{fx/fx}$) and 6 mutant ($Dcr^{fx/fx}$;*MisCre*) testes at P0 and P5 were extracted using the RNeasy Micro Kit (Qiagen) according to the manufacturer's protocol. Total RNAs for each of the 24 independent samples were reverse transcribed and 1/40th of the cDNA was used as template for PCR amplification as previously described (Cederroth et al., 2007). The statistical significance of fold-changes was determined by a paired Student's *t*-test. Primers used for qRT-PCR are listed in Supplementary Table 5.

Results

Complete and specific elimination of miRNAs in Sertoli cells of $Dcr^{fx/fx}$;MisCre testes

To investigate the *in vivo* role of *Dicer* in SCs, mice bearing two loxP-flanked alleles of *Dicer* ($Dcr^{fx/fx}$) (Harfe et al., 2005) were crossed with mice carrying the Mis-Cre transgene (Lecureuil et al., 2002), and their progeny were then intercrossed to obtain males in which *Dicer* was specifically inactivated in SCs ($Dcr^{fx/fx}$;*MisCre*), as well as control $Dcr^{fx/wt}$;*MisCre* and $Dcr^{fx/fx}$ littermates. *Mis*-driven *Cre* recombinase has been reported to efficiently delete floxed alleles specifically in SCs from E14.5 onwards (Lecureuil et al., 2002; Vernet et al., 2006). We ourselves confirmed the specificity of Dicer ablation first by β-gal immunohistochemistry (IHC) (Fig. S1). To further analyze the efficiency of Dicer removal in SCs we sought to assess whether the

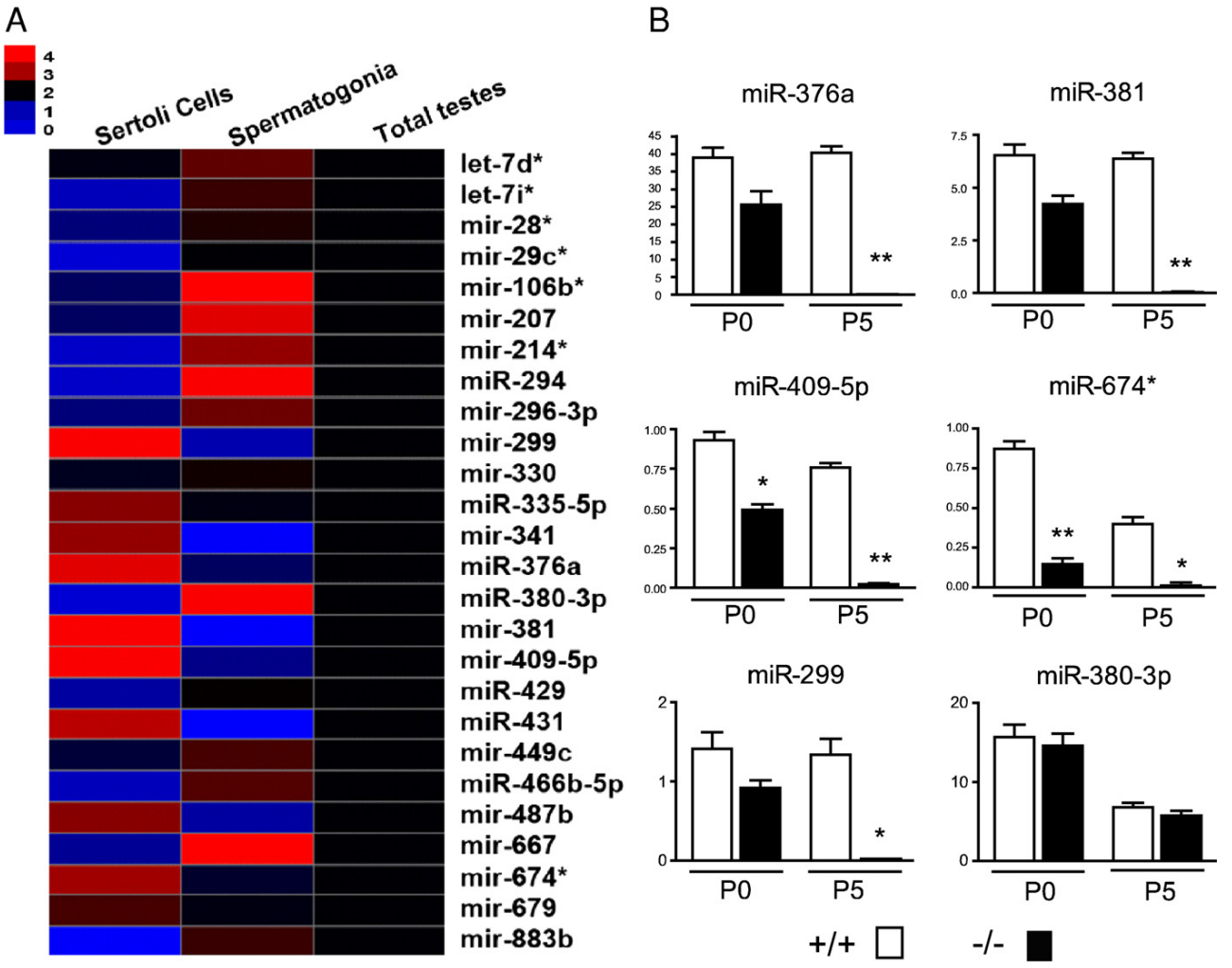


Fig. 1. Complete SC-specific *Dicer* ablation and subsequent SC-specific miRNA suppression using the *MisCre* recombinase. (A) Shown in this heat map are the expression levels of 26 representative miRNAs from purified mouse Sertoli cells, primitive type A spermatogonia and total testes at P6, as determined by quantitative RT-PCR. mir-299, mir-376a, mir-381, mir-409-5p, mir-674*, mir-431, mir-341 and mir-487b appear to be predominantly expressed in P6 Sertoli cells, with the first five being exclusively expressed in SCs. (B) Expression levels of 5 Sertoli cell-specific miRNAs (miR-376a, miR-381, miR-409-5p, miR-674*, and miR-299) were completely suppressed in P5 $Dcr^{fx/fx}$;*MisCre* (−/−) testes whereas those of a spermatogonia-specific miRNA (miR-380-3p) were unaffected. *$p<0.05$, **$p<0.01$, ***$p<0.001$ versus controls.

M.D. Papaioannou et al. / Developmental Biology 326 (2009) 250–259

downstream products of Dicer—microRNAs—were eliminated too. For this purpose, we first sequenced the small RNA-ome from purified mouse P6 SCs using a method previously described (Ro et al., 2007b) and identified a total of 248 miRNAs present in SCs (Supplementary Table 1), of which 5 were exclusively or predominantly expressed in SCs (miR-299, miR-376a, miR-381, miR-409-5p, and miR-674*, see Fig. 1B). Real-time PCR showed that the levels of these 5 SC-specific miRNAs were completely suppressed in P5 $Dcr^{fx/fx}$;*MisCre* testes, whereas spermatogonia-specific miRNAs were unaffected (Fig. 1B), thus further confirming the efficiency and specificity of *Dcr* excision. Interestingly, SC-specific miRNAs were reduced by about 2-fold in P0 $Dcr^{fx/fx}$;*MisCre* testes even though the Cre recombinase activity starts at E14.5. This suggests that pre-existing miRNAs in SCs are quite stable molecules and that they remain present within the cell for several days.

Ablation of Dicer in Sertoli cells results in reduced testis size and infertility

$Dcr^{fx/fx}$;*MisCre* males were viable, grew to adulthood normally and appeared to have normal sexual behavior and external genitalia when compared to control littermates. At P60, testes lacking Dicer in SCs showed a dramatic, 90%, mass reduction compared to control $Dcr^{fx/wt}$;*MisCre* and $Dcr^{fx/fx}$ littermates (10±3 mg versus 100±20 mg and 100±20 mg respectively, Figs. 2A–D). Testicular descent had occurred normally in $Dcr^{fx/fx}$;*MisCre* males and internal reproductive organs such as the seminal vesicles and the prostate were normally masculinized (data not shown). $Dcr^{fx/fx}$;*MisCre* males exhibited virile, androgen-dependent behavior including normal mounting and copulatory activity (data not shown), and importantly, their testicular testosterone levels at P0, P5 and P60 were not significantly different from those in controls, although interstitial hyperplasia was observed from P5 onwards (Fig. S1). Despite this, $Dcr^{fx/fx}$;*MisCre* males were infertile: during a 6-month mating period with wild type C57BL/6J females, *they* (n=8) failed to produce any offspring, whereas control littermates systematically sired offspring (n=8–12 pups/litter, 6–9 litters per mating cage). Sperm count analysis revealed that no spermatozoa were present in P60 $Dcr^{fx/fx}$;*MisCre* caudal epididymides, whereas normal sperm counts were found for control littermates (Fig. 2G). Histological analysis confirmed the complete absence of spermatozoa in testes (Fig. 2F) and epididymal ducts (Fig. 2I) of $Dcr^{fx/fx}$;*MisCre* males. Instead, numerous exfoliated germ cells were found in the lumen of mutant epididymal ducts (arrowhead in Fig. 2I).

Ablation of Dicer in SCs results in severely impaired spermatogenesis and age-dependent testis degeneration

Histological analysis revealed a severely impaired spermatogenesis and testis degeneration in P60 $Dcr^{fx/fx}$;*MisCre* mice; with rare exceptions, elongated spermatids were completely absent. More

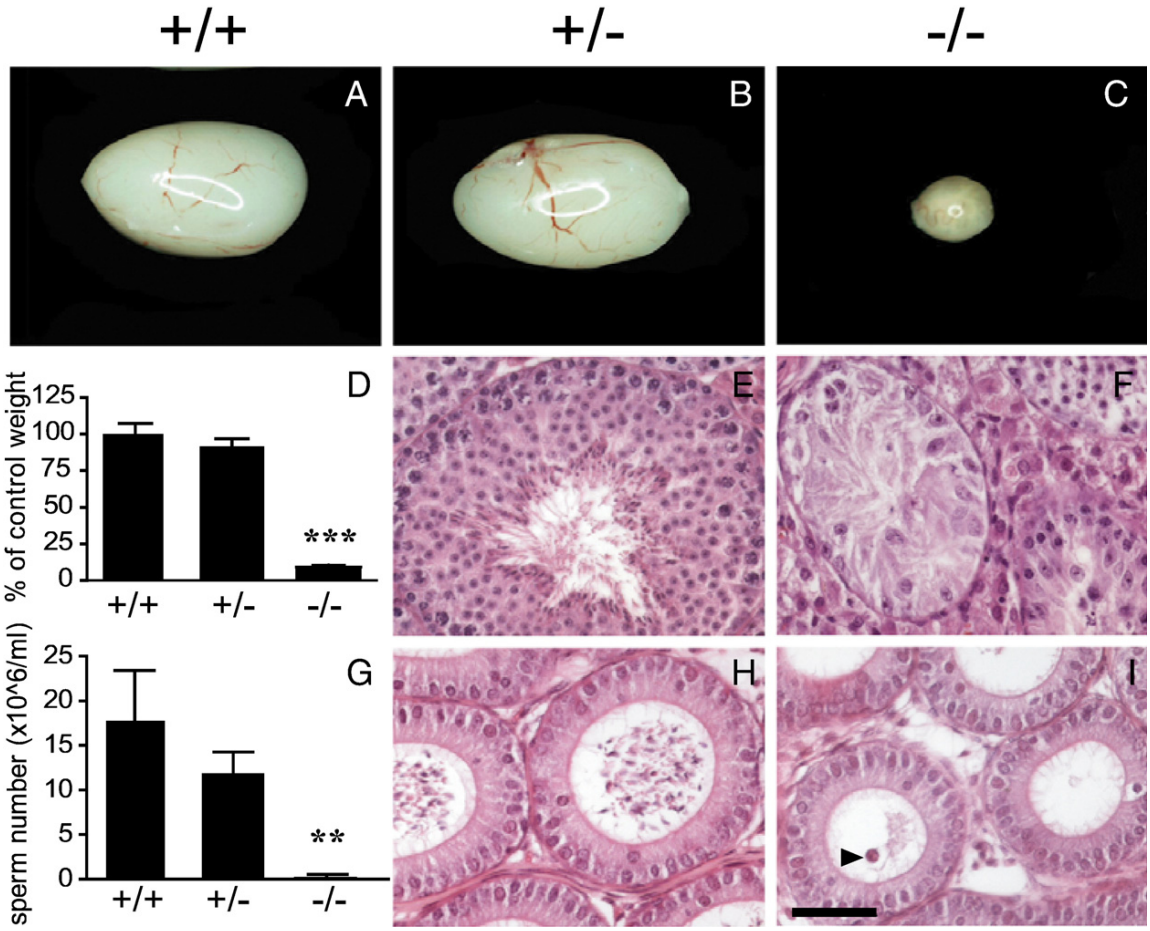


Fig. 2. Dramatic size reduction and complete absence of spermatozoa in $Dcr^{fx/fx}$;*MisCre* testes. At P60, testes from $Dcr^{fx/fx}$;*MisCre* mice (C) showed a drastic, 90% reduction (D, n=8–14 animals per genotype) in size compared to control $Dcr^{fx/fx}$ (A) and $Dcr^{fx/wt}$;*MisCre* (B) littermates. H&E staining of testes (E, F) and epidydimides (H, I) revealed complete absence of mature spermatozoa. Exfoliated germ cells were present in mutant epidydimal ducts (arrowhead in I). (G) Results of sperm count analysis (n=4–9 animals per genotype). +/+, +/− and −/− are abbreviations for $Dcr^{fx/fx}$, $Dcr^{fx/wt}$;*MisCre* and $Dcr^{fx/fx}$;*MisCre* animals respectively. Results are mean±SEM, *p<0.05, **p<0.01, ***p<0.001 versus controls. Scale bar: 50 μm.

precisely, the seminiferous epithelium of mutant testes displayed a wide range of defects, including vacuolization (Fig. 3A), presence of Sertoli-cell-only (SCO) tubes (Fig. 3B), tubes arrested at an early post-meiotic stage (Fig. 3C), as well as tubes which were not only smaller in diameter and devoid of a lumen, but also showed complete disorganization of the typical cell layering (Fig. 3D). The latter was confirmed by IHC: anti-GATA4 staining revealed the abnormal presence of SC nuclei in the center of tubes (Fig. 3F), whereas an anti-MVH antibody, which specifically labels the germ cell cytoplasm, revealed severe germ cell disorganization (Fig. 3H).

By 3 months of age (P90), degeneration was even more severe in *Dcr*$^{fx/fx}$;*MisCre* testes, with most of the tubes having formed their lumen, but with the majority of them having degenerated into SCOs with severely impaired SC morphology (Fig. 4D). In fact, staining with tight-junction-associated protein (TJP1, formerly known as zonula occludens 1, ZO-1)—a marker of SC tight junctions— was not only present at the basal lamina of mutant tubes but in the adluminal compartment too (SI Fig. 2F), suggesting defective SC cyto-architecture and polarity. By 6 months of age (P180), mutant testes' size was reduced to 5% of that of a control (Fig. S3A) and had degenerated into a mass of interstitial cells containing only rare remaining tubes (Fig. 4E), thus showing that testis degeneration aggravates upon aging. In fact, these sparse remaining tubes were almost completely devoid of germ cells, as confirmed by anti-MVH staining (Fig. 4F). These results show that SC Dicer is required both for SC survival and their capacity to support germ cell development.

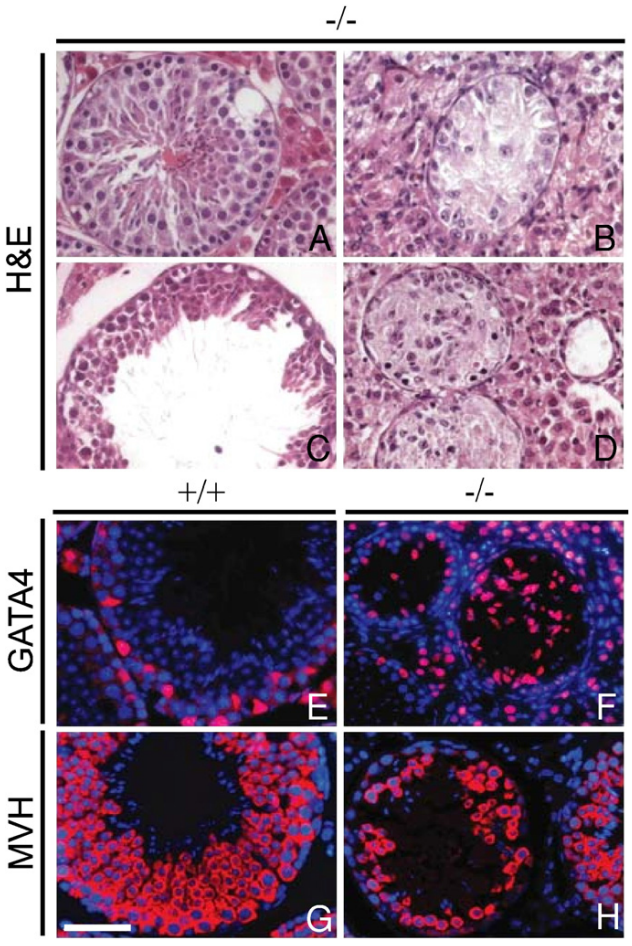


Fig. 3. Numerous spermatogenic defects in adult *Dcr*$^{fx/fx}$;*MisCre* testes. H&E staining of representative P60 *Dcr*$^{fx/fx}$;*MisCre* testis sections. Spermatogenic defects included vacuolization (A), Sertoli-cell-only (SCO) tubes (B), tubes with spermatogenic arrest (C) or disorganization of the seminiferous epithelium (D). Anti-GATA4 (red) staining (E,F) revealed abnormal positioning of SC nuclei (F), whereas anti-MVH (red) staining (G,H) revealed disorganization and reduction in germ cell number (H) in *Dcr*$^{fx/fx}$;*MisCre* testes. DAPI (blue) was used for nuclear staining (E–H). +/+ and −/− are abbreviations for *Dcr*$^{fx/fx}$, and *Dcr*$^{fx/fx}$;*MisCre* animals respectively. Scale bar: 50 μm.

Tubular abnormalities appear as early as P5 in Dcr$^{fx/fx}$*;MisCre males*

To unravel the events that led to infertility, we compared development of mutant and control testes from P0 to P42, when the first spermatogenic wave is completed. At birth (P0), *Dcr*$^{fx/fx}$;*MisCre* and control testes were morphologically indistinguishable (Figs. S4A–F). The first abnormalities began to appear at P5, about 8–9 days after the onset of the *Mis-Cre* transgene expression and when miRNAs were completely eliminated from SCs. At this stage, gonocytes resume proliferation, while moving towards the basement membrane (arrowhead in Fig. 5A). In *Dcr*$^{fx/fx}$;*MisCre* testes, two major defects were observed: first, instead of lying against the basement membrane, SC nuclei were mislocalized in the center of the tubes, as confirmed by anti-GATA4 staining (arrows in Figs. 5D, E). Second, numerous pycnotic cells were present within mutant tubes (arrowheads in Fig. 5D). By P15, an 80% reduction in the size of *Dcr*$^{fx/fx}$;*MisCre* testes compared to controls was observed (2.0±1 mg versus 20±2 mg respectively, $p<0.0001$, $n=6$–16 animals per genotype). Pachytene spermatocytes had not yet appeared and germ cell layering was severely perturbed (Fig. 5L). No tube had formed a lumen, and numerous pycnotic cells were found within (arrowheads in Fig. 5J). Several aspects of SC morphology suggested that these cells had remained immature: their nucleus did not display the characteristic irregular shape (compare insets in Figs. 5J, G), they were mostly mislocalized in the center of the tubes instead of the periphery (Figs. 5J, K) and TJP1 staining was more diffuse and discontinuous around the baso-lateral site of SCs compared to controls (Fig. S2D). At P21 (Figs. S4M–R), when secondary spermatocytes had appeared in control testes, and the first round spermatids were seen in some tubes, mutant testes continued to show a severe delay in meiosis and SC maturation, to harbor numerous pycnotic cells, and to display cellular disorganization. However, a few tubes had begun to form a lumen, suggesting that some SCs were partially functional. By P42, round spermatids were drastically reduced in number, and elongated spermatids were completely absent in *Dcr*$^{fx/fx}$;*MisCre* testes. Tubular disorganization and vacuolization had become severe (Figs. 5P–R), germ cells showed extensive apoptosis, and most of the tubes were SCOs, although notably, those which were less severely affected displayed a lumen.

Taken together, these data show that SC loss of Dicer severely impairs the prepubertal spermatogenic wave due to defective SC maturation, which includes dysfunctional secretory activity (absence of lumen), abnormal nuclear positioning and morphology and incapacity to properly support meiosis and spermiogenesis.

Increased SC proliferation followed by massive cell apoptosis in prepubertal Dcr$^{fx/fx}$*;MisCre testes*

The dramatic testis size reduction we observed as early as P15 led us to the assumption that the balance between cell proliferation and death had been perturbed. Surprisingly, as evidenced by double anti-GATA4/anti-Ki67 IHC, there was a significant increase in SC proliferation of P0, P5 and P15 *Dcr*$^{fx/fx}$;*MisCre* testes. At P0 and P5, we found respectively a 1.25-fold (63.9% versus 51.2%) and 1.5-fold (56.2% versus 36.4%) increase of SC proliferation in mutant testes compared to controls (Fig. 6A). At P15, when SCs normally cease dividing and become mature, we still observed a 2.6-fold (1.8% versus 4.6%) increase in proliferation in mutant testes, thus further suggesting a delayed SC maturation.

These results prompted us to assess cell apoptosis in mutant testes. Whereas at P0 no difference between controls and mutants was

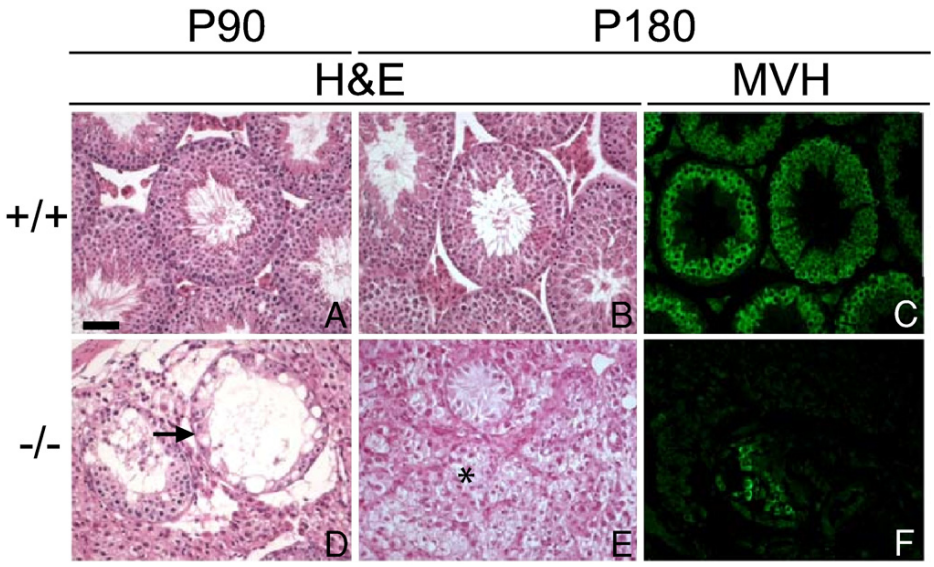


Fig. 4. Age-dependent testis degeneration in *Dcr*$^{fx/fx}$;*MisCre* mice. Control (A, B) and mutant (D, E) testis sections at P90 (A, D) and P180 (B, E) were stained with H&E. At P90, most tubes had become SCOs (tube pointed with arrow in panel D); by P180, an almost complete testicular degeneration was observed, with only 5–6 (per transverse section) tubes remaining, surrounded by a mass of interstitial cells (asterisk in panel E); in addition, these few remaining tubes were almost completely devoid of germ cells (F), as evidenced by anti-MVH staining (C, F). +/+ and −/− are abbreviations for *Dcr*$^{fx/fx}$ and *Dcr*$^{fx/fx}$;*MisCre* animals respectively. Scale bar: 50 μm.

found, a 26- and 7-fold increase in apoptosis was observed in P5 and P15 mutant testes respectively, compared to controls (Fig. 6B). To reveal the identity of these apoptotic cells, we performed double anti-cleaved caspase3/anti-GCNA1 IHC, so as to label apoptotic cells and germ cell nuclei respectively. At P5, cleaved caspase-3 positive cells (Fig. 6C) were almost exclusively GCNA1 negative (Fig. 6D), suggesting that the large majority of apoptotic cells in *Dcr*$^{fx/fx}$;*MisCre* testes were SCs (arrows in Fig. 6E). However at P15, both germ and SCs were found to be apoptotic (Figs. 6F–H), while at P21, apoptotic cells were almost exclusively germ cells (Figs. 6I–K). These findings suggest that the striking testis size reduction is a consequence of both the incapacity of SCs to sustain spermatogenesis and the subsequent germ cell death.

Ablation of Dicer in SCs causes alterations in the testicular transcriptome

To determine whether Dicer affects SC function at the transcriptional level, we performed a microarray analysis on control and mutant testes, just prior to (P0) and when the first morphological changes appear (P5). At these early postnatal stages, more than 80% of the tubular cells are SCs (Bellve et al., 1977), which minimizes the tissue's heterogeneity.

Of the 29,000 probe sets defined as present in mutant or control testes at P0, 77 were up-regulated (+2 fold) and 68 were down-regulated (−2 fold) (SI Table 2), whereas 787 and 796 probe sets were found to be up- or down-regulated respectively in P5 mutant testes (Supplementary Table 3). The variation in abundance of several key transcripts was further confirmed by quantitative RT-PCR (Fig. S7). Among the genes down-regulated in P0 mutant testes were *Glial-derived nerve factor* (*Gdnf*), *mannosidase IIx* (*Man2a2*), *serpin peptidase inhibitor* (*Serpin*A5), *Claudin11* and *Sox9*, all of which, when inactivated in mice, lead to diverse spermatogenic defects resulting in infertility. To this group of transcripts was added another set of down-regulated genes in P5 mutant testes including *Gata1*, *Kitl* (*SCF*), *Bclw*, *Dhh*, *Stra6*, *Wt1*, *Inhibin-β*, *connexin43*, *Jam2*, *Tjp2* and *Amh*, all known to be major regulators of testicular development or spermatogenesis (reviewed in Matzuk and Lamb, 2002). In contrast, when looking at the genes up-regulated in mutant testes—that could be direct targets of miRNAs—we found none that has been reported to be involved in testicular function, with the notable exception of Bcl2L11, a member of the proapoptotic Bcl-2 homology3-only protein family required for the elimination of supernumerary germ cells during the first spermatogenic wave (Coultas et al., 2005).

A hierarchical clustering (Fig. 7A) of the profile of the 145 probe sets showing a≥2-fold change in expression in P0 mutant testes revealed that the majority of the genes down-regulated at P0 remained so at P5. In contrast, the expression level of most of the genes upregulated in P0 mutant testes—with the exception of Bcl2L11—returned to normal levels at P5, a finding indicative of a transient upregulation. This also suggests that the "permanent" gene down-regulation is more likely to be responsible for the observed phenotype. We therefore used the Ingenuity Pathway Analysis (IPA) software to classify the differentially expressed genes in groups of common function. We found that genes critical for "cell signaling", "cell death", "organ development", "cellular development" and "tissue development" were the five most statistically significant functional groups affected by the SC loss of Dicer (Fig. 7B). Interestingly, each of these groups contained both up- and down-regulated genes—in approximately equal numbers—suggesting that the effect of Dicer loss is probably the result of a deregulation of both miRNAs and other factors downstream of Dicer itself.

Several studies have reported an enrichment of sequences complementary to miRNA seeds in the 3′-untranslated region (3′-UTR) of mRNAs that are upregulated upon *Dicer*—or specific miRNA—deletion. To assess if this was also the case for *Dcr*$^{fx/fx}$;*MisCre* testes, we compared the distribution of predicted miRNA target sites among transcripts that were up- or down-regulated in P0 and P5 *Dcr*$^{fx/fx}$;*MisCre* testes for the 248 miRNAs we cloned from purified P6 SCs. To predict likely functional miRNAs targets, we used 3 different bioinformatic approaches: when considering either all potential sequences complementary to the seed, or 1 seed per miRNA/per transcript, distributions between up- or down-regulated transcripts in P0/P5 mutant testes were not significantly different (Mann-Whitney test *p*-values 0.238 and 0.181 respectively). We also performed the test using a selection of potential functional seeds based on a thermodynamic model (Kertesz et al., 2007); distributions were again not significantly different (*p*-values of 0.155 and 0.112 respectively) although a slightly higher significance was found. These findings show that the 3′-UTR sequences of transcripts up-

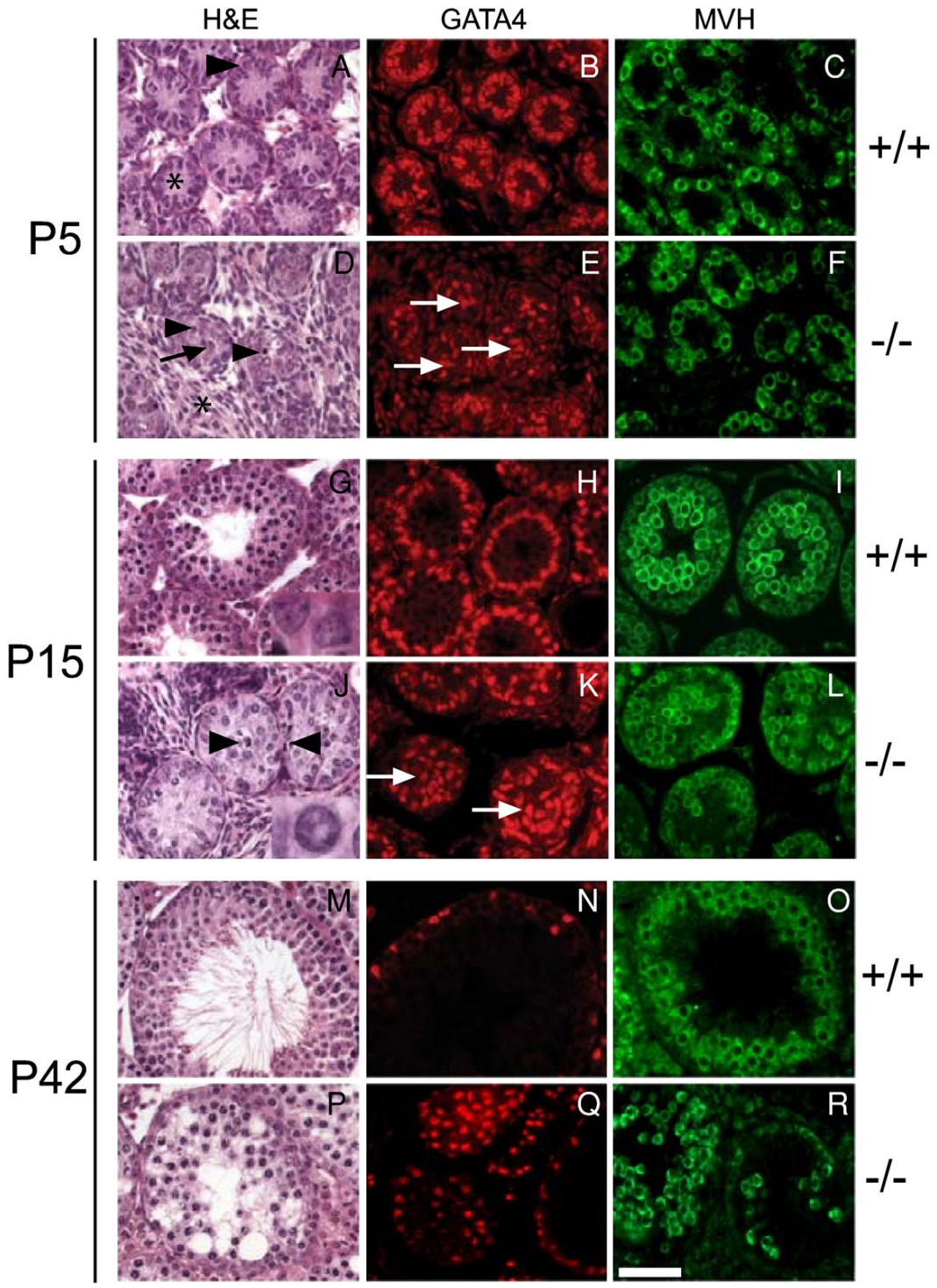


Fig. 5. Tubular defects appear as early as P5 in *Dcr*$^{fx/fx}$;*MisCre* mice. At P5, three major abnormalities were evident in *Dcr*$^{fx/fx}$;*MisCre* testes: interstitial hyperplasia (asterisk in panel D), pycnotic cells (arrowheads in panel D) within tubes, SC nuclear mislocalization (arrows in panels D and E) with almost complete absence of cytoplasm (compare tubule structure in panel A and D, SC cytoplasm in A marked with asterisk). At P15, SC nuclear mislocalization persisted (arrows in panel K) with the majority of them having an abnormal circular (inset in panel J) rather than flattened triangular (inset in panel G) shape, and germ cell disorganization was remarkable (L). By P42 almost all tubes had become severely vacuolized (P) and germ cell loss had become striking (R). +/+ and −/− are abbreviations for *Dcr*$^{fx/fx}$ and *Dcr*$^{fx/fx}$;*MisCre* animals respectively. Scale bar: 50 μm.

regulated in P0 and P5 *Dcr*$^{fx/fx}$;*MisCre* testes are not significantly enriched for miRNA-binding sites, and therefore suggest that—at least in SCs—miRNAs may have limited direct effects on target gene expression at the RNA level.

Finally, we assessed the expression levels of certain repetitive elements, since there is evidence for the implication of Dicer-dependent small RNAs in the repression of repetitive parasitic sequences (Murchison et al., 2007; Svoboda et al., 2004). Quantitative

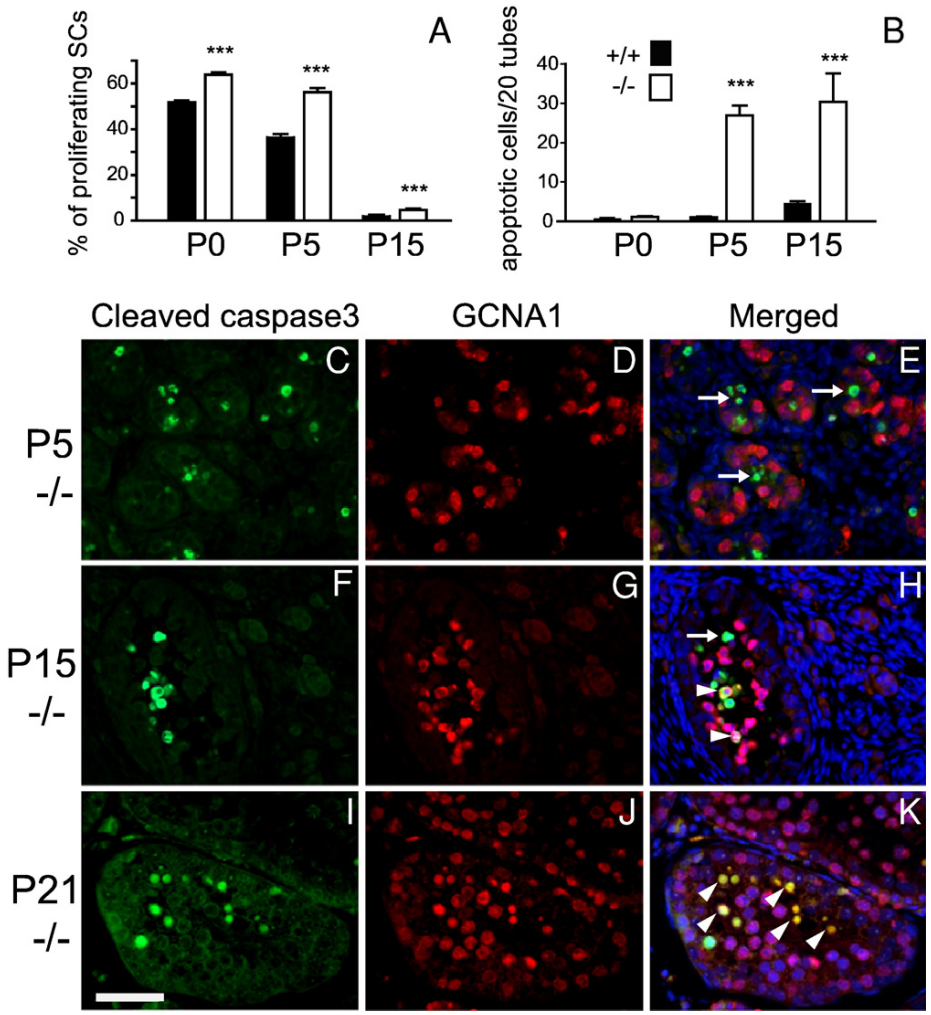


Fig. 6. Increased SC proliferation and massive cell apoptosis in prepubertal *Dcr*$^{fx/fx}$;*MisCre* testes. Percentages of proliferating SCs as revealed by double IHC staining with anti-GATA4 and anti-Ki67 at P0, P5 and P15 are plotted in panel A. Quantifications of TUNEL labeled cells revealed a dramatic increase in apoptosis in P5 and P15 mutant testes (B). Double IHC staining (C–K) with anti-GCNA (red) and anti-cleaved caspase3 (green) antibodies was performed on control (data not shown) and mutant (−/−) testis sections at P5 (C–E), P15 (F–H) and P21 (I–K). At P5, absence of GCNA1 and cleaved caspase3 co-localization is indicative of SC apoptosis (white arrows in panel E), whereas at P15 and P21, double GCNA1-cleaved caspase3 cells appeared (white arrowheads), thus revealing apoptotic germ cells. DAPI (blue) was used for nuclear staining (E, H, K). +/+ and −/− are abbreviations for *Dcr*$^{fx/fx}$ and *Dcr*$^{fx/fx}$;*MisCre* animals respectively. Results are mean±SEM (n=3 animals/ genotype/ stage), *p<0.05, **p<0.01, ***p<0.001 versus controls. Scale bar: 50 µm.

RT-PCR analysis revealed no significant differences in the abundance of transcripts for MT (mouse transcript), IAP (Intracisternal A particle element), Line1 (long interspersed nuclear elements), SineB1 and SineB2 (short interspersed repetitive elements) in P0 and P5 mutant testes compared to controls (Fig. S6). Thus, the *Dcr*$^{fx/fx}$;*MisCre* phenotype is probably not due to the derepression of these specific repetitive elements.

Discussion

Given the fundamental importance of gamete production for the perpetuation of a species, it comes as no surprise that spermatogenesis is a process tightly regulated at multiple levels, including the transcriptional and post-transcriptional levels. An issue that has not yet been addressed is whether post-transcriptional regulation of spermatogenesis also occurs in the somatic compartment of the testis. Here, we report that Dicer, a central component of the RNAi machinery, is essential for SC maturation, function and survival. Specific deletion of Dicer in SCs leads to infertility due to absence of mature spermatozoa and testis degeneration, thus highlighting the absolute necessity of Dicer for the development of fully competent SCs.

The first histological signs of the phenotype appeared a few days after birth, and led already before puberty to a gradual degeneration of the seminiferous epithelium's architecture. In fact, our expression profiling analyses revealed that although P0 control and mutant testes are morphologically indistinguishable, miRNA levels in SCs are reduced by approximately 50% and at the same time significant transcriptional alterations have already occurred in mutants. At P0, SC loss of Dicer affected the expression of not only numerous mRNAs specifically expressed in SCs, but also mRNAs in gonocytes (e.g. Serpin5a, TSLC1) and in LCs (e.g. Amhr2, PNMT). This suggests that despite no apparent testicular phenotype at birth, damages in SCs have already had important secondary effects on gene expression in the adjacent cell lineages, but more importantly, it confirms the long-standing notion of a close crosstalk between SCs and germ cells [reviewed in (Jegou, 1993)]. From P5 onwards, numerous testicular abnormalities appeared, the most prominent being a delay in SC maturation and a delayed entry into meiosis. Testicular degeneration worsened upon aging; by 6 months of age, mutant testes were

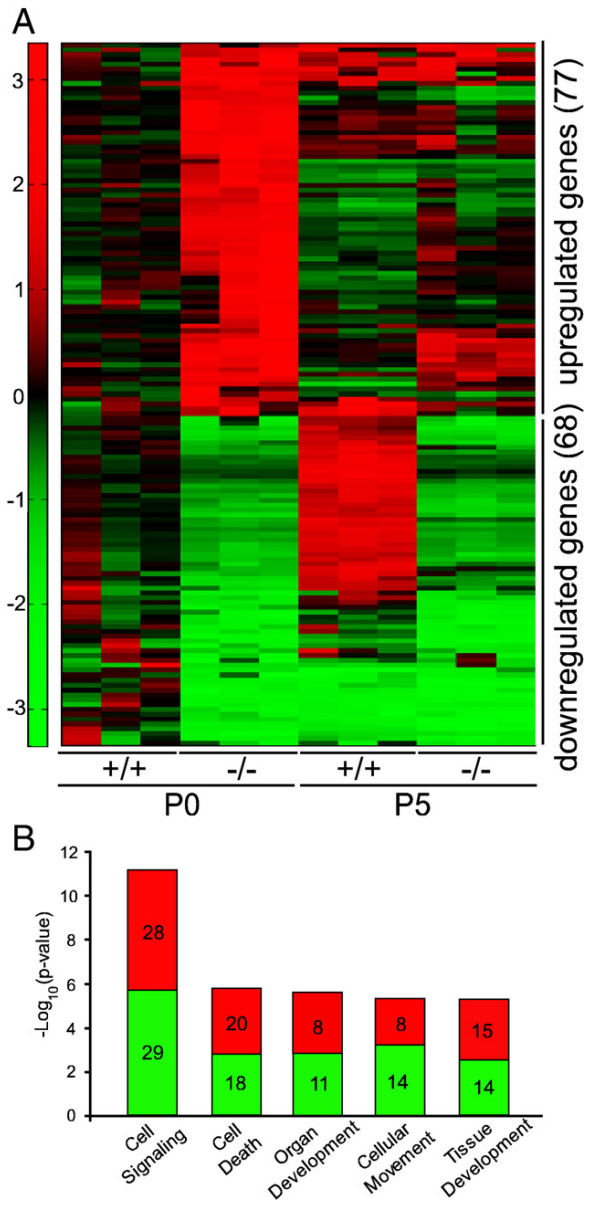


Fig. 7. Global expression analysis of the transcriptomes of control and mutant testes at P0 and P5. (A) Hierarchical clustering of the 145 probe-sets exhibiting a =|2| fold change in expression in P0 *Dcr*$^{fx/fx}$;*MisCre* (−/−) compared to *Dcr*$^{fx/fx}$ (+/+) testes. Each line represents a probe set and each column corresponds to a specific stage and genotype as indicated. Red and green colors indicate increased and decreased expression respectively. (B) IPA analysis revealed the five most statistically significant functional groups affected by the SC loss of Dicer. Each bar corresponding to a gene group is split in two (red for up-, green for down-regulated genes); numbers within are the numbers of modulated genes in each group.

composed almost exclusively of Leydig and fibroblast-like cells interspersed with rare—completely disrupted—tubes. Overall, our results suggest that the striking testis size reduction is very likely to have been the result of both (i) SC inability to support germ cell survival and spermatogenesis and (ii) SC death.

Increased cell death has in fact been observed in numerous conditional *Dicer* knockouts (Chen et al., 2008; Harfe et al., 2005; Harris et al., 2006), thereby raising the possibility that Dicer might be a "universal" regulator of cell survival. Specific ablation of *Dicer* in SCs was no exception since increased levels of SC apoptosis were detected as early as P5. At first glance, the dramatic phenotype of aging mutant testes could be simply explained by massive SC apoptosis leading to subsequent germ cell death. However, a closer examination of the phenotype suggests that Dicer has additional roles in regulating SC function, independent of its requirement for cell survival. Several of our findings are in support of this notion: (1) the significant alterations in gene expression at a time (P0) when histological defects and apoptosis are not yet detectable in mutant testes indicate that expression of genes essential for SC function is already affected and is not a consequence of cell death. (2) Apoptosis in mutant SCs appears gradually: some SCs die as early as P5, while some remain viable for several months and are at some level capable of supporting spermatogenesis, although they display other abnormalities (defective maturation, abnormal cellular architecture and polarization—as evidenced by TJP1 staining). Finally, (3) by 2 months of age, most remaining tubes are SCOs and composed of mutant, yet *viable* SCs which are *Dicer*-deficient, as confirmed by recombination of the Rosa26 stop LacZ marker (Fig. S1), thus showing that mutant SCs have lost their capacity to support germ cell survival, but remain viable themselves. These data reinforce our belief that Dicer should not be merely viewed as a global regulator of cell survival, and that the effects caused by its absence should not be interpreted solely on the basis of cell death.

Among the genes whose expression was downregulated in P0 and/or P5 mutant testes, were *Gdnf*, *Kitl*, *Man2a2*, *Gata1*, *Dhh*, *Serpin*A5, *Wt1*, and *Sox9*, all known to result in diverse spermatogenic defects when deleted *in vivo*. The significant reduction of *Kitl* and *Gdnf* levels was of particular interest; a balance between the Kitl/c-kit and GDNF/Ret signaling pathways is known to control the choice between spermatogonial differentiation and renewal (reviewed in Wong et al., 2005). It is thereby reasonable to assume that deregulation of this balance, notably a 5- and 3-fold reduction of *Gdnf* and *Kitl* respectively, could perturb the initial phase of spermatogenesis. It is possible that around P5, when spermatogonia resume proliferation and either renew themselves or differentiate, the reduction of Gdnf impairs their capacity to renew, whereas the reduction of Kitl negatively affects their capacity to differentiate. Defective spermatogonial renewal could lead to gradual germ cell loss, and could thereby explain the tubular degeneration we observed upon aging.

An essential question that emerges with the findings presented here is which biological activities mediated by Dicer are essential for SC function. Dicer is involved in a variety of gene-silencing phenomena at the transcriptional or translational level—through the activity of small RNAs—but is also required for the maintenance of chromatin structure (Kanellopoulou et al., 2005). Therefore, it is likely that the *Dcr*$^{fx/fx}$;*MisCre* phenotype is the result of (1) the deregulation of genetic elements that are directly under the control of Dicer itself, and/or (2) the deregulation of —direct or indirect—miRNA target genes. In the first case, loss of Dicer could lead to the up-regulation of genetic elements that are normally transcriptionally silent, which in its turn could affect the expression of other genes essential for spermatogenesis. In the second case, loss of miRNAs—subsequent to the loss of Dicer—could result in the up-regulation of direct miRNA target genes, which in their turn could deregulate the expression of other downstream factors. For the moment, our data favor the second option, since we found no significant change in the expression of a *selected* set of repetitive elements in mutant testes. A tempting hypothesis is raised by the fact that most genes that could actually be responsible for the observed phenotype were down-regulated in mutant testes. Taken together, our data suggest a model in which *Dicer* deletion leads to the gradual—but ultimately complete—disappearance of miRNAs in Sertoli cells, followed by a major transcriptome deregulation that could be the result of an alteration in the fine tuning of protein synthesis, such as the upregulation of potential transcriptional repressors. This hypothesis is supported by recent high-throughput proteomic analyses revealing that a single miRNA can

repress the production of hundreds of proteins but that this repression is relatively mild, rarely exceeding 4-fold (Baek et al., 2008; Selbach et al., 2008). We therefore hypothesize that the infertility phenotype observed in $Dcr^{fx/fx}$;*MisCre* mice is the indirect consequence of the downregulation of genes essential for Sertoli cell capacity to support germ cell survival and differentiation, such as *Gdnf*, *Kitl*, *Wt1*, and *Sox9*.

In conclusion, a better understanding of spermatogenesis is essential in order to become able to treat a rapidly increasing number of cases of male infertility. By demonstrating with our study that Dicer is essential for spermatogenesis, not only do we unravel a novel role for this gene, but we also provide new insights on the mechanisms controlling SC function and germ stem cell niche regulation in mammals.

Acknowledgments

We would like to thank Laurence Tropia for excellent technical assistance, Christelle Borel and Jean-Dominique Vassalli for critical reading of the manuscript. We are grateful to G. Enders, I. Mason and T. Noce for antibodies. S.N. has received a Swiss National Science Foundation Grant no. 3100A0-119862, and is also supported by the Prof. Dr. Max Cloëtta Foundation and the Société Académique de Genève. Wei Yan was supported in part by grants from the National Institute of Health (HD048855 and HD050281).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ydbio.2008.11.011.

References

Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., Bartel, D.P., 2008. The impact of microRNAs on protein output. Nature 455 (7209), 64–71.
Bellve, A.R., Cavicchia, J.C., Millette, C.F., O, Brien, D.A., Bhatnagar, Y.M., Dym, M., 1977. Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization. J Cell Biol. 74, 68–85.
Braun, R.E., 1998. Post-transcriptional control of gene expression during spermatogenesis. Semin. Cell Dev. Biol. 9, 483–489.
Brennan, J., Capel, B., 2004. One tissue, two fates: molecular genetic events that underlie testis versus ovary development. Nat. Rev. Genet. 5, 509–521.
Cederroth, C.R., Schaad, O., Descombes, P., Chambon, P., Vassalli, J.D., Nef, S., 2007. ER {alpha} is a major contributor to estrogen-mediated fetal testis dysgenesis and cryptorchidism. Endocrinology 148 (11), 5507–5519.
Chen, J.F., Murchison, E.P., Tang, R., Callis, T.E., Tatsuguchi, M., Deng, Z., Rojas, M., Hammond, S.M., Schneider, M.D., Selzman, C.H., Meissner, G., Patterson, C., Hannon, G.J., Wang, D.Z., 2008. Targeted deletion of Dicer in the heart leads to dilated cardiomyopathy and heart failure. Proc. Natl. Acad. Sci. U. S. A. 105, 2111–2116.
Coultas, L., Bouillet, P., Loveland, K.L., Meachem, S., Perlman, H., Adams, J.M., Strasser, A., 2005. Concomitant loss of proapoptotic BH3-only Bcl-2 antagonists Bik and Bim arrests spermatogenesis. EMBO J. 24, 3963–3973.
Gonzalez-Gonzalez, E., Lopez-Casas, P.P., Del Mazo, J., 2008. The expression patterns of genes involved in the RNAi pathways are tissue-dependent and differ in the germ and somatic cells of mouse testis. Biochim. Biophys. Acta 1779 (5), 306–311.
Guerif, F., Cadoret, V., Plat, M., Magistrini, M., Lansac, J., Hochereau-De Reviers, M.T., Royere, D., 2002. Characterization of the fertility of Kit haplodeficient male mice. Int. J. Androl. 25, 358–368.
Harfe, B.D., McManus, M.T., Mansfield, J.H., Hornstein, E., Tabin, C.J., 2005. The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb. Proc. Natl. Acad. Sci. U. S. A. 102, 10898–10903.
Harris, K.S., Zhang, Z., McManus, M.T., Harfe, B.D., Sun, X., 2006. Dicer function is essential for lung epithelium morphogenesis. Proc. Natl. Acad. Sci. U. S. A. 103, 2208–2213.
Hayashi, K., Chuva de Sousa Lopes, S.M., Kaneda, M., Tang, F., Hajkova, P., Lao, K., O, Carroll, D., Das, P.P., Tarakhovsky, A., Miska, E.A., Surani, M.A., 2008. MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis. PLoS ONE 3, e1738.
Jegou, B., 1992. The Sertoli cell. Baillieres Clin. Endocrinol. Metab. 6, 273–311.
Jegou, B., 1993. The Sertoli-germ cell communication network in mammals. Int. Rev. Cytol. 147, 25–96.
Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., Rajewsky, K., 2005. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev. 19, 489–501.
Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., Segal, E., 2007. The role of site accessibility in microRNA target recognition. Nat. Genet. 39, 1278–1284.
Kotaja, N., Bhattacharyya, S.N., Jaskiewicz, L., Kimmins, S., Parvinen, M., Filipowicz, W., Sassone-Corsi, P., 2006. The chromatoid body of male germ cells: similarity with processing bodies and presence of Dicer and microRNA pathway components. Proc. Natl. Acad. Sci. U. S. A. 103, 2647–2652.
Lecureuil, C., Fontaine, I., Crepieux, P., Guillou, F., 2002. Sertoli and granulosa cell-specific Cre recombinase activity in transgenic mice. Genesis 33, 114–118.
Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., Johnson, J.M., 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433, 769–773.
Matzuk, M.M., Lamb, D.J., 2002. Genetic dissection of mammalian fertility pathways. Nat. Cell Biol. 4, s41–s49 (Suppl.).
Mruk, D.D., Cheng, C.Y., 2004. Sertoli–Sertoli and Sertoli–germ cell interactions and their significance in germ cell movement in the seminiferous epithelium during spermatogenesis. Endocr. Rev. 25, 747–806.
Murchison, E.P., Stein, P., Xuan, Z., Pan, H., Zhang, M.Q., Schultz, R.M., Hannon, G.J., 2007. Critical roles for Dicer in the female germline. Genes Dev. 21, 682–693.
Orth, J.M., Gunsalus, G.L., Lamperti, A.A., 1988. Evidence from Sertoli cell-depleted rats indicates that spermatid number in adults depends on numbers of Sertoli cells produced during perinatal development. Endocrinology 122, 787–794.
Pillai, R.S., Bhattacharyya, S.N., Filipowicz, W., 2007. Repression of protein synthesis by miRNAs: how many mechanisms? Trends Cell Biol. 17, 118–126.
Ro, S., Park, C., Jin, J., Sanders, K.M., Yan, W., 2006. A PCR-based method for detection and quantification of small RNAs. Biochem. Biophys. Res. Commun. 351, 756–763.
Ro, S., Park, C., Sanders, K.M., McCarrey, J.R., Yan, W., 2007a. Cloning and expression profiling of testis-expressed microRNAs. Dev. Biol. 311, 592–602.
Ro, S., Song, R., Park, C., Zheng, H., Sanders, K.M., Yan, W., 2007b. Cloning and expression profiling of small RNAs expressed in the mouse ovary. Rna 13, 2366–2380.
Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R., Rajewsky, N., 2008. Widespread changes in protein synthesis induced by microRNAs. Nature 455 (7209), 58–63.
Sharpe, R.M., McKinnell, C., Kivlin, C., Fisher, J.S., 2003. Proliferation and functional maturation of Sertoli cells, and their relevance to disorders of testis function in adulthood. Reproduction 125, 769–784.
Svoboda, P., Stein, P., Anger, M., Bernstein, E., Hannon, G.J., Schultz, R.M., 2004. RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. Dev. Biol. 269, 276–285.
Vernet, N., Dennefeld, C., Guillou, F., Chambon, P., Ghyselinck, N.B., Mark, M., 2006. Prepubertal testis development relies on retinoic acid but not rexinoid receptors in Sertoli cells. EMBO J. 25, 5816–5825.
Wong, M.D., Jin, Z., Xie, T., 2005. Molecular mechanisms of germline stem cell regulation. Annu. Rev. Genet. 39, 173–195.
Yan, N., Lu, Y., Sun, H., Tao, D., Zhang, S., Liu, W., Ma, Y., 2007. A microarray for microRNA profiling in mouse testis tissues. Reproduction 134, 73–79.

Research

© 2011 by The American Society for Biochemistry and Molecular Biology, Inc.
This paper is available on line at http://www.mcponline.org

Loss of Dicer in Sertoli Cells Has a Major Impact on the Testicular Proteome of Mice

Marilena D. Papaioannou‡¶¶, Mélanie Lagarrigue§¶¶, Charles E. Vejnar‡¶, Antoine D. Rolland‖, Françoise Kühne‡, Florence Aubry‖, Olivier Schaad, Alexandre Fort‡, Patrick Descombes**, Marguerite Neerman-Arbez‡, Florian Guillou‡‡, Evgeny M. Zdobnov‡¶, Charles Pineau§‖, and Serge Nef‡§§**

Sertoli cells (SCs) are the central, essential coordinators of spermatogenesis, without which germ cell development cannot occur. We previously showed that Dicer, an RNaseIII endonuclease required for microRNA (miRNA) biogenesis, is absolutely essential for Sertoli cells to mature, survive, and ultimately sustain germ cell development. Here, using isotope-coded protein labeling, a technique for protein relative quantification by mass spectrometry, we investigated the impact of Sertoli cell-Dicer and subsequent miRNA loss on the testicular proteome. We found that, a large proportion of proteins (50 out of 130) are up-regulated by more that 1.3-fold in testes lacking Sertoli cell-Dicer, yet that this protein up-regulation is mild, never exceeding a 2-fold change, and is not preceeded by alterations of the corresponding mRNAs. Of note, the expression levels of six proteins of interest were further validated using the *A*bsolute *Quan*tification (AQUA) peptide technology. Furthermore, through 3′UTR luciferase assays we identified one up-regulated protein, SOD-1, a Cu/Zn superoxide dismutase whose overexpression has been linked to enhanced cell death through apoptosis, as a likely direct target of three Sertoli cell-expressed miRNAs, miR-125a-3p, miR-872 and miR-24. Altogether, our study, which is one of the few *in vivo* analyses of miRNA effects on protein output, suggests that, at least in our system, miRNAs play a significant role in translation control. ***Molecular & Cellular Proteomics 10: 10.1074/mcp.M900587-MCP200, 1–14, 2011.***

In all sexually reproducing organisms, germ cells (GCs)[1], in contrast to somatic cells, are the only cells that can give rise to a new organism; GCs give rise to the gametes—egg in females and sperm in males. Spermatogenesis refers to the development of mature haploid sperm from diploid spermatogonial cells within the testis of the male reproductive tract. It is typically divided in three strictly regulated phases, the mitotic, the meiotic, and the phase of spermiogenesis, which culminates with spermiation, the release of spermatozoa in the testicular seminiferous tubule's lumen. Spermatogenesis ensures continuous gamete production and occurs throughout adulthood in consecutive waves within the seminiferous tubules of the testis (reviewed in (1)). Apart from the GCs, which undergo spermatogenesis, the supporting cells of the testis called Sertoli cells (SCs), play a central role in the coordination of this process (for review see (2, 3)). SCs structurally and nutritionally support GCs and secrete factors that control, among other events, the survival and progression of GCs through the sequential steps of spermatogenesis (for example see (4–6)).

Post-transcriptional control plays an essential role in the regulation of spermatogenesis. During GC development, transcription and translation are un-coupled: transcription occurs massively following meiosis, with postmeiotic transcripts accumulating in large amounts, becoming deadenylated and stored in a repressed, dormant form in the spermatid cytoplasm for 4–5 days, whereas translation occurs at later stages (7). In addition to this "classic" mechanism, a novel system of post-transcriptional control mediated by microRNAs (miRNAs) is lately emerging with an important role during spermatogenesis ((8–10), and reviewed in (11)). miRNAs are endogenous, single-stranded, noncoding RNAs of ~22 nucleotides that act as post-transcriptional regulators of gene expression. They are generated through a multistep enzymatic process that involves the function of Dicer (Dcr), an RNaseIII endonuclease essential for the production of mature miRNAs (reviewed in (12)). miRNAs bind most frequently to the 3′UTR (3′ untranslated region) of target mRNAs, although recent studies show that some can also bind within the coding sequence (CDS) of mRNAs (reviewed in (13)), and depending on sequence complementarity, induce either mRNA degradation or translational repression of their target (for review see (14)). Importantly though, it has been reported that in some

From the ‡Department of Genetic Medicine and Development, University of Geneva Medical School and ¶Swiss Institute of Bioinformatics and **Genomics Platform, National Center of Competence in Research "Frontiers in Genetics," University of Geneva, 1211 Geneva 4, Switzerland; §Proteomics Core Facility Biogenouest, Inserm, U625, Campus de Beaulieu, Rennes, F-35042, France; ‖Inserm, U625, Univ Rennes I, IFR-140, GERHM, Campus de Beaulieu, Rennes, F-35042, France; ‡‡Unité PRC, UMR 6175 INRA-CNRS-Université de Tours-Haras Nationaux, 37380 Nouzilly, France

Received November 30, 2009, and in revised form, April 27, 2010
Published, MCP Papers in Press, May 12, 2010, DOI 10.1074/mcp.M900587-MCP200

[1] The abbreviations used are: GCs, germ cells; DCR, Dicer; ITMS, ion trap mass spectrometry; ICPL, isotope-coded protein labeling; miRNA, microRNA; SCs, Sertoli cells.

Effect of Sertoli-cell Dicer on Testicular Proteome

cases, miRNAs can also promote gene expression (15, 16), thus broadening even more their range of effects.

Although miRNA effects at the mRNA level have been frequently evaluated (for example (17–19)), their impact on protein output, which is thought to be the primary effect of animal miRNAs, has been, technically, more difficult to assess. One study used stable isotope labeling by amino acids in cell culture (SILAC) technology to investigate the effect of a single miRNA on protein output and reported that miR-1 can regulate a substantial percentage of the HeLa proteome (20). Only recently though two groups performed a *large-scale* protein analysis that unraveled the impact of miRNAs on protein output; both concluded that, in addition to down-regulating mRNA levels, a single miRNA can repress the production of hundreds of proteins, but that this repression is relatively mild (21, 22).

We previously generated a mouse model in which Dcr -and miRNAs- are eliminated uniquely in the SCs of the testis (10). We found that this ablation leads to complete infertility because of severe spermatogenic defects and gradual testicular degeneration; importantly, significant transcriptome (mRNA) down-regulation of genes such as *Gdnf*, *KitL*, *Man2a2,* and *Wt1*, all with essential roles during spermatogenesis, was detected upon SC-Dcr loss (10). Here, in order to investigate the impact of SC-Dcr loss at the proteome level, we performed ICPL (isotope-coded protein label) analysis (23), which allowed us, by means of MS, to relatively quantify proteins whose expression was altered between SC-Dcr-depleted ($Dcr^{fx/fx}$;MisCre, hereafter referred to as mutant) and wild-type ($Dcr^{fx/fx}$, hereafter referred to as control) testes. We found that more than a third of 130 quantified testicular proteins are up-regulated in mutant testes, yet at a relatively mild level, and that, importantly, this up-regulation does not reflect detectable changes in their respective mRNA levels. Of note, protein absolute quantification was achieved in independent experiments using the AQUA (Absolute QUAntification) peptide strategy (24) on a selected set of proteins and thus validated the results obtained through ICPL analysis. In addition, we identified *Sod-1*, a gene up-regulated at the protein level, as a direct *in vitro* post-transcriptional target of three SC-expressed miRNAs, miR-125a-3p, miR-872, and miR-24, we hypothesize that its up-regulation upon SC-Dcr and miRNA loss could account, partially, for the observed testicular degeneration. Globally, our findings further reinforce the current notion of animal miRNAs exerting one of their primary negative effects at the translational level, but most importantly, open new perspectives in studying the testicular proteome and its relation to the miRNA machinery.

EXPERIMENTAL PROCEDURES

Affymetrix Microarray Analysis—Microarray analysis is described in (10). All microarray data are available through ArrayExpress (http://www.ebi.ac.uk/arrayexpress/, accession number: E-TABM-426).

Protein Extraction and ICPL Labeling—Performing differential proteomics analysis using extremely small micro-dissected tissue samples is indeed a challenge. Considering in addition, the cost of knockout animals, experiments were only performed once. Protein extracts were prepared from 20 pairs of control and 20 pairs of mutant P0 (postnatal day 0) testes; for additional information on the generation of the *$Dcr^{fx/fx}$;MisCre* mouse strain, refer to (10). Tissues were homogenized by sonication on dry ice in a lysis buffer (6 M guanidine HCl, pH 8.5, tissue/buffer: 1/1.5(w/v)) and were then placed for 1 h at 4 °C before being centrifuged (15,000 × *g*, 30min, 4 °C). The resulting supernatants were then ultracentrifuged (105,000 × *g*, 1 h, 4 °C). Protein concentration of the resulting supernatants was measured with a bicinchoninic acid assay (Sigma-Aldrich) and was adjusted to 5 mg/ml by addition of lysis buffer. Disulfide bonds were reduced with 0.2 M tris(2-carboxyethyl)phosphine and then alkylated with 0.4 mM iodoacetamide. For each sample, 100 μg of proteins were labeled using the ICPL-kit (Serva Electrophoresis, Heidelberg, Germany) according to the manufacturer's instructions. Briefly, free amino groups (lysine residues and N-terminal NH_2) of proteins from control and mutant extracts were labeled at room temperature for 2 h with the light (^{12}C- nicotinoyloxysuccinimide) and the heavy (^{13}C- nicotinoyloxysuccinimide) ICPL reagents, respectively. Following quenching excess reagent with 6 M hydroxylamine, the two labeled samples were mixed, purified by acetone-precipitation (−20 °C, overnight), and subsequently dissolved in 20 mM HEPES. Labeled proteins (50 μg) were separated by SDS-PAGE on a 12% precast gel (Gebagel, Gene Bio-Applications, Interchim, Montluçon, France). The gel was subsequently stained with Coomasie blue R-350 using the EZBlue gel staining reagent (Sigma-Aldrich, Saint-Quentin Fallavier, France). The entire gel lane was cut into 20 bands, which were washed with different acetonitrile (ACN)/100 mM NH_4HCO_3 solutions. In-gel digestion was performed overnight at 37 °C with modified trypsin (Promega, Charbonnières-les-Bains, France). Proteolytic peptides were then extracted from the gel by sequential incubation in the following solutions: ACN/H_2O/TFA, 70:30:0.1 (v/v/v), 100% ACN and ACN/H_2O/TFA, 70:30:0.1 (v/v/v), and extracts were eventually concentrated by evaporation to a final volume of 30 μl.

Nano-LC-MS-MS Analysis—Proteolytic mixtures were separated on a nano-high performance liquid chromatography system (Ultimate 3000, Dionex, Jouy-en-Josas, France), with an injection volume of 22 μl: first, they were concentrated on a C18-PepMap trapping reverse phase column (5 μm, 300 Å/300 μm i.d. × 5 mm, Dionex), and were then eluted with a 75-min, 2–90% ACN gradient in 0.05% formic acid, at a flow rate of 250 nL/min. The nano-LC apparatus was coupled on-line with an Esquire HCT Ultra PTM Discovery mass spectrometer (Bruker Daltonik, GmbH, Bremen, Germany), equipped with a nanoflow electrospray ionization (ESI) source and an ion trap analyzer (ITMS). The mass spectrometer was operated in the positive ionization mode. The EsquireControl™ software (Bruker Daltonik, GmbH) automatically alternated MS and MS-MS acquisitions and was tuned to preferentially subject ICPL labeled peptides to MS-MS acquisitions.

Protein Identification and Relative Quantification—The DataAnalysis 3.4 software (Bruker Daltonik, GmbH) was used to create the peak lists from raw data. For each acquisition, a maximum of 500 compounds were detected with an intensity threshold of 250,000 and the charge state of precursor ions was automatically determined by resolved-isotope deconvolution. The Biotools 3.1 software (Bruker Daltonik, GmbH) was used to submit MS/MS data to the Swiss-Prot database (version 47, Mus musculus taxonomy, 568,851 sequence entries) using the Mascot algorithm (Mascot server v2.2; *http://www.matrixscience.com*). Given that modification of lysine residues by ICPL labeling prevents their cleavage by trypsin, arginine C was selected as *enzyme* with one allowed missed cleavage. In addition, carbamidomethylation of cysteins was set *as fixed modifications*, and labeling of lysine residues and of the N-terminal NH_2 group of proteins

by light or heavy ICPL reagents, as well as methionine oxidation were considered as *variable modifications*. The mass tolerance for parent and fragment ions was set to 1.2 and 0.5 Da, respectively. Peptide identifications were accepted if the individual ion scores were above 25 (the ion score is −10*log(P), where P is the probability that the observed match is a random event). Protein identifications were accepted if the score indicated identity or extensive homology, *i.e.* the probability that the identification is a random match was lower than 5%. Matches corresponding to the heavy and the light labeled forms of the same peptide counted for one peptide. Single peptide-based identifications were accepted because missed cleavages of labeled lysine residues leads to a global reduction in the number of peptides produced in comparison to "classical" trypsin digestion and to the formation of relatively long peptides that can single-handedly represent a sufficient percentage of protein sequence coverage.

MS/MS spectra were searched against a randomized sequence (decoy) database using Mascot to determine the false discovery rate defined as the number of validated decoy hits/(number of validated target hits + number of decoy hits)*100. Thus, a satisfactory false discovery rate of 1.15% was obtained for the totality of identifications acquired during ICPL analysis.

Relative protein quantification was obtained using the WarpLC 1.1 software (Bruker Daltonik, GmbH). This automatically calculates the heavy-to-light (H/L) ratios by comparing the relative intensities of the extracted ion chromatograms (EIC) that are reconstituted by extraction of the intensities of m/z ratios corresponding to the labeled peptides observed on MS spectra. For each EIC, the contribution of 1+, 2+, and 3+ charge states of the peptide was considered and smoothing was applied (one smoothing cycle with Gauss algorithm and a smoothing width of 3 s). For each protein, the H/L ratio was calculated by averaging the different H/L ratios obtained for each pair of labeled peptides.

In the present study, the amount of peptides obtained following peptide extraction from both sample pools was not enough to perform technical replicates. Reproducibility and accuracy of ICPL experiments performed by ESI-ITMS were evaluated in five independent technical replicates using a standard mixture of ICPL labeled proteins containing bovine serum albumin with a heavy-to-light ratio of 1:1 (ICPL-kit, Serva Electrophoresis). For bovine serum albumin, an average H/L ratio of 0.94 was obtained, very close to the theoretical value of 1, and corresponding to a variation coefficient of 8%.

AQUA Peptide Analysis—AQUA [$^{13}C_6$, $^{15}N_2$] K-Lysine-labeled and [$^{13}C_6$, $^{15}N_4$] R-Arginine-labeled peptides (listed in supplemental Table 1) were synthesized and quality- and quantity-controlled by Sigma-Aldrich. All AQUA peptide standard solutions were prepared from stock solutions at 5 pmol/μl according to the manufacturer's instructions. All samples used for AQUA peptide experiments were systematically prepared in low adsorption tubes (LoBind tubes, Eppendorf, Le Pecq, France) to minimize errors because of peptide adsorption (25). AQUA peptide standard solutions were prepared at 0.1, 0.2, 0.5, 1, 2 and 5 fmol/μl and were analyzed by nano-liquid chromatography (LC)-ESI-ITMS with an injection volume of 10 μl in three analytical replicates for calibration. For each AQUA peptide, the corresponding EIC area was automatically determined by the QuantAnalysis 1.8 software (Bruker Daltonik, GmbH) and plotted against the injected amount to obtain the calibration curves. The linearity of the response was verified for all AQUA peptides with correlation coefficients ranging from 0.988 to 0.996. In addition, the analytical repeatability of measurement of EIC areas corresponding to the different AQUA peptides was evaluated: satisfactory coefficient of variations (CV) ranging from 2% to 9% ($n = 6$) were obtained for an AQUA peptide concentration of 2 fmol/μl.

Protein extracts used for the ICPL experiment, that is, one sample of 20 pairs of control and one sample of 20 pairs of mutant testes, were also used for AQUA peptide analysis. Proteins from the control and mutant sample were precipitated with acetone overnight at −20 °C and dissolved in Laemmli buffer (Gene Bio-Application). The two samples were then independently separated by SDS-PAGE on a 12% precast gel (Gebagel). Following fixation, washing, and staining, both entire gel lanes were manually cut into 20 pieces. Disulfide bonds were then reduced with dithiotreitol and alkylated with iodoacetamide. Protein in-gel digestion and proteolytic peptide extraction from each gel band was then performed. In order to precisely control the final volume of proteolytic peptides, extracts were completely dried by evaporation and dissolved with 20 μl of H_2O/formic acid (95/5, v/v) solution, then with 20 μl of a H_2O/ACN/formic acid (95/5/0.2, v/v/v) solution and vigorously sonicated and vortexed. AQUA peptide standards were added in precise amounts to the samples just before nano-LC-MS analysis. For all AQUA peptide analysis, the EsquireControl software was operated in the Multiple Reaction Monitoring mode to specifically subject the labeled (AQUA peptides) and unlabeled (peptides from the sample) peptides to MS/MS fragmentations. Then, for each fragmented peptide, an EIC was reconstituted by extracting the signals corresponding to fragment ions specific to the peptide of interest. Absolute quantification was obtained by comparing the EIC areas of the unlabeled peptide and its corresponding AQUA peptide added in precise amount.

Real-Time Quantitative PCR—Total RNAs from six control and six mutant P0 testes were extracted using the RNeasy Micro Kit (Qiagen, Basel, Switzerland) according to the manufacturer's protocol. For each of the 12 individual samples, 1 μg of total RNA was reverse transcribed with the Superscript II Reverse Transcriptase (Invitrogen, Basel, Switzerland) according to the manufacturer's instructions, and 1/40 of the cDNA was used as template for Real-Time PCR amplification on a Freedom Evo 150 System (Tecan, Männedorf, Switzerland) using the Power SYBR Green PCR master mix (ABI, Foster City, CA). Raw threshold-cycle (*Ct*) values were obtained with the SDS 2.0 software (ABI). Relative quantities (*RQ*) were calculated with the formula $RQ = E^{-Ct}$, using efficiencies (*E*) calculated with the DART-PCR algorithm, as described (26). Mean quantities were calculated from triplicate PCR reactions for each sample, and were normalized to two similarly measured quantities of *Gapdh* and *Trf1R* as described (27). Normalized quantities were averaged for three replicates for each data point and represented as the mean ± S.D. The highest normalized relative quantity was arbitrarily designated as a value of 100.0. Fold changes were calculated from the quotient of means of these normalized quantities and reported as values ± S.E. The statistical significance of fold-changes was determined by an unpaired Student's *t* test. Primers used are listed in supplemental Table 2.

Spermatogenic Cell Purification—Mature Sertoli and peritubular myoid cells were prepared from 10 C57BL/6J males aged P16, whereas immature Sertoli cells were prepared from 16 C57BL/6J animals aged P6, as previously described (28). Germ cells were prepared using the STAPUT technique according to (29); spermatogonia were prepared from 40 C57BL/6J males aged P6–8, whereas pachytene spermatocytes and spermatids were prepared from six adult (P60) C57BL/6J mice. To verify cell purity, 5×10^5 cells were fixed in phosphate-buffered saline (PBS)/PAF 1% for 10′ at room temperature, washed in PBS and then conserved overnight at 4 °C in PBS/FCS 1%. Cells were then marked with propidium iodide (100 μg/ml) in PBS/0.2% saponin (30′, RT) and were sorted on a FacsCalibur machine (Beckton Dickinson, France), in order to quantify their contamination. Leydig cells were prepared from 16 adult (12-week-old) mice, as previously described (30); their purity was assessed by incubating cells with NAD (in Nitro Blue Tetrazolium, NBT, N-6876, Sigma-Aldrich) for 90' and quantifying the percentage of cells having acquired a violet color, indicative of the presence of 3β-HSD.

Effect of Sertoli-cell Dicer on Testicular Proteome

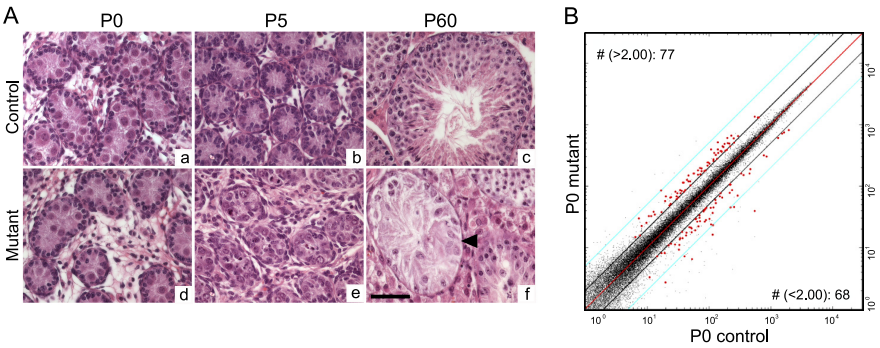


FIG. 1. **Morphological abnormalities appear by postnatal (P) day 5, yet mRNA transcripts are affected upon SC-Dcr loss already by P0.** *A*, Hematoxylin and eosin-stained paraffin sections from P0 (a, d), P5 (b, e) and P60 (c, f) testis sections of control ($Dcr^{fx/fx}$) (a, b, c) and mutant ($Dcr^{fx/fx}$;*MisCre*) (d, e, f) mice; note the dramatic spermatogenic defects (arrowhead points to a tube containing only SCs) observed in adult P60 mutant testes. Scale bar: 50 μm. *B*, Scatterplot depicting genes showing at P0 differential expression between control and mutant whole testes. Each dot (black or red) represents a gene; genes represented as red dots are those which are either up-regulated (77 genes) or down-regulated (68 genes) >2-fold in mutant testes. Diagonal black bars represent a 2-fold threshold.

microRNA Expression Profiling with Illumina Arrays—Total RNA was isolated with Trizol (Invitrogen, Basel, Switzerland) and quality controlled for RNA integrity by capillary electrophoresis on an Agilent 2100 Bioanalyzer. miRNA profiling was performed according to the manufacturer's protocol using the Illumina MicroRNA Expression Profiling Mouse Panel (Illumina, Hayward, CA), which contains 656 assays for miRNAs described in *miRBase v12.* Briefly, for each sample, 500 ng of total RNA was polyadenylated and converted into cDNA using an oligo dT-Reverse PCR primer. miRNA-specific oligos (extended with specific address sequences and Forward PCR primer sequences) were then hybridized to cDNAs. Following extension using DNA polymerase, products were PCR-amplified using Cy3-labeled forward and unlabeled reverse primers, then purified and eventually hybridized onto a Sentrix Array Matrix overnight. The Sentrix Array Matrix was washed and scanned on a BeadArrays reader. Data were normalized and analyzed using the Illumina Beadstudio 3.1.3 (background correction and quantile normalization without scaling). Expression profiles for each sample were imported into GeneSpringGX 7.3.1 (Agilent Technologies) and MatLab (MathWorks, Inc) and further analyzed in order to identify differentially expressed miRNAs. MicroRNAs were considered as being expressed when the expression was above 1000 (arbitrary units).

MicroRNA Target Recognition Analysis—Target sites were initially identified by the presence of miRNA *seeds* (the minimum sequence of nucleotides required for successful miRNA binding on its target, see supplemental Fig. 1A), and their biological relevance was estimated using the following three models: The first model relies on the thermodynamics of the miRNA-mRNA interactions. The energy balance of these interactions ($\Delta\Delta G$) was computed with a method similar to that of (31): It includes the free energy gain resulting from the formation of the miRNA-target duplex (ΔG duplex) and the free energy required for the unfolding of the target site and of at least 10 nucleotides upstream and 15 downstream of the target site (ΔG open). The second model relies on sequence features as described in (32). The third model relies on conservation of the seed sequences among placental species: the Phastcons scores of the seed sequence bases, provided by the UCSC genome browser (33), were summed (the sum allows to include the effect of the seed length). For seed identification, we used standard parameters, requiring seed length to be 6–8 bases from position 2 of the miRNA, and not allowing mismatches except for a single G:U wobble in 7-mers and two G:U wobbles in 8-mers. Of note, for each model, the first quartile of the ranked predictions was considered as biologically significant in our target site enrichment analysis. A schematic representation of the described strategy is shown in supplemental Fig. 1B.

In vitro Luciferase Assays—The 3′UTR of *Sod-1* was PCR-amplified from genomic DNA using the following oligos: F:5′-ATATGGTCTAGAACATTCCCTGTGTGGTCTGAG-3′, R:5′-ATATGGCCGGCCGTCACACAGTTACAA-3′, and was subcloned in a TOPOII vector (Invitrogen, Basel, Switzerland). The insert was then digested out and directionally inserted downstream of the Firefly luciferase coding sequence in the XbaI and FseI sites of the pTal-Luc vector (Clontech, Sait-Germain-en-Laye, France). Mutated Sod-1 3′UTR constructs were generated with the QuikChange II Site-Directed Mutagenesis kit (Stratagene, Agilent Technologies, Schweiz AG), as described in the manufacturer's protocol, using oligos carrying a fully mutated seed sequence. The day before transfection, 10^4 HEK293T cells/well were seeded in 96-well plates; transfection was performed with (i) 100 ng of the pTal-Luc-Sod1–3′UTR (wild-type or mutant) Firefly plasmid, (ii) 5 ng of the transfection control pRL-SV40 Renilla luciferase plasmid (Promega AG, Dübendorf)) and (iii) 10 nM of the pre-miR-125a-3p (#PM12378), pre-miR-872 (#PM12800), or pre-miR-24 (#PM10737) (Ambion, Applied Biosystems Europe BV), using Lipofectamine 2000 (Invitrogen, Basel, Switzerland) according to the manufacturer's instructions. The Firefly and Renilla luciferase activities were measured 48 h post-transfection using the Dual Luciferase Assay system (Promega AG, Dübendorf) as described in the manufacturer's protocol. All experiments were performed three times, with each experimental condition being performed in four technical replicates. A schematic representation of the Luciferase assays' strategy is shown in Fig. 6*A*.

RESULTS

SC-expressed miRNAs Affect Testicular Transcription—We previously generated a mouse model in which Dicer (Dcr), and subsequently miRNAs, are specifically eliminated in the Sertoli cells (SCs) of the testis ($Dcr^{fx/fx}$;*MisCre*), and found that this loss leads to complete infertility (10). We were able to detect already by postnatal day 5 (P5), a delay in SC maturation and an initial increase in SC proliferation followed by highly elevated levels of SC and GC apoptosis, events that ultimately led to a dramatic testicular degeneration during adulthood (Fig. 1*A*). Importantly, although at birth (P0) no morphological (histological) defects were detected (Fig. 1*A*),

we measured several alterations of the testicular transcriptome. More precisely, we found 77 and 68 genes to be ≥ 2 -fold up- and down-regulated respectively in P0 testes lacking Dcr in SCs (Fig. 1*B*). Deregulated genes included among others *Gdnf, Kitl, Serpin5a, Sox9, Wt1,* and *Cldn11,* all of which have key roles during spermatogenesis. However, the *in vivo* effect of SC-miRNA depletion on protein output was not addressed.

Sertoli-cell Loss of Dicer Causes Significant Proteome Alterations—Here, to assess the impact of SC-miRNA loss on the testicular proteome, we performed relative quantification of proteins on P0 whole testis protein extracts of control and mutant mice using ICPL: Proteins extracted from P0 control and mutant testes were labeled with the light (L) and heavy (H) ICPL reagent respectively, mixed, prefractionated by gel electrophoresis, excised, and trypsin digested. The obtained peptide mixtures were analyzed by nano-ESI-ITMS for protein identification and relative quantification (Fig. 2*A*).

By querying the Swiss-Prot database with the Mascot algorithm, we obtained 240 protein identifications showing a score superior to the identity or the extensive homology threshold. These 240 identifications actually corresponded to 168 proteins, each associated with a nonredundant Entrez-Gene (EG) identifier. Out of these 168 nonredundant proteins, a mutant/control (H/L) protein ratio was calculated for 130 of them (all 130 proteins are listed in supplementary Table 3, and all peptide sequences for the identified proteins are listed in supplementary Table 4); for the remaining 38, this was not possible because of the absence of detected labeled peptides from either the control or the mutant sample. The minimum variation of H/L ratios associated to significant variation of protein expression was determined similarly to (34). The average of the CV (coefficient of variation) obtained for H/L ratios of all proteins for which at least two peptides were quantified was 11.1%. We thus considered that a variation of 30% (>2 CV) was significant. This significant variation of 30% largely overcomes technical variability in our experiments that was demonstrated to be 8%. Of these 130 quantified proteins, 50 were up-regulated (H/L ratio ≥1.3), whereas only 3 were down-regulated (H/L ≤0.7) in mutant testes (Table I). The remaining 77 showed no significant difference in abundance between control and mutant testes (0.7<H/L<1.3). More precisely, of the 50 up-regulated proteins, 23 showed a mild (1.3–1.5-fold change) up-regulation, and the remaining 27 displayed an H/L ratio between 1.5- and 2- (Fig. 2*B* and Table I). These findings are in agreement with two recent studies that reported a relatively *mild* repression of hundreds of proteins upon miRNA overexpression (21, 22).

Independent Validation of Testicular Protein Levels by AQUA Peptide Analysis—To confirm the differential expression levels we detected through ICPL analysis, we selected six proteins, namely four up-regulated (Vimentin, Atp5d, Anxa2, and Sod1) in mutant testes and two unaffected (Prdx1 and Gstm1), for further validation by means of AQUA peptide

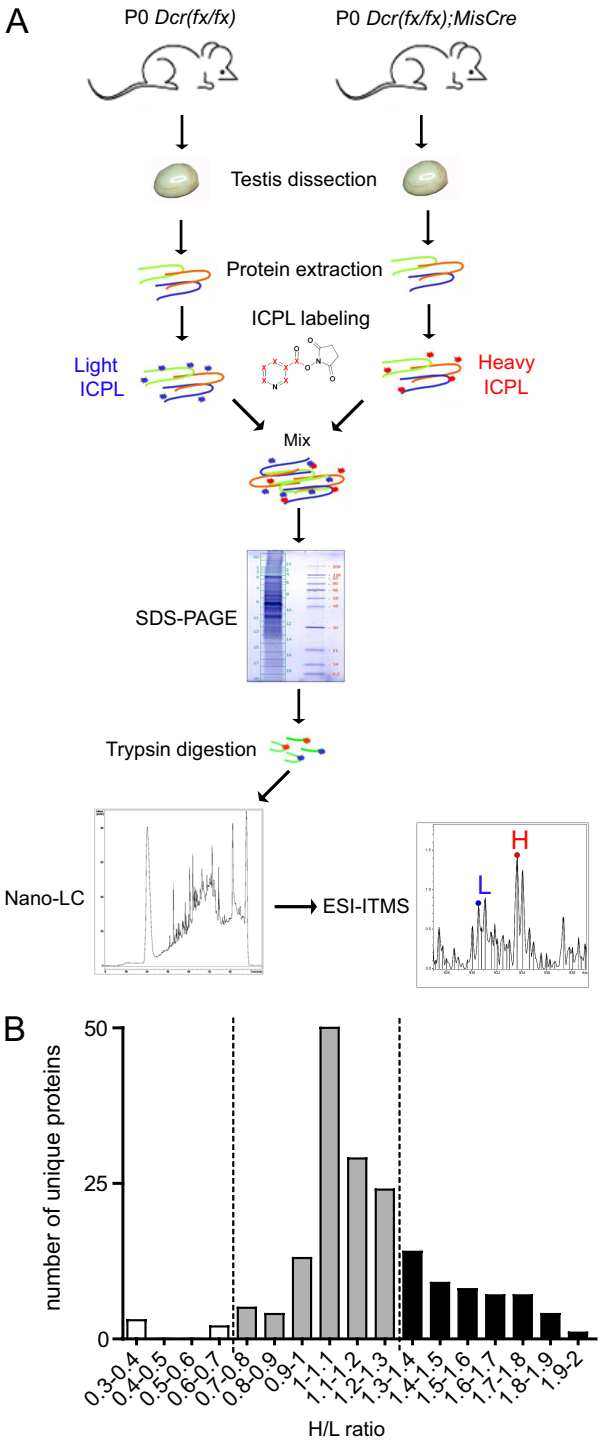


FIG. 2. **SC-loss of Dcr causes significant proteome alterations at birth.** *A*, Schematic representation of the experimental design employed for relative quantification of proteins using ICPL. *B*, Shown here are proteins up-regulated (black bars), unaffected (gray bars) or down-regulated (white bars) upon SC-Dcr loss. H/L (mutant/control) ratios between 0.7 and 1.3 are indicative of no difference between control and mutant testes.

Effect of Sertoli-cell Dicer on Testicular Proteome

TABLE I

List of proteins whose mutant/control (H/L) ratio, as revealed by ICPL analysis, is >1.3 (50 proteins, Sfrs-Stip1) or <0.7 (3 proteins, Hbb-b1, Fasn, Hbb-b2). Also shown are two unaffected (0.7<H/L<1.3) proteins (Gstm1, Prdx1) that were used for AQUA validation analysis

Protein Name	Gene Name	H/L (ICPL)[a]	H/L (AQUA)[b]	mRNA (Affy)[c]	mRNA (qPCR)[d]
Splicing factor, arginine/serine-rich 1	Sfrs	1.97		−1.08	
Heat shock protein HSP 90-alpha	Hsp90aa1	1.86		1.12	0.75
Heat shock 70 kDa protein 1L	Hspa1l	1.84		−1.17	
Heat shock protein HSP 90-beta	Hsp90ab1	1.82		1.02	
40S ribosomal protein S11	Rps11	1.8		1.04	
Tubulin α-1B chain	Tuba1b	1.79		1.08	
Annexin A2	Anxa2	1.78	1.87	1	ns
ADP/ATP translocase 1	Slc25a4	1.77		1.13	
Poly(rC)-binding protein 1	Pcbp1	1.77		1.01	
Profilin-2	Pfn2	1.76		1.09	
Endoplasmin	Hsp90b1	1.74		1.05	
Apolipoprotein A-I	Apoa1	1.72		−1.22	
Vimentin	Vim	1.69	1.93	1.03	0.75
Superoxide dismutase [Cu-Zn]	Sod1	1.68	1.44	−1.04	ns
Elongation factor 2	Eef2	1.66		1.06	
Rho GDP-dissociation inhibitor 1	Arhgdia	1.62		1.09	
Lamin-B1	Lmnb1	1.61		1.09	
Actin, cytoplasmic 2	Actg1	1.6		1.1	
Serum albumin	Alb	1.6		−1.21	
Heat shock 70 kDa protein 1B	Hspa1b	1.58		–	
60S ribosomal protein L14	Rpl14	1.57		−1.15	0.75
40S ribosomal protein S14	Rps14	1.56		1.05	
Histone H4	Histh4	1.56		–	
Protein disulfide-isomerase	P4hb	1.56		1.08	
ADP/ATP translocase 2	Slc25a5	1.55		−0.99	
Splicing factor, proline- and glutamine-rich	Sfpq	1.53		1.03	ns
ATP synthase subunit delta, mitochondrial	Atp5d	1.52	1.78	1.02	ns
Ig κchain C region	Igk-C	1.48		–	
Tubulin β-5 chain	Tubb5	1.47		1.11	
Actin, γ-enteric smooth muscle	Actg2	1.46		1.13	
60S ribosomal protein L10	Rpl10	1.45		1.04	
60S ribosomal protein L28	Rpl28	1.45		1.02	
Elongation factor 1-α1	Eef1a1	1.42		1.07	
Phosphoglycerate mutase 1	Pgam1	1.42		1.01	
Peptidyl-prolyl *cis-trans* isomerase A	Ppia	1.41		1.08	
Nucleophosmin	Npm1	1.4		1.06	
Calmodulin	Calm3;Calm1;Calm2	1.38		−1.04;1.07;1.03	
Heat shock cognate 71 kDa protein	Hspa8	1.37		−1.01	
40S ribosomal protein S20	Rps20	1.35		1.01	
60 kDa heat shock protein, mitochondrial	Hspd1	1.35		1.12	
60S ribosomal protein L18	Rpl18	1.35		1.09	
ATP-citrate synthase	Acly	1.35		1.17	
40S ribosomal protein S3	Rps3	1.34		1.04	
ATP synthase subunit β, mitochondrial	Atp5b	1.34		1.01	
60S ribosomal protein L13	Rpl13	1.33		1.02	
40S ribosomal protein S8	Rps8	1.32		1.01	
ATP synthase subunit α, mitochondrial	Atp5a1	1.32		1.06	
Heterogeneous nuclear ribonucleoprotein A3	Hnrnpa3	1.32		1.11	
Histone H1.2	Hist1h1c	1.3		1	
Stress-induced-phosphoprotein 1	Stip1	1.3		−0.99	
Hemoglobin subunit β-1	Hbb-b1	0.36		1.06	
Fatty acid synthase	Fasn	0.37		−1.18	
Hemoglobin subunit β-2	Hbb-b2	0.38		1.06	
Glutathione S-transferase Mu 1	Gstm1	0.85	0.86	−1.38	0.5
Peroxiredoxin-1	Prdx1	1.18	1.3	0.99	

[a] Marked here is the H/L protein ratio, as measured through the ICPL analysis.

[b] Marked here is the H/L ratio of six selected proteins, as measured through AQUA peptide analysis.

[c] Shown in this column are the mutant/control mRNA ratios revealed by our Affymetrix (Affy) analysis (Ref (10)). Note that no statistically significant difference in abundance of mRNAs is observed between mutants and controls. The mark (–) indicates that the mentioned protein corresponds to multiple EG identifiers and thus, we were not able to sort out the correct corresponding Affymetrix probeset value.

[d] Marked here is the mutant/control mRNA ratio as measured through quantitative Real-Time PCR (also see Fig. 4). 'ns' indicates no significant difference.

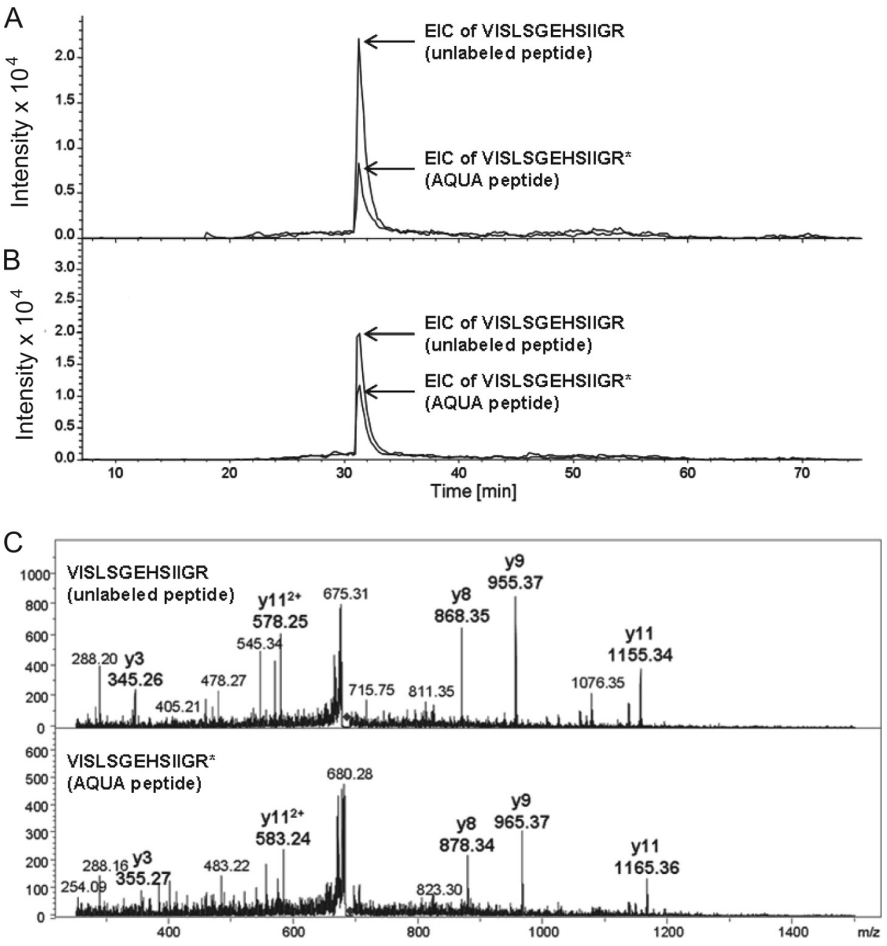


FIG. 3. **AQUA peptide analysis validated the ICPL results: shown here is the example of SOD-1.** Extracted ion chromatograms (EICs) were reconstituted by extracting the signals of y3, y8, y9, y11, and $y11^{2+}$ sequence ions of the VISLSGEHSIIGR unlabeled peptide and the VISLSGEHSIIGR* AQUA peptide used for the quantification of SOD1 by multiple reaction monitoring in mutant (*A*) and control (*B*) testes and their corresponding MS/MS spectra (*C*). The EICs obtained for AQUA peptides used as internal standards in (*A*) and (*B*) correspond to an injected amount of 30 fmol.

analysis, an MS-based technique for the absolute quantification of proteins. AQUA peptides are chemically synthesized isotope-labeled peptides whose sequences correspond to proteolytic peptides of the proteins to be quantified. They were spiked into the sample in known quantities before LC-MS/MS analysis. Absolute quantification was achieved by comparing the signals corresponding to AQUA and proteolytic peptides (an example for SOD-1 is shown in Fig. 3). Absolute quantities determined in mutant and control samples were used to calculate mutant-to-control ratios, which we found to be close to those obtained with relative quantification by ICPL for all six proteins, a result that validated our ICPL results. Note that all AQUA values are indicated in Table I.

Protein Up-regulation Upon Sertoli-cell Loss of Dcr is not Accompanied by mRNA Alterations—Next, we went on to assess whether the changes we measured in protein output were the result of changes in mRNA expression levels. Forty-seven out of 50 up-regulated proteins showed no difference in their mRNA expression levels between controls and mutants, as evidenced by their expression levels measured on the Affymetrix microarray (for the remaining three proteins, we were not able to sort out the corresponding Affymetrix probeset value, because they match to multiple EG identifiers) (Table I), thus suggesting that they represent genes whose expression is controlled at the translational level. In fact, we further selected seven up-regulated proteins (among them were those that had been validated by AQUA peptide analysis) and performed RealTime qPCR for their respective genes on P0 control and mutant whole testes: these genes showed either no difference (*Anxa2, Sod1, Sfpq,* and *Atp5d*), or, interestingly, a ~25% reduction in their mRNA levels (*Vimentin, Rpl14,* and *Hsp90aa1*), whereas their respective proteins were >1.3 times more abundant in mutant testes (Fig. 4*A*). We also selected three unaffected proteins (Gstm1, Cyp11a1, and Prdx1) to evaluate their mRNA levels, and found that the mRNA expression levels of *Cyp11a1* and *Prdx1* remained unaffected, whereas that of *Gstm1* showed a ~50% reduction in mutant testes (Fig. 4*B*). Taken together, these findings demonstrate that loss of *Dcr* and miRNAs in SCs has a significant impact on testicular proteins, without however affecting the amounts of the corresponding mRNAs.

Several microRNAs are Expressed in Immature SCs—The fact that the protein up-regulation we measured is not accompanied by an mRNA up-regulation in mutant testes strongly

Effect of Sertoli-cell Dicer on Testicular Proteome

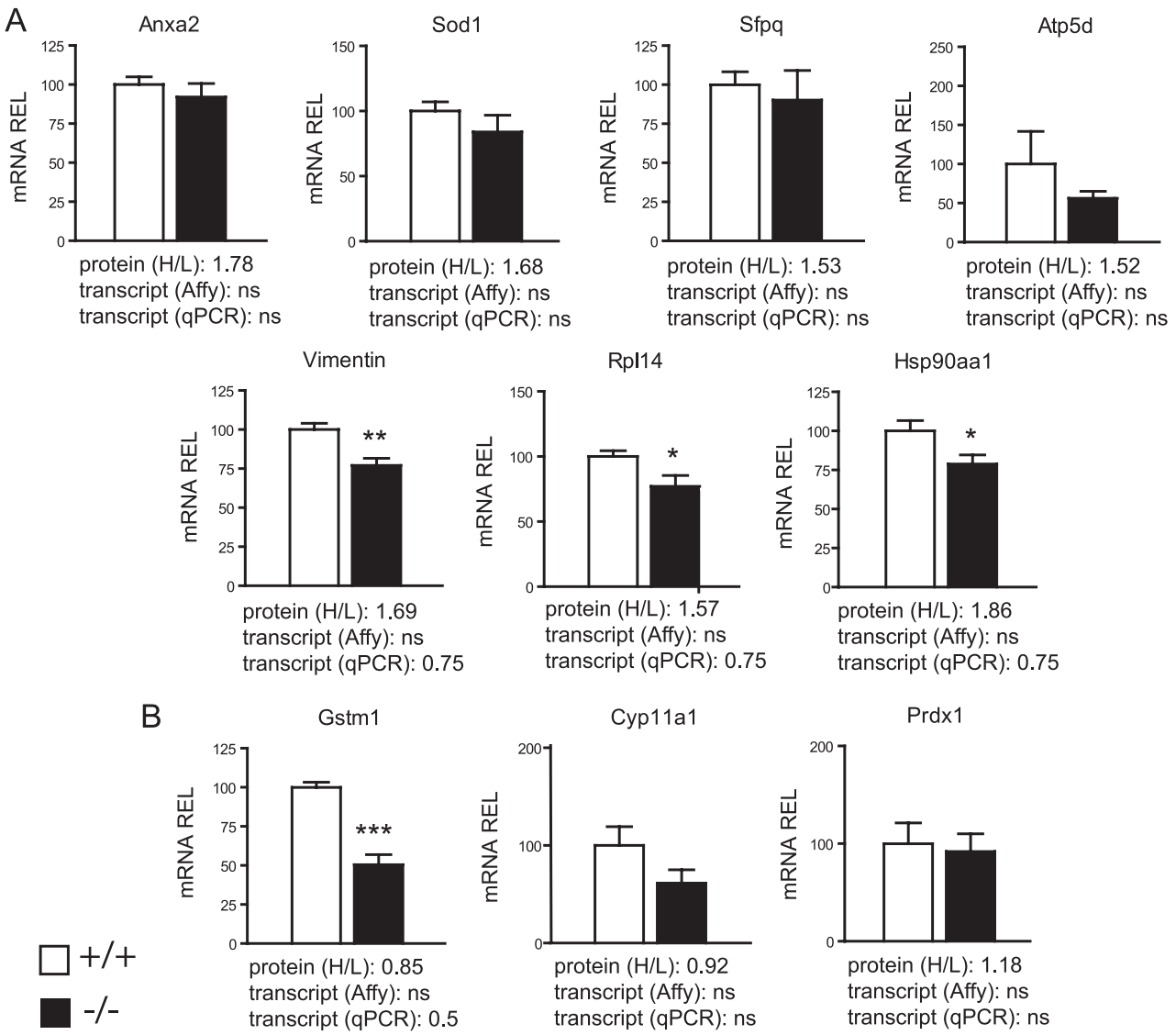


FIG. 4. **Real-Time qPCR revealed that protein up-regulation upon SC-Dcr loss is not accompanied by alterations at the mRNA level.** *A*, Shown here are four genes (*Anxa2, Sod1, Sfpq, and Atp5d*) whose mRNA levels are not significantly different between controls and mutants (qPCR and Affy), but whose respective proteins are up-regulated in mutant testes (H/L>1.3), and three genes (*Vimentin, Rpl14, Hsp90aa1*) whose mRNA levels show a ~25% reduction (qPCR), but whose respective proteins are up-regulated in mutant testes (H/L>1.3). *B*, Gstm1, Cyp11a1 and Prdx1 protein levels were unaffected and so were their transcript levels, except for *Gstm1,* which showed a 50% reduction in mutant testes. For each gene, the mutant/control (H/L) protein ratio revealed by ICPL analysis, as well as their mutant/control mRNA (transcript) ratio revealed either through microarray (Affy) or qPCR is given beneath the graph. Results are mean ± S.E. ($n = 6$ animals/group), $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ *versus* controls; ns: not significant. Primers used are listed on supplemental Table S2. +/+ and −/− are abbreviations for control $Dcr^{fx/fx}$, and mutant $Dcr^{fx/fx}$;*MisCre* animals respectively. REL: relative expression levels.

suggested that their respective genes are most likely to represent direct SC-miRNA targets regulated at the translational, and not the transcriptional level. To further investigate this hypothesis, we first set out to characterize the miRNA expression profile of SCs using a miRNA microarray. For this purpose, we analyzed the expression of 656 murine miRNAs in purified populations of testicular cells, namely immature P6 and mature P17 SCs, in adult Leydig cells, in different types of germ cells (spermatogonia A, spermatogonia B, and intermediate, pachytene spermatocytes and spermatids), as well as in immature P6 whole testes, using the Illumina microRNA Expression Profiling System. We identified a set of 382 miRNAs expressed in SCs (supplemental Table 5). Of these, we found that 50 are expressed more than two times more abundantly in immature P6 SCs in comparison to P6 whole testes (shown in two graphs in Fig. 5*A*), a finding that *could* suggest a potential

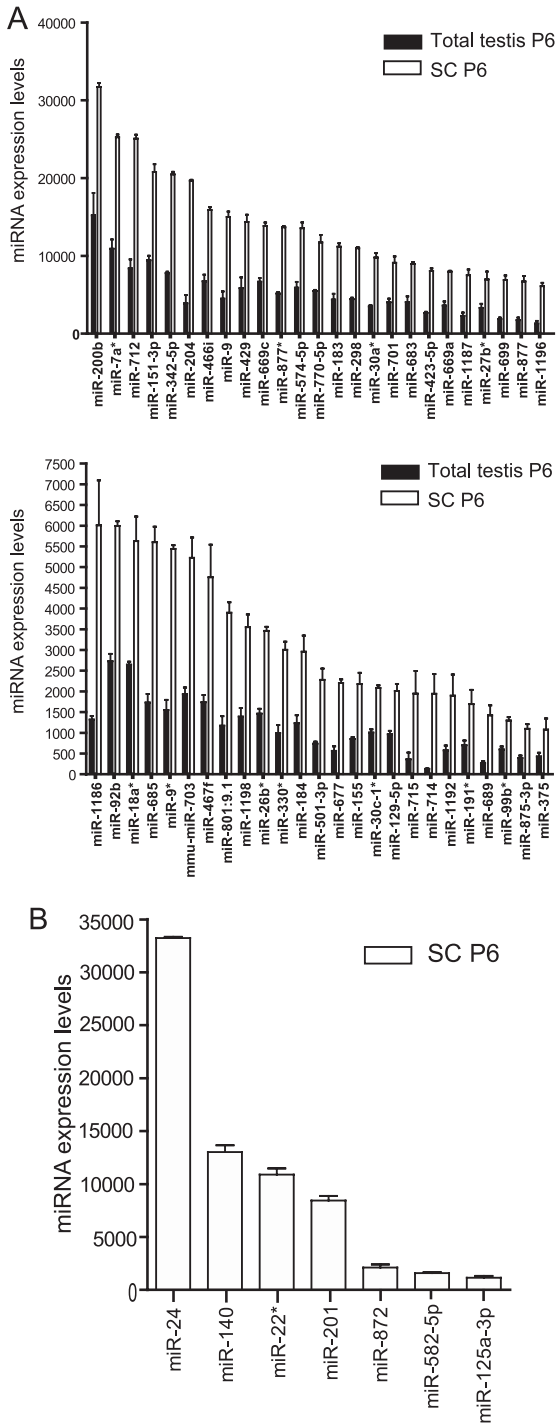


FIG. 5. **miRNAs expressed in SCs.** *A*, Shown here in two graphs are the expression profiles of the 50 SC-enriched miRNAs, in descending order of expression in P6 SCs. The entire list of 382 miRNAs expressed in SCs, along with their sequences is provided in supplemental Table 5. *B*, Shown here are the expression profiles of 7 SC-expressed miRNAs predicted to have a putative binding site in the 3′UTR of *Sod-1*. Results are represented as mean values of three biological replicates ±s.d. miRNA nomenclature is based on the Sanger miRBase, v12.

TABLE II

SC-miRNA target site enrichment among genes coding for proteins deregulated upon SC-Dcr loss. Shown here are the % of up-regulated proteins (up) versus those unaffected (equal) upon SC-Dcr loss that were enriched in SC-miRNA target sites, when considering either only the presence of a seed sequence (seeds), an energetically favorable miRNA-mRNA duplex (ΔG duplex), a favorable target site sequence context (Targetscan) or the conservation of the seed sequence (conservation). The numbers in parenthesis indicate the number of genes taken into consideration in each fraction

Fractions	Proteome fractions with an H/L threshold of 1.3		
	Up (59)		Equal (96)
	% targeted	Enrichment (*p* value)	% targeted
Seeds	93.2	0.40	90.6
ΔG duplex	77.8	0.072	65.2
Targetscan	74.6	0.71	77.1
Conservation	78	0.013	59.4

biological role for these miRNAs in SCs, without however neglecting the potential role of other SC-expressed miRNAs.

The 3′UTRs of Genes Up-regulated Upon SC-Dcr Loss at the Protein Level are Enriched for SC-miRNA Target Sites—Having in hand these SC-expressed miRNAs, and given that miRNAs most frequently bind on regions of an mRNA's 3′UTR, we went on to assess whether the 3′UTRs of transcripts coding for up-regulated proteins are actually enriched for SC-miRNA target sites. First, we considered as a criterion for successful miRNA binding only the presence of a seed (the minimum sequence of nucleotides required for successful miRNA binding on its target) in the transcripts' 3′UTR (supplemental Fig. 1), but found that none of the seed sequences were significantly enriched in genes coding for up-regulated proteins in mutant testes, when compared with the unaffected proteins (One-sided Fisher test p value = 0.40, Table II). However, when we further refined our query to target sites bearing a conserved-in-placental-species seed sequence, we found that genes coding for up-regulated proteins in mutant testes were *slightly*, yet significantly, enriched in SC-miRNA target sites (One-sided Fisher test p value = 0.0128, Table II). Interestingly, nonsignificant enrichments were observed when considering either energetically favorable miRNA-mRNA duplexes (ΔG duplex, One-sided Fisher test p value = 0.0721, Table II) or a favorable target site sequence context (One-sided Fisher test p value = 0.71, Table II). Altogether, although the low statistical significance of the enrichment is acknowledged, these findings tend to suggest that genes coding for up-regulated proteins upon SC-Dcr loss are likely to represent direct SC-miRNA targets.

SOD-1, a Protein Up-regulated Upon SC-Dcr Loss, is a Direct Target of Three SC-expressed miRNAs—Next, we selected *Sod-1*, one of the genes we found to be up-regulated at the protein level, in order to evaluate its direct post-transcriptional targeting by SC-miRNAs. *Sod-1* was selected because of its potential biological interest: SOD-1 is a Cu-Zn

Effect of Sertoli-cell Dicer on Testicular Proteome

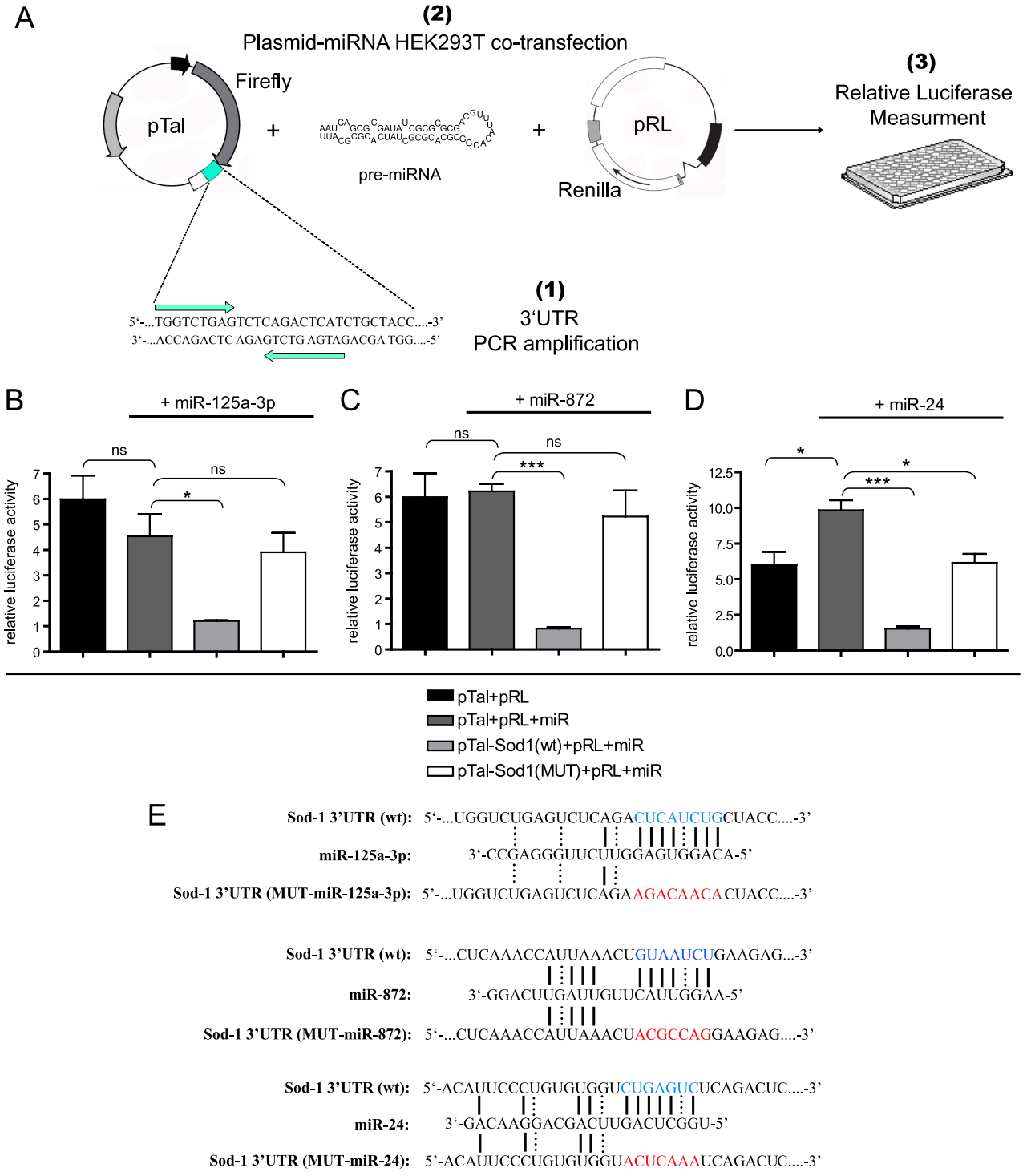


FIG. 6. **Testing the binding of the Sod-1 3′UTR by SC-expressed miRNAs.** *A*, The *Sod-1* 3′UTR was PCR-amplified (1) and cloned just downstream of the Firefly luciferase CDS in the pTal plasmid. The pTal-Sod1–3′UTR plasmid was cotransfected with a pRL transfection control vector expressing the Renilla luciferase CDS and 10 nM of a miRNA precursor in HEK293T cells (2) and the relative luciferase activity was measured (3). Cotransfection of pTal (Firefly) and pRL (Renilla) (black bars) yielded a certain Firefly/Renilla luciferase level; addition of either pre-miR-125a-3p or pre-miR-872 did not alter the relative luciferase levels (*B*, *C*, dark gray bars), however that of pre-miR-24 did (see text) (*D*, dark gray bar); cloning the wt *Sod-1* 3′UTR in pTal in the presence of either one of three miRNAs caused a significant reduction in the relative

superoxide dismutase that catalyzes the reaction $2O_2^- + 2H^+ \rightarrow O_2 + H_2O_2$, thus protecting cells from oxidative damage (reviewed in (35)). Perturbation of this reaction's equilibrium may lead to oxidative damage; indeed, it has been shown that overproduction of SOD-1 can cause increased oxidative damage resulting in enhanced cell death by apoptosis (for example see (36, 37)). We thus reckoned that *Sod-1* would be an interesting candidate gene for the explanation of the testicular degeneration phenotype we observed. Using the three prediction models described in the Experimental Procedures, we predicted one putative target site for each of seven SC-expressed miRNAs, namely miR-125a-3p, miR-140, miR-24, miR-201, miR-22*, miR-872, and miR-582–5p, on the 3′UTR of *Sod-1*. Of these, miR-24 showed the strongest expression in P6 SCs, miR-872 much lower, and miR-125a-3p the lowest (Fig. 5*B*), although the latter was slightly enriched in P6 SCs in comparison to P6 whole testis (data not shown). These three miRNAs were selected for further analysis: The *Sod-1* 3′UTR was cloned downstream of the Firefly luciferase coding sequence in the pTal plasmid. A dual-luciferase assay was then performed in HEK293T cells by transfecting the pTal plasmid harboring or not the 3′UTR of *Sod-1*, along with one of the three precursor miRNAs and a transfection control vector expressing the Renilla luciferase coding sequence (pRL) (Fig. 6*A*). Transfection of the empty pTal and pRL plasmids in the presence of pre-miR-125a-3p or pre-miR-872 did not significantly alter the relative luciferase levels (Figs. 6*B* and 6*C*, dark gray bars). However cotransfection with pre-miR-24 caused an increase in the relative luciferase levels (Fig. 6*D*, dark gray bar), which is most likely due to the fact that the Renilla CDS harbors one putative miR-24 binding site: binding of miR-24 causes a decrease in the Renilla levels, thus an increase in the Firefly/Renilla ratio. When cloning the wild-type (wt) *Sod-1* 3′UTR downstream of the Firefly luciferase coding sequence, the Firefly/Renilla luciferase levels showed a significant decrease for all three miRNAs (Figs. 6*B*, 6*C* and 6*D*, light gray bars). To further confirm the specificity of these effects, we generated three *Sod-1* 3′UTR constructs, each harboring a mutated seed sequence for the three SC-expressed miRNAs (Fig. 6*E*) and repeated the luciferase assays as described above. For all three miRNAs, seed mutation abolished the miRNA repressive effect on the *Sod-1* 3′UTR (Figs. 6*B*, 6*C*, and 6*D*, white bars), thus strongly suggesting that the 3′UTR of *Sod-1* is a direct target of these three SC-expressed miRNAs.

DISCUSSION

Although the primary effect of animal miRNAs is thought to occur at the level of translational repression, most studies until today have measured their effect at the mRNA level, mostly through DNA microarrays. We ourselves measured the impact of an SC-*Dcr* loss at the mRNA level by performing an Affymetrix microarray on P0 and P5 whole testes and measured several transcriptome alterations occurring in mutant testes (10). The effect of miRNAs on protein output has been more difficult to study in a high-throughput manner, and only recently two groups performed such an analysis, showing that a single miRNA can repress the production of hundreds of proteins, yet at a relatively mild level (21, 22). Additional studies have unraveled similar results, although at a smaller scale (for example see (38, 39)). Here, in an effort to reveal the molecular factors whose deregulation causes infertility in mice lacking SC-Dcr, we have used quantitative mass spectrometry to measure the effect of SC-Dcr and subsequent miRNA loss on the testicular proteome. To our knowledge, this is the first report using ICPL technology to study *in vivo* differential protein expression in the testis. We report the quantification of 130 nonredundant proteins, of which 50 are up-regulated more than 30% in mutant testes, whereas the remaining are unaffected (77 proteins) or down-regulated (3 proteins). We find that protein up-regulation is mild, that is, never exceeding a 2-fold change, yet the fact that from 53 differentially expressed proteins, the vast majority (50 proteins) is up-regulated in mutant testes further reinforces the notion that translational repression, at least in our system, is one of the primary effects of animal miRNAs.

Indeed, one striking finding of our study is the large proportion of proteins up-regulated upon SC-Dcr loss (50/130 or ~38%). Certainly, we acknowledge that extrapolation to the whole proteome would be speculative, because our mass spectrometry analysis quantified only a small set of *highly expressed* proteins, however, it would tend to suggest that SC-miRNAs have a significant impact on testicular translational control. In itself, this is not unexpected: recently, the global impact of miRNAs on protein output was investigated by quantitative mass spectrometry and showed that a single miRNA can directly repress translation of hundreds of genes (21, 22). Taking into account that SCs express hundreds of miRNAs and that *in silico* analyses have predicted several thousands of protein-coding genes to be potential targets of hundreds of miRNAs, our own findings suggest that miRNAs play a broad role in the fine-tuning of protein synthesis in SCs. They are thereby essential for their survival and maturation and eventually for the entire male reproductive function.

A careful comparison of our own data to those of the two above-mentioned studies, (21, 22) reveals some interesting

luciferase activity (*B*, *C*, *D*, light gray bars), whereas mutation of the seed sequences abolished the miRNA repressive effect (*B*, *C*, *D*, white bars); $*p < 0.05$, $**p < 0.01$, $***p < 0.001$ *versus* controls; ns: not significant. (*E*) Schematic representation of the binding of miR-125a-3p, miR-872 and miR-24 on either the wt or the mutated *Sod-1* 3′UTR sequence. Wt seeds are marked in blue, mutated seeds in red, Watson-Crick base pairing with a straight line and U-G wobbles with a dotted line.

Effect of Sertoli-cell Dicer on Testicular Proteome

points worth of discussion. First, the Selbach and Baek papers conclude that only targets translationally repressed by more than a third also display detectable mRNA alterations, whereas those modestly repressed show little or no change at the mRNA level. The protein up-regulation we measure in our system, which falls within a 1.3–2-fold range, is actually not accompanied at all by alterations at the mRNA level. This finding could be interpreted as a more significant miRNA-mediated translational control in SCs than in other systems, although, again, extrapolation to the whole proteome must be done with caution. We should also mention here that, in comparison to our study, the fact that the Selbach and Baek papers report the identification of ~3000 proteins, is likely because of the following reasons: (a) They used stable isotope labeling with amino acids in cell culture (SILAC) a technique *not* applicable to tissues, in which trypsin cleavage occurs following both lysine and arginine residues, thus generating numerous peptides and increasing the chances of protein identification and quantification; in our study, ICPL labeling prevents trypsin cleavage following lysine residues, allowing cleavage to occur only next to arginine residues, therefore the number of generated peptides is significantly smaller; (b) The mass spectrometer used in the Selbach and Baek papers was an LTQ-Orbitrap instrument, which is of higher performance and resolution than the one we used.

It would be worth noting at this point that, for purely technical reasons, our starting material for the mass spectrometry analysis was whole testis protein extracts, whereas depletion of miRNAs was performed uniquely in SCs. Thus, the presence of a heterogeneous population of cells in P0 testes probably masks the true impact on the protein output of SCs because of the dilution by proteins originating from other cell populations. The use of purified SCs would certainly further refine our results and thereby unravel novel, or additional, molecular targets that could explain the observed testicular degeneration and eventual infertility caused by the loss of SC-Dcr and miRNAs. In fact, as a first step toward this direction, we used purified wild-type SCs to perform a miRNA expression profiling analysis. This allowed us to unravel several SC-expressed miRNAs that we then used to assess whether the transcripts coding for up-regulated proteins upon SC-Dcr loss are enriched for SC-miRNA target sites. The enrichment was significant only when taking into consideration seed conservation among placental species. However, neither an energetically favorable miRNA-mRNA duplex, nor a favorable target site sequence context yielded a significant SC-miRNA target site enrichment. This could be explained by the fact that the ΔG duplex feature is based on an RNA-only model and that more importantly, the sequence context parameters are evaluated based on RNA-microarrays (32). As described above, the differences in protein level we measured here are because of translational differences and not mRNA alterations. Prediction features were described to be differentially relevant at each step of the RNAi pathway (40), therefore, models for target prediction trained on the mRNA level are expected to be less accurate when no mRNA degradation is involved. This might thus explain our insignificant target site enrichment when considering a favorable sequence context or miRNA-mRNA duplex. In contrast, the conservation feature captures a "blinder" information (*i.e.* without a regulation model), which allows us to significantly isolate miRNA effects on the measured proteome.

Among the proteins up-regulated in mutant testes, SOD-1 retained our attention. We reckoned that because SOD-1 is a Cu/Zn-superoxide dismutase whose overproduction causes increased oxidative damage resulting in enhanced cell death through apoptosis (for example see (36, 37)), its up-regulation could be detrimental for cell survival, and thereby account, at least partially, for the testicular degeneration we observed upon SC-Dcr loss. By performing an *in vitro* dual-luciferase assay, we found the 3′UTR of *Sod-1* to be directly targeted by three SC-expressed miRNAs: miR-125a-3p, miR-872 and miR-24. Of note, because SOD-1 is present in both SCs and in all types of GCs (41), the effect could be Sertoli-cell autonomous or not. In either case, taking also into consideration the ~3-fold mRNA up-regulation of Bcl2l11, a facilitator of apoptosis we previously detected at P0 (10), we are tempted to believe that two independent, miRNA-mediated, cell-death molecular mechanisms are at the origin of -at least part of- the observed testicular degeneration. It would certainly be interesting to find out whether the observed SOD-1 increase upon SC-Dcr loss at birth is maintained at later stages of testis development. If this were indeed true, a chronic oxidative damage could most likely explain the almost complete loss of testicular structures upon aging.

An additional interesting issue raised by our findings is whether the effects on protein output of mutant testes are because of direct SC-miRNA-mediated inhibition of protein synthesis, or because of indirect repressive mechanisms. Several indications suggest that a direct effect of miRNAs on target genes may account for most of the proteome alterations. First, we performed all of our analyses at an early stage of testis development (P0), when miRNAs are beginning to be depleted from SCs and when no morphological and histological alterations are yet detected, a fact that would tend to suggest a direct miRNA effect because of a rather restricted time window for secondary, indirect effects to occur. Second, although we did observe transcriptional alterations in mutant testes, the deregulated genes at P0 represent only 0.5% of the total transcriptome (145 deregulated probe sets *versus* 29.000 probe sets considered to be expressed in our Affymetrix analysis (10)), and most importantly, do not account for any protein deregulation, thus suggesting that, any alterations at the protein level are most likely to represent direct effects in our system. Finally, the fact that the 3′UTR of transcripts coding for up-regulated proteins are enriched, although slightly, for SC-miRNA target sites, and that one of

these proteins, SOD-1, is, at least *in vitro*, directly targeted by three SC-expressed miRNAs, points toward a direct negative miRNA effect on protein synthesis.

Overall, with this study, we unravel a molecular mechanism that could partially explain the observed testicular degeneration caused by SC-Dcr and miRNA loss. Most importantly, we show, for the first time to our knowledge, that miRNAs have quite a significant impact on the testicular protein output and thus further reinforce the current notion of animal miRNAs exerting their primary negative effect at the translational level.

Acknowledgments—We would like to thank Nicolas Veillard for excellent technical assistance; all the members of the Nef Laboratory for critical comments and discussion on the manuscript; all members of the Proteomics Core Facility Biogenouest for valuable assistance during the mass spectrometry experiments.

[S] This article contains supplemental Fig. S1 and Tables S1 to S5.

§§ To whom correspondence should be addressed: Department of Genetic Medicine and Development, University of Geneva Medical School, 1, rue Michel-Servet, CH 1211 Geneva 4, Switzerland. Phone: +41 22 379 5193; Fax: +41 22 379 5260; E-mail: Serge.Nef@unige.ch.

¶¶ Authors contributed equally to this work.

REFERENCES

1. Cooke, H. J., and Saunders, P. T. (2002) Mouse models of male infertility. *Nat. Rev. Genet* **3,** 790–801
2. Jegou, B. (1992) The Sertoli cell. *Clin. Endocrinol Metab.* **6,** 273–311
3. Jégou, B. (1993) The Sertoli-germ cell communication network in mammals. *Int. Rev. Cytol.* **147,** 25–96
4. Chen, C., Ouyang, W., Grigura, V., Zhou, Q., Carnes, K., Lim, H., Zhao, G. Q., Arber, S., Kurpios, N., Murphy, T. L., Cheng, A. M., Hassell, J. A., Chandrashekar, V., Hofmann, M. C., Hess, R. A., and Murphy, K. M. (2005) ERM is required for transcriptional control of the spermatogonial stem cell niche. *Nature* **436,** 1030–1034
5. Costoya, J. A., Hobbs, R. M., Barna, M., Cattoretti, G., Manova, K., Sukhwani, M., Orwig, K. E., Wolgemuth, D. J., and Pandolfi, P. P. (2004) Essential role of Plzf in maintenance of spermatogonial stem cells. *Nat. Gen.* **36,** 653–659
6. Meng, X., Lindahl, M., Hyvönen, M. E., Parvinen, M., de Rooij, D. G., Hess, M. W., Raatikainen-Ahokas, A., Sainio, K., Rauvala, H., Lakso, M., Pichel, J. G., Westphal, H., Saarma, M., and Sariola, H. (2000) Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. *Science* **287,** 1489–1493
7. Braun, R. E. (1998) Post-transcriptional control of gene expression during spermatogenesis. *Sem. Cell Develop. Biol.* **9,** 483–489
8. Hayashi, K., Chuva de Sousa Lopes, S. M., Kaneda, M., Tang, F., Hajkova, P., Lao, K., O'Carroll, D., Das, P. P., Tarakhovsky, A., Miska, E. A., and Surani, M. A. (2008) MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis. *PLoS ONE* **3,** e1738
9. Maatouk, D. M., Loveland, K. L., McManus, M. T., Moore, K., and Harfe, B. D. (2008) Dicer1 is required for differentiation of the mouse male germline. *Biol. Reprod.* **79,** 696–703
10. Papaioannou, M. D., Pitetti, J. L., Ro, S., Park, C., Aubry, F., Schaad, O., Vejnar, C. E., Kühne, F., Descombes, P., Zdobnov, E. M., McManus, M. T., Guillou, F., Harfe, B. D., Yan, W., Jégou, B., and Nef, S. (2009) Sertoli cell Dicer is essential for spermatogenesis in mice. *Develop. Biol.* **326,** 250–259
11. Papaioannou, M. D., and Nef, S. (2010) microRNAs in the Testis: Building Up Male Fertility. *J. Androl.* **31,** 26–33
12. Kim, V. N., Han, J., and Siomi, M. C. (2009) Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell. Biol.* **10,** 126–139
13. Rigoutsos, I. (2009) New tricks for animal microRNAS: targeting of amino acid coding regions at conserved and nonconserved sites. *Cancer Res.* **69,** 3245–3248
14. Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.* **9,** 102–114
15. Vasudevan, S., Tong, Y., and Steitz, J. A. (2007) Switching from repression to activation: microRNAs can up-regulate translation. *Science* **318,** 1931–1934
16. Place, R. F., Li, L. C., Pookot, D., Noonan, E. J., and Dahiya, R. (2008) MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 1608–1613
17. Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R., and Pasquinelli, A. E. (2005) Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* **122,** 553–563
18. Giraldez, A. J., Mishima, Y., Rihel, J., Grocock, R. J., Van Dongen, S., Inoue, K., Enright, A. J., and Schier, A. F. (2006) Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312,** 75–79
19. Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., Bartel, D. P., Linsley, P. S., and Johnson, J. M. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433,** 769–773
20. Vinther, J., Hedegaard, M. M., Gardner, P. P., Andersen, J. S., and Arctander, P. (2006) Identification of miRNA targets with stable isotope labeling by amino acids in cell culture. *Nucl. Acids Res.* **34,** e107
21. Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P. (2008) The impact of microRNAs on protein output. *Nature* **455,** 64–71
22. Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature* **455,** 58–63
23. Schmidt, A., Kellermann, J., and Lottspeich, F. (2005) A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **5,** 4–15
24. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 6940–6945
25. Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., Vandenesch, F., and Garin, J. (2007) Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol. Cell Proteomics* **6,** 2139–2149
26. Peirson, S. N., Butler, J. N., and Foster, R. G. (2003) Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. *Nucl. Acids Res.* **31,** e73
27. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3, RESEARCH0034
28. Toebosch, A. M., Robertson, D. M., Klaij, I. A., de Jong, F. H., and Grootegoed, J. A. (1989) Effects of FSH and testosterone on highly purified rat Sertoli cells: inhibin alpha-subunit mRNA expression and inhibin secretion are enhanced by FSH but not by testosterone. *J. Endocrinol.* **122,** 757–762
29. Bellvé, A. R. (1993) Purification, culture, and fractionation of spermatogenic cells. *Meth. Enzymol.* **225,** 84–113
30. Manna, P. R., Tena-Sempere, M., and Huhtaniemi, I. T. (1999) Molecular mechanisms of thyroid hormone-stimulated steroidogenesis in mouse leydig tumor cells. Involvement of the steroidogenic acute regulatory (StAR) protein. *J. Biol. Chem.* **274,** 5909–5918
31. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.* **39,** 1278–1284
32. Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., Lim, L. P., and Bartel, D. P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27,** 91–105
33. Karolchik, D., Kuhn, R. M., Baertsch, R., Barber, G. P., Clawson, H., Diekhans, M., Giardine, B., Harte, R. A., Hinrichs, A. S., Hsu, F., Kober, K. M., Miller, W., Pedersen, J. S., Pohl, A., Raney, B. J., Rhead, B., Rosenbloom, K. R., Smith, K. E., Stanke, M., Thakkapallayil, A., Trumbower, H., Wang, T., Zweig, A. S., Haussler, D., and Kent, W. J. (2008) The UCSC Genome Browser Database: 2008 update. *Nucl. Acids Res.* **36,** D773–779
34. Sarioglu, H., Brandner, S., Jacobsen, C., Meindl, T., Schmidt, A., Kellermann, J., Lottspeich, F., and Andrae, U. (2006) Quantitative analysis of 2,3,7,8-tetrachlorodibenzo-p-dioxin-induced proteome alterations in 5L rat hepatoma cells using isotope-coded protein labels. *Proteomics* **6,**

Effect of Sertoli-cell Dicer on Testicular Proteome

2407–2421
35. Turner, T. T., and Lysiak, J. J. (2008) Oxidative stress: a common factor in testicular dysfunction. *J. Androl.* **29,** 488–498
36. Peled-Kamar, M., Lotem, J., Okon, E., Sachs, L., and Groner, Y. (1995) Thymic abnormalities and enhanced apoptosis of thymocytes and bone marrow cells in transgenic mice overexpressing Cu/Zn-superoxide dismutase: implications for Down syndrome. *The EMBO J.* **14,** 4985–4993
37. Sanij, E., Hatzistavrou, T., Hertzog, P., Kola, I., and Wolvetang, E. J. (2001) Ets-2 is induced by oxidative stress and sensitizes cells to H(2)O(2)-induced apoptosis: implications for Down's syndrome. *Biochem. Biophys. Res. Comm.* **287,** 1003–1008
38. Taguchi, A., Yanagisawa, K., Tanaka, M., Cao, K., Matsuyama, Y., Goto, H., and Takahashi, T. (2008) Identification of hypoxia-inducible factor-1 alpha as a novel target for miR-17–92 microRNA cluster. *Cancer Res.* **68,** 5540–5545
39. Yang, Y., Chaerkady, R., Beer, M. A., Mendell, J. T., and Pandey, A. (2009) Identification of miR-21 targets in breast cancer cells using a quantitative proteomic approach. *Proteomics* **9,** 1374–1384
40. Hausser, J., Landthaler, M., Jaskiewicz, L., Gaidatzis, D., and Zavolan, M. (2009) Relative contribution of sequence and structure features to the mRNA binding of Argonaute/EIF2C-miRNA complexes and the degradation of miRNA targets. *Genome Res.* **19,** 2009–2020
41. Gu, W., Morales, C., and Hecht, N. B. (1995) In male mouse germ cells, copper-zinc superoxide dismutase utilizes alternative promoters that produce multiple transcripts with different translation potential. *J. Biol. Chem.* **270,** 236–243

4.2 Integration of microRNA miR-122 in hepatic circadian gene expression

To synchronize gene expression with geophysical time, light-sensitive organisms have a molecular clock (Gachon et al. [70]). While the molecular clock has similar components in different tissue, its output varies substantially between tissues (Storch et al. [71]). Tissue specific regulation by miRNAs might explain these differences. The involvement of miRNA-122 that constitutes more than 70% of all miRNA molecules in hepatocytes (Lagos-Quintana et al. [66]) was investigated. While miR-122 is transcribed in a circadian fashion, mature miR-122 remains nearly constant over the day. The miR-122 locus is regulated by the circadian clock component REV-ERBα.

	% targeted genes			Fisher test p-value	
Fraction / Test	**Up**	**Unchanged**	**Down**	**Up vs Unchanged balanced**	**Down vs Unchanged balanced**
Genes	329	12521	197		
All seeds	67.1	38.5	29.1	10^{-22}	0.99
ΔG duplex	52.3	20.7	15.9	10^{-24}	0.92
Seed conservation	26.0	10.8	6.98	10^{-7}	1
TargetScan context score	56.7	31.0	26.4	10^{-18}	0.93
Above 3 filters	5.87	1.23	0.78	10^{-6}	0.81

Table 4 *Target site enrichment with different prediction filters for miR-122 KD at the mRNA level.* See **Table 1** for the column description.

The possible role of miR-122 was first investigated by a genome-wide identification of miR-122 targets by knocking-down (KD) its expression in hepatocytes at two time-points (Zeitgeber time ZT0, and ZT12) with two different controls. Triplicates of mRNA profiling following antisense oligonucleotides (ASOs) injection were performed on Affymetrix arrays. With a threshold of 1.5 on the fold-changes, I found 1.7 fold more genes in the up-regulated fraction compared to the unchanged fraction, and 2.3 fold more compared to the down-regulated fraction (**Table 4**), indicating strong direct effects of the miR-122 KD and limited indirect effects. I then tested if up-regulated genes were enriched for miR-122 seed-matches, and found a significant enrichment ($p=10^{-22}$ with Fisher's exact test). With target predictions refined with an energy-based model, conservation, or the TargetScan features (computed with miRmap with the 3 features of TargetScan), I also found significant enrichments. On the contrary, I found no enrichment between the down-regulated and unchanged fractions.

Similarly to the analysis of the Dicer KO, I computed the target site enrichment between the up-regulated and unchanged fractions for all miRNAs (**Figure 2**). Interestingly, miR-122 had the highest enrichment for all methods, expect for the energy based methods. While the enrichment for miR-122 was higher for the energy-based methods, miR-122 enrichment was at the third and tenth positions for these methods. However, the miRNAs that have an higher enrichment than miR-122 were highly AU-rich, introducing a bias in my analysis. For example, the first miRNA for *ΔG duplex* (miR-137) has an 82% AU-rich 5' half-part. Moreover, in the case of a KD experiment, the miRNA is expressed at endogenous level, which is rarely the case for overexpression experiments. The effect of a less abundant miRNA could possibly be more efficiently modeled by integrating kinetic effects.

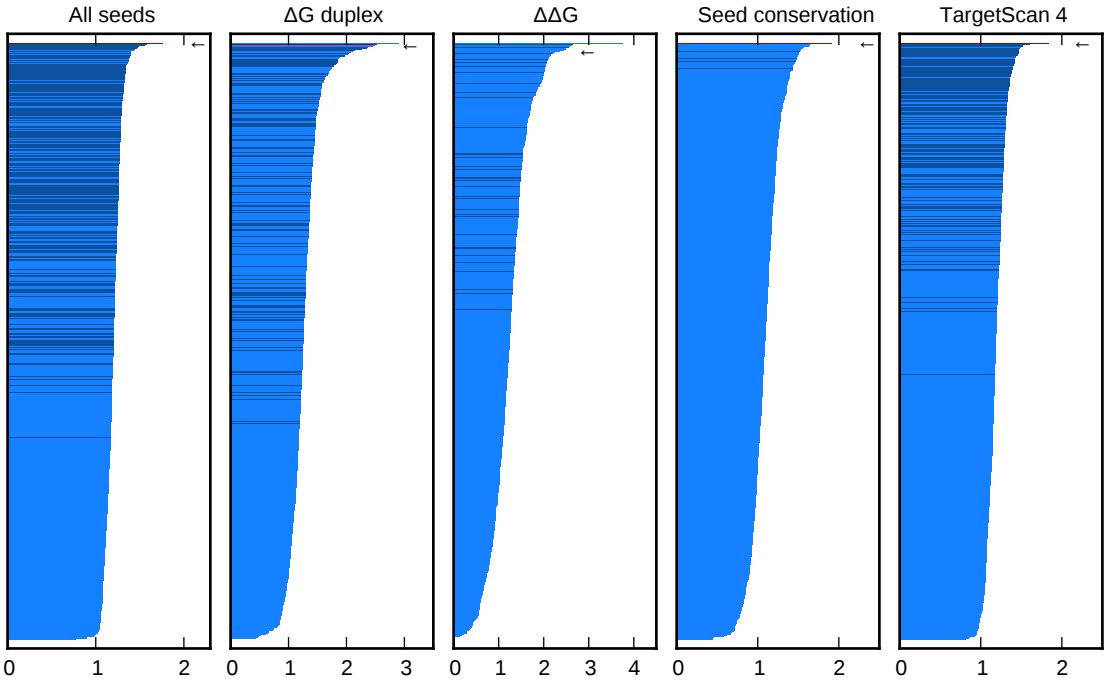


Figure 2 *Target site enrichment for miR-122 target sites in the up-regulated mRNA fraction.* For each miRNA annotated in the mouse genome, ratios of targeted mRNA in up-regulated and unchanged fractions are represented. Significant ratios were determined with one-sided Fisher test with Dunn-Šidák multiple test correction at 10%. The color code is explained on **Table 2**. An arrow indicates the position of miR-122.

Our half-life estimate of RISC-bound miR-122 exceeded 24h. Three scenarios were proposed to explain the role of circadian miR-122 production. First, a gene expressed in a circadian manner would increase its ratio between high and low expressed levels if a miRNA is permanently repressing a basal level. Second, different miR-122 sub-populations can exist. For example, RNA editing could make a subpopulation inefficient to target circadian genes. Third, newly synthesized miR-122 are ready to bind mRNAs, whereas old RISCs might already be bound and therefore less available to repress newly synthesized mRNAs.

Downloaded from genesdev.cshlp.org on September 4, 2009 - Published by Cold Spring Harbor Laboratory Press

Integration of microRNA miR-122 in hepatic circadian gene expression

David Gatfield,[1,10] Gwendal Le Martelot,[1,8] Charles E. Vejnar,[2,3,8] Daniel Gerlach,[2,3] Olivier Schaad,[4] Fabienne Fleury-Olela,[1] Anna-Liisa Ruskeepää,[5] Matej Oresic,[5] Christine C. Esau,[6] Evgeny M. Zdobnov,[2,3,7] and Ueli Schibler[1,9]

[1]Department of Molecular Biology, Sciences III, University of Geneva, 30, CH-1211 Geneva, Switzerland; [2]Department of Genetic Medicine and Development, University of Geneva Medical School, CH-1211 Geneva, Switzerland; [3]Swiss Institute of Bioinformatics, CH-1211 Geneva, Switzerland; [4]Genomics Platform, University of Geneva Medical School, CH-1211 Geneva, Switzerland; [5]VTT Technical Research Centre of Finland, FI-02044 VTT, Finland; [6]Regulus Therapeutics, Carlsbad, California 92008, USA; [7]Imperial College London, SW7 2AZ London, United Kingdom

In liver, most metabolic pathways are under circadian control, and hundreds of protein-encoding genes are thus transcribed in a cyclic fashion. Here we show that rhythmic transcription extends to the locus specifying miR-122, a highly abundant, hepatocyte-specific microRNA. Genetic loss-of-function and gain-of-function experiments have identified the orphan nuclear receptor REV-ERBα as the major circadian regulator of mir-122 transcription. Although due to its long half-life mature miR-122 accumulates at nearly constant rates throughout the day, this miRNA is tightly associated with control mechanisms governing circadian gene expression. Thus, the knockdown of miR-122 expression via an antisense oligonucleotide (ASO) strategy resulted in the up- and down-regulation of hundreds of mRNAs, of which a disproportionately high fraction accumulates in a circadian fashion. miR-122 has previously been linked to the regulation of cholesterol and lipid metabolism. The transcripts associated with these pathways indeed show the strongest time point-specific changes upon miR-122 depletion. The identification of *Pparβ/δ* and the peroxisome proliferator-activated receptor α (PPARα) coactivator *Smarcd1/Baf60a* as novel miR-122 targets suggests an involvement of the circadian metabolic regulators of the PPAR family in miR-122-mediated metabolic control.

[*Keywords*: Circadian; miRNA; miR-122; metabolism; clock; PPAR]

Supplemental material is available at http://www.genesdev.org.

Received January 12, 2009; revised version accepted April 20, 2009.

Light-sensitive organisms possess a circadian timekeeping system that serves to synchronize gene expression and physiology with geophysical time (Reppert and Weaver 2002; Gachon et al. 2004). Current models of the mammalian molecular clocks are based on two interlocked transcriptional feedback loops (Sato et al. 2006): a positive limb, in which the heterodimeric BMAL1:CLOCK transcription factor mediates the transcriptional activation of cryptochrome (*Cry1* and *Cry2*) and period genes (*Per1* and *Per2*), and a negative limb, in which PER:CRY complexes repress the BMAL1:CLOCK-mediated transcription of their own genes. Coordination between the two limbs is accomplished by nuclear receptors of the REV-ERB and ROR families (Preitner et al. 2002; Reppert and Weaver 2002; Sato et al. 2004). Cyclic *Rev-erbα* transcription is regulated by the mechanisms described above for *Cry* and *Per* genes, and the circadian accumulation of the repressor REV-ERBα results in the rhythmic repression of target genes, such as *Bmal1*, carrying retinoid-related orphan receptor elements (ROREs) (Ueda et al. 2002). In addition to these transcriptional feedback loops, numerous post-translational modifications of core clock proteins are known to contribute to the rhythm-generating clockwork circuitry (Gallego and Virshup 2007).

The cyclic expression of clock output genes can be governed directly by core clock components via E-box or RORE sequences (Ueda et al. 2002), or transcription factors such as PAR bZip proteins whose genes are regulated by these mechanisms (Gachon et al. 2004). However, despite the similar molecular makeup of the core oscillator in different organs, its outputs vary substantially between tissues (e.g., Storch et al. 2002). Gene expression profiling in liver has suggested that, depending on the algorithms used for the identification of cyclically expressed genes, 2%–10% of the transcriptome may be under circadian control (Panda et al. 2002; Storch et al.

[8]**These authors contributed equally to this work.**
Corresponding authors.
[9]**E-MAIL ueli.schibler@unige.ch; FAX 41-22-3796868.**
[10]**E-MAIL david.gatfield@unige.ch; FAX 41-22-3796868.**
Article is online at http://www.genesdev.org/cgi/doi/10.1101/gad.1781009.

Downloaded from genesdev.cshlp.org on September 4, 2009 - Published by Cold Spring Harbor Laboratory Press

2002; Kornmann et al. 2007a; Miller et al. 2007). Many of these genes are involved in hepatocyte-specific metabolic pathways.

In part, the synergistic activation of genes by circadian and tissue-specific transcription factors may account for the rhythmic expression of cell type-specific transcripts. However, tissue-specific post-transcriptional regulation of gene expression may also participate in this endeavor. It is estimated that in mammals ~30% of all mRNAs are subject to regulation by microRNAs (miRNAs) (Lewis et al. 2005), and miRNAs have been implicated in the post-transcriptional control of cellular proliferation, development, and differentiation (Bushati and Cohen 2007). miRNAs are short (~22 nucleotides [nt]), endogenous RNAs that promote translational repression and/or destabilization of target mRNAs (Bushati and Cohen 2007; Liu 2008). Target recognition occurs via base-pairing interactions with the 3′ untranslated region (UTR). Usually the 5′ portion of the miRNA forms a perfect hybrid with a 6- to 8-nt seed site, whereas the remainder of the miRNA undergoes interactions of only partial complementarity with the 3′UTR of its target mRNA (Lewis et al. 2005). The mismatches and gaps between miRNA and mRNA duplexes render the de novo prediction of miRNA targets challenging. Generally, a given miRNA can be expected to fine-tune the production of large sets of proteins within the cell (Baek et al. 2008; Liu 2008; Selbach et al. 2008).

Given the large fraction of mRNAs targeted by miRNAs, it is likely that miRNAs also modulate clock and clock output functions (Cheng et al. 2007; Xu et al. 2007; Yang et al. 2008). We wished to examine this conjecture and initiated our studies with miR-122, a miRNA that has been proposed to constitute up to 70% of all miRNA molecules in hepatocytes (Lagos-Quintana et al. 2002). The knockdown of miR-122 expression in mice and monkeys has previously been recognized to result in a down-regulation of cholesterol and lipid metabolizing enzymes and a reduction in plasma cholesterol levels (Krutzfeldt et al. 2005; Esau et al. 2006; Elmen et al. 2008a,b). Both lipid and cholesterol metabolism are well known for their daytime-dependent regulation, similar to many other hepatic functions that require coordination of food intake with nutrient-processing and energy homeostasis (Panda et al. 2002).

Here, we show that transcription of the miR-122 locus is under circadian control, involving the transcriptional repressor REV-ERBα. Thus, pri- and pre-miRNA precursors oscillated about fourfold to 10-fold in abundance during the day but accumulated at nearly constant levels in the livers of *Rev-erbα* knockout mice. However, due to its high stability mature miR-122 levels were virtually constant throughout the day. Despite the apparent invariable temporal accumulation of miR-122, the identification of its target mRNAs suggested that miR-122 nevertheless participates in the circadian control of many transcripts involved in hepatic metabolism. Among the miR-122 targets we found the mRNAs encoding peroxisome proliferator-activated receptor β/δ (PPARβ/δ) and SMARCD1/BAF60a, which are themselves circadian regulators of metabolism (Yang et al. 2006; Seedorf and Aberle 2007; Li et al. 2008).

Results

The miR-122 locus is transcribed in a circadian fashion

In a search for miRNAs that could shape the circadian expression of target mRNAs, we analyzed the expression of various miRNAs in mouse liver at different time points (Zeitgeber time, ZT) around the day. Several miRNAs (miR-19, miR-20, miR-22, miR-24, miR-30, miR-92, miR-126-3p), some of which had been predicted to target clock components (Lewis et al. 2005), only showed modest, if any, circadian changes in expression, as judged by Northern blot analysis (Supplemental Fig. 1). However, analysis of miR-122, the most abundant miRNA in liver, revealed that pre-mir-122 oscillated with an approximately fivefold daily amplitude in abundance, whereas mature miR-122 levels remained nearly constant over the day (Fig. 1A,B). Pre-mir-122 is a 66-nt hairpin-shaped precursor molecule from which the endonuclease Dicer cleaves the mature 22-nt miR-122. The mature miRNA is then incorporated into the RNA-induced silencing complex (RISC). The same expression pattern for pre-mir-122 was detected with a probe recognizing the strand complementary to the miRNA (known as the miRNA* sequence) (Fig. 1). The observed circadian changes in pre-mir-122 levels could be the result of either circadian synthesis or circadian processing into mature miRNA. To distinguish between these possibilities we analyzed the circadian levels of the miR-122 primary transcript, pri-mir-122, a ~5-kb precursor (Chang et al. 2004), from which the pre-miRNA is cleaved by the Drosha-containing microprocessor complex.

As shown in Figure 1A (bottom panels), pri-mir-122 accumulation was highly circadian (~10-fold amplitude), showing a similar phase as pre-mir-122 (i.e., minimal levels at ZT8-12 and maximal levels at ZT24). We wanted to test if high-amplitude circadian precursors were specific for miR-122 or were a common feature of miRNAs. Two other loci tested, pri-mir-17-92 and pri-mir-22, did not show the circadian pattern observed for pri-mir-122 (Supplemental Fig. 1C,D). This suggested that specifically the miR-122 locus was transcribed in a circadian fashion. The two intermediates in miR-122 biogenesis can be expected to be short-lived and reflect the rate at which the gene is transcribed. In contrast, the absence of cyclic expression at the level of mature miR-122 was probably due to its high metabolic stability. Indeed, based on Northern blot experiments, we estimated that the ratio of miR-122/pre-mir-122 steady-state levels (which is largely determined by the ratio of the half-lives of the two species) is in the range of 400:1. If one assumes that the pre-mir-122 half-life is a few minutes, this means that the miR-122 half-life is probably well beyond 24 h.

The orphan nuclear receptor REV-ERBα drives circadian mir-122 transcription

We wished to study the molecular mechanism accounting for circadian mir-122 transcription. The phase of

Downloaded from genesdev.cshlp.org on September 4, 2009 - Published by Cold Spring Harbor Laboratory Press

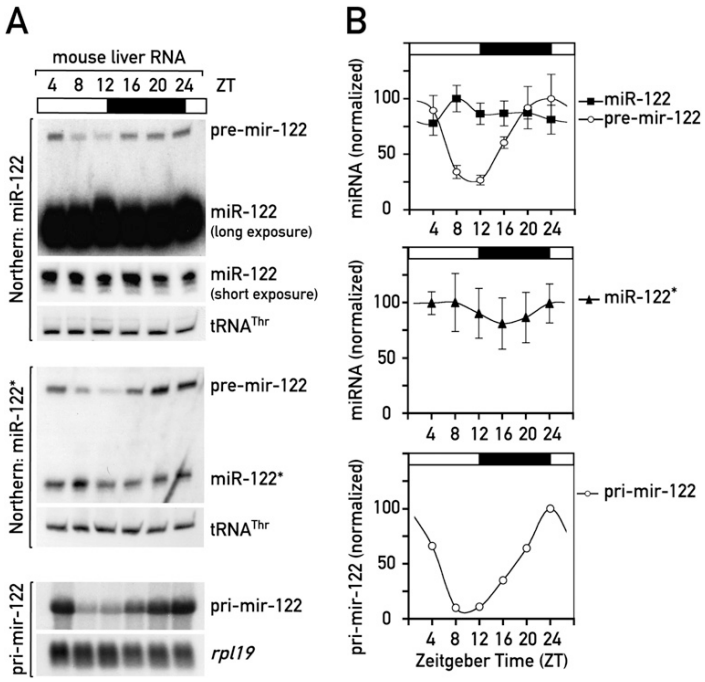


Figure 1. miR-122 precursors are circadian in mouse liver. (*A*) Northern blot analysis of miR-122 and its precursor RNAs using whole-cell RNA from male C57BL/6 mice sacrificed at the indicated ZT values around the clock. An RNA pool from three mice was used per time point, $tRNA^{Thr}$ and *rpl19* mRNA served as loading controls in denaturing polyacrylamide (*top* and *middle* panels) and agarose gel electrophoresis (*bottom* panels), respectively. (*Top* panels) miR-122 and pre-mir-122. (*Middle* panels) pre-mir-122 and miR-122*. miR-122* is the antisense "passenger strand" that is incorporated into RISC at low levels. (*Bottom* panels) pri-mir-122. (*B*, *top* and *middle* panels) miR-122, miR-122* and pre-mir-122 levels, normalized to $tRNA^{Thr}$, from Northern blots in which single animals were analyzed (data not shown). Mean values ± SEM. (*Bottom* panel) Quantification of pri-mir-122 levels, normalized to the circadianly invariant *rpl19*, from the Northern blot shown in *A*.

pri-/pre-mir-122 expression suggested that the circadian transcriptional repressor REV-ERBα might be involved: REV-ERBα protein expression peaks at around ZT8, leading to minimal transcript levels for REV-ERBα target genes at around ZT12 (Preitner et al. 2002; Ueda et al. 2002). Consistent with the hypothesis of miR-122 being a REV-ERBα target gene, the mir-122 promoter contains two conserved ROREs ~120–160 base pairs (bp) upstream of the transcriptional start site (Fig. 2A; see also Supplemental Fig. 2 for an alignment of the promoter region in 32 mammalian species). More importantly, the amplitude of cyclic pri-mir-122 accumulation was severely blunted in the livers of *Rev-erbα* knockout animals (Fig. 2B,C), and mature miR-122 accumulated to 1.6-fold higher levels (Fig. 2D). The residual amplitude in mir-122 transcription was possibly caused by REV-ERBβ, a highly related paralog of REV-ERBα (Preitner et al. 2002). A second mouse model, in which REV-ERBα was overexpressed specifically in hepatocytes (Kornmann et al. 2007a), showed the converse effect (i.e., 1.7-fold reduced miR-122 levels). In summary, these findings supported a model according to which the miR-122 locus is regulated by the circadian clock component REV-ERBα.

Does the miR-122 locus specify multiple functional RNAs?

Since the accumulation of pri-mir-122, but not that of mature miR-122, was rhythmic, we considered that this locus produced additional biologically active RNAs with shorter half-lives than miR-122. In fact, several pri-miRNAs are polycistronic and produce multiple miRNAs (Sewer et al. 2005). Although mature miR-122 shows a sequence conservation of 100% from fish to humans (Gerlach et al. 2009), its pri-miRNA gene structure is conserved only in mammals. In these organisms, the transcription start site is flanked by elements of a classical RNA polymerase II (PolII)-dependent promoter, which drives transcription of the ~5-kb capped and polyadenylated pri-mir-122 containing the pre-mir-122 hairpin at its 3′-end (Fig. 2A; Supplemental Figs. 2, 3A; Chang et al. 2004). Overall, the pri-mir-122 sequence is poorly conserved, and we did not detect additional potential miRNAs (or conserved open reading frames) within the primary transcript. A thorough bioinformatics search for conserved RNA secondary structures within the pri-mir-122 genomic locus in the genomes of six mammalian species also failed to identify additional RNA structures that could carry a function (Supplemental Fig. 3). Thus, it appeared likely that a potential biological function associated with the circadian control of pri-mir-122 transcription was mediated by miR-122 itself.

Genome-wide identification of miR-122 targets

As miR-122 was produced in a circadian fashion, we wondered whether it might assume rhythmic functions despite its long half-life. We decided to approach this question in an unbiased way by identifying putative miR-122 targets. In particular, we wished to determine whether there are targets whose daily rhythms are influenced by miR-122. To deplete miR-122, we injected antisense oligonucleotides (ASOs) intraperitoneally into mice (termed 122ASO in the following sections) and used genome-wide Affymetrix oligonucleotide arrays to determine the impact this had on hepatic mRNA levels. As

Downloaded from genesdev.cshlp.org on September 4, 2009 - Published by Cold Spring Harbor Laboratory Press

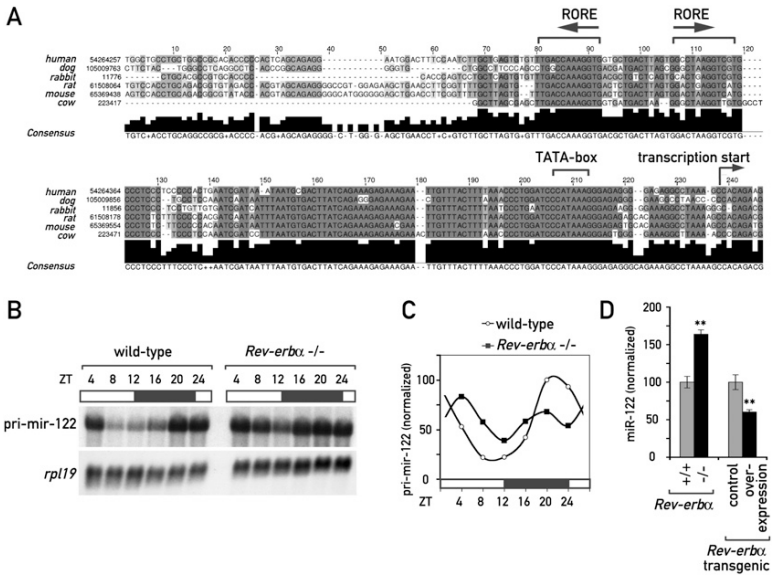


Figure 2. REV-ERBα is involved in circadian control of the miR-122 locus. (*A*) Alignment of the genomic sequence upstream of the predicted transcriptional start site of pri-mir-122 in six mammalian species (extracted from the University of California at Santa Cruz alignment; see Supplemental Fig. 3). The predicted ROREs, TATA-box, and transcriptional start site are indicated. (*B*) Northern blot analysis of pri-mir-122 in total RNA samples from *Rev-erbα* knockout and littermate control mice sacrificed at the indicated ZT values around the clock. For each time point, an RNA pool of three female mice was used. (*C*) Quantification of the Northern blot shown in *B*; values are pri-mir-122 normalized to *rpl19*. (*D*) miR-122 levels in total liver RNA from individual animals (mixed ZTs) of the indicated genotypes were quantified by Northern blot (data not shown). Control animals were set to 100%. Data are mean ± SEM ($n = 36$ for *Rev-erbα*$^{-/-}$ vs. *Rev-erbα*$^{+/+}$ and $n = 18$ for REV-ERBα overexpression vs. control); (**) $P < 5*10^{-5}$ (two-tailed Student's *t*-test).

controls, we used animals treated with ASOs targeting a miRNA not expressed in liver (miR-124; samples 124ASO in the following sections) or with PBS alone. Mice were sacrificed at time-points ZT0 and ZT12, when pri-mir-122 transcription was highest and lowest, respectively.

The efficiency of miR-122 depletion was between 89% and 99% as judged by Northern blot hybridization (Fig. 3A,B). Residual miR-122 levels were consistently lower for mice sacrificed at ZT12, when miR-122 production was low, suggesting that miRNA stability was decreased by 122ASOs. Importantly, the abundance of the unrelated miRNA let-7a remained unchanged (Fig. 3A), demonstrating the specificity of the ASO. To functionally assess if miR-122 was sufficiently depleted to derepress its targets, we determined the mRNA levels of the formerly suggested targets *glycogen synthase 1* (*Gys1*) and *aldolase A* (*AldoA*) by quantitative RT–PCR (qPCR). Similar to what had been observed previously (Krutzfeldt et al. 2005; Esau et al. 2006), these mRNAs were up-regulated twofold to sevenfold (Fig. 3C).

miRNAs initially have been proposed to mediate translational repression of their target mRNAs. This is often accompanied by a decrease in mRNA abundance (Baek et al. 2008; Selbach et al. 2008). Transcriptomal profiling using microarrays is therefore a convenient means to

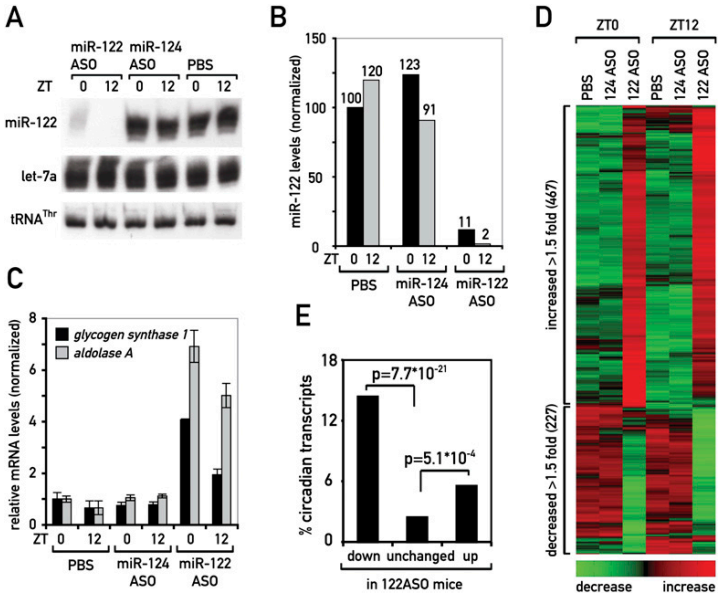


Figure 3. Analysis of miR-122 targets at two time points ZT0 and ZT12. (*A*) Northern blot analysis of miR-122, let-7a, and tRNA^{Thr} of mice treated with miR-122 ASO, miR-124 ASO, or PBS. Pools of RNA of three mice were loaded per lane. (*B*) Quantification of Northern blot shown in *A*. (*C*) qPCR analysis of RNAs from individual mice treated with the ASOs or PBS, as indicated. Probes used were for the known miR-122 targets *glycogen synthase 1* and *aldolase A*, normalized to 45S pre-rRNA. Values are mean ± SEM ($n = 3$). (*D*) Heat map of the probe sets up- and down-regulated in 122ASO-treated animals relative to both control groups, 124ASO- and PBS-treated animals (cutoff 1.5). The heat scale at the *bottom* of the panel represents changes on a linear scale, where green and red represent minimal and maximal expression, respectively. (*E*) Enrichment for circadian transcripts in the up- and down-regulated fractions in 122ASO mice. *P*-values were determined by a χ^2 test.

Downloaded from genesdev.cshlp.org on September 4, 2009 - Published by Cold Spring Harbor Laboratory Press

identify potential miRNA targets. Obviously, this technology is unable to detect miRNA targets whose translational attenuation is not accompanied by increased degradation.

Using Affymetrix microarray hybridization, we detected signals for a total of 22,384 probe sets, representing 11,638 transcripts. Among these, we found 343 transcripts (represented by 467 probe sets) that were up-regulated, and 188 transcripts (227 probe sets) that were down-regulated at at least one of the two time points in 122ASO-treated animals, when we applied a 1.5-fold expression change cutoff (Fig. 3D). We next analyzed whether transcripts up-regulated in 122ASO livers were enriched for potential miR-122 targets. For the prediction of potential miR-122-binding sites we applied a model that takes into account both the presence of miRNA seed sites and the energy of miRNA:mRNA duplexes, ensuring that energetically stable miRNA–target interactions are considered. Using this thermodynamic model (with an energy cutoff of −15 kcal/mol), we observed that 52% of transcripts in the up-regulated fraction contained a predicted miR-122-binding site (Supplemental Fig. 4). With only 22% of transcripts in the unchanged and 14% in the down-regulated fraction, this enrichment in the up-regulated fraction was statistically highly significant (up vs. unchanged: P-value $\sim 10^{-39}$; up vs. down: P-value $\sim 10^{-17}$). The differences between the unchanged and down-regulated fractions, however, were barely significant (P-value ~0.02). With a less elaborate model that only considers seed site presence, the enrichment for potential miR-122 targets in the up-regulated fraction was significant as well (Supplemental Fig. 4).

We next wished to determine, whether transcripts showing a time point-specific regulation by miR-122 could be clustered into particular metabolic pathways. To this end, we selected the transcripts that showed regulation upon 122ASO treatment exclusively at one of the two time points. Genome ontology (GO) analyses in the down-regulated fraction revealed that the genes involved in lipid and cholesterol metabolism (which had been reported previously to be most responsive to miR-122 depletion) also showed the strongest temporal regulation ($P \sim 10^{-10}$). Thus, the down-regulation of these mRNAs was significantly stronger at ZT12 than at ZT0 (Supplemental Fig. 5A). For up-regulated genes, transcripts belonging to GO:9607 "response to biotic stimuli" were most overrepresented ($P \sim 10^{-7}$). Their up-regulation occurred mainly at ZT0 and less so at ZT12 (Supplemental Fig. 5B). These observations suggested a considerable amount of cross-talk between circadian gene expression and miR-122, and encouraged us to analyze the effect of miR-122 depletion on circadian gene expression in greater detail.

Circadian transcripts are highly enriched among miR-122 targets

We wished to focus on transcripts that were direct potential targets of miR-122 and that showed circadian expression. For the genome-wide analysis of cyclic transcripts, we used previously reported transcriptome profiling experiments (Kornmann et al. 2007a). This work analyzes the hepatic transcriptome in a transgenic mouse model in which REV-ERBα can be conditionally overexpressed in liver in a doxycycline-dependent manner (tet-off system). In the presence of doxycycline, the hepatic circadian clock is functional in these animals, as the *Rev-erbα* transgene is constitutively repressed. The gene expression profiles from these animals, sampled over a 48-h period (with a resolution of 4 h), have been used to identify the circadian hepatic transcriptome using stringent algorithms (Kornmann et al. 2007a,b). In the absence of doxycycline, REV-ERBα overexpression arrests the endogenous liver clock. Thus, most circadian genes lose rhythmicity, with the notable exception of a small fraction of transcripts whose rhythms are driven by systemic cues rather than local oscillators (Kornmann et al. 2007a,b). In these mice, REV-ERBα overexpression also led to reduced miR-122 levels (Fig. 2D). It may thus be assumed that the derepression of miR-122 targets contributed to the gene expression changes observed upon REV-ERBα overexpression. We therefore compared the gene expression changes common to REV-ERBα overexpression and 122ASO administration. Of the transcripts whose abundance changed under both conditions, the majority (79.2%) indeed showed regulation in the same direction and only few (20.8%) showed reverse regulation (Supplemental Fig. 6). These observations lend further support to a role of REV-ERBα in miR-122 regulation.

We next analyzed the probe sets representing transcripts with circadian accumulation. Using stringent algorithms, these corresponded to ~2.8% of the liver transcriptome (Kornmann et al. 2007a,b). We found that the up- and down-regulated fractions in the 122ASO mice were significantly enriched for circadian transcripts: 14.4% of the down-regulated, 5.5% of the up-regulated, but only 2.4% of the unchanged mRNAs were among those classified as circadian (Fig. 3E). We thus concluded that the effects of depleting miR-122 were biased toward a misregulation of circadian transcripts. Since the enrichment was particularly high in the down-regulated fraction that contained indirect miR-122 targets, there were possibly common circadian regulatory mechanisms in control of this group. Indeed, almost a quarter of the transcripts in this fraction belonged to lipid/cholesterol metabolizing enzymes.

Identification of circadian mRNAs with functional miR-122-binding sites

We next investigated in more detail the group of transcripts with circadian accumulation that were up-regulated upon miR-122 depletion, as this subset was likely to contain the direct miR-122 targets (Fig. 4A). Within this group, 16 transcripts (specified by 19 probe sets) contained potential miR-122-binding sites in their 3′UTRs and were therefore candidates for circadianly expressed miR-122 targets (Fig. 4A, bold type). Many of them were also up-regulated in REV-ERBα-overexpressing

Downloaded from genesdev.cshlp.org on September 4, 2009 - Published by Cold Spring Harbor Laboratory Press

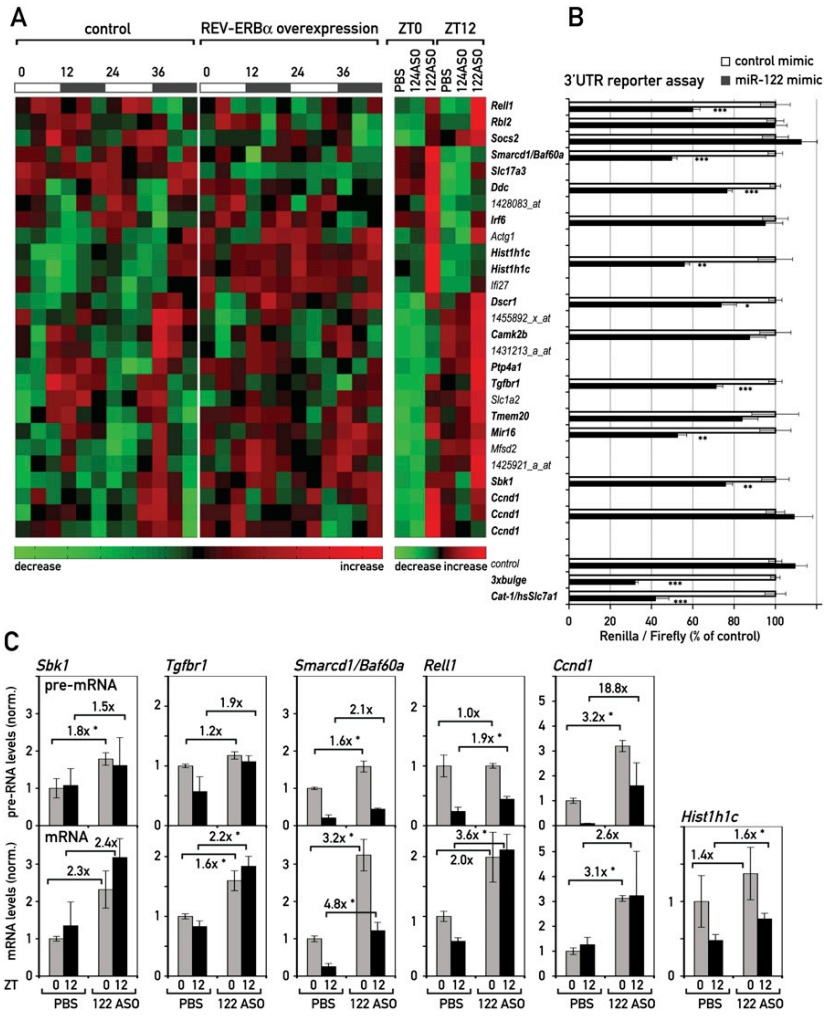


Figure 4. Circadian genes are miR-122 targets. (*A*) Heat map of the circadian probe sets (*left* and *middle* panel; taken from Kornmann et al. 2007b) that are up-regulated in 122ASO mice (*right* panel). *Smarcd1/Baf60a* was just below the stringent criteria used for circadian expression in the microarray data of Kornmann et al. (2007b), but was also included in the figure as it was confirmed as robustly circadian by qPCR (see Fig. 5). Heat scales at the *bottom* of the panels represent changes on a linear scale with green and red representing minimal and maximal expression, respectively. Transcripts in bold type contain potential miR-122 seed sites in their 3′UTRs. (*B*) The effect of miR-122 mimics in a 3′UTR luciferase assay. Control has only the vector 3′UTR, containing no seed sites. 3xbulge and *Cat-1/hsSlc7a1* are positive controls for 3′UTRs known to be regulated by miR-122. Values are mean ± SEM ($n \geq 6$ per transfection). (*) $P < 10^{-2}$; (**) $P < 10^{-3}$; (***) $P < 10^{-4}$ (two-tailed Student's *t*-test). (*C*) qPCR analysis in 122ASO mice and PBS controls of pre-mRNA (*top* panels) and mRNA (*bottom* panels) levels of selected transcripts from *A*. *Hist1h1c* is an intron-less gene; hence, pre-mRNA levels were not measured. Note that *Ccnd1* is also changed on the pre-mRNA level and is hence probably up-regulated by an indirect, transcriptional effect. Data are mean values of three mice per condition ±SEM. (*) $P < 10^{-2}$ (two-tailed Student's *t*-test).

animals (Fig. 4A, middle panel). We wished to verify that the changes in mRNA abundance detected by microarray analysis were potentially the direct result of miR-122 derepression, as opposed to more complicated indirect effects. Therefore, we tested the impact of miR-122 on the 3′UTRs of several candidate transcripts in cotransfection experiments. To this end, we cloned the candidate 3′UTRs into a vector carrying a renilla luciferase reporter gene, and transfected these constructs together with synthetic miRNA mimics into Hela cells, which do not express endogenous miR-122. We then measured the ability of a miR-122 mimic to inhibit the expression of luciferase when its open reading frame was followed by a particular 3′UTR. Two 3′UTRs known to be regulated by miR-122 served as positive controls: an artificial 3′UTR containing three optimized miR-122-binding sites (3xbulge) (Pillai et al. 2005), and the 3′UTR of *Cat-1/human Slc7a1*, a well-known miR-122 target (Chang et al. 2004; Bhattacharyya et al. 2006). These two 3′UTRs mediated a miR-122-dependent repression by about 68% and 58%, respectively. In contrast, luciferase expression from reporters harboring the vector-based 3′UTR devoid of miR-122 seed sites was not affected (Fig. 4B; Supplemental Fig. 7). Of the circadian transcripts up-regulated in 122ASO mice, we found that the 3′UTRs of *Rell1* (receptor expressed in lymphoid tissues-like 1), *Smarcd1/Baf60a* (SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily d, member 1/BRG1-associated factor 60a), *Ddc* (dopa decarboxylase), *Hist1h1c* (histone cluster 1, H1c), *Dscr1* (down syndrome critical region protein 1), *Tgfbr1* (TGF-β receptor type 1), *Mir16* (membrane-interacting protein of RGS16), and *Sbk1* (SH3-binding kinase 1) conferred sensitivity toward miR-122 (Fig. 4B). A complete compilation of the >30 3′UTRs we tested, including those of several newly identified miR-122 targets, is given in Supplemental Figure 7.

miR-122 contributes to circadian mRNA expression

For some selected targets, we wanted to verify that their up-regulation in the 122ASO mice was indeed caused by

Downloaded from genesdev.cshlp.org on September 4, 2009 - Published by Cold Spring Harbor Laboratory Press

post-transcriptional, rather than indirect transcriptional mechanisms. Since miRNAs are thought to act on processed mRNAs, a derepression mediated by the 122ASO should manifest itself on the level of the mature mRNA, but not on that of its pre-mRNA. Indirect effects, however, can be expected to occur through changes in transcription rates, caused by the up-regulation of activators or repressors whose production depends on miR-122. These changes should also be visible on the pre-mRNA level. Hence, we designed qPCR probes enabling us to measure mRNA and intron-containing pre-mRNA levels of several of the identified targets. Our analyses showed that the up-regulation of mature mRNA levels for the transcripts *Sbk1, Tgfbr1, Smarcd1/Baf60a, Rell1,* and *Hist1h1c* was similar, or even greater, than assessed by the microarray analysis. The effects of the 122ASO on pre-mRNA levels, however, were less pronounced (Fig. 4C). In contrast, a transcript such as *Ccnd1* fulfills the criteria for being indirectly affected. Thus, while *Ccnd1* was also circadian and up-regulated in 122ASO mice (Fig. 4A), it did not confer sensitivity to miR-122 in the 3′UTR assay (Fig. 4B). In keeping with this observation, the changes in *Ccnd1* expression were already observed on the level of *pre-Ccnd1* mRNA accumulation (Fig. 4C).

To evaluate more precisely which influence miR-122 had on shaping the rhythmic accumulation of these transcripts, we extended our analyses to 122ASO mice that had been sacrificed at six time points around the clock. Using RNA pools from three to four animals per time point and for both control and 122ASO mice (see Supplemental Fig. 8), we observed similar increases in target mRNA accumulation as in the previous two time point experiments (Fig. 5A; Supplemental Fig. 8D). In addition, it was apparent that miR-122 depletion had striking effects on the circadian amplitude (*Smarcd1/Baf60a, Ddc, Hist1h1c*), magnitude (*Rell1*) and phase (*Smarcd1/Baf60a, Hist1h1c,* and *Ddc*) of accumulation (Fig. 5A, bottom panels). For several transcripts (*Smarcd1/Baf60a, Ddc,* and *Hist1h1c*) we also observed that derepression caused an especially strong up-regulation at around ZT4 (Fig. 5A, bottom panels). This time point corresponds to a few hours after maximal mir-122 transcription (see Fig. 1B). Moreover, despite a particularly efficient miR-122 depletion at ZT12 (Fig. 3A,B; Supplemental Fig. 8), derepression clearly had a milder effect at this time point (Fig. 5A, bottom panels). For some of the miR-122 targets, these time-dependent effects were already observed in the microarray data (Fig. 4A). Due to their low abundance, the detection of the corresponding pre-mRNAs was less robust than that of the mature transcripts (Fig. 5A, top panels). Nevertheless, it was evident that (with the exception of *Ccnd1*) differences between 122ASO and control mice could not be accounted for by different transcription rates. These findings indicated that miR-122 probably assumes rhythmic functions despite its constant levels (see the Discussion). Importantly, the circadian clock per se did not appear to be affected by 122ASO treatment, as the mRNA levels of core clock and clock output genes were

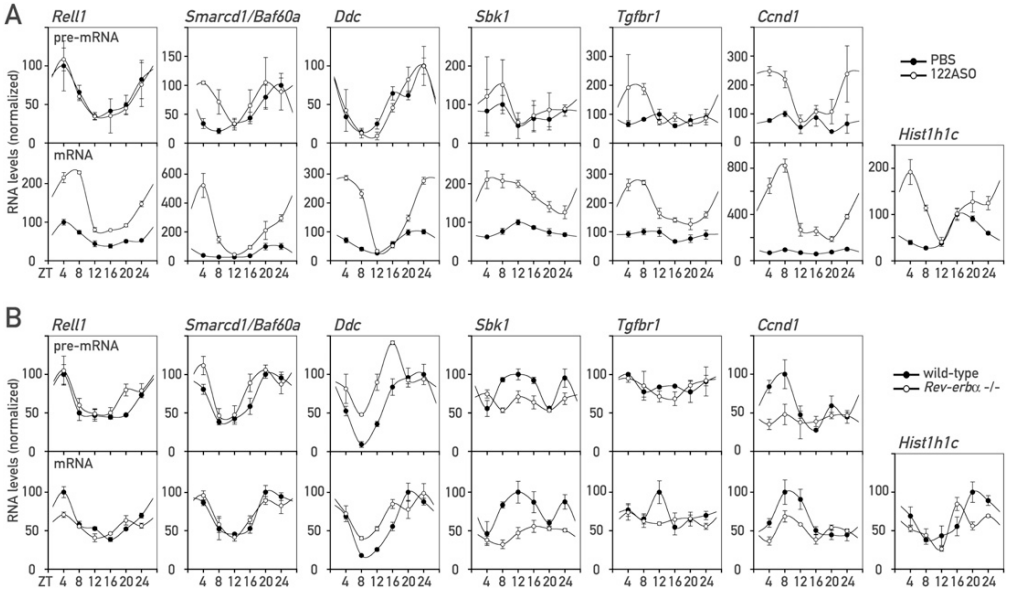


Figure 5. miR-122 targets in 122ASO and *Rev-erbα* knockout mice around the clock. (*A*) pre-mRNA (*top* panels) and mRNA (*bottom* panels) levels for the indicated transcripts in 122ASO and PBS-injected control mice around the clock. For each data point, transcript levels were measured in triplicate by qPCR using a pool of total liver RNA isolated from three to four mice. Due to low abundance, the detection of pre-mRNA levels was less robust, as indicated by generally larger error bars (standard deviations) in the qPCR analysis. (*B*) As in *A*, pre-mRNA (*top* panels) and mRNA (*bottom* panels) levels measured around the clock in *Rev-erbα* knockout and wild-type littermate animals, using a pool of whole-cell liver RNA isolated from five female mice.

Downloaded from genesdev.cshlp.org on September 4, 2009 - Published by Cold Spring Harbor Laboratory Press

essentially unchanged in 122ASO mice (Supplemental Fig. 9).

As shown in Figure 2, the circadian amplitude of mir-122 transcription was blunted in *Rev-erbα* knockout mice (Fig. 2B,C), leading to ~1.6-fold higher miR-122 levels (Fig. 2D). We wanted to examine whether these alterations in miR-122 production were sufficient to perturb the rhythmic expression of any of the targets analyzed above. When measuring mRNA and pre-mRNA abundances around the clock in *Rev-erbα* knockout and control animals, we observed the strongest post-transcriptional perturbations of rhythms for *Rell1*, a ubiquitously expressed member of the tumor necrosis factor (TNF) receptor family (Cusick et al. 2006). The amplitude of *Rell1* mRNA, but not that of its pre-mRNA, was blunted in the *Rev-erbα* knockout (1.7-fold amplitude) as compared with the control mice (2.6-fold amplitude) (Fig. 5B). This further supported our conclusion that *Rell1* was an example for an mRNA whose circadian rhythm was partially shaped by post-transcriptional mechanisms. In view of its up-regulation in 122ASO mice on the mRNA but not the pre-mRNA level (Figs. 4C, 5A) and the fact that the *Rell1* 3′UTR conferred sensitivity to miR-122 in the reporter assay (Fig. 4B), it is likely that miR-122 was directly implicated in this process. The cyclic accumulation of other miR-122 targets, such as *Smarcd1/Baf60a*, was unchanged in *Rev-erbα* knockout mice, whereas some changes already occurred on the pre-mRNA level (Fig. 5B).

Cross-talk between miR-122 and PPARs

The down-regulation of enzymes associated with lipid and cholesterol metabolism (see Krutzfeldt et al. 2005; Esau et al. 2006; Elmen et al. 2008a; this study) in miR-122-depleted mice implies that the corresponding mRNAs are regulated by indirect mechanisms. However, the direct miR-122 targets responsible for these control mechanisms remained to be identified. We suspected that these direct targets were also expressed in a circadian manner, since the down-regulation of mRNAs encoding lipid and cholesterol enzymes was daytime-dependent (Supplemental Fig. 5A). Interestingly, recent work has suggested that SMARCD1/BAF60a, a component of the SWI/SNF chromatin-remodeling complex, specifically regulates hepatic lipid metabolizing genes (Li et al. 2008). In our experiments, *Smarcd1/Baf60a* mRNA appeared as a circadian and direct miR-122 target, as it was robustly up-regulated in 122ASO mice (Figs. 4A,C, 5A) and as its 3′UTR was responsive to miR-122 in our cotransfection experiments (Figs. 4B, 6A). Li et al. (2008) further demonstrated that SMARCD1/BAF60a interacts and cooperates with the metabolic regulator PPARα, and that SMARCD1/BAF60a and PPARα share a large number of target genes.

PPARs belong to the nuclear hormone receptor superfamily and are well-known metabolic regulators. They are activated upon binding to their mainly amphipathic ligands, which are mostly derived from dietary fat or endogenous fatty acid metabolism. Of the three PPAR

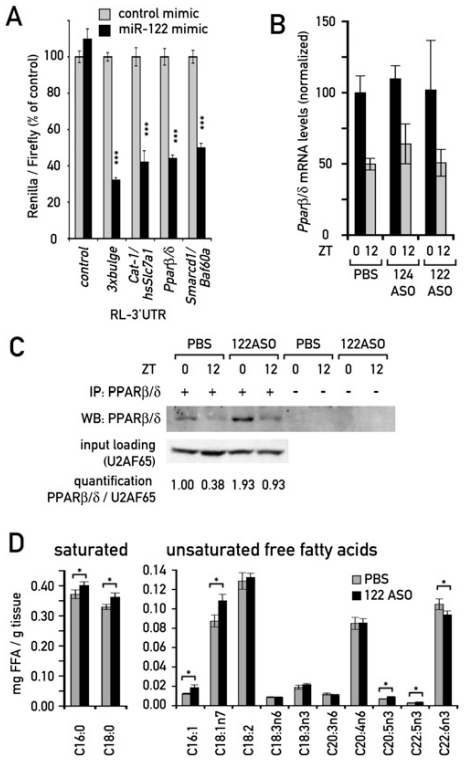


Figure 6. Cross-talk between miR-122 and PPAR receptors. (*A*) The effect of the miR-122 mimic in a 3′UTR luciferase assay as in Fig. 4B, using the *Pparβ/δ* and *Smarcd1/Baf60a* 3′UTRs. Values are mean ± SEM ($n \geq 9$ per transfection). (***) $P < 10^{-5}$ (two-tailed Student's *t*-test). (*B*) Expression levels of *Pparβ/δ* mRNA quantified from Northern blots. Data are mean ± SEM ($n = 3$ animals per condition). (*C*) Immunoprecipitation-Western blot of PPARβ/δ protein from 122ASO and PBS-treated mice, as described in the Materials and Methods. Each immunoprecipitation was performed from a pool of extracts from three mice. U2AF65 protein levels in the input of the same pool served as a loading control. (*D*) FFA levels in liver pieces from 122ASO- and PBS-injected animals, as determined by GC/MS. Values are mean ± SEM ($n = 6$). (*) $P < 0.05$ (two-tailed Student's *t*-test).

isotypes, PPARα, and the less-studied PPARβ/δ, serve predominantly catabolic functions, whereas PPARγ mainly promotes lipid storage in adipose tissue. In liver, all PPARs show circadian expression (Yang et al. 2006). Although we did not find PPAR transcripts misregulated using microarrays with RNA from 122ASO mice, we noticed that the *Pparβ/δ* 3′UTR contained four miR-122 seed sites that could be predicted to confer strong targeting by miR-122. We therefore tested if the *Pparβ/δ* 3′UTR showed sensitivity to miR-122 in our cotransfection experiments. Indeed, this 3′UTR caused a miR-122-dependent reduction of luciferase activity by 56%, which was among the highest down-regulation effects we observed in these assays. Only the two positive controls, the artificial 3xbulge and the *Cat-1/human Slc7a1* 3′UTR showed a slightly stronger repression (Fig. 6A; Supplemental Fig. 7). Consistent with these findings, we observed that whereas *Pparβ/δ* mRNA levels remained

Downloaded from genesdev.cshlp.org on September 4, 2009 - Published by Cold Spring Harbor Laboratory Press

unchanged upon miR-122 depletion (Fig. 6B), the protein was up-regulated around twofold to threefold, as judged by Western blot experiments with 122ASO liver extracts (Fig. 6C). These findings strongly suggested that *Pparβ/δ* was a bona fide miR-122 target that thus far had been overlooked, supposedly because it is not regulated on the level of mRNA stability.

Unsaturated fatty acids are probably the most important endogenous PPAR ligands, and their levels are known to be tightly regulated in vivo. The perturbation of lipid metabolism associated with miR-122 depletion may thus also lead to changes in PPAR ligand availability. We therefore determined the concentrations of free fatty acids (FFAs) in livers from 122ASO and control mice by GC/MS. Several unsaturated FFA species were indeed significantly changed, including palmitoleic acid (C16:1; up by 51% in 122ASO mice) and vaccenic acid (C18:1n7; up by 24%) (Fig. 6D). The latter constitutes a significant proportion of the total unsaturated FFA pool and has previously been proposed as a PPARβ/δ ligand (Fyffe et al. 2006). We therefore deemed it likely that PPAR activity in 122ASO mice was additionally modulated by changes in ligand concentration.

We conclude that miR-122 has several ties to the PPAR family of nuclear receptors, via *Pparβ/δ*, *Smarcd1/Baf60a*, and possibly ligand availability. Given the important functions PPARs possess in regulating metabolism in liver, these connections are very likely to contribute to the overall metabolic phenotype observed in 122ASO mice.

Discussion

Circadian mir-122 transcription and function

In the present study, we show that miRNA miR-122 expression and function are embedded in the output system of the circadian clock. Thus, we found that the miR-122 locus was transcribed in a circadian manner, manifesting itself in rhythmic pri-mir-122 and pre-mir-122 expression. Based on genetic loss-of-function and gain-of-function experiments we concluded that the orphan receptor REV-ERBα is the dominant regulator of circadian mir-122 transcription. On a genome-wide scale, we observed that the portion of the transcriptome sensitive to miR-122 depletion was highly enriched for circadian mRNAs, and it appeared that these were biased toward specific circadian phases (Supplemental Fig. 10). This temporal gating was particularly evident for mRNAs encoding cholesterol and lipid metabolizing enzymes, which were identified previously as indirectly regulated miR-122 targets (Krutzfeldt et al. 2005; Esau et al. 2006; Elmen et al. 2008a,b). Further analyses of individual up-regulated transcripts around the clock enabled us to identify several circadian transcripts that were likely candidates for direct miR-122 targets. The rhythmic accumulation of these mRNAs showed changes in amplitude (*Smarcd1/Baf60a, Ddc,* and *Hist1h1c*), magnitude (*Rell1*), and phase (*Smarcd1/Baf60a* and *Hist1h1c*) upon miR-122 depletion. In *Rev-erbα* knockout animals, miR-122 synthesis was nearly constant over the day and steady-state miR-122 levels were 1.7-fold elevated. REV-ERBα regulates many clock-controlled genes directly by repressing their transcription in a cyclic manner (see also Supplemental Fig. 6; G Le Martelot, T Claudel, O Schaad, B Kornmann, G Lo Sasso, A Moschetta, and U Schibler, in prep.). Irrefutable evidence that changes in miR-122 levels and/or production account for the circadian misregulation of target transcripts in *Rev-erbα* knockout mice is therefore difficult to obtain. However, as indicated by our analysis of pre-mRNA and mRNA expression, miR-122 misregulation is likely to be responsible for the altered circadian amplitude of *Rell1* mRNA accumulation in *Rev-erbα* knockout mice. Interestingly, *Rev-erbα* knockout mice show a cholesterol- and lipid-related phenotype opposite to 122ASO mice (G Le Martelot, T Claudel, O Schaad, B Kornmann, G Lo Sasso, A Moschetta, and U Schibler, in prep.). Again, REV-ERBα probably regulates these pathways mainly by more direct, transcriptional mechanisms, but miR-122 up-regulation is likely to contribute to these phenotypes as well.

The regulation of lipid metabolism by miR-122 may involve PPAR receptors

The direct miR-122 targets involved in hepatic lipid metabolism have not yet been identified. The decrease in hepatic fatty acid and cholesterol synthesis and the increase in hepatic fatty acid oxidation are paralleled by an increased activation of AMP-activated protein kinase (AMPK) in 122ASO mice (Esau et al. 2006). Thus, miR-122 may act through the modulation of this central sensor of metabolism. Our experiments also uncovered several connections of miR-122 to the nuclear receptors of the PPAR family, which are well-known regulators of metabolism. Specifically, we found that upon miR-122 inactivation, PPARβ/δ protein was up-regulated by around twofold to threefold. The *Pparβ/δ* 3′UTR contains seed sites for miR-122, and among the >30 3′UTRs we tested, it conferred one of the strongest levels of miR-122-mediated repression. In liver, PPARα and PPARβ/δ, the two PPARs executing catabolic functions, are both expressed in a circadian manner with a phase difference of ~8 h (Yang et al. 2006). Since PPARα is the predominant isoform in this organ, hepatic functions of PPARβ/δ have not yet been studied in detail. Although PPAR functions can vary in different tissues, it is interesting to note that recently an interaction between the PPARβ/δ and AMPK pathways was shown in muscle. Thus, a constantly active VP16-PPARβ/δ transgene led to constitutive AMPK stimulation (Narkar et al. 2008). Therefore, it is tempting to speculate that at least in part the AMPK activation (Esau et al. 2006) could be the result of higher PPARβ/δ protein levels in the livers of miR-122-depleted mice.

The newly identified miR-122 target *Smarcd1/Baf60a* provides a second link to PPARs. *Smarcd1/Baf60a* is a core subunit of the SWI/SNF chromatin remodeling complexes and was very recently identified in a screen for transcription factors whose activity is augmented by

Downloaded from genesdev.cshlp.org on September 4, 2009 - Published by Cold Spring Harbor Laboratory Press

PPARγ coactivator-1α (PGC-1α) (Li et al. 2008). In this study, SMARCD1/BAF60a overexpression in hepatocytes was shown to have surprisingly specific effects on the transcriptional activation of genes involved in fatty acid oxidation, and many of these were also activated by a synthetic PPARα agonist. In addition, SMARCD1/BAF60a was found to physically interact with PPARα and to be required for its function. Both proteins are corecruited with PGC-1α to PPAR response element (PPRE)-containing promoters.

A third connection of miR-122 to PPARs was provided by the observation that the livers of miR-122-depleted animals contained higher levels of FFAs, known to serve as PPAR ligands. All in all, our results suggest that PPARs might act as mediators to link miR-122 function to the control of circadian gene expression and hepatic lipid metabolism, although the detailed genetic and biochemical dissection of this network will require many additional experiments.

Speculations about the circadian action of miR-122

Generally, RISC-bound miRNAs are thought to be long-lived (Lee et al. 2003; Lund et al. 2004), and we estimated the miR-122 half-life to exceed 24 h. In the light of the large, stable steady-state pool of miR-122 the question thus arises of how circadian miR-122 production could nevertheless have an impact on its targets. Below we present three possible mechanisms through which miR-122 could modulate circadian gene expression on the post-transcriptional level. The first involves the attenuation of basal mRNA accumulation and/or translation by invariant miR-122 activity, the second implies chemically different subpopulations of miR-122 with distinct purposes, and the third a different availability of newly assembled and old RISC complexes for being loaded on target mRNAs.

Strictly speaking, there is no requirement for a miRNA to be circadian itself in order to contribute to the circadian accumulation of a target transcript. For example, the constant repression of basal levels of translation from such mRNA could strongly increase the circadian amplitude of the produced protein, as depicted in Figure 7A. Such a mechanism could have the additional benefit of conferring robustness to low protein expression in the trough: Rather than relying on very low transcription rates, which inevitably contain a stochastic component, low expression levels might thus be achieved more

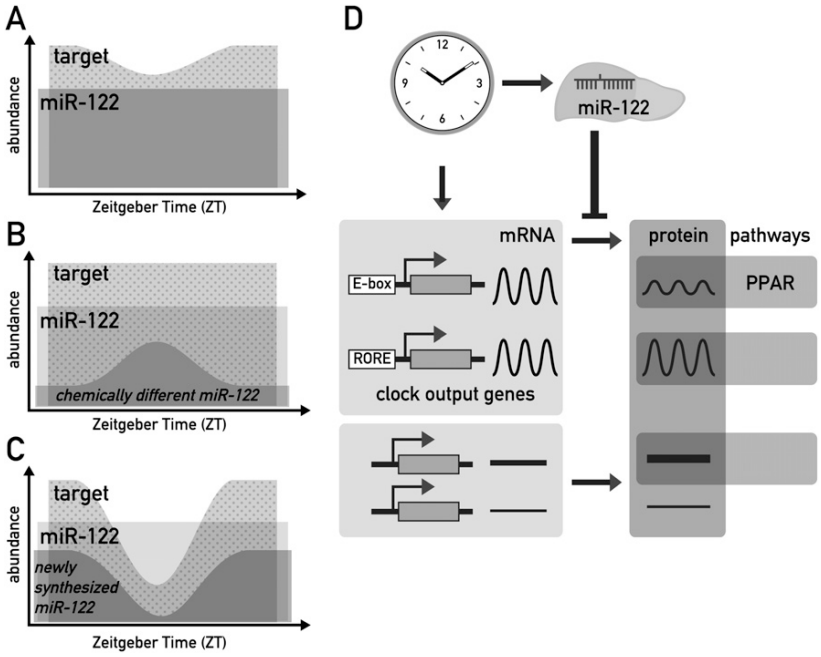


Figure 7. Models for how miR-122 could impart on circadian gene expression of its targets. (*A*) Even constant miR-122 levels (dark gray) could shape the circadian rhythm of a target (light gray) by constantly repressing basal levels of translation (represented by the overlap of the two areas). Only the amount of target mRNA represented by the dotted area would be available for translation into protein. As shown in the cartoon, this mechanism could increase the amplitude of cycling and convert a low-amplitude mRNA rhythm into a higher amplitude protein rhythm. In addition, this regulation could confer robustness to low protein expression levels in the trough, as described in the Discussion. The mechanism depicted in this cartoon would require a high affinity of the miRNA–target interaction and an excess of targets over the miRNA. Considering that miR-122 probably has hundreds of targets, of which many contain several seed sites, this assumption is quite plausible even for this highly abundant miRNA (*B*) Conceivably, chemically distinct, short-lived miR-122 subpopulations (dark gray) could exist within the pool of bulk miR-122. If these distinct miR-122 species also had specific functional properties, this speculative model would imply that target mRNAs would be subject to circadian repression. Consequently, the transcript available to produce protein (dotted area) would show circadian oscillations. (*C*) Conceivably, newly assembled RISC complexes could immediately get committed to their target mRNAs and remain stably associated with them, as described in the Discussion. The availability of such newly assembled miR-122 RISC would be expected to closely follow circadian miR-122 production (dark gray). If targets are transcribed circadianly as well, the phase relationship of the two rhythmic processes will determine to what extent a target will encounter miR-122 RISCs in the cell, and what influence this has on the circadian amplitude, magnitude, and phase of the produced protein (dotted area). As in *A*, this model would demand that the miRISC–mRNA affinity be high and that the targets are in excess. (*D*) Model for the integration of miR-122 in circadian hepatic gene expression. MiR-122 is depicted as a tissue-specific modulator of circadian output genes, with PPAR-dependent regulation of gene expression as one of the regulated output pathways.

Downloaded from genesdev.cshlp.org on September 4, 2009 - Published by Cold Spring Harbor Laboratory Press

precisely by simultaneously producing a mRNA and its inhibiting miRNA, which will partially annul each other. A related role for miRNAs in denoising and conferring robustness to gene expression has previously been suggested in developmental timing (Stark et al. 2005; Cohen et al. 2006; Li et al. 2006, 2009). With regard to liver-specific miR-122, this mechanism could also represent a way of modulating the circadian rhythm of outputs in a tissue-specific manner.

As mentioned above the high metabolic stability of miR-122 prevents its cyclic accumulation. However, our experiments do not exclude that functionally distinct, less stable subpopulations exist within the large pool of miR-122 molecules. This speculative scenario is schematically depicted in Figure 7B (see the figure legend for explanation). Although we currently have no direct evidence for such distinct miR-122-containing RISC subpopulations, they could be produced by miRNA editing, RISC protein composition or subcellular localization. It is interesting to note in this context that miR-122 was recently shown to undergo cytoplasmic 3′ adenylation, affecting miR-122 stability (Katoh et al. 2009). Hence, different miR-122 subpopulations with varying metabolic stabilities may indeed coexist. *Nocturnin*, a rhythmically expressed deadenylase, is also involved in the regulation of lipid metabolism (Green et al. 2007). Although bona fide target mRNAs have not yet been identified for this enzyme, the regulation of poly(A) length is known to contribute to translational repression also in the case of miRNA-mediated mechanisms (Liu 2008; Eulalio et al. 2009). Circadian deadenylation may thus also contribute to the post-transcriptional control of protein synthesis. It should be emphasized in this context that almost half of the cycling liver proteins identified by mass spectrometry are translated from stably expressed mRNAs (Reddy et al. 2006).

Recent work has suggested that the ternary RISC–miRNA–target complex is remarkably stable, allowing for the immunopurification of RISC-bound targets (e.g., Beitzinger et al. 2007; Karginov et al. 2007). One might therefore speculate that mainly uncommitted, "fresh" miRNA-loaded RISCs are available for the silencing of newly synthesized targets, whereas "old" RISCs, which are already engaged in silencing, are less so (Fig. 7C). Since "fresh" miR-122 RISC is produced in a circadian fashion, the extent of target capture and silencing may well be daytime-specific. If targets are transcribed circadianly as well, it becomes evident that the phase relationship of the two rhythmic processes will determine to what extent a target will encounter miR-122 RISCs in the cell. For several of the circadian miR-122 target profiles we determined in 122ASO livers (e.g., *Smarcd1/Baf60a* and *Ddc*; see Fig. 5A), the factor of up-regulation upon miR-122 depletion was indeed especially high around ZT4, just after the peak of miR-122 production. These transcripts were transcribed in phase with miR-122, and miR-122 could function to buffer against and counteract too extreme target oscillations. Future experiments will need to address whether and to what extent the three miR-122-related mechanisms contribute to the post-transcriptional regulation of circadian output rhythms.

Materials and methods

Animal care and treatment

Animal studies were conducted in accordance with the regulations of the veterinary office of the State of Geneva. Mice were maintained under standard animal housing conditions (12-h light/12-h dark cycles; free access to food/water). *Rev-erbα* knockout/transgenic mice have been described (Preitner et al. 2002; Kornmann et al. 2007a). ASO treatment was performed in 11-wk male C57BL/6 mice (Elevage Janvier) by intraperitoneal injection. ASOs were chimeric 2′-fluoro/2′-O-methoxyethyl-modified oligonucleotides with a completely modified phosphorothioate backbone. The exact chemistry is available on request. Mice received four doses of 20 mg of ASO per kilogram of body weight in 150 μL, or 150 μL of saline alone (PBS control), over the course of 2 wk. Two days to 3 d after the last injection, animals were sacrificed at the respective ZTs, and livers were snap-frozen in liquid nitrogen.

RNA analysis

RNA was prepared as in Kornmann et al. (2007a), except that the LiCl wash was omitted to prevent loss of small RNAs. mRNA Northern blots were performed as in Kornmann et al. (2007a). Single-stranded ^{32}P-labeled DNA probes were generated by linear PCR using standard methods. Templates were obtained by PCR amplification from liver cDNA or genomic DNA using gene-specific oligonucleotides (Supplemental Table 1). For miRNA Northern blots 10–30 μg of total RNA per sample were separated by 15% denaturing PAGE/1× TBE, electroblotted (36 min; 3.3 mA/cm^2; 0.5× TBE; 4°C) to Genescreen Plus (NEN) membrane, and immobilized by UV and baking. Hybridizations with radioactively labeled oligonucleotide probes were performed overnight in 5× SSC, 20 mM Na phosphate at pH 7.2, 7% SDS, 2× Denhardt's solution at 50°C, followed by four 15-min washes (3× SSC, 25 mM Na phosphate at pH 7.5, 5% SDS, 10× Denhardt's) and a 5-min wash with 1× SSC and 1% SDS. The sequences of oligonucleotide probes are listed in Supplemental Table 1. Quantification of Northern blots was performed by phosphorimaging using Quantity One Software (Bio-Rad).

Global transcriptome profiling using Affymetrix oligonucleotide microarrays

Whole-cell liver RNAs from ASO-injected mice (ZT0 and ZT12) were analyzed individually on a total of 18 microarrays. Five micrograms of RNA were employed for the synthesis of biotinylated cRNA, of which 8.75 μg were hybridized to Affymetrix Mouse Genome 430 2.0 arrays according to the supplier's instructions. To identify differentially expressed transcripts, pairwise comparisons were carried out using Affymetrix GCOS 1.2 software. Transcripts were considered as expressed if they were detectable in at least two of three replicates in at least one of the experimental conditions. To compare two experimental conditions, each of the triplicates of one condition was compared with the triplicates of the other condition, resulting in nine pairwise comparisons. This approach is based on the Mann-Whitney pairwise comparison test, and allows the ranking of results by concordance and the calculation of significance (*P*-value) for each identified change in gene expression (Hubbell et al. 2002; Liu et al. 2002). Genes whose concordance in the pairwise comparisons exceeded the imposed threshold of 77% (seven of nine comparisons) were considered to be statistically significant. Transcripts were considered as up- or down-regulated in 122ASO samples when their accumulation had an average

Downloaded from genesdev.cshlp.org on September 4, 2009 - Published by Cold Spring Harbor Laboratory Press

change of at least 1.5-fold with regard to both control samples, 124ASO, and PBS. The extraction of circadian genes from Affymetrix data sets (Fig. 4A) has been described previously (Kornmann et al. 2007b). The ArrayExpress repository (http://www.ebi.ac.uk/arrayexpress) accession number for the microarray data is E-TABM-692.

qPCR analysis

cDNA was synthesized from 2 μg of DNase-digested whole-cell RNA using random hexamers and SuperScript II reverse transcriptase (Invitrogen) following the supplier's instructions. cDNAs were PCR-amplified (7900HT Sequence Detection Systems, Applied Biosystems) using TaqMan Universal Master Mix, No AmpErase UNG (Applied Biosystems), and raw threshold cycle (Ct) values were calculated with SDS 2.0 software (Applied Biosystems). Mean levels were calculated from triplicate PCR assays for each sample and normalized to those obtained for the control transcripts *Eef1a1*, *Gapdh*, *GusB*, and 45S pre-rRNA. RT^- samples were included to exclude contaminations with genomic DNA. For primers and probes, see Supplemental Table 1.

miR-122 target predictions and enrichment statistics

We relied on the Ensembl version 50 mapping of the Affymetrix probes to transcripts. Up-regulated, down-regulated, and unchanged transcripts were selected as described above. The seed sequence of the miRNA was defined as 6–8 bases from the second position of the miRNA 5′-end, not allowing mismatches except a single G:U in 7-mers and 2 G:U in 8-mers. Duplex energies were computed with the cofolding function from the RNA Vienna Package (Hofacker 2003). The statistical significance of the putative miR-122 target site enrichment in the up, equal, and down fractions was evaluated using a χ^2 test.

Plasmids, clonings, and analysis of 3′UTRs

3′UTR sequences were amplified by PCR from mouse liver cDNA or genomic DNA with specific oligonucleotides (Supplemental Table 1) and cloned 3′ to the renilla luciferase (RL) sequence in vector pRL-control. The identity of the UTRs was verified by sequencing. Plasmids pRL-control and pRL-Cat-1 are as in Bhattacharyya et al. (2006) and pRL-3xbulge is similar to the homonymous plasmid in Pillai et al. (2005), except that bulges match miR-122 instead of let-7. For normalization, a CMV-driven firefly luciferase-expressing plasmid on the basis of pEGFP-C1 was used. Details on all plasmids are available on request. For 3′UTR assays, 2 ng of pRL, 40 ng of FL plasmid, and 10 pmol of miRNA mimic (miR-122 and control mimic cel-miR-67 from Dharmacon) were transfected into 10^4 HeLa cells per well of a 96-well plate by reverse transfection using Lipofectamine 2000 (Invitrogen) according to the supplier's instructions. Transfection mixes were replaced by normal growth medium after 6 h. Luciferase activities were measured 28 h after transfection with the Dual-Glo Luciferase Assay System (Promega). Renilla luciferase signals were normalized to firefly luciferase and for each 3′UTR construct set to 100% for the cotransfection with the control mimic. Each transfection was repeated at least six times. Growth medium was DMEM, 10% FCS, 1% PSG (Gibco).

Immunoprecipitation-Western blotting

Liver pieces of 122ASO and control mice, ZT0 and ZT12 (triplicates) were homogenized in three volumes of RIPA (150 mM NaCl, 1% NP40, 0.5% Na-deoxycholate, 0.1% SDS, 50 mM Tris-HCl at pH 8.0, protease inhibitors) using a motorized hand tool (Xenox). Insoluble material was removed by centrifugation (15 min, 20,000*g*, 0°C). Supernatants were kept at −80°C. For the immunoprecipitation, extracts were further diluted to 5 vol of RIPA per volume of liver and adjusted to 0.2% SDS. After another spin (as above), equal amounts of protein extract from the triplicates of the same experimental condition were pooled (~600 μg protein/liver). An aliquot was kept for the input sample, and immunoprecipitation was performed from the remaining pool using standard protocols with a rabbit polyclonal antibody to PPARβ/δ (ab8937, Abcam) and protein A-agarose (Roche). Immunoprecipitated complexes and inputs were analyzed by SDS-PAGE/Western blotting using antibodies to PPARβ/δ and U2AF65 (Sigma, U4758). Semiquantitative analysis of Western blots was performed using Quantity One Software (Bio-Rad).

FFA analysis

Liver homogenates in MeOH (0.1% BHT) were spiked with heptadecatrienoate, TAG (17:0/17:0/17:0) and heptadecanoic acid (FFA C17:0) and extracted by chloroform after addition of 0.9% sodium chloride. The lower organic phase was separated, evaporated under nitrogen flow, and dissolved into petroleum ether (bp 40°C–60°C). The samples were transesterified with sodium methoxide (NaOMe, 0.5 M in MeOH), acidified (15% $NaHSO_4$ in H_2O) and extracted with petroleum ether. The organic phase containing fatty acid methyl esters (FAME) from bound fatty acids and FFAs was separated, evaporated under nitrogen flow, and redissolved into hexane. Two-microliter aliquots were used for GC injection (splitless 1 min) at 280°C and the analyses were performed on an FFAP fused silica capillary column (25 m, i.d. 0.32 mm) by using helium as the carrier gas (pressure program). The oven temperature was increased from 70°C to 240°C at 7°C per minute, and the fatty acids were detected by flame ionization detector (FID, 300°C). Identification was based on retention times and GC/MS spectra of reference substances.

Acknowledgments

We thank Suvendra Bhattacharyya and Witek Filipowicz for plasmids, staff at the NCCR Genomics Platform for help with microarray/qPCR experiments, Tuulikki Seppänen-Laakso for help with FFA measurements, Nicolas Leuenberger and Walter Wahli for discussions and communication of unpublished data, members of the Schibler laboratory for comments on this manuscript, and Nicolas Roggli for help with artwork. This research was supported by the Swiss National Science Foundation (through an individual research grant to U.S., and the National Center of Competence in Research Program Frontiers in Genetics, and grant SNSF PDFM33-118375 to E.Z.), the State of Geneva, the Louis Jeantet Foundation of Medicine, the Bonizzi-Theler-Stiftung, and the 6th European Framework Project EUCLOCK. D. Gatfield received and gratefully acknowledges long-term fellowships from The Federation of European Biochemical Societies (FEBS) and The International Human Frontier Science Program Organization (HFSP).

References

Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. 2008. The impact of microRNAs on protein output. *Nature* **455:** 64–71.

Beitzinger M, Peters L, Zhu JY, Kremmer E, Meister G. 2007. Identification of human microRNA targets from isolated argonaute protein complexes. *RNA Biol* **4:** 76–84.

Downloaded from genesdev.cshlp.org on September 4, 2009 - Published by Cold Spring Harbor Laboratory Press

Bhattacharyya SN, Habermacher R, Martine U, Closs EI, Filipowicz W. 2006. Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* **125:** 1111–1124.

Bushati N, Cohen SM. 2007. microRNA functions. *Annu Rev Cell Dev Biol* **23:** 175–205.

Chang J, Nicolas E, Marks D, Sander C, Lerro A, Buendia MA, Xu C, Mason WS, Moloshok T, Bort R, et al. 2004. miR-122, a mammalian liver-specific microRNA, is processed from hcr mRNA and may downregulate the high affinity cationic amino acid transporter CAT-1. *RNA Biol* **1:** 106–113.

Cheng HY, Papp JW, Varlamova O, Dziema H, Russell B, Curfman JP, Nakazawa T, Shimizu K, Okamura H, Impey S, et al. 2007. microRNA modulation of circadian-clock period and entrainment. *Neuron* **54:** 813–829.

Cohen SM, Brennecke J, Stark A. 2006. Denoising feedback loops by thresholding—A new role for microRNAs. *Genes & Dev* **20:** 2769–2772.

Cusick JK, Xu LG, Bin LH, Han KJ, Shu HB. 2006. Identification of RELT homologues that associate with RELT and are phosphorylated by OSR1. *Biochem Biophys Res Commun* **340:** 535–543.

Elmen J, Lindow M, Schutz S, Lawrence M, Petri A, Obad S, Lindholm M, Hedtjarn M, Hansen HF, Berger U, et al. 2008a. LNA-mediated microRNA silencing in non-human primates. *Nature* **452:** 896–899.

Elmen J, Lindow M, Silahtaroglu A, Bak M, Christensen M, Lind-Thomsen A, Hedtjarn M, Hansen JB, Hansen HF, Straarup EM, et al. 2008b. Antagonism of microRNA-122 in mice by systemically administered LNA-antimiR leads to up-regulation of a large set of predicted target mRNAs in the liver. *Nucleic Acids Res* **36:** 1153–1162.

Esau C, Davis S, Murray SF, Yu XX, Pandey SK, Pear M, Watts L, Booten SL, Graham M, McKay R, et al. 2006. miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. *Cell Metab* **3:** 87–98.

Eulalio A, Huntzinger E, Nishihara T, Rehwinkel J, Fauser M, Izaurralde E. 2009. Deadenylation is a widespread effect of miRNA regulation. *RNA* **15:** 21–32.

Fyffe SA, Alphey MS, Buetow L, Smith TK, Ferguson MA, Sorensen MD, Bjorkling F, Hunter WN. 2006. Recombinant human PPAR-β/δ ligand-binding domain is locked in an activated conformation by endogenous fatty acids. *J Mol Biol* **356:** 1005–1013.

Gachon F, Nagoshi E, Brown SA, Ripperger J, Schibler U. 2004. The mammalian circadian timing system: From gene expression to physiology. *Chromosoma* **113:** 103–112.

Gallego M, Virshup DM. 2007. Post-translational modifications regulate the ticking of the circadian clock. *Nat Rev Mol Cell Biol* **8:** 139–148.

Gerlach D, Kriventseva EV, Rahman N, Vejnar CE, Zdobnov EM. 2009. miROrtho: Computational survey of microRNA genes. *Nucleic Acids Res* **37:** D111–D117. doi: 10.1093/nar/gkn707.

Green CB, Douris N, Kojima S, Strayer CA, Fogerty J, Lourim D, Keller SR, Besharse JC. 2007. Loss of Nocturnin, a circadian deadenylase, confers resistance to hepatic steatosis and diet-induced obesity. *Proc Natl Acad Sci* **104:** 9888–9893.

Hofacker IL. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res* **31:** 3429–3431.

Hubbell E, Liu WM, Mei R. 2002. Robust estimators for expression analysis. *Bioinformatics* **18:** 1585–1592.

Karginov FV, Conaco C, Xuan Z, Schmidt BH, Parker JS, Mandel G, Hannon GJ. 2007. A biochemical approach to identifying microRNA targets. *Proc Natl Acad Sci* **104:** 19291–19296.

Katoh T, Sakaguchi Y, Miyauchi K, Suzuki T, Kashiwabara S, Baba T, Suzuki T. 2009. Selective stabilization of mammalian microRNAs by 3′ adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. *Genes & Dev* **23:** 433–438.

Kornmann B, Schaad O, Bujard H, Takahashi JS, Schibler U. 2007a. System-driven and oscillator-dependent circadian transcription in mice with a conditionally active liver clock. *PLoS Biol* **5:** e34. doi: 10.1371/journal.pbio.0050034.

Kornmann B, Schaad O, Reinke H, Saini C, Schibler U. 2007b. Regulation of circadian gene expression in liver by systemic signals and hepatocyte oscillators. *Cold Spring Harb Symp Quant Biol* **72:** 319–330.

Krutzfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, Manoharan M, Stoffel M. 2005. Silencing of microRNAs in vivo with 'antagomirs.' *Nature* **438:** 685–689.

Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. 2002. Identification of tissue-specific microRNAs from mouse. *Curr Biol* **12:** 735–739.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, et al. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425:** 415–419.

Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120:** 15–20.

Li Y, Wang F, Lee JA, Gao FB. 2006. MicroRNA-9a ensures the precise specification of sensory organ precursors in *Drosophila*. *Genes & Dev* **20:** 2793–2805.

Li S, Liu C, Li N, Hao T, Han T, Hill DE, Vidal M, Lin JD. 2008. Genome-wide coactivation analysis of PGC-1α identifies BAF60a as a regulator of hepatic lipid metabolism. *Cell Metab* **8:** 105–117.

Li X, Cassidy JJ, Reinke CA, Fischboeck S, Carthew RW. 2009. A microRNA imparts robustness against environmental fluctuation during development. *Cell* **137:** 273–282.

Liu J. 2008. Control of protein synthesis and mRNA degradation by microRNAs. *Curr Opin Cell Biol* **20:** 214–221.

Liu WM, Mei R, Di X, Ryder TB, Hubbell E, Dee S, Webster TA, Harrington CA, Ho MH, Baid J, et al. 2002. Analysis of high density expression microarrays with signed-rank call algorithms. *Bioinformatics* **18:** 1593–1599.

Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U. 2004. Nuclear export of microRNA precursors. *Science* **303:** 95–98.

Miller BH, McDearmon EL, Panda S, Hayes KR, Zhang J, Andrews JL, Antoch MP, Walker JR, Esser KA, Hogenesch JB, et al. 2007. Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation. *Proc Natl Acad Sci* **104:** 3342–3347.

Narkar VA, Downes M, Yu RT, Embler E, Wang YX, Banayo E, Mihaylova MM, Nelson MC, Zou Y, Juguilon H, et al. 2008. AMPK and PPARδ agonists are exercise mimetics. *Cell* **134:** 405–415.

Panda S, Antoch MP, Miller BH, Su AI, Schook AB, Straume M, Schultz PG, Kay SA, Takahashi JS, Hogenesch JB. 2002. Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* **109:** 307–320.

Pillai RS, Bhattacharyya SN, Artus CG, Zoller T, Cougot N, Basyuk E, Bertrand E, Filipowicz W. 2005. Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* **309:** 1573–1576.

Preitner N, Damiola F, Lopez-Molina L, Zakany J, Duboule D, Albrecht U, Schibler U. 2002. The orphan nuclear receptor REV-ERBα controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell* **110:** 251–260.

Reddy AB, Karp NA, Maywood ES, Sage EA, Deery M, O'Neill JS, Wong GK, Chesham J, Odell M, Lilley KS, et al. 2006.

Downloaded from genesdev.cshlp.org on September 4, 2009 - Published by Cold Spring Harbor Laboratory Press

Circadian orchestration of the hepatic proteome. *Curr Biol* **16:** 1107–1115.

Reppert SM, Weaver DR. 2002. Coordination of circadian timing in mammals. *Nature* **418:** 935–941.

Sato TK, Panda S, Miraglia LJ, Reyes TM, Rudic RD, McNamara P, Naik KA, FitzGerald GA, Kay SA, Hogenesch JB. 2004. A functional genomics strategy reveals Rora as a component of the mammalian circadian clock. *Neuron* **43:** 527–537.

Sato TK, Yamada RG, Ukai H, Baggs JE, Miraglia LJ, Kobayashi TJ, Welsh DK, Kay SA, Ueda HR, Hogenesch JB. 2006. Feedback repression is required for mammalian circadian clock function. *Nat Genet* **38:** 312–319.

Seedorf U, Aberle J. 2007. Emerging roles of PPARδ in metabolism. *Biochim Biophys Acta* **1771:** 1125–1131.

Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455:** 58–63.

Sewer A, Paul N, Landgraf P, Aravin A, Pfeffer S, Brownstein MJ, Tuschl T, van Nimwegen E, Zavolan M. 2005. Identification of clustered microRNAs using an ab initio prediction method. *BMC Bioinformatics* **6:** 267. doi: 10.1186/1471-2105-6-267.

Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM. 2005. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution. *Cell* **123:** 1133–1146.

Storch KF, Lipan O, Leykin I, Viswanathan N, Davis FC, Wong WH, Weitz CJ. 2002. Extensive and divergent circadian gene expression in liver and heart. *Nature* **417:** 78–83.

Ueda HR, Chen W, Adachi A, Wakamatsu H, Hayashi S, Takasugi T, Nagano M, Nakahama K, Suzuki Y, Sugano S, et al. 2002. A transcription factor response element for gene expression during circadian night. *Nature* **418:** 534–539.

Xu S, Witmer PD, Lumayag S, Kovacs B, Valle D. 2007. MicroRNA (miRNA) transcriptome of mouse retina and identification of a sensory organ-specific miRNA cluster. *J Biol Chem* **282:** 25053–25066.

Yang X, Downes M, Yu RT, Bookout AL, He W, Straume M, Mangelsdorf DJ, Evans RM. 2006. Nuclear receptor expression links the circadian clock to metabolism. *Cell* **126:** 801–810.

Yang M, Lee JE, Padgett RW, Edery I. 2008. Circadian regulation of a limited set of conserved microRNAs in *Drosophila*. *BMC Genomics* **9:** 83. doi: 10.1186/1471-2164-9-83.

4.3 Silencing of c-Fos expression by microRNA-155 is critical for dendritic cell maturation and function

The regulation mediated by miRNAs is involved in immune system functions (Xiao and Rajewsky [72]). In particular, miR-155 expression, derived from an exon of the B-cell integration cluster (BIC), is induced during the activation of immune cells, notably the dendritic cells (DCs). To investigate the role of miR-155 in DCs maturation, miR-155 homozygous KO mouse were used (Rodriguez et al. [73]). In the miR-155 KO mouse, bone marrow–derived DCs (BM-DCs) functions were significantly impaired. In particular, the repression of c-Fos expression, mediated by miR-155, is essential for DC maturation. Moreover, a de-regulated expression of c-Fos led to the same phenotype as a miR-155 KO, confirming the essential role of miR-155 in DC maturation and function.

Involvement of miR-155 in the maturation of BM-DCs was investigated by mRNA profiling of immature and mature BM-DCs for wild-type and miR-155 KO. I analyzed the Illumina arrays and examined the target enrichment similarly to my previous analysis (Gatfield et al. [74]). I found the effect of miR-155 to be significant in mature BM-DCs (Table 2 in Dunand-Sauthier et al. [75]). I also calculated the enrichment for all mouse miRNAs (Figure 2 in Dunand-Sauthier et al. [75]), and confirmed this effect was only attributable to miR-155.

IMMUNOBIOLOGY

Silencing of c-Fos expression by microRNA-155 is critical for dendritic cell maturation and function

Isabelle Dunand-Sauthier,[1] Marie-Laure Santiago-Raber,[1] Leonardo Capponi,[1] Charles E. Vejnar,[2] Olivier Schaad,[3] Magali Irla,[1] Queralt Seguín-Estévez,[1] Patrick Descombes,[3] Evgeny M. Zdobnov,[2] Hans Acha-Orbea,[4] and Walter Reith[1]

[1]Department of Pathology and Immunology, Faculty of Medicine, [2]Swiss Institute of Bioinformatics, Department of Medical Genetics and Development, Faculty of Medicine, and [3]Genomics Platform, National Centre of Competence in Research Frontiers in Genetics, University of Geneva, Geneva, Switzerland; and [4]Department of Biochemistry, Faculty of Biology and Medicine, University of Lausanne, Epalinges, Switzerland

MicroRNAs (miRNAs) are small, noncoding RNAs that regulate target mRNAs by binding to their 3′ untranslated regions. There is growing evidence that microRNA-155 (miR155) modulates gene expression in various cell types of the immune system and is a prominent player in the regulation of innate and adaptive immune responses. To define the role of miR155 in dendritic cells (DCs) we performed a detailed analysis of its expression and function in human and mouse DCs. A strong increase in miR155 expression was found to be a general and evolutionarily conserved feature associated with the activation of DCs by diverse maturation stimuli in all DC subtypes tested. Analysis of miR155-deficient DCs demonstrated that miR155 induction is required for efficient DC maturation and is critical for the ability of DCs to promote antigen-specific T-cell activation. Expression-profiling studies performed with miR155$^{-/-}$ DCs and DCs overexpressing miR155, combined with functional assays, revealed that the mRNA encoding the transcription factor c-Fos is a direct target of miR155. Finally, all of the phenotypic and functional defects exhibited by miR155$^{-/-}$ DCs could be reproduced by deregulated c-Fos expression. These results indicate that silencing of c-Fos expression by miR155 is a conserved process that is required for DC maturation and function. (*Blood*. 2011;117(17):4490-4500)

Introduction

MicroRNAs (miRNAs) are small, single-stranded, noncoding RNAs that regulate mRNAs by binding to their 3′ untranslated (3′UTR) regions.[1,2] More than 9000 miRNAs have been identified in more than 100 species. Most miRNA genes are transcribed by RNA polymerase II into primary miRNA transcripts that are processed in the nucleus by a complex containing the RNase III endonuclease Drosha.[1] The resulting precursor miRNAs are transported to the cytoplasm, where the mature miRNAs are excised by a complex containing the endonuclease Dicer.[1] Mature miRNAs are incorporated into the RNA-induced silencing complex, which binds to the 3′UTRs of target mRNAs, inducing their degradation and/or repressing their translation. Posttranscriptional regulation of gene expression by miRNAs is critical for a wide range of physiologic and pathologic processes, including cell proliferation, apoptosis, differentiation, morphogenesis, development, and oncogenesis.[1-4]

Several miRNAs play pivotal roles in the immune system.[5-7] MicroRNA-155 (miR155) has emerged as a particularly prominent player in innate and adaptive immune responses.[5,7] miR155 is derived from an exon of the B-cell integration cluster (*BIC*) gene, which was identified as a common integration site of avian leucosis virus in chicken B-cell lymphomas.[8,9] *BIC* is a non-protein-coding gene for which the only known function is the production of miR155. Subsequent studies revealed that miR155 expression is deregulated in diverse cancers.[10,11] The molecular mechanisms underlying the oncogenic role of miR155 remain unclear.

miR155 expression is induced during the activation of T cells, B cells, monocytes, macrophages, and dendritic cells (DCs), suggesting that it plays multiple roles in the immune system.[5] In agreement with this, the immune system of miR155-deficient mice is compromised by defects in several cell types.[12,13] Activated T cells from miR155$^{-/-}$ mice exhibit a bias toward Th2 differentiation and express elevated levels of IL4, IL5, and IL10. This was attributed to the fact that miR155 targets the mRNA coding for c-Maf, a transcription factor implicated in IL-4 expression and Th2 differentiation.[12] The B-cell compartment in miR155$^{-/-}$ mice exhibits defects in germinal center development and in the generation of efficient antibody responses. miR155 is critical for affinity maturation because the generation of plasma cells produces high-affinity isotype-switched antibodies and the development of memory B cells.[12-14] The B-cell defects in miR155$^{-/-}$ mice result at least in part from miR155 repressing the expression of the transcription factor PU.1[14] and activation-induced cytidine deaminase.[15,16] Lastly, bone marrow–derived DCs (BM-DCs) from miR155$^{-/-}$ mice are impaired in their ability to activate T cells.[12]

We recently reported that the induction of miR155 expression in human monocyte–derived DCs (Mo-DCs) exposed to the TLR4 ligand lipopolysaccharide (LPS) leads to modulation of the IL1 signal transduction pathway.[17] Another study found that miR155 induces down-regulation of DC-specific intercellular adhesion molecule-3 grabbing nonintegrin in human Mo-DCs by inhibiting the expression of PU.1.[18] Neither study elucidated the T cell–activation defect exhibited by miR155-deficient BM-DCs.

Submitted September 17, 2010; accepted February 20, 2011. Prepublished online as *Blood* First Edition paper, March 8, 2011; DOI 10.1182/blood-2010-09-308064.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

© 2011 by The American Society of Hematology

To define the role of miR155 in DCs, we analyzed its expression and function in human and mouse DCs exposed to various stimuli. Activation of miR155 expression was found to be a general evolutionarily conserved process that was correlated with maturation induced by diverse stimuli in all DC subtypes tested. Microarray experiments revealed that silencing of c-Fos expression is a key function of miR155 in DCs. Finally, functional experiments performed with miR155-deficient DCs and DCs in which c-Fos expression was deregulated demonstrated that the repression of c-Fos by miR155 is required for DC maturation and the ability of DCs to promote antigen-specific T-cell activation.

Methods

Mice

2D2, OTII, and miR155$^{-/-}$ mice have been described previously.[12,19,20] Mice were bred under specific pathogen–free conditions and used for experiments at 8-10 weeks of age. Animal experiments were performed with permission of the cantonal and national veterinary authorities.

Cells

The mouse DC^{2114} cell line[21] was cultured in IMDM supplemented with 10% FCS, 0.05mM β-mercaptoethanol, 1000 units/mL of penicillin, and 1000 μg/mL of streptomycin. 293T cells were cultured in DMEM supplemented with 10% FCS, 1000 units/mL of penicillin, and 1000 μg/mL of streptomycin. Cells were cultured under 5% CO_2 in a humidified incubator.

Human Mo-DCs were prepared as described previously.[22] Bone marrow–derived plasmacytoid DCs (BM-pDCs) were derived from tibia and femur bone marrow suspensions from 8- to 10-week-old mice, as described previously.[23] BM-DC differentiation was performed by incubation of 1×10^6 bone marrow cells per milliliter in DMEM medium supplemented with 10% FCS and 5% of a supernatant from a hybridoma-producing GM-CSF. CD11c$^+$ BM-DCs, CD11c$^+$B220$^+$ BM-pDCs, and splenic CD11c$^+$CD8α$^+$ and CD11c$^+$CD8α$^-$ DCs were purified by sorting with a FACSVantage SE (Becton Dickinson). DC maturation was induced with 25 ng/mL of LPS (Alexis), 0.05 mg/mL of poly (I:C) (Amersham Biosciences), 0.2nM CpG oligodeoxynucleotide 1826 (TriLink BioTechnologies), 10 μg/mL of peptidoglycan (PGN; Sigma), 500 ng/mL of Pam3CysSerLys4 (PAM3CSK4; InvivoGen), 200 ng/mL of flagellin (InvivoGen), 100 ng/mL of TNFα, 100 ng/mL of fibroblast-stimulating lipopeptide-1 (FSL-1), 10 μg/mL of muramyl dipeptide (MDP; Calbiochem), 3 μg/mL of imiquimod (InvivoGen), or CpG plus anti-CD40 antibodies (rat FGK45 hybridoma). Splenic CD4$^+$ T cells were purified from 2D2 or OTII mice using a CD4$^+$ T cell–isolation kit (Miltenyi Biotec).

Lentiviral transductions

A fragment of the *BIC* gene encoding miR155 was amplified by PCR from mouse genomic DNA using the primers 5′-GTGCTGCAAACCAGGAAG-3′ and 5′-CCTTACAAAGAGTTGTTCATC-3′. This *BIC* fragment was cloned into the pDONR221 vector using the Gateway BP Clonase Enzyme Mix (Invitrogen). This vector was recombined with pDONRP4-P1R into the 2K7 green fluorescent protein (GFP) lentiviral vector[24] using the Gateway LR Plus Clonase Enzyme Mix (Invitrogen) to generate a vector expressing the *BIC* precursor under control of the EF1α promoter. This vector also expresses GFP to permit the evaluation of transduction efficiencies and the purification of transduced cells. The mutated BIC expression vector was generated by mutating the miR155 sequence in the BIC expression vector. *c-Fos* cDNA was amplified using the primers 5′-ATGACGTTTAAACGCCACCATGATGTTCTCGGGTTTC-3′ and 5′-ATGACGTTTAAACTCACAGGGCCAGCAGCGT-3′. This *c-Fos* cDNA was cloned into the lentiviral pWPI vector. Transduction of mouse DC^{2114} cells was performed as described previously.[25]

Microarray experiments

Microarray experiments and miRNA target site analyses were performed as detailed in supplemental Methods (available on the *Blood* Web site; see the Supplemental Materials link at the top of the online article). Microarray data reported in our study have been deposited in the ArrayExpress database under accession numbers E-MTAB-497 (Figure 2B array) and E-MTAB-498 (Figure 2C array).

Quantitative RT-PCR

RNA was extracted with TRIzol. Human and mouse miR155 cDNAs were generated using specific primers and MultiScribe Reverse Transcriptase, and real-time PCR was performed using hsa-miR155 and Mmu-miR155 TaqMan MicroRNA Assays (Applied Biosystems). Mouse and human mRNAs were quantified by real-time RT-PCR using the iCycler iQ Real-Time PCR Detection System (Bio-Rad) and iQ SYBR Green Supermix (Bio-Rad). Expression levels were normalized using β-actin mRNA, TATA-binding protein mRNA, or 18S rRNA. Results were quantified using a standard curve generated with serial dilutions of input cDNA. Primers are listed in supplemental Table 1.

Luciferase reporter assays

The complete 3′UTRs of human (762 bp) and mouse (800 bp) c-Fos mRNAs were amplified by PCR and inserted downstream of the Renilla luciferase gene in the dual luciferase reporter plasmid psiCHECK-2 (Promega). The QuikChange Multi Site-Directed Mutagenesis Kit (Stratagene) was used to mutate the putative miR155-binding sites. Luciferase reporter assays were performed as described previously.[17]

Flow cytometry

Flow cytometry was performed with a FACSCalibur (Becton Dickinson) and analyzed with WinMDI 2.8 software. Staining was performed in the presence of saturating concentrations of 2.4G2 anti-FcγRII/III monoclonal antibodies. Antibodies are listed in supplemental Table 2.

Western blotting

Protein extracts were fractionated by SDS-PAGE and Western blotting was performed with the antibodies listed in supplemental Table 2.

T-cell stimulation

LPS-treated BM-DCs or CpG-treated DC^{2114} cells were loaded with 20 μg/mL of myelin oligodendrocyte glycoprotein (MOG_{35-55}) peptide, 1 μg/mL of OVA peptide, or 1 mg/mL of OVA protein, and cocultured with 10^5 2D2 or OTII T cells. Control cultures contained equal numbers of unloaded DCs. T-cell activation was assessed after 18 (BM-DC cocultures) or 6 (DC^{2114} cocultures) hours. CD69$^+$ cells were quantified by flow cytometry. Secretion of IL2 was measured by ELISA according to the manufacturer's instructions (eBioscience). T-cell proliferation was assessed by [^{3}H]-thymidine incorporation.

Immunofluorescence microscopy

Cells were seeded on glass coverslips, cultured for 24 hours in the absence or presence of LPS (BM-DCs) or CpG (DC^{2114} cells), and fixed for 10 minutes at room temperature with 1% paraformaldehyde in PBS. BM-DCs were stained using the antibodies indicated in supplemental Table 2. Nuclei were stained with DAPI. DC^{2114} cells were visualized on the basis of endogenous GFP expression. Cells were observed in a Zeiss Axiocam microscope using Axiovision LE software.

Statistical analysis

P values were calculated using the Student *t* test with 2-tailed distribution and 2-sample unequal variance parameters.

Results

Induction of miR155 expression in human Mo-DCs

miRNA expression profiles were compared between immature Mo-DCs and Mo-DCs activated with LPS using a human miRNA microarray and a multispecies miRNA microarray (supplemental Figure 1A-B). Maturation was verified by examining the up-regulation of MHC class II (MHCII), CD83, CD86, and CD40 expression (supplemental Figure 1C). The miRNA that was up-regulated the most strongly and reproducibly was miR155.

Real-time RT-PCR experiments confirmed that miR155 expression was increased strongly in mature Mo-DCs relative to monocytes and immature Mo-DCs (supplemental Figure 2A). Maturation was controlled by assessing the induction of IL6 and IL12-p40 mRNA expression and the reduced expression of the mRNA coding for the MHCII transactivator CIITA.[22] Time-course experiments indicated that miR155 accumulation was induced rapidly and increased progressively, reaching maximal levels after 24-48 hours of stimulation with LPS (supplemental Figure 2B). Quantification of its corresponding precursor miRNA indicated that miR155 accumulation resulted from a rapid and strong transcriptional activation of the *BIC* gene (supplemental Figure 2B). In addition, time-course experiments indicated that miR155 and *BIC* expression were induced by poly (I:C) with kinetics similar to those observed for stimulation with LPS (supplemental Figure 2B).

miR155 and BIC transcripts were next quantified in Mo-DCs exposed to the TLR2 ligand PGN, the TLR3 ligand poly (I:C), the TR5 ligand flagellin, and TNFα (supplemental Figure 3A). Maturation was again controlled by examining IL12-p40, IL6, and CIITA mRNA expression. miR155 and *BIC* expression were induced by all 4 stimuli, although the increase was weaker than that observed for exposure to LPS. *BIC* expression was also induced in Mo-DCs by IFNα (supplemental Figure 3B). These results indicate that ligands that trigger DC maturation induce an increase in miR155 expression.

Maturation-induced miR155 expression is conserved in mouse DCs

To determine whether the induction of miR155 expression during DC activation is a conserved process, we extended our analysis to mouse DCs. miR155 expression was first studied in a DC cell line (DC^{2114}) derived from a transgenic mouse–expressing SV40 T-antigen under control of the CD11c promoter.[21] DC^{2114} cells correspond to $CD8\alpha^+$ DCs and reproduce faithfully most of the key features of their in vivo counterparts: they can capture, process, and present antigens to $CD4^+$ T cells; cross-present antigens to $CD8^+$ T cells; be activated by classic maturation stimuli; and produce a pattern of chemokines and proinflammatory cytokines typical of primary DCs. A potent maturation stimulus for DC^{2114} cells is the TLR9 ligand CpG in combination with anti-CD40 (αCD40) antibodies. Maturation induced by CpG + αCD40 was assessed by examining cell-surface MHCII ($I\text{-}A^b$), CD80, CD86, and CD40 expression (supplemental Figure 4A). We performed miRNA expression-profiling experiments to identify miRNAs that undergo changes in expression in DC^{2114} cells stimulated with CpG + αCD40, and miR155 was found to be up-regulated strongly and reproducibly (supplemental Figure 4B).

Real-time RT-PCR experiments confirmed that miR155 expression increased dramatically in CpG + αCD40–stimulated DC^{2114} cells (supplemental Figure 4C). This induction was correlated with efficient maturation, as assessed by quantifying IL6, IL12-p40, and CIITA mRNA expression (supplemental Figure 4C). Time-course experiments indicated that miR155 accumulation in DC^{2114} cells was induced rapidly and increased progressively, reaching maximal levels after 12-24 hours of stimulation with CpG + αCD40 (supplemental Figure 4D). Induction of miR155 expression resulted from a rapid and strong activation of the *BIC* gene (supplemental Figure 4D). We also studied miR155 and *BIC* expression in DC^{2114} cells subjected to other signals (supplemental Figure 4E). Stimuli that promoted efficient maturation (PGN and poly (I:C) induced a strong increase in miR155 and *BIC* expression. Conversely, neither miR155 nor *BIC* expression was induced by stimuli that failed to promote efficient maturation (PAM3CSK4, LPS, and flagellin).

We next studied the activation of miR155 and *BIC* expression in primary mouse DCs. miR155 expression was induced strongly upon the maturation of conventional BM-DCs, BM-pDCs, and splenic $CD8\alpha^-$ and $CD8\alpha^+$ DCs (supplemental Figure 5A). Furthermore, BIC expression was induced in BM-DCs by all maturation stimuli tested, including LPS, poly (I:C), PGN, PAM3CSK4, flagellin, FSL-1, imiquimod, CpG, MDP, and IFNα (supplemental Figure 5B-C).

Table 1. DC numbers in $miR155^{-/-}$ mice

	WT C57BL/6	$miR155^{-/-}$	*P*
Total splenocytes	$189.10^6 \pm 57.10^6$	$128.10^6 \pm 27.10^6$	.1752
$CD11c^+CD8\alpha^-$ cDCs, %	1.06 ± 0.13	0.86 ± 0.17	.1730
$CD11c^+CD8\alpha^+$ cDCs, %	0.37 ± 0.07	0.34 ± 0.06	.5257
$CD11c^+B220^+$ pDCs, %	1.00 ± 0.13	1.33 ± 0.18	.0610

Means ± SDs were derived from 3 mice of each genotype.

miR155 is required for DC maturation and function

Numbers of conventional DCs ($CD11c^+CD8\alpha^+$, $CD11c^+CD8\alpha^-$) and pDCs ($CD11c^+B220^+$) were unaffected in $miR155^{-/-}$ mice (Table 1). The differentiation of BM-DCs and BM-pDCs was also unchanged in BM cultures from $miR155^{-/-}$ mice. The deficiency in miR155 does therefore not lead to a general defect in DC development.

Unstimulated wild-type (WT) and $miR155^{-/-}$ BM-DCs exhibited similar frequencies of cells displaying a dendritic morphology characterized by long cellular protrusions (Figure 1A). However, the fraction of cells exhibiting this morphology after LPS treatment was significantly lower for $miR155^{-/-}$ BM-DCs (Figure 1A). LPS-induced increases in the cell-surface expression of MHCII ($I\text{-}A^b$) and costimulatory molecules (CD86, CD40, and CD80) were attenuated in BM-DCs from $miR155^{-/-}$ mice (Figure 1B). This was particularly evident for CD86 and CD40 (Figure 1C). BM-DCs from both WT and miR155-deficient mice exhibited strong induction of IL12-p40, IL12-p35, IL1β, IL6, and TNFα mRNA expression after exposure to LPS (supplemental Figure 6). However, the induction of IL12-p40, IL12-p35, and TNFα mRNAs tended to be slightly attenuated in $miR155^{-/-}$ BM-DCs. These findings indicate that $miR155^{-/-}$ BM-DCs exhibit selective defects in key processes associated with DC maturation.

Antigen-specific T cell–activation assays were performed to determine the functional consequences of the impaired maturation of miR155-deficient BM-DCs. OVA-specific $CD4^+$ T cells from TCR-transgenic OTII mice were stimulated with LPS-treated WT or $miR155^{-/-}$ BM-DCs that had been loaded or not with OVA peptide or OVA protein. OTII T-cell activation was assessed by examining CD69 expression and IL2 secretion. OVA-specific OTII

BLOOD, 28 APRIL 2011 • VOLUME 117, NUMBER 17

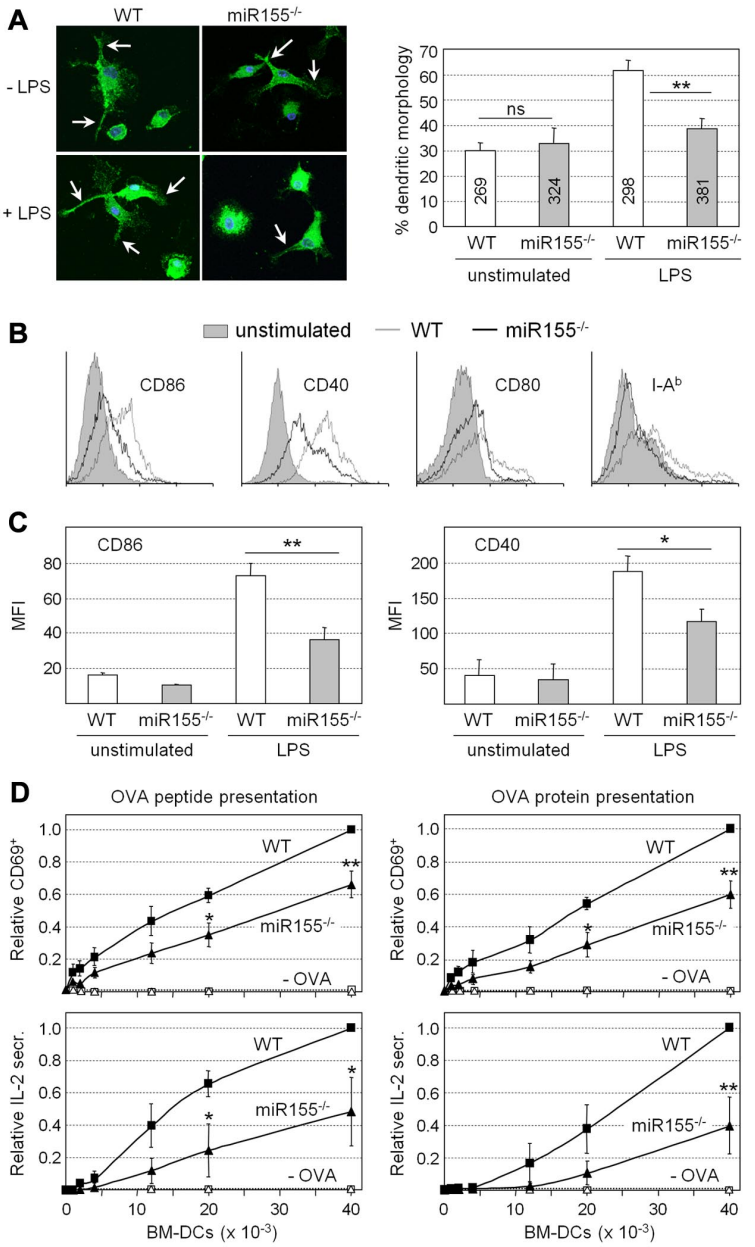


Figure 1. Phenotypic and functional defects exhibited by miR155$^{-/-}$ DCs. (A) Unstimulated and LPS-treated BM-DCs prepared from WT and miR155$^{-/-}$ mice were stained with antibodies against CD11c, and the frequencies of cells exhibiting a characteristic dendritic morphology were determined. Representative images are shown at the left; dendritic protrusions are indicated with arrows. The bar graph represents the means and SDs derived from 3 independent BM-DC preparations; ns, not significant; $^{**}P < .01$. The numbers of cells examined are indicated for each bar. (B) Cell-surface CD86, CD40, CD80, and I-A^b expression was analyzed by flow cytometry for unstimulated and LPS-treated BM-DCs from WT and miR155$^{-/-}$ mice. Histograms are representative of 3 experiments. (C) The mean fluorescence intensity (MFI) for cell-surface CD86 and CD40 expression was determined by flow cytometry for unstimulated and LPS-treated BM-DCs from WT and miR155$^{-/-}$ mice. The means and SDs derived from 3 independent experiments are shown; $^{*}P < .05$; $^{**}P < .01$. (D) LPS-treated BM-DCs from WT and miR155$^{-/-}$ mice were loaded with OVA peptide (left panels) or OVA protein (right panels) and cocultured with OVA-specific CD4$^+$ T cells purified from TCR-transgenic OTII mice. BM-DCs that had not been loaded with antigen (−OVA) were used as negative controls. T-cell activation was determined by the analysis of cell-surface CD69 expression (top panels, relative frequencies of CD69$^+$ cells) or secretion of IL2 into the supernatants (bottom panels, relative IL2 secretion). The means and SDs derived from 3 independent experiments are shown; $^{*}P < .05$; $^{**}P < .01$.

T-cell activation induced by miR155-deficient BM-DCs was significantly impaired (Figure 1D). Similar results were obtained by assessing T-cell activation and proliferation induced by the presentation of MOG$_{35-55}$ to MOG-specific CD4$^+$ T cells purified from TCR-transgenic 2D2 mice. MOG$_{35-55}$-specfic 2D2 T-cell activation and proliferation induced by miR155-deficient BM-DCs were significantly reduced (supplemental Figure 7). BM-DCs from miR155$^{-/-}$ mice therefore exhibit marked defects in their ability to promote antigen-specific T-cell activation.

Identification of mRNAs regulated by miR155 in DCs

Microarray experiments were performed to document differences between the global gene-expression profiles of mature WT and miR155$^{-/-}$ BM-DCs. Direct targets of miR155 were expected to be enriched among mRNAs that are up-regulated in the miR155-deficient BM-DCs relative to WT DCs. mRNAs that were significantly increased in mature miR155$^{-/-}$ BM-DCs were therefore analyzed for the presence of potential miR155-binding sites. Six to 8 nucleotide sequences showing complementarity to the "seed" region situated at the 5′ end (positions 2-9) of miR155 were significantly enriched in the 3′UTRs of mRNAs that were up-regulated in mature miR155$^{-/-}$ BM-cDCs (Table 2). The enrichment of miR155-binding sites was also confirmed using 3 additional prediction models relying on favorable binding energy, sequence context, or evolutionary conservation (Table 2). As expected, no significant enrichment of miR155 targets was observed in mRNAs that were down-regulated in mature miR155$^{-/-}$ BM-DCs or in mRNAs that were increased or decreased in immature miR155$^{-/-}$ BM-DCs (Table 2). We next scanned the mRNAs that were up-regulated in miR155$^{-/-}$ BM-DCs for the presence of target sites for all mouse miRNAs included in the miRBase database (Version 14). miR155 was the only miRNA for

Table 2. Analysis of miR155 signature in miR155$^{-/-}$ BM-DCs

	Immature BM-DCs				Mature BM-DCs*			
	Targets in up-regulated mRNAs†		Targets in down-regulated mRNAs†		Targets in up-regulated mRNAs†		Targets in down-regulated mRNAs†	
	%	*P*	%	*P*	%	*P*‡	%	*P*
5% significance threshold for differential mRNA expression§								
Seeds	22	.9	22	.9	37	**10^{-6}**	19	.98
Conserved seeds	9.7	.6	4.7	1.0	15	**.004**	4	1.00
Targetscan 4	4.5	1.0	6.5	.8	12	**.003**	4	.99
ΔG duplex	1.2	.7	1.0	.7	2.4	.082	.4	.97
1% significance threshold for differential mRNA expression¶								
Seeds	22	.7	28	.4	42	**.001**	27	.33
Conserved seeds	4.2	.9	4.8	.9	17	**.024**	5.1	.9
Targetscan 4	5.3	.7	7.1	.6	13	.08	10	.24
ΔG duplex	.0	1.0	1.6	.7	4.3	.14	0.0	1.0

*Treated with LPS for 24 hours.
†Expression in miR155$^{-/-}$ BM-DCs relative to WT BM-DCs.
‡One-sided Fisher test (bold indicates $P < .05$).
§541 up-regulated mRNAs in miR155$^{-/-}$ DCs (*t* test).
¶139 up-regulated mRNAs in miR155$^{-/-}$ DCs (*t* test).

which target site enrichment was observed using each of the 4 prediction models (Figure 2A). These results indicate that the altered mRNA expression profile of miR155$^{-/-}$ BM-DCs exhibits a clear miR155 signature.

The expression levels of several mRNAs that were known or suspected to be regulated by miR155 or a viral ortholog (miR-K12-1) encoded by Kaposi-sarcoma–associated herpes virus[14,18,26,27] were increased in mature miR155$^{-/-}$ BM-DCs (Figure 2B). Increased expression was statistically significant for some of these mRNAs (Picalm, Pu.1, and Smad5) but not for others (c-Fos and Ship), suggesting that down-regulation of target mRNAs by miR155 is variable in efficiency. This is consistent with the fact that repression by miRNAs can occur mainly at the translational level, with little or no reduction in mRNA abundance.

As a complementary approach to identifying targets of miR155, we studied the impact of overexpressing miR155 in DCs. DC^{2114} cells were transduced with a lentiviral *BIC* expression vector that drives miR155 expression to a level

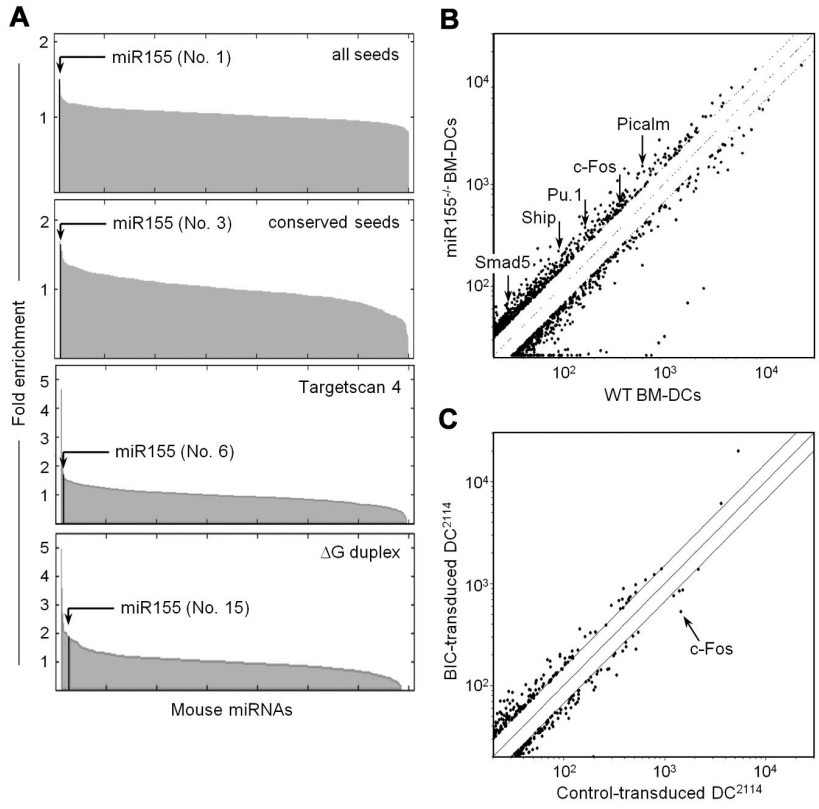


Figure 2. Identification of mRNAs that are regulated by miR155 in DCs. (A) Microarray experiments were performed to compare the gene-expression profiles of mature LPS-treated BM-DCs from WT and miR155$^{-/-}$ mice. The 3′UTRs of mRNAs that were significantly up-regulated in miR155$^{-/-}$ DCs were analyzed for the presence of potential target sites of all mouse miRNAs using 4 prediction models (all seeds, conserved seeds, Targetscan 4, and ΔG duplex). The graphs represent the fold enrichment of target sites for each miRNA. The position of miR155 and its ranking with respect to target-site enrichment are indicated for each graph. (B) Microarray data for mature LPS-treated BM-DCs from WT and miR155$^{-/-}$ mice are represented as a scatter plot showing average normalized signal intensities derived from 3 independent experiments. Each dot represents a probe set corresponding to one mRNA. Only dots corresponding to mRNAs exhibiting greater than a 1.5-fold difference in expression between the 2 genotypes are shown. Dots corresponding to Picalm, c-Fos, Pu.1, Ship, and Smad5 mRNAs are indicated. (C) Microarray experiments were performed to compare the gene-expression profiles of control and BIC-transduced DC^{2114} cells. Results are represented as a scatter plot showing average normalized signal intensities derived from 3 independent experiments. Each dot represents a probe set corresponding to one mRNA. Only dots corresponding to mRNAs exhibiting greater than a 1.5-fold difference in expression between control and BIC-transduced cells are indicated. The dot corresponding to c-Fos mRNA is indicated.

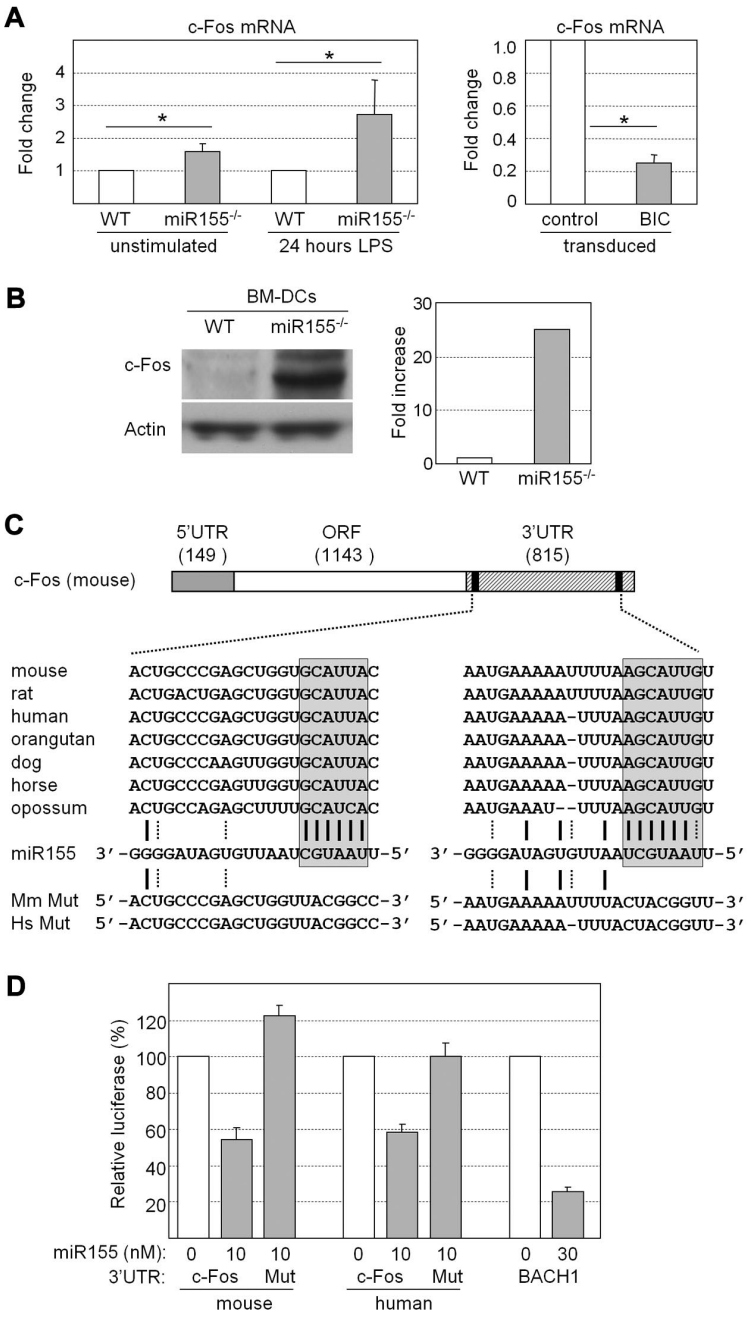


Figure 3. c-Fos expression is regulated by miR155. (A) Expression of c-Fos mRNA was analyzed by real-time RT-PCR in unstimulated and LPS-stimulated BM-DCs from WT or miR155$^{-/-}$ mice (left) and in DC^{2114} cells transduced with empty vector or the BIC expression vector (right). Results are represented as relative c-Fos mRNA expression. The means and SDs derived from 3 independent experiments are shown; *$P < .05$. (B) Expression of c-Fos protein was analyzed by Western blotting in BM-DCs prepared from WT and miR155$^{-/-}$ mice. Actin was used as internal control. A gel representative of 3 independent experiments is shown (left). c-Fos signals were quantified and normalized relative to actin (right). The results represent the mean fold increase derived from 2 independent experiments. (C) Schematic representation of mouse c-Fos mRNA. The sizes in nucleotides of the 5′UTR, open reading frame (ORF), and 3′UTR are indicated. The 3′UTR contains 2 predicted binding sites (black boxes) for miR155. The sequence of mouse miR155 is shown aligned with its predicted target sites in the 3′UTR of c-Fos mRNAs from the indicated species. A-U and G-C base pairs are represented by solid lines; G-U base pairs are represented by dotted lines. The miR155 seed region and its complementary sequences in c-Fos mRNAs are enclosed by boxes. Sequences of the mutated (Mut) 3′UTRs of human (Hs) and mouse (Mm) c-Fos mRNA are indicated. (D) Luciferase reporter constructs containing the WT or mutated 3′UTRs of human or mouse c-Fos mRNA were transfected into 293T cells. A reporter construct containing the 3′UTR of BACH1 mRNA, a known target of miR155, was used as positive control. The constructs were transfected together with the indicated amounts of human or mouse miR155. Luciferase activity was measured 24 hours after transfection, normalized with respect to the activity obtained with a control reporter vector, and is expressed as relative luciferase activity. The means and SDs derived from 3 independent experiments are shown.

comparable to that observed in DC^{2114} cells stimulated with CpG + αCD40 (supplemental Figure 8A). Examinations of cell-surface maturation markers revealed that enforced miR155 expression in DC^{2114} cells did not trigger spontaneous maturation or hinder maturation induced by CpG + αCD40 (supplemental Figure 8B). Microarray experiments were performed to document differences between the global gene-expression patterns exhibited by DC^{2114} cells transduced with the BIC expression vector and a control vector. Only minor changes in gene expression were induced by miR155 overexpression (Figure 2C). Among the target mRNAs that were up-regulated in miR155$^{-/-}$ BM-DCs (Figure 2B), only c-Fos mRNA was reduced in DC^{2114} cells overexpressing miR155 (Figure 2C).

Repression of c-Fos expression by miR155

The microarray data suggested that c-Fos mRNA could be a critical target of miR155 in DCs. Real-time RT-PCR experiments confirmed that c-Fos mRNA abundance was indeed significantly increased in miR155$^{-/-}$ BM-DCs, whereas it was markedly decreased in DC^{2114} cells transduced with the BIC expression vector (Figure 3A). Furthermore, transduction with the BIC expression vector also induced a decrease in c-Fos mRNA levels in WT and miR155$^{-/-}$ BM-DCs, as well as in a control c-Fos–expressing mouse epithelial cell line (MLE12) (supplemental Figure 9A-C).

Western blot experiments demonstrated that the low levels of c-Fos detected in WT BM-DCs was strongly increased in miR155$^{-/-}$

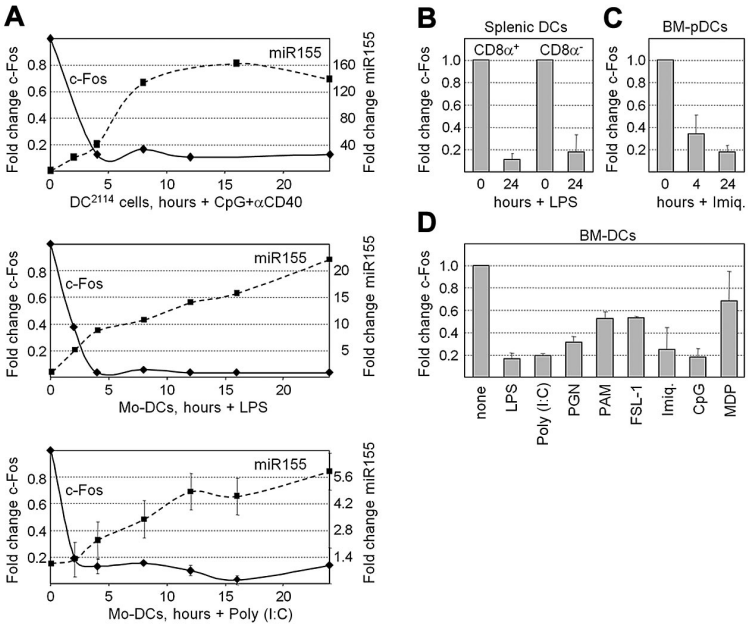


Figure 4. c-Fos expression is silenced during DC maturation. (A) Time-course RT-PCR experiments were performed to quantify the expression of c-Fos mRNA and miR155 in mouse DC^{2114} cells stimulated with CpG + αCD40 and human Mo-DCs stimulated with LPS or poly (I:C). Results are represented as the fold change in the expression of c-Fos mRNA (left axes) and miR155 (right axes). Representative experiments are shown for the top 2 panels. The means and SDs derived from 2 experiments are shown for the bottom panel. (B) c-Fos mRNA was quantified by real-time RT-PCR in unstimulated and LPS-treated $CD8\alpha^+$ and $CD8\alpha^-$ splenic DCs. Results are represented as the fold change in c-Fos mRNA expression. The means and SDs derived from 2 independent experiments are shown. (C) c-Fos expression was analyzed in microarray data derived from BM-pDCs stimulated for 0, 4, and 24 hours with imiquimod. Results are represented as the fold change in signal intensities for c-Fos mRNA. The means and SDs derived from 3 experiments are shown. (D) c-Fos mRNA was quantified by real-time RT-PCR in unstimulated mouse BM-DCs and BM-DCs stimulated with LPS, poly (I:C), PGN, PAM3CSK4, FSL-1, imiquimod, CpG, or MDP. Results are represented as the fold change in c-Fos mRNA expression. The means and SDs derived from 2 independent experiments are shown.

BM-DCs (Figure 3B). The fact that c-Fos protein was increased to a greater extent than c-Fos mRNA (25-fold versus 2- to 3-fold) suggested that miR155 represses c-Fos expression mainly at the level of translation.

Computational approaches for identifying miRNA target sequences based on well-established criteria, including complementarity to the miRNA seed region, favorable sequence context, stability of the miRNA-mRNA duplex, and conservation across multiple species,[28-31] predicted 2 miR155-binding sites in the 3′UTR of c-Fos mRNA in all species examined (Figure 3C). Conservation was strongest in the segments showing complementary to the seed region of miR155 (Figure 3C).

The complete 3′UTRs of mouse and human c-Fos mRNA were inserted into reporter vectors downstream of the Renilla luciferase gene. As controls we used vectors in which the seed regions of the 2 predicted miR155-binding sites within the c-Fos 3′UTRs were mutated. These constructs were transfected into 293T cells with or without the corresponding human or mouse miR155 precursors. Cotransfection of the nonmutated construct with the miR155 precursors resulted in a significant reduction in luciferase activity (Figure 3D). In contrast, no reduction was observed when the mutated constructs were cotransfected with the miR155 precursors. These results confirmed that the predicted miR155-binding sites in the 3′UTR of c-Fos mRNA were indeed targeted directly by miR155.

c-Fos expression is silenced during DC maturation

Time-course RT-PCR experiments were performed to quantify the changes in c-Fos mRNA expression that occur in mouse DC^{2114} cells stimulated with CpG + αCD40 and in human Mo-DCs treated with LPS or poly (I:C). In each system, maturation was accompanied by a rapid decrease in c-Fos mRNA abundance that was correlated with the concomitant increase in miR155 expression (Figure 4A). c-Fos expression was also decreased in $CD8\alpha^+$ and $CD8\alpha^-$ splenic DCs after exposure to LPS (Figure 4B). Analysis of microarray expression-profiling experiments indicated that c-Fos expression is also silenced in BM-pDCs stimulated with imiquimod (Figure 4C). Finally, a reduction in c-Fos mRNA expression was in BM-DCs by all stimuli that activated miR155 expression, including LPS, poly (I:C), PGN, PAM3CSK4, FSL-1, imiquimod, CpG, and MDP (Figure 4D). Silencing of c-Fos expression is thus a conserved and general feature of DC maturation.

Deregulated c-Fos expression recapitulates miR155 deficiency in DCs

To determine whether impaired silencing of c-Fos expression might be responsible for the defects exhibited by $miR155^{-/-}$ DCs (Figure 1), we analyzed the consequences of deregulated c-Fos expression in DCs. DC^{2114} cells were transduced with a lentiviral vector expressing c-Fos under the control of heterologous promoter and 3′UTR regions. High transduction frequencies (supplemental Figure 10A) and efficient c-Fos expression (Figure 5A) were obtained. c-Fos overexpression in the transduced cells (Figure 5A) attained a level similar to that observed in $miR155^{-/-}$ BM-DCs (Figure 3B). In contrast to the endogenous c-Fos mRNA in control vector–transduced DC^{2114} cells, c-Fos mRNA levels were not decreased during maturation in c-Fos–transduced cells (Figure 5B). Unstimulated DC^{2114} cells transduced with c-Fos and control vectors exhibited similar frequencies of cells displaying a dendritic morphology (Figure 5C). However, the increase in the fraction of cells carrying dendritic protrusions after stimulation with CpG was significantly lower for c-Fos–transduced cells (Figure 5C). Enhanced CD86 and CD40 expression after activation with CpG was also significantly reduced in c-Fos–transduced DC^{2114} cells (Figure 5D). Furthermore, c-Fos–transduced DC^{2114} cells exhibited a strongly reduced ability to induce antigen-specific T-cell activation. CD69 expression and IL2 secretion by OTII cells was strongly reduced when c-Fos–transduced DC^{2114} cells were used as stimulators (Figure 5E). This was true irrespective of whether the DC^{2114} cells were loaded with OVA peptide or OVA protein (Figure 5E). Similarly, CD69 expression and IL2 secretion by 2D2 cells was strongly reduced when they were cocultured with c-Fos–transduced DC^{2114} cells loaded with MOG_{35-55} peptide (supplemental Figure 10B). Finally, although c-Fos–transduced DC^{2114} cells retained the ability to up-regulate the expression of proinflammatory cytokine mRNAs during maturation, this induction tended to be reduced 2- to 3-fold relative to untransduced DC^{2114} cells (supplemental Figure

BLOOD, 28 APRIL 2011 • VOLUME 117, NUMBER 17

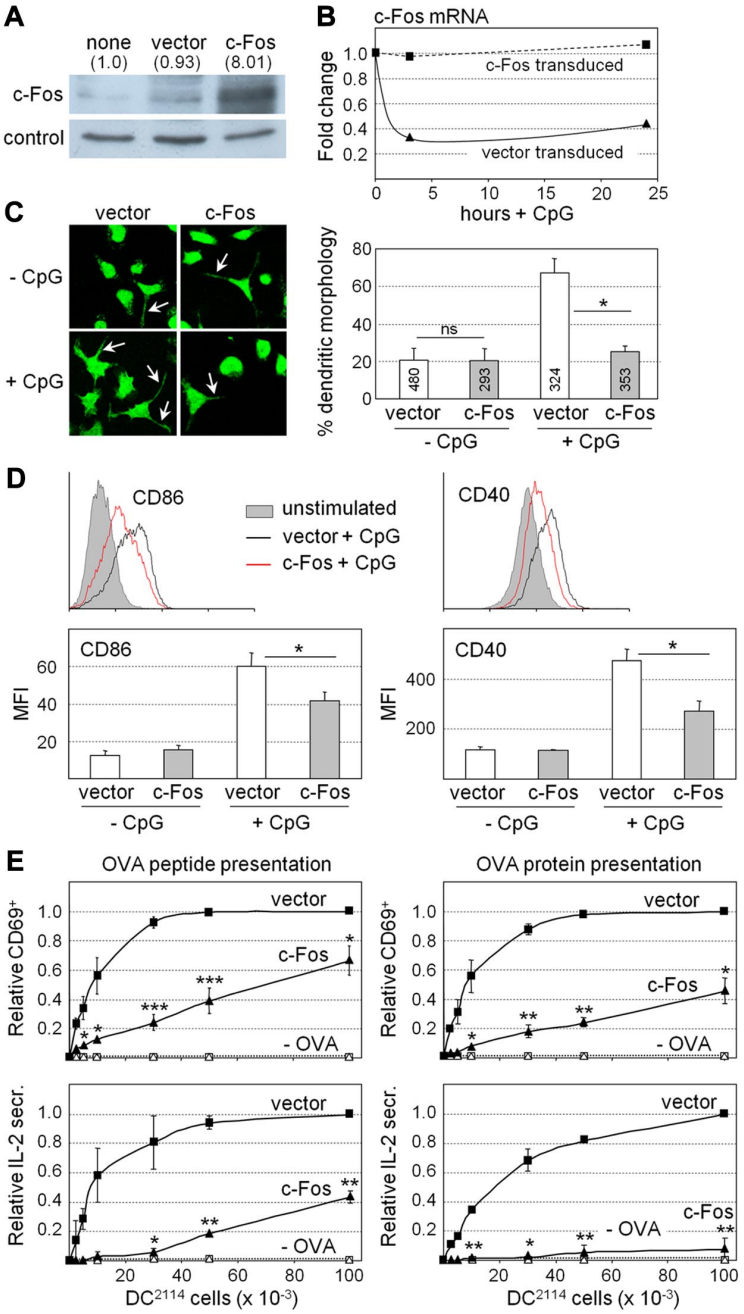


Figure 5. Deregulation of c-Fos expression induces phenotypic and functional defects in DCs. (A) Expression of c-Fos protein was analyzed by Western blotting in untransduced DC^{2114} cells and DC^{2114} cells transduced with an empty expression vector or a c-Fos expression vector. Actin was used as internal control. A representative gel is shown. c-Fos signals were quantified and normalized relative to actin. Changes (-fold) are indicated above the gel. (B) c-Fos mRNA was quantified by real-time RT-PCR in CpG-treated DC^{2114} cells transduced with an empty expression vector or a c-Fos expression vector. A representative graph of 2 independent experiments is shown. (C) Unstimulated and CpG-treated DC^{2114} cells transduced with an empty expression vector or a c-Fos expression vector were examined by immunofluoresence microscopy, and frequencies of cells exhibiting a characteristic dendritic morphology were determined. Representative images are shown at the left; dendritic protrusions are indicated with arrows. The bar graph represents the means and SDs derived from 3 independent transductions; ns, not significant; $^*P < .05$. (D) Cell-surface CD86 and CD40 expression was analyzed by flow cytometry for unstimulated and CpG-stimulated DC^{2114} cells. The histograms (top) are representative of at least 3 experiments. The graphs (bottom) represent the mean fluorescence intensities (MFI) for cell-surface CD86 and CD40 expression, and show the means and SDs derived from at least 3 independent experiments; $^*P < .05$. (E) CpG-treated DC^{2114} cells transduced with an empty vector or a c-Fos expression vector were loaded with OVA peptide (left panels) or OVA protein (right panels) and cocultured with OVA-specific CD4$^+$ T cells purified from TCR-transgenic OTII mice. DC^{2114} cells that had not been loaded with antigen (−OVA) were used as negative controls. T-cell activation was determined by the analysis of cell-surface CD69 expression (top panels, relative frequencies of CD69$^+$ cells) or secretion of IL2 into the supernatants (bottom panels, relative IL2 secretion). The means and SDs derived from 3 independent experiments are shown; $^*P < .05$; $^{**}P < .01$; $^{***}P < .001$.

11). These results indicate that deregulated c-Fos expression in DC^{2114} cells leads to phenotypic and functional defects that are strikingly similar to those observed in miR155$^{-/-}$ BM-DCs.

Discussion

We demonstrate here that miR155 expression is induced by diverse maturation stimuli in human Mo-DCs and in all mouse DC subsets examined. Up-regulated miR155 expression thus appears to be a general feature of DC activation. miR155 was also found to be required for DC maturation. This function was emphasized by the finding that miR155$^{-/-}$ DCs exhibited marked defects in their acquisition of phenotypic and functional properties of mature DCs. These defects included a block in the appearance of a typical dendritic morphology, a reduction in the up-regulation of costimulatory molecules, particularly CD40 and CD86, and a strongly impaired ability to promote antigen-specific CD4$^+$ T-cell activation and proliferation. Not all DC functions were perturbed, because the production of the proinflammatory cytokines IL1β, TNFα, IL12, and IL6 was affected only modestly. These findings extend the previous observation that miR155$^{-/-}$ BM-DCs are less efficient at inducing CD4$^+$ T-cell activation.[12] They are also consistent with the finding that conditional deletion of the gene coding for Dicer in Langerhans cells leads to impaired maturation of this DC subset. As observed for miR155$^{-/-}$ DCs, Dicer-deficient Langerhans cells

exhibited reduced up-regulation of CD40 and CD86 expression and an impaired ability to induce antigen-specific $CD4^+$ T-cell activation.[32]

c-Fos mRNA levels were found to decrease in a manner that was closely correlated with increased miR155 expression during the activation of human Mo-DCs and various mouse DC subtypes by diverse stimuli. Analysis of data derived from published microarray experiments[33-35] confirmed that c-Fos mRNA levels are down-regulated during DC maturation. Furthermore, c-Fos expression was de-repressed in $miR155^{-/-}$ DCs, whereas it was reduced in DCs overexpressing miR155. Finally, the 3′UTR of c-Fos mRNA contains 2 miR155-binding sites and is a direct target of miR155. These results indicate that silencing of c-Fos expression by miR155 is a general mechanism associated with DC maturation.

The importance of miR155-mediated silencing of c-Fos expression in DCs was underscored by the striking similarity between the phenotypic and functional defects displayed by $miR155^{-/-}$ DCs and DCs in which c-Fos expression was removed from its endogenous regulatory controls. Constitutive c-Fos expression in DCs was sufficient to reproduce all of the defects documented for $miR155^{-/-}$ DCs. These findings imply that the abrogation of c-Fos repression by miR155 accounts for many of the functional defects exhibited by $miR155^{-/-}$ DCs.

Earlier studies suggested that c-Fos could inhibit proinflammatory cytokine production by DCs. siRNA-mediated c-Fos knockdown experiments and analyses performed with $c\text{-}Fos^{-/-}$ DCs indicated that c-Fos dampens IL12-p70 secretion by human Mo-DCs and mouse splenic DCs stimulated with the TLR2 ligand PAM3CSK4.[36-38] TNFα, IL12-p70, and IL6 secretion by mature BM-DCs was also attenuated in a c-Fos–dependent manner by stimuli that raise the intracellular concentration of cAMP.[39] Finally, overexpression of c-Fos in BM-DCs dampened IFNβ, IL12-p40, and IL12-p70 production in response to stimulation with CpG.[40] The inhibition of proinflammatory cytokine production by c-Fos was of variable magnitude in different systems. It was not evident in mouse splenic DCs activated with zymosan and only weak in splenic and BM-DCs treated with LPS.[37-39] Our results are consistent with these findings, because proinflammatory cytokine mRNA induction was attenuated only slightly in $miR155^{-/-}$ DCs activated with LPS, but was reduced more markedly in CpG-treated DC^{2114} cells overexpressing c-Fos. The strength of the repressive effect of c-Fos on proinflammatory cytokine production could be influenced by various parameters, including the type of DC, the nature and potency of the maturation signal, and the level of c-Fos expression. We have observed that there is a correlation between the level of c-Fos overexpression in DC^{2114} cells and the extent of inhibition of proinflammatory cytokine mRNA induction (data not shown).

Our results indicate that miR155 regulates both the stability and translation of c-Fos mRNA. Repression at the level of translation appears to be the dominant mechanism, because the increase in c-Fos protein in $miR155^{-/-}$ DCs was considerably greater than the increase in c-Fos mRNA abundance. We have also found that transcription of the c-Fos gene is silenced in Mo-DCs treated with LPS (unpublished data). c-Fos expression in DCs is thus regulated at the levels of transcription, mRNA stability, and translation, suggesting that tightly controlled silencing of c-Fos expression is critical for DC maturation and function.

It remains to be determined whether the repression of c-Fos by miR155 is a mechanism that is specific to DCs. miR155 expression is also induced during the activation of other cell types, including B cells and macrophages. We have observed that c-Fos mRNA levels are strongly decreased in B cells activated with CpG (unpublished data). Down-regulation of c-Fos mRNA expression has also been documented in macrophages stimulated with LPS.[41] Repression of c-Fos expression by miR155 may thus be of more general importance for the activation of various cell types.

Two additional miRNAs have been implicated in the regulation of c-Fos expression. miR101 promotes apoptosis by inhibiting c-Fos expression in human hepatocellular carcinoma cells.[42] c-Fos translation was also reported to be repressed by miR7b in the hypothalamus after chronic hyperosmolar stimulation.[43] Our miRNA expression–profiling experiments indicated that miR101 and miR7b are not expressed at significant levels in immature human or mouse DCs and are not induced upon maturation. It is therefore unlikely that these 2 miRNAs collaborate with miR155 in regulating c-Fos expression in DCs. However, the identification of 3 miRNAs capable of targeting c-Fos mRNA in different systems suggests that the regulation of c-Fos expression by cell-type–restricted miRNAs is a widespread mechanism.

The 3′UTR of c-Fos mRNA contains adenylate- and uridylate-rich elements (ARE) known to promote mRNA degradation.[44] A link was recently established between ARE-mediated mRNA decay and regulation by miRNAs.[45] This raises the attractive possibility that c-Fos expression could be controlled in a cell-type–specific manner by collaboration between specific miRNAs and the ARE-mediated mRNA decay machinery.

c-Fos functions as one subunit of a group of dimeric transcription factors collectively referred to as activating protein 1 (AP-1). AP-1 proteins constitute a population of homo- and heterodimeric complexes containing subunits belonging to the Fos (c-Fos, FosB, Fra-1, and Fra-2), Jun (c-Jun, JunB, and JunD), and activating transcription factor families. AP-1 factors have been implicated in a wide range of processes, including proliferation, differentiation, apoptosis, responses to stress and environmental cues, oncogenic transformation, and metastasis.[46,47] The composition of the population of AP-1 dimers is one of the critical parameters determining AP-1 function in different cell types. Our finding that continued c-Fos expression is detrimental for DC maturation suggests that AP-1 complexes containing c-Fos might repress genes that are implicated in DC maturation. This interpretation is consistent with the recent finding that c-Fos can inhibit TNFα expression by binding to the p65 subunit of NF-kB and inhibiting its recruitment to the *Tnf* promoter.[39] However, a change in the expression of a specific AP-1 subunit could also affect the overall balance between the relative abundance of different AP-1 complexes. An alternative possibility could therefore be that silencing of c-Fos expression induces a shift in the equilibrium between different AP-1 dimers such that the formation of specific AP-1 complexes required for maturation is favored. Distinguishing between these and other possibilities will require a detailed characterization of the composition of the AP-1 population present in immature DCs and the changes that occur in this population when maturation is induced.

The *BIC* gene was first identified as a frequent integration site of avian leucosis virus in chicken B-cell lymphomas.[8,9] It was subsequently found that miR155 expression is frequently up-regulated in B-cell lymphomas and in a wide range of other cancers.[10,11] Furthermore, mice carrying a transgene that enforces miR155 expression in B cells are characterized by a preleukemic pre–B-cell proliferation that eventually develops into B-cell malignancy.[48] The mechanisms underlying the oncogenic role of miR155 remain obscure, but our results raise the intriguing possibility that these mechanisms could involve deregulated c-Fos expression. Indeed, although oncogenic transformation and tumor progression

were initially believed to be associated with increased c-Fos expression, there is growing evidence that c-Fos can also exert tumor-suppressive functions.[49] It is therefore tempting to speculate that abrogation of the tumor suppressor influence of c-Fos could contribute to the oncogenic role of miR155 in the development of tumors.

Acknowledgments

We are grateful to Emmanuèle Barras, Antoine Geinoz, Grégory Schneiter, Mahdia Benkhoucha, Patrice Lalive, David Suter, Bertrand Huard, and Mylène Docquier for technical help and to all members of the laboratory for valuable discussions. 2D2 mice were provided by B. Becher (Zurich, Switzerland) with the permission of V. Kuchroo (Boston, MA). OTII mice were provided by S. Amigorena (Paris, France).

Work in the laboratory of W.R. was supported by the Swiss National Science Foundation, the Geneva Cancer League, The Ernst and Lucy Schmidheiny Foundation, the Swiss Multiple Sclerosis Society, the National Center of Competence in Research on Neural Plasticity and Repair (NCCR-NEURO), and the European Union FP6 consortium Dendritic Cells for Novel Immunotherapies (DC-THERA). Work in the laboratory of H.A.-O. and E.M.Z. was supported by the Swiss National Science Foundation. M.-L.S.-R. was supported by the Alliance for Lupus Research.

Authorship

Contribution: I.D.-S. conceived, designed, and performed the experiments and wrote the paper; W.R. conceived and designed the experiments and wrote the paper; L.C. performed experiments; M.I. and Q.S.-E. contributed to the experimental design; H.A.-O. shared reagents and materials; C.E.V., O.S., P.D., and E.M.Z. performed bioinformatical analysis; and M.-L.S.-R. helped with the design of experiments and analysis of data.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Walter Reith, Department of Pathology and Immunology, Faculty of Medicine, University of Geneva, 1 rue Michel-Servet, CH-1211 Geneva, Switzerland; e-mail: walter.reith@unige.ch.

References

1. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell.* 2004;116(2):281-297.
2. Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet.* 2008;9(2):102-114.
3. Ambros V. The functions of animal microRNAs. *Nature.* 2004;431(7006):350-355.
4. Lu J, Getz G, Miska EA, et al. MicroRNA expression profiles classify human cancers. *Nature.* 2005;435(7043):834-838.
5. Lindsay MA. microRNAs and the immune response. *Trends Immunol.* 2008;29(7):343-351.
6. Taganov KD, Boldin MP, Baltimore D. MicroRNAs and immunity: tiny players in a big field. *Immunity.* 2007;26(2):133-137.
7. Xiao C, Rajewsky K. MicroRNA control in the immune system: basic principles. *Cell.* 2009;136(1):26-36.
8. Clurman BE, Hayward WS. Multiple proto-oncogene activations in avian leukosis virus-induced lymphomas: evidence for stage-specific events. *Mol Cell Biol.* 1989;9(6):2657-2664.
9. Tam W, Ben-Yehuda D, Hayward WS. bic, a novel gene activated by proviral insertions in avian leukosis virus-induced lymphomas, is likely to function through its noncoding RNA. *Mol Cell Biol.* 1997;17(3):1490-1502.
10. Garzon R, Calin GA, Croce CM. MicroRNAs in Cancer. *Annu Rev Med.* 2009;60:167-179.
11. Tili E, Croce CM, Michaille JJ. miR-155: on the crosstalk between inflammation and cancer. *Int Rev Immunol.* 2009;28(5):264-284.
12. Rodriguez A, Vigorito E, Clare S, et al. Requirement of bic/microRNA-155 for normal immune function. *Science.* 2007;316(5824):608-611.
13. Thai TH, Calado DP, Casola S, et al. Regulation of the germinal center response by microRNA-155. *Science.* 2007;316(5824):604-608.
14. Vigorito E, Perks KL, Abreu-Goodger C, et al. microRNA-155 regulates the generation of immunoglobulin class-switched plasma cells. *Immunity.* 2007;27(6):847-859.
15. Dorsett Y, McBride KM, Jankovic M, et al. MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. *Immunity.* 2008;28(5):630-638.
16. Teng G, Hakimpour P, Landgraf P, et al. MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. *Immunity.* 2008;28(5):621-629.
17. Ceppi M, Pereira PM, Dunand-Sauthier I, et al. MicroRNA-155 modulates the interleukin-1 signaling pathway in activated human monocyte-derived dendritic cells. *Proc Natl Acad Sci U S A.* 2009;106(8):2735-2740.
18. Martinez-Nunez RT, Louafi F, Friedmann PS, Sanchez-Elsner T. MicroRNA-155 modulates the pathogen binding ability of dendritic cells (DCs) by down-regulation of DC-specific intercellular adhesion molecule-3 grabbing non-integrin (DC-SIGN). *J Biol Chem.* 2009;284(24):16334-16342.
19. Barnden MJ, Allison J, Heath WR, Carbone FR. Defective TCR expression in transgenic mice constructed using cDNA-based alpha- and beta-chain genes under the control of heterologous regulatory elements. *Immunol Cell Biol.* 1998;76(1):34-40.
20. Bettelli E, Pagany M, Weiner HL, Linington C, Sobel RA, Kuchroo VK. Myelin oligodendrocyte glycoprotein-specific T cell receptor transgenic mice develop spontaneous autoimmune optic neuritis. *J Exp Med.* 2003;197(9):1073-1081.
21. Steiner QG, Otten LA, Hicks MJ, et al. In vivo transformation of mouse conventional CD8{alpha}+ dendritic cells leads to progressive multisystem histiocytosis. *Blood.* 2008;111(4):2073-2082.
22. Landmann S, Muhlethaler-Mottet A, Bernasconi L, et al. Maturation of dendritic cells is accompanied by rapid transcriptional silencing of class II transactivator (CIITA) expression. *J Exp Med.* 2001;194(4):379-391.
23. Santiago-Raber ML, Dunand-Sauthier I, Wu T, et al. Critical role of TLR7 in the acceleration of systemic lupus erythematosus in TLR9-deficient mice. *J Autoimmun.* 2010;34(4):339-348.
24. Suter D, Montet X, Tirefort D, Cartier L, Krause K. A tetracycline-inducible lentivector system based on EF1-alpha promoter and native tetracycline repressor allows in vivo gene induction in implanted ES cells. *J Stem Cells.* 2007;2(2):63-72.
25. Salmon P, Trono D. Production and titration of lentiviral vectors. *Curr Protoc Hum Genet.* 2007;Chapter 12:Unit 12.10.
26. Faraoni I, Antonetti FR, Cardone J, Bonmassar E. miR-155 gene: A typical multifunctional microRNA. *Biochim Biophys Acta.* 2009;1792(6):497-505.
27. Gottwein E, Mukherjee N, Sachse C, et al. A viral microRNA functions as an orthologue of cellular miR-155. *Nature.* 2007;450(7172):1096-1099.
28. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell.* 2009;136(2):215-233.
29. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. *Cell.* 2003;115(7):787-798.
30. Brennecke J, Stark A, Russell RB, Cohen SM. Principles of microRNA-target recognition. *PLoS Biol.* 2005;3(3):e85.
31. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell.* 2007;27(1):91-105.
32. Kuipers H, Schnorfeil FM, Fehling HJ, Bartels H, Brocker T. Dicer-dependent microRNAs control maturation, function, and maintenance of Langerhans cells in vivo. *J Immunol.* 2010;185(1):400-409.
33. Ebstein F, Lange N, Urban S, Seifert U, Kruger E, Kloetzel PM. Maturation of human dendritic cells is accompanied by functional remodelling of the ubiquitin-proteasome system. *Int J Biochem Cell Biol.* 2009;41(5):1205-1215.
34. Fulcher JA, Hashimi ST, Levroney EL, et al. Galectin-1-matured human monocyte-derived dendritic cells have enhanced migration through extracellular matrix. *J Immunol.* 2006;177(1):216-226.
35. Messmer D, Messmer B, Chiorazzi N. The global transcriptional maturation program and stimuli-specific gene expression profiles of human myeloid dendritic cells. *Int Immunol.* 2003;15(4):491-503.
36. Agrawal S, Agrawal A, Doughty B, et al. Cutting edge: different Toll-like receptor agonists instruct dendritic cells to induce distinct Th responses via differential modulation of extracellular signal-regulated kinase-mitogen-activated protein kinase and c-Fos. *J Immunol.* 2003;171(10):4984-4989.
37. Dillon S, Agrawal A, Van Dyke T, et al. A Toll-like receptor 2 ligand stimulates Th2 responses in vivo, via induction of extracellular signal-regulated kinase mitogen-activated protein kinase and

c-Fos in dendritic cells. *J Immunol.* 2004;172(8): 4733-4743.

38. Dillon S, Agrawal S, Banerjee K, et al. Yeast zymosan, a stimulus for TLR2 and dectin-1, induces regulatory antigen-presenting cells and immunological tolerance. *J Clin Invest.* 2006;116(4):916-928.
39. Koga K, Takaesu G, Yoshida R, et al. Cyclic adenosine monophosphate suppresses the transcription of proinflammatory cytokines via the phosphorylated c-Fos protein. *Immunity.* 2009; 30(3):372-383.
40. Kaiser F, Cook D, Papoutsopoulou S, et al. TPL-2 negatively regulates interferon-beta production in macrophages and myeloid dendritic cells. *J Exp Med.* 2009;206(9):1863-1871.
41. Brogdon JL, Xu Y, Szabo SJ, et al. Histone deacetylase activities are required for innate immune cell control of Th1 but not Th2 effector cell function. *Blood.* 2007;109(3):1123-1130.
42. Li S, Fu H, Wang Y, et al. MicroRNA-101 regulates expression of the v-fos FBJ murine osteosarcoma viral oncogene homolog (FOS) oncogene in human hepatocellular carcinoma. *Hepatology.* 2009;49(4):1194-1202.
43. Lee HJ, Palkovits M, Young WS. 3rd. miR-7b, a microRNA up-regulated in the hypothalamus after chronic hyperosmolar stimulation, inhibits Fos translation. *Proc Natl Acad Sci U S A.* 2006; 103(42):15669-15674.
44. Chen CY, Xu N, Shyu AB. mRNA decay mediated by two distinct AU-rich elements from c-fos and granulocyte-macrophage colony-stimulating factor transcripts: different deadenylation kinetics and uncoupling from translation. *Mol Cell Biol.* 1995;15(10):5777-5788.
45. Jing Q, Huang S, Guth S, et al. Involvement of microRNA in AU-rich element-mediated mRNA instability. *Cell.* 2005;120(5):623-634.
46. Wagner EF. AP-1–Introductory remarks. *Oncogene.* 2001;20(19):2334-2335.
47. Eferl R, Wagner EF. AP-1: a double-edged sword in tumorigenesis. *Nat Rev Cancer.* 2003;3(11): 859-868.
48. Costinean S, Zanesi N, Pekarsky Y, et al. Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. *Proc Natl Acad Sci U S A.* 2006; 103(18):7024-7029.
49. Durchdewald M, Angel P, Hess J. The transcription factor Fos: a Janus-type regulator in health and disease. *Histol Histopathol.* 2009;24(11): 1451-1461.

4.4 miRmap: Comprehensive prediction of microRNA target repression strength

Nucleic Acids Research Advance Access published October 2, 2012

Nucleic Acids Research, 2012, 1–11
doi:10.1093/nar/gks901

miRmap: Comprehensive prediction of microRNA target repression strength

Charles E. Vejnar[1,2] and Evgeny M. Zdobnov[1,2,3,*]

[1]Department of Genetic Medicine and Development, University of Geneva, Rue Michel-Servet 1, 1211 Geneva 4, [2]Swiss Institute of Bioinformatics, 1211 Geneva, Switzerland and [3]Imperial College London, South Kensington Campus, London, SW7 2AZ, UK

Received February 2, 2012; Revised September 5, 2012; Accepted September 7, 2012

ABSTRACT

MicroRNAs, or miRNAs, post-transcriptionally repress the expression of protein-coding genes. The human genome encodes over 1000 miRNA genes that collectively target the majority of messenger RNAs (mRNAs). Base pairing of the so-called miRNA 'seed' region with mRNAs identifies many thousands of putative targets. Evaluating the strength of the resulting mRNA repression remains challenging, but is essential for a biologically informative ranking of potential miRNA targets. To address these challenges, predictors may use thermodynamic, evolutionary, probabilistic or sequence-based features. We developed an open-source software library, miRmap, which for the first time comprehensively covers all four approaches using 11 predictor features, 3 of which are novel. This allowed us to examine feature correlations and to compare their predictive power in an unbiased way using high-throughput experimental data from immunopurification, transcriptomics, proteomics and polysome fractionation experiments. Overall, target site accessibility appears to be the most predictive feature. Our novel feature based on PhyloP, which evaluates the significance of negative selection, is the best performing predictor in the evolutionary category. We combined all the features into an integrated model that almost doubles the predictive power of TargetScan. miRmap is freely available from http://cegg.unige.ch/mirmap.

INTRODUCTION

MicroRNAs (miRNAs) are short (~22 nt) non-coding RNAs that guide the RNA-induced silencing complex (RISC) to post-transcriptionally repress the expression of protein-coding genes by binding to targeted messenger RNAs (mRNAs) (1–3). The detailed mechanism of this guidance is not yet resolved, but exact pairing between the so-called 'seed' region, positions from 2 to 7 (or 8) from the 5′-end of the miRNA, and the 3′-UTR of the mRNA is believed to be necessary for most animal miRNA–mRNA interactions (4). Such miRNA seed pairing with a 3′-UTR of an mRNA, however, is not always sufficient for a functional interaction (4), and in a few specific cases, non-canonical pairing (non-Watson–Crick pairing) with G:U wobbles or mismatches may be acceptable (4,5). Nevertheless, in all recent large-scale miRNA experiments (6–9), the strongest prediction signal remains the presence of seed matching sites in regulated mRNAs, and therefore, it is commonly used as a mandatory signal in functional assays. Since the seed match spans only six or seven nucleotides, many of such matches may occur simply by chance. Searching for longer seed matches, which are less likely to occur by chance but also yield stronger repression, therefore increases the specificity while reducing the sensitivity of the target search. Indeed, the seed definition has a prominent effect on the sensitivity (10). Even with a stringent seed definition, there are still many potential miRNA targets, and experimentally testing all miRNA–mRNA combinations having a seed match is practically not feasible. Prioritization of targets for any miRNA functional analysis is therefore of critical importance. This necessitates the ranking of potential miRNA targets bearing a seed, not only predicting in a binary manner if an mRNA is a target or not. A biologically meaningful ranking criterion is the miRNA-mediated repression strength that can be experimentally measured as the effect on mRNA or protein levels. We used a collection features to computationally predict the miRNA repression strength from additional information beyond the seed match, and thereby rank putative miRNA–mRNA interactions in a biologically relevant manner.

The interaction between a miRNA and its mRNA target site can be considered from (i) a thermodynamic, (ii) a probabilistic, (iii) an evolutionary or (iv) a sequence-based point of view. Several computational tools (11) for

*To whom correspondence should be addressed. Tel: +41 223795973; Email: evgeny.zdobnov@unige.ch

© The Author(s) 2012. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from http://nar.oxfordjournals.org/ by Charles Vejnar on October 15, 2012

Table 1. Approaches used by miRNA target prediction software tools

	Thermodynamic	Evolutionary	Probabilistic	Sequence-based	References
miRmap	✓	✓	✓	✓	
TargetScan		✓[a]		✓	Grimson *et al.* (6)
PITA	✓				Kertesz *et al.* (12)
PicTar	✓	✓	✓		Krek *et al.* (13)
miRanda	✓				John *et al.* (14)
RNAhybrid	✓				Rehmsmeier *et al.* (15)
DIANA-microT	✓	✓			Kiriakidou *et al.* (16)
ElMMo		✓	✓		Gaidatzis *et al.* (17)
PACMIT	✓		✓		Marín and Vanícek (18)

[a]We used the TargetScan context score (6). An evolutionary approach was latter added in TargetScan (19), but it is a separated filter not included in the context score model.

miRNA target site prediction have been developed that use one or more of these aspects (Table 1). This overlap hinders effective comparisons of individual predictor performances, which may use overlapping sets of prediction features and variable implementations of the same approaches. Moreover, most of these programs are not freely available, complicating direct comparisons (20). To avoid this type of benchmarking bias, more recent studies (21–23) have recomputed predictions with limited sets of features focusing on binary predictions of target or non-target instead of considering the strength of repression. Ignoring the fact that miRNA repression strength displays a continuous range of strong-to-weak effects makes the distinction between target and non-target a matter of choosing an arbitrary cutoff. Here, we present a thorough comparison of the power of individual approaches to predict the repression strength of miRNA–mRNA pairs, assessed using data from transcriptomics, immunopurification (IP), proteomics and polysome fractionation high-throughput experiments. This was achieved using miRmap, our implementation of a comprehensive set of prediction features that we have made available as an open-source Python library. The features encompass the thermodynamic, conservation, probabilistic and sequence-based categories; eight features have been described previously in the literature, while three are novel features, each from a different category. We examined correlations among the features, measured the predictive power and combined all of them into an integrated prediction model.

MATERIALS AND METHODS

Experimental data

Expression microarrays of miRNA-transfected HeLa cells from Grimson *et al.* (6) were downloaded from GEO (GSE8501) for miRNAs 122 a, 128 a, 132, 133 a, 142, 148 b, 181 a, 7 and 9. We used expression data at 24 h post-transfection and, similar to Grimson *et al.* (6), only selected probes with signal intensities above the median in the control transfection experiments to retain only the transcripts expressed enough to observe miRNA silencing.

Similarly, we downloaded the expression microarrays from Linsley *et al.* (24) from GEO (GSE6838). We used the experiments GSM156522, GSM156523, GSM156524, GSM156545, GSM156546, GSM156547, GSM156548, GSM156553, GSM156557, GSM156559, GSM156576, GSM156577, GSM156578, GSM156579 and GSM156581, measured at 24 h with the same experimental conditions. We applied the same selection filter as above (6).

We downloaded the Selbach *et al.* (7) proteomics over-expression data directly from the web site dedicated to the article. We included expression fold-changes measured at 32 h for miR-1, miR-155 and miR-16 but excluded let-7 b and miR-30 a as these miRNAs exert a negative feedback effect on the RNA silencing pathway (7,21).

HITS-CLIP data from the Chi *et al.* (9) study were also downloaded from the web site dedicated to the article. After cross-linking Argonaute (Ago) with its neighbouring RNAs, the authors immunopurified Ago and sequenced the pulled-down RNAs. We used the peak height as a measure of miRNA targeting for the 20 available most abundant miRNAs and filtered the relevance of the peaks using a biological complexity (BC, a measure of reproducibility between biological replicates) criterion strictly superior to 1 for medium stringency.

Hendrickson *et al.* (25) injected miR-124 into HEK293T cells and measured (i) the miR-RISC association with Ago IP, (ii) transcriptome expression with microarrays and (iii) translation activity with polysome fractionation. We used dataset number 5 from the Supplementary Information which includes all measurements for each transcript.

Sequence data

RefSeq 47 (26) mapped on the human (hg19) and mouse (mm9) genomes by the UCSC (27) were used to define mRNA annotations, restricted to 'NM_' transcripts. miRBase 16 (28) was used for miRNA annotations.

Target prediction features

Thermodynamics of miRNA–mRNA interactions

The miRNA–mRNA pair forms an RNA duplex. Using the Vienna RNA Secondary Structure library (29), we computed the minimum free folding energy (MFE) of this duplex (with the 'cofold' function), and named it

Downloaded from http://nar.oxfordjournals.org/ by Charles Vejnar on October 15, 2012

'ΔG duplex'. While the structure with the lowest predicted energy or MFE is the most stable structure, populations of RNAs adopt different sub-optimal structures *in vivo*. We computed the ensemble free energy of the binding (with the 'co_pf_fold' function), and named this feature 'ΔG binding'. We used the 'cofold' function for the 'ΔG duplex' computation as this function of the Vienna RNA Secondary Structure library is more appropriate than the modified 'duplexfold' function used in PITA (12) to compute this feature. The 'duplexfold' function was written to quickly scan for possible hybridization sites, whereas the 'cofold' function, albeit being more computationally intensive, was specifically designed to compute the duplex free energy taking into account intra-molecular and inter-molecular pairs.

The RISC is much larger than the miRNA (30) and must bind to the mRNA in an extended single-stranded form. We computed the energy required to unfold the 3′-UTR region of the target site (this area can be optionally extended), named 'ΔG open', similarly to PITA (12), with the 'pf_fold' function from the Vienna Library (29). The computation of 'ΔG open' requires two energy calculations; the free energy of the mRNA constrained to maintain the target site single stranded is subtracted from the free energy of the same unconstrained mRNA. The single-strand constraint was placed on a segment of 70 nucleotides centred on the target site. Finally, 'ΔG open' summed with 'ΔG duplex' or 'ΔG binding' gives the total system energy: we named it 'ΔG total' (named ΔΔG in PITA (12)).

Probability of the motif occurrence

We modelled the 3′-UTR sequence as a Markov process (order 1, as 3′-UTR sequences are too short to parameterize higher orders) and determined the expected probability of finding at least *n* occurrences of the motif defined as either an exact seed match or the full miRNA binding site, using two different methods. In the first method, the probability distribution was approximated with a binomial distribution, as in Marín and Vanícek (18), while in the second method, we computed the exact probability distribution based on the theoretical work of Nuel *et al.* (31).

Conservation of the target site

Using the UCSC (27) MultiZ multiple genome sequence alignments (hg19, MultiZ 46-way; mm9, MultiZ 30-way), we searched for conserved miRNA target sites in the alignment blocks defined by the 3′-UTRs of the reference species (human or mouse for the HITS-CLIP data). From a mammalian species tree (UCSC (27)), we first pruned all the species that did not contain the target site. We then summed the lengths of the remaining branches (as in (32)) to obtain the branch length score (BLS). As implemented by Friedman *et al.* (19), we summed the branch lengths of the species topology fitted for each 3′-UTR alignment with the REV model using the PhyloFit program from the PHAST suite (33). Tree manipulations were done with the DendroPy (34) library.

To test for evidence of negative selection acting on miRNA target sites, we used the Siepel, Pollard and Haussler (SPH) test implemented in the PhyloP program of the PHAST suite (33). This test evaluates if the branch lengths of the tree built from the target sites are significantly shorter (less divergent because of negative selection) than the background (the 3′-UTR as for the previous method). The reported values in the text are the test $-\log(P\text{-value})$.

PhastCons 46-way run data from UCSC (27) were used to compute the average seed match probability to be a conserved element. The PhastCons scores of each base in the seed were averaged to obtain the seed score (23,35).

Sequence features

We implemented the three sequence features of the TargetScan context score (6): (i) the A and U nucleotide ratio over G and C, weighted around the seed match, (ii) the 3′-compensatory pairing feature and (iii) the distance between the target site and the nearest 3′-UTR end.

Relative importance of features

We computed the relative importance of features in the multiple linear models with the CAR method (36) which decomposes the proportion of the variance explained by each variable of a model while taking the correlations among variables into account.

RESULTS

miRNA target prediction library

We developed a comprehensive prediction model implemented as the miRmap open-source Python library (Figure 1) with a total of 11 features covering a wide range of published and novel methods (Table 2). With our own implementation, we compared the different features without the biases inherent to comparison of pre-computed predictions. We evaluated the features' individual predictive power, measured their intercorrelations and examined different combinations of methods. Additionally, in order to facilitate the library usage, five features are implemented in pure Python.

Novel methods include (i) a more accurate way to compute the binding energy between the miRNA and the mRNA based on the ensemble free energy instead of the minimum free energy, (ii) an exact method to compute the probability that the seed match is an over-represented motif in the 3′-UTR and (iii) a non-empirical statistical test to assess the significance of target site evolutionary conservation.

ΔG binding

miRNAs bind to their targeted mRNAs forming a helix. The minimum free folding energy (MFE) of these duplexes can be computed ('ΔG duplex') but the structure with the MFE only represents a fraction of the possible and existing structures. Additionally, 'ΔG duplex' is a measure of the energy of the entire double-stranded structure, it does not describe the binding energy itself. This is captured by the 'ΔG binding' measurement, which

Downloaded from http://nar.oxfordjournals.org/ by Charles Vejnar on October 15, 2012

represents only the binding energy computed from the ensemble free energy.

P exact

Within 3′-UTRs, only certain sequence regions have regulatory or structural roles. These regions can therefore be considered as islands of natural selection in a sea of mostly neutrally evolving sequence; ~5% of the human 3′-UTR bases are constrained (37). This distinction can be exploited within a probabilistic (or evolutionary, see next paragraph) framework to distinguish the background sequence composition from the target site composition. Having modelled the background sequence composition (with a Markov process, see 'Materials and Methods' section), it is possible to compute a probability distribution of motif occurrences in order to assess the significance of the site presence. Several approximations (e.g. Gaussian, Poisson, binomial or large deviation) can be used to compute the probability distribution depending on the sequence length and the expected number of motif occurrences. As 3′-UTR sequences are relatively short, we computed not only an approximate distribution ('P.over binomial') but also an exact distribution ('P.over exact').

```
1  >>> import mirmap
2  >>> import mirmap_library_link
3  >>>
4  >>> mimset = mirmap.mm(<ENST00000354719 sequence>,
   'UUGUGCUUGAUCUAACCAUGU')
5  >>> mimset.find_potential_targets_with_seed( \
   ... allowed_lengths=[7])
6  >>>
7  >>> mimset.libs = \
   ... mircap_library_link.LibraryLink(<path>)
8  >>>
9  >>> mimset.eval_tgs_au()
10 >>> mimset.eval_tgs_position()
11 >>> mimset.eval_tgs_pairing3p()
12 >>> mimset.eval_dg_duplex()
13 >>> mimset.eval_dg_open()
14 >>> mimset.eval_dg_total()
15 >>> mimset.eval_prob_exact()
16 >>> mimset.eval_prob_binomial()
17 >>>
18 >>> print mimset.report()
19 1707                          1737
20 |                             |
21 GUCCUGUAAUCUGUUUCUAGGUGAAGCAUACUCCAGUGUUU
22                        |||||||.
23            UGUACCAAUCUAGUUCGUGUU
24   AU content                       0.02802
25   UTR position                    -0.03272
26   3' pairing                      -0.01050
27   ΔG duplex (kcal/mol)           -14.10000
28   ΔG binding (kcal/mol)          -14.30905
29   ΔG open (kcal/mol)              16.87775
30   ΔG total (kcal/mol)              2.77775
31   Probability (Exact)              0.20204
32   Probability (Binomial)           0.15780
33 >>>
```

Figure 1. miRmap library usage: after importing the library (lines 1 and 2), a 'mimset' object is created containing the mRNA and miRNA sequences. We then call a method of the mimset object to search (line 5) for seeds with a length of 7 (all parameters have defaults that can be changed this way). The link with the C libraries is initalized on line 7. We then manually evaluate the repression strength with differents methods (lines 9–16). Each of these methods have modifiable parameters. We finally print a report (line 18).

PhyloP

Empirical distributions described previously (19,32) can be used to assess the statistical significance of the 'BLS' (see 'Materials and Methods' section). Alternatively, a theoretical framework (33) may be used to test for significant natural selection; the SPH test evaluates the probability that part of a sequence is under selection, in our case negative selection. This framework relies on a comparison of the reference tree built from the complete 3′-UTR multiple sequence alignment and the tree built from the target site (the sequence region delineated by the seed match or the full target site) multiple sequence alignment.

For a meaningful comparison of a potential target site to the complete 3′-UTR, each of the sequences in the target site alignment should be a recognizable miRNA binding site. In other words, for the 'PhyloP' feature to produce meaningful results, target site positions should be conserved among species. To test this condition, potential target sites were identified by searching the 3′-UTR alignments of all human mRNAs for matches to all known human miRNA seeds. Positions are conserved for the majority of human seed matches; on average, 76% of the human seed matches are found at the same position in the alignment for the other mammalian species. For this analysis, sequences of species in the alignment without any seed match were discarded. According to this analysis, the turn-over of miRNA target sites in mammals seems to be low. The conservation of target site positions in the alignment supports our usage of PhyloP. Moreover, the

Table 2. miRNA target prediction features of the miRmap library

Category	Feature	Description	Python-only	Remarks
Thermodynamic	ΔG duplex	MFE with RNAcofold		
	ΔG binding	Binding energy based on ensemble free energy		*New feature*
	ΔG open	mRNA opening free energy—Accessibility		As in PITA (12)
	ΔG total	ΔG Duplex + ΔG open		Similar to ΔΔG in PITA (12)
Probabilistic	P.over binomial	Site over-representation prob. (binomial dist.)	✓	As in PACMIT (18)
	P.over exact	Site over-representation prob. (exact dist.)		*New feature*
Conservation	BLS	Branch length score on 3′-UTR fitted tree	✓	Similar to Stark *et al.* (32)
	PhyloP	SPH test from PhyloP		*New feature*
Sequence	AU content	AU nucleotide composition around the seed	✓	As in TargetScan (6)
	UTR position	Distance from the nearest 3′-UTR end	✓	As in TargetScan (6)
	3′-pairing	3′-compensatory pairing	✓	As in TargetScan (6)

Downloaded from http://nar.oxfordjournals.org/ by Charles Vejnar on October 15, 2012

percentages vary from 47 to 99% if we analyse each miRNA individually. The miRNAs with low complexity sequences tend to have low percentages, which also support the choice of this test as low complexity miRNAs have less specific target sites.

Correlation among features

We identified potential miRNA target sites by searching for matches to canonical 7-mer seeds on all 3′-UTRs of the human transcripts and predicted their strengths using the 11 methods of our miRmap library (see above and 'Materials and Methods' section). We focused our analysis on 7-mer seeds rather than shorter 6-mer seeds as stronger mRNA repression is associated with longer seeds. While this choice results in greater confidence in our feature performance analysis, target prediction with increased sensitivity could be easily obtained by integrating shorter seeds (see below). To evaluate the target site rankings computed with each feature and ignore other differences, e.g. their variances, we computed the Spearman rank correlation between feature pairs (Supplementary Table S1). The absolute values are plotted in Figure 2A.

The three most highly correlated feature pairs are those that measure the same underlying parameters using slightly different approaches: 'ΔG duplex' and 'ΔG binding' with 0.962, 'P.over exact' and 'P.over binomial' with 0.806 and 'ΔG open' and 'ΔG total' with 0.725. 'ΔG open' and 'AU content' show a correlation of −0.635; as folding algorithms rely on pairing and stacking energies that are stronger for GC than AU pairs, AU-rich sequences form potentially less stable structures, which explains the inverse correlation between 'ΔG open' and 'AU content'. Since these two features evaluate the accessibility of mRNA to miRNA repression, we grouped them in an 'accessibility group' together with 'ΔG total'.

As only the top miRNA target predictions are often used in experimental studies, we measured the overlap among features for their best quartiles. On the first Venn diagram (Figure 2B), we present one feature per group (accessibility, conservation and probabilistic), revealing the low overlap among these methods. The second Venn diagram (Figure 2C) confirms that 'ΔG open' and 'AU content' features belong to the same accessibility group whereas 'ΔG duplex' is a distinct feature not related to the target accessibility. However, target prediction program comparisons (see 'Introduction' section) often include PITA (12) which combines both 'ΔG open' and 'ΔG duplex', making any conclusions made in these comparisons about individual feature performance inaccurate.

Individual feature performance

We evaluated the performance of each feature using data from seven experiments coming from five studies (Table 3) that cover different aspects of miRNA repression and use different assay techniques. (i) Chi *et al.* (9) performed an Ago-RNA cross-linking experiment followed by IP and sequencing from which miRNA binding sites

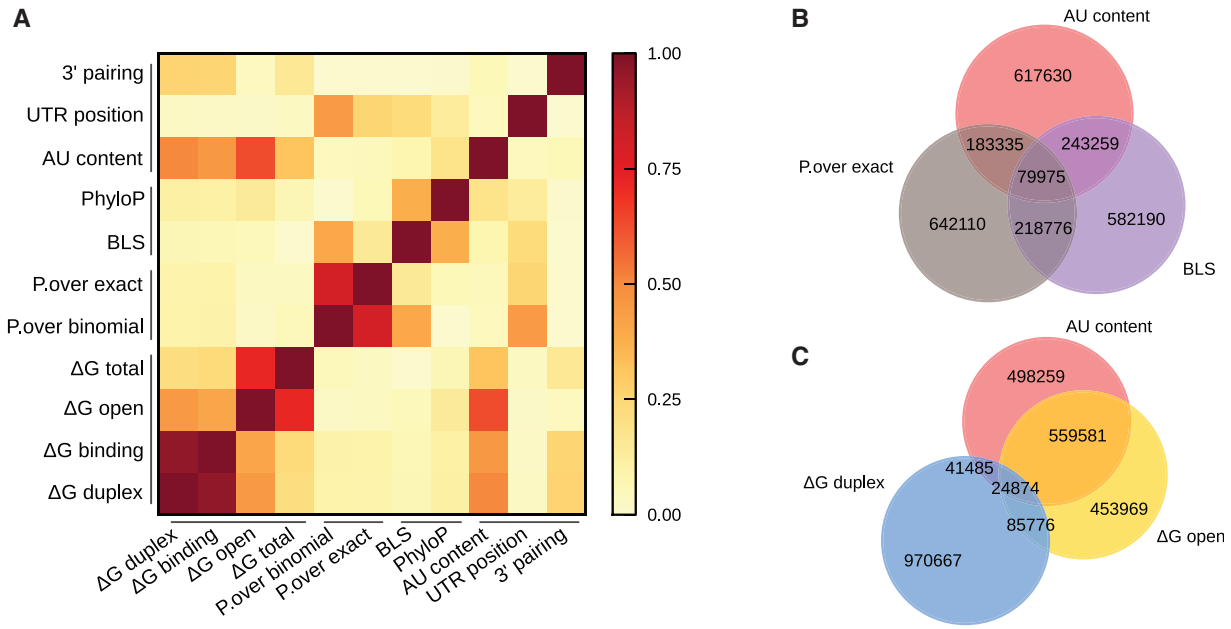


Figure 2. Correlation among features based on prediction for human miRNAs and mRNAs. (**A**) A heatmap of the absolute values of Spearman correlation coefficients between pairs of features classified in methods categories. Venn diagrams (**B**) and (**C**) show the overlaps among the first best prediction quartiles of selected features. One feature per category (sequence-based with 'AU content', conservation with the 'BLS' and probabilistic with 'P.over exact') is shown on (A). Venn diagram (C) underlines the high overlap between 'AU content' and 'ΔG open' that we grouped in the 'accessibility group', whereas 'ΔG duplex' has a very low overlap with these two features. We grouped 'ΔG duplex' with 'ΔG binding' in the 'binding energy' group. Numbers of predicted relationships between human miRNA and mRNA are written in the corresponding overlaps of the Venn diagrams.

Downloaded from http://nar.oxfordjournals.org/ by Charles Vejnar on October 15, 2012

were assayed. (ii) Hendrickson *et al.* (25) performed an Ago-IP without cross-linking that we included to underline the effect of the cross-linking step. To measure the effect on mRNA levels, we used studies based on miRNA transfections followed by microarray measurements from (iii) Grimson *et al.* (6), (iv) Linsley *et al.* (24) and (v) Hendrickson *et al.* (25). To assess the effect of miRNA on translation, we took advantage of polysome fractionation experiments from (vi) Hendrickson *et al.* (25), and of proteomics experiments from (vii) Selbach *et al.* (7) based on the pSILAC technology to obtain the final translation output.

We identified potential miRNA target sites by searching for matches to canonical 7-mer seeds on the transcripts involved in each experiment and predicted their strength with the 11 methods implemented in our miRmap library, and an additional feature derived from the PhastCons UCSC track (see 'Materials and Methods' section) to facilitate comparisons with Wen *et al.* (23) results. We then evaluated the correlations between the measured and predicted miRNA repression strengths.

We focused our first analysis on the transcriptomics data, as these experiments measure a predominant effect of miRNA repression (38,39) and have the largest scale ('Trans.Grimson', 'Trans.Linsley' and 'Trans.Hendrickson' involve a total of 24 miRNAs). Figure 3 shows the linear regressions and correlations between each feature and the observed reductions in mRNA levels for the 'Trans.Grimson' dataset (Supplementary Table S2). The correlation coefficients range from 0.000 for the worst performing feature, 'ΔG duplex', to −0.229 for the best feature, 'AU content'. The next best features are 'PhyloP', 'PhastCons', 'ΔG total', 'ΔG open', followed by 'P.over exact' and 'BLS'. Two of our novel features show better correlations than their related features: (i) 'PhyloP' is the best performing conservation method (−0.205) and (ii) 'P.over exact' performs better than 'P.over binomial', i.e. computing the exact probability distribution is better than using the binomial approximation (0.170 versus 0.147). In addition, (iii) considering the ensemble energy outperforms using only the MFE ('ΔG binding': 0.023 versus 'ΔG duplex': 0.000).

In our second analysis, we examined all the datasets in order to compare the performance of each feature across additional aspects of miRNA repression, assessed through IP, proteomics and polysome fractionation experiments. Correlations for each feature and each experimental

Table 3. Experimental studies used to evaluate miRNA target prediction features

Dataset name	Type	Publication
Trans.Grimson	Microarray	Grimson *et al.* (6)
Trans.Linsley	Microarray	Linsley *et al.* (24)
Prot.Selbach	pSILAC	Selbach *et al.* (7)
IPcross.Chi	HITS-CLIP	Chi *et al.* (9)
IP.Hendrickson	Immunopurification	Hendrickson *et al.* (25)
Trans.Hendrickson	Microarray	Hendrickson *et al.* (25)
RibN.Hendrickson	Polysome fractionation	Hendrickson *et al.* (25)

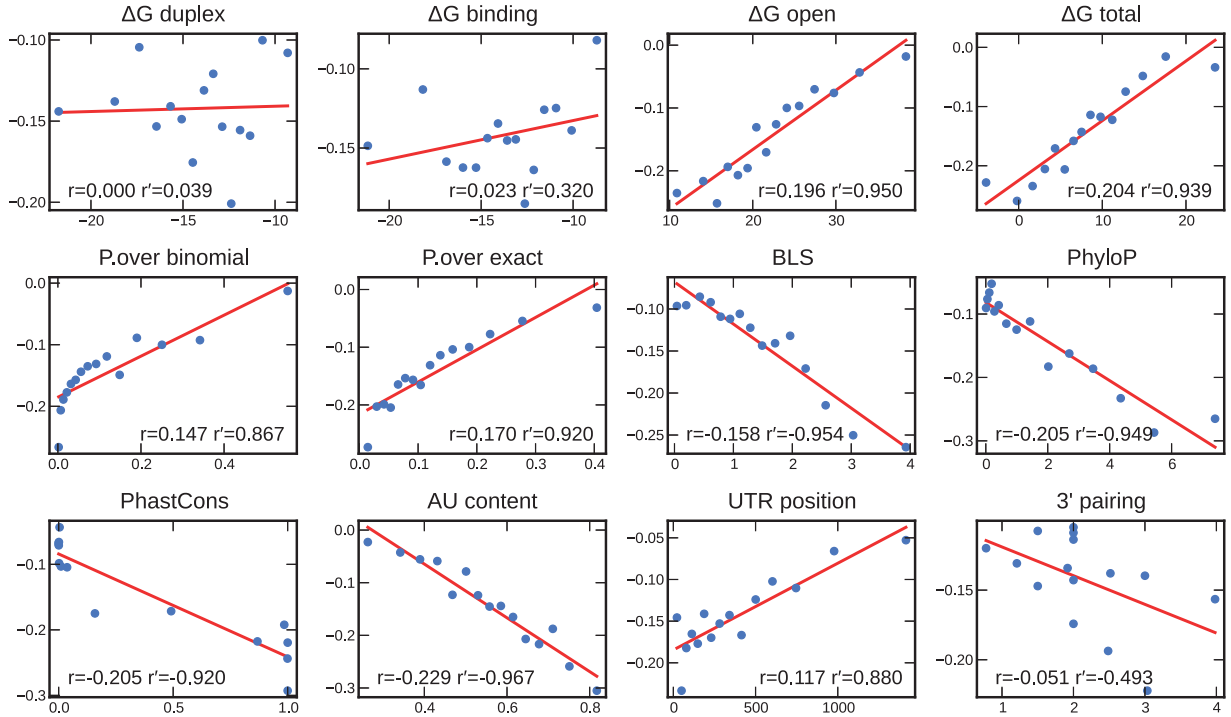


Figure 3. Correlation between each feature and the expression fold-changes of mRNAs following miRNA injection ('Trans.Grimson' dataset). Data points were binned in 15 equally sized bins. The average in each bin is represented by a blue dot. We fitted a linear regression model (red line) on the blue dots. r is the correlation on the full dataset; r' is the correlation on the binned dataset. P-values can be found in Supplementary Table S2.

Downloaded from http://nar.oxfordjournals.org/ by Charles Vejnar on October 15, 2012

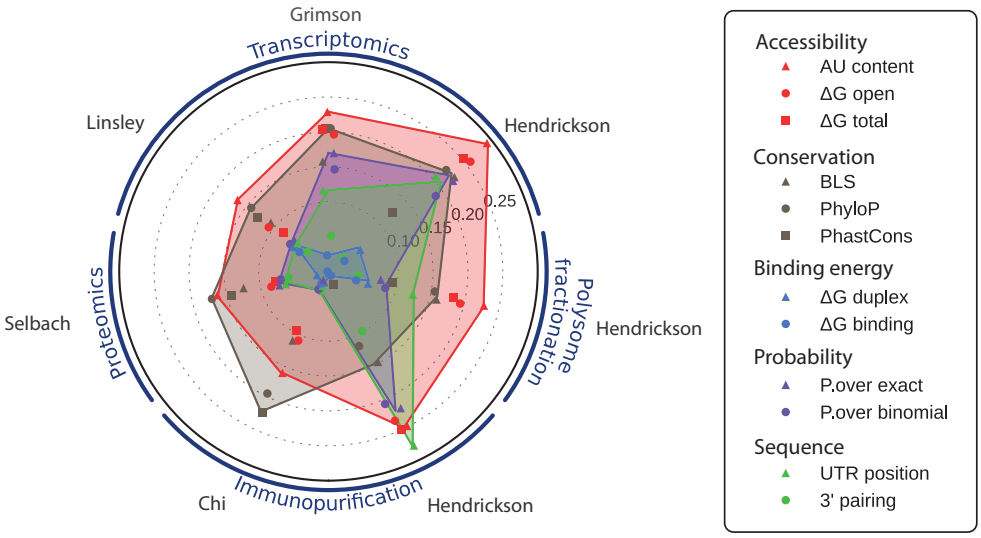


Figure 4. Correlation between each feature and the seven experimental miRNA repression measures (the name of the first author of each dataset is shown in grey) classified in transcriptomics, proteomics, IP and polysome fractionation experiment types. Target prediction features are organized into groups that aim to evaluate the same type of information. The radial axis represents the correlation coefficient (the highest correlations are the furthest from the centre of the circle).

dataset are plotted in Figure 4 (Supplementary Table S2). Remarkably, feature performances show high consistency between each of the experimental datasets: accessibility features (red) always perform well, while binding energies (light blue) are always poorly predictive. As target sites in our study contain a seed, the part of the binding energy discriminating target sites is due to the seed nucleotide composition and to the pairing outside the seed. This energy does not drive the miRNA repression strength, as confirmed by the low performance of '3′-pairing'. Moreover, the ranking of each feature performance is very similar between datasets using the same experimental techniques, e.g. the 'Trans.Grimson' and 'Trans.Linsley' datasets. While based on only a single miRNA, the 'Trans.Hendrickson' dataset shows better overall performance with only minor differences: 'UTR position' improved its ranking while 'PhastCons' is outperformed by 'BLS'.

'AU content' consistently provides the best measure of target site accessibility. This is in agreement with findings from Wen *et al.* (23), but in contrast to results from Hausser *et al.* (21), which described better performance with 'ΔG open' for an IP experiment. However, for the 'IP.Hendrickson' dataset, which, like Hausser *et al.* (21) involved IP without cross-linking, 'AU content' and 'ΔG open' perform equally well. The 'IP.Hendrickson' experiment is also distinguished by the probabilistic (purple) and 'UTR position' (green) features that outperform the conservation features (grey), which may be explained by the lower precision of this method (i.e. IP without cross-linking), performed with a single miRNA.

The best conservation feature performance is generally slightly lower than the best accessibility feature, but it outperforms 'AU content' for the proteomics and HITS-CLIP datasets. 'PhastCons' performance on the HITS-CLIP dataset is consistent with findings from Wen *et al.* (23). Our novel conservation feature, 'PhyloP', shows the best or tied-best performance for five out of the seven datasets. When outperformed, it is only marginally outperformed implying that 'PhyloP' is the best overall conservation feature.

Hendrickson *et al.* (25) polysome fractionation measured the miRNA effects as ribosome occupancy (fraction of a given gene's transcripts associated with ribosomes) and ribosome density (the average number of ribosomes bound per unit length of coding sequence). Effects caused by the miRNA on both parameters were detected by the authors, but were substantially higher on the ribosome density, in agreement with the absence of correlation with the ribosome occupancy that we observed, i.e. this measurement is not quantitative. However, the ribosome density is a quantitative measure of the miRNA effect, as the correlations were as high or higher than those of the large-scale transcriptomics experiments. We observed again, as for all Hendrickson *et al.* (25) datasets, a higher correlation for the 'UTR position' feature, probably caused by the experimental setup.

Combining prediction features

The features correlate linearly with experimentally measured miRNA repression levels. We combined 10 features of our miRmap library (we excluded 'ΔG total', as this feature is simply the sum of 'ΔG duplex' and 'ΔG open') with a multiple linear regression on the 'Trans.Grimson' dataset ($P = 4.9 \times 10^{-110}$; Supplementary Figure S7). This model explains 12.7% of the variance, close to a 2-fold increase over TargetScan context score (6): with the same type of regression, the three features of TargetScan context score ('AU content', '3′-pairing' and 'UTR position') explain only 7.49% of the variance. This improved performance of our model is confirmed by the

Downloaded from http://nar.oxfordjournals.org/ by Charles Vejnar on October 15, 2012

higher correlations with the experimental measurements, computed in the same manner as the individual feature correlations (Figure 5A). The contribution of each feature (on 'Trans.Grimson' dataset, Figure 5B) generally mirrors the rankings based on individual feature correlations: 'AU content' is the most explanatory feature, but 'P.over exact' contributes more in the regression model than its correlation rank suggests. Interestingly, the conservation features 'PhyloP' and 'BLS' contribute ~14 and ~11%, respectively, despite using the same input data (multiple genome sequence alignments) both contribute substantially to the explanation of the variance. Among the accessibility features, 'ΔG open' contributes only half as much as 'AU content' (15 and 30%, respectively). A model limited to the five features with the greatest contributions in the model with all the features (they represent 90.5% of the variance explanation of the full model) still explains 11.6% of the variance.

Instead of evaluating the model directly in terms of the explained variance, the quality of the ranking can be estimated by ordering the target sites by predicted strength, binning them and computing the mean expression fold-change of each bin. This approach, also used in (40) to evaluate the ranking of different tools for predicting miRNA repression strength on translation with proteomics data, was applied to 10 quantiles of the ordered predictions (Supplementary Figure S2). The overall distribution was shifted to lower fold-changes for miRmap compared with TargetScan context score, indicating a better ranking as a decrease in fold-change corresponds to greater repression. For the first quantile, the mean fold-change was reduced from −0.32 to −0.39 with miRmap.

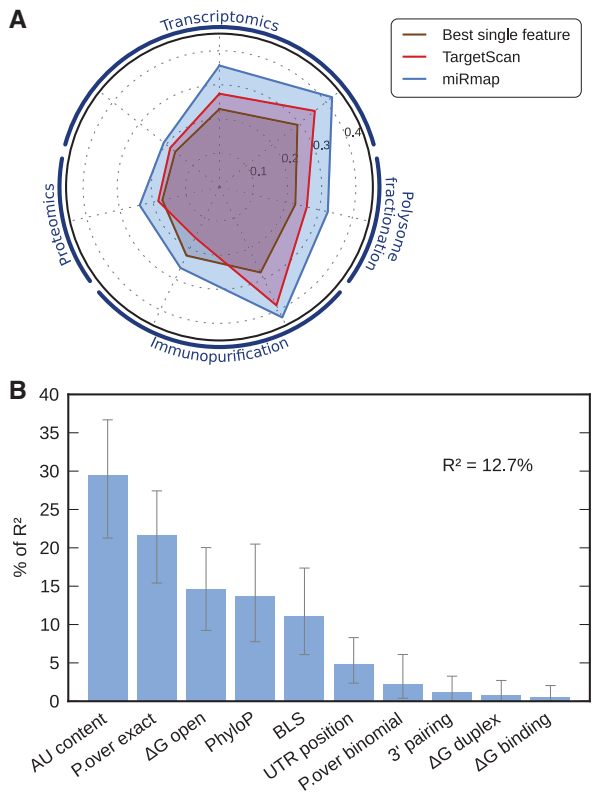


Figure 5. (**A**) Performance comparison (as coefficient correlations with experimental miRNA repression measures; order of the experiments is the same as Figure 4) of the best performing feature (brown), TargetScan context score (red) and miRmap (blue). (**B**) Feature relative importance in the miRmap multiple linear regression model predicting miRNA repression strength. R^2 is the proportion of variance explained by the model. 'AU content' is the most explanatory variable with 29% of R^2.

Multiple linear regressions with the other datasets further support the conclusions from the analyses of individual feature performance (Supplementary Figures S1 and S3). They confirmed (i) the importance of 'PhyloP' for the 'IPcross.Chi' dataset (64% of R^2) over 24% for 'AU content', (ii) the similar importance of 'PhyloP' and 'AU content' for proteomics (31% and 39% of R^2, respectively) and (iii) the relevance of polysome fractionation experiment ('RibN.Hendrickson' dataset) to measure miRNA repression strength compared with proteomics as 10.6% of the variance was explained by the model (5.75% for proteomics). We also observed that the model computed on the 'Trans.Linsley' dataset explains only 4.36% of the variance even though this dataset is larger and based on the same techniques as the 'Trans.Grimson' dataset (R^2 = 12.7%).

Shorter seeds may also promote miRNA repression, but usually with lower efficiencies (4). We therefore tested our approach on canonical 6-mer seeds by computing a model with these seed matches on the 'Trans.Grimson' dataset. While the global importance of each feature remained generally similar, with accessibility features being the most explanatory, R^2 dropped to 8.31% of the variance (Supplementary Figure S4A), which still outperforms TargetScan context score (R^2 = 4.70%). Interestingly, the importance of the 'P.over exact' probabilistic feature was reduced from 22 to 7%—falling from second position to fifth—as expected with shorter seeds where matches occur more frequently by chance and are therefore less statistically distinguishable from the background. We also evaluated the model by computing the distribution of fold-changes (Supplementary Figure S4B). As expected, the mean fold-changes were not as low as with the 7-mer seeds, nevertheless they confirmed the better ranking achieved with miRmap compared with TargetScan context score, e.g. the mean fold-change of the first quantile was reduced from −0.16 to −0.21. These results were further supported by the analysis of the other datasets (Supplementary Figures S5 and S6).

Combining multiple target sites

Each mRNA can contain many miRNA target sites. Although most experimental datasets focus on a single miRNA at a time (or all miRNAs for the 'IPcross.Chi' dataset), a framework that can capture the multiplicity of these interactions should improve the predictive power. We examined three simple functions to combine the individual scores of target sites into a global metric at the mRNA level: the best (minimum or maximum depending on the sign of the correlation), the sum and the log of the sum of

Downloaded from http://nar.oxfordjournals.org/ by Charles Vejnar on October 15, 2012

the exponentials. For this analysis, we selected transcripts from the 'Trans.Grimson' dataset with exactly two target sites, resulting in a sample size of 370 mRNAs (only 53 mRNAs have exactly three target sites). For this study, only features predicting different strengths for each target site in a 3′-UTR are appropriate as they would show different correlations for each function, thereby allowing function comparison. As the probabilistic features compute the probability of a fixed number of seed matches in the 3′-UTR, and as the 'BLS' score is also computed for the entire 3′-UTR, they could not be used.

The log of the sum of the exponentials function is designed to approximate interaction kinetics on the principle that stronger sites would drive the observed repression at the mRNA level. However, this function performed poorly for every feature as opposed to the sum (Supplementary Figure S8), which means that every target site has the same importance, indicating that the quantity of miRNA molecules is not limiting the repression reaction in this experiment. Regarding the binding energy features, 'ΔG duplex' and 'ΔG binding', the minimum energies provided the best predictors, i.e. the best site drives the repression for these two features. In contrast to their relatively poor performance with single site predictions, their performance was substantially increased (with correlations from 0 to 0.094 ($P = 0.072$) and 0.023 to 0.119 ($P = 0.022$) for 'ΔG duplex' and 'ΔG binding', respectively) but they still did not outperform the other features. The performance ranking among the remaining features was not substantially different to the single site predictions and, as already observed before (7), summing was the best option for the majority of them.

DISCUSSION

We examined the performance of 12 features designed to predict the strength of miRNA repression on targeted mRNAs independently, and combined them into a linear model. This approach allowed us to assess feature accuracy to rank miRNA targets and avoid the choice of a threshold or the definition of a negative dataset (see 'Introduction' section). Overall, our combined features predict the strength of miRNA target repression more accurately: on the 'Trans.Grimson' dataset, our model explains 12.7% of the variance whereas TargetScan context score ('AU content', '3′-pairing' and 'UTR position') explains 7.49% with the same type of linear model. We tested a more elaborate method than linear regression, the ensemble rule fitting, but it did not improve the predictions (data not shown). In our linear model, the feature explaining the largest part of the variance is the 'AU content' (29% of R^2) which measures the accessibility of the miRNA target sites to the RISC. This result is consistent with TargetScan, but the proportion of the variance explained by this feature decreased from 74% of R^2 in TargetScan to 29% in our model, as we included an additional method to compute the accessibility ('ΔG open'). Indeed, the correlations among the features, and their individual performance across different datasets, revealed five distinct groups of prediction features. In particular, the accessibility group includes the thermodynamic evaluation of the cost to open the target site and the neighbouring structures ('ΔG open'), and the 'AU content' feature, which are well correlated and performed similarly across all experimental datasets. Interestingly, 'ΔG open' is outperformed by 'AU content': computing a weighted partial (the stacking energy is ignored in the 'AU content' feature) accessibility feature is better than the allegedly more accurate feature that attempts to compute the 'true' accessibility.

Other miRNA target prediction tools (Table 1) consider a single or subset of our features. For example, PITA (13) considers only 'ΔG total', and PACMIT (18) a combination of 'ΔG open' and 'P.over binomial'. As the performance of each of these features is lower than the combined approach of miRmap (Figure 4), these tools have less predictive power. While assessment of tools with different seed lengths, features and annotation sets have its caveats (see 'Introduction' section), TargetScan context score was the best performing tool according to large-scale proteomics experiments (7,8). As miRmap's ranking of miRNA targets outperforms that of the TargetScan context score, we can speculate that our approach is the most predictive. Although we concentrated on 7-mer seeds, we showed that the same approach can be applied to 6-mer seeds, and it may also be used for the rarer centred seeds (41) to increase the overall prediction sensitivity (10).

The natural selection measured by either the 'BLS' or our 'PhyloP' feature is remarkably well correlated with the strength of repression: selected target sites are also sites of stronger repression. It is also known that older miRNAs have higher expression levels (42). Natural selection is acting on both the miRNA expression level and the repression strength to maximize the repression efficiency. Furthermore, a correlation between the mRNA accessibility and the target site conservation has been shown in *Drosophila* (43) which can partially explain the good performance of the accessibility features ('ΔG open' and 'AU content') as this parameter is naturally selected. This dependence among the features partially explains why their individual performance is not additive in the global model. The probabilistic features also correlate with the conservation features but they are usually outperformed by the conservation features, even if they sometimes have similar performance (e.g. the probabilistic features are similar to the 'BLS' performance for the 'Trans.Grimson' dataset). In terms of computation, and more importantly of input data (multiple alignments, etc.), the probabilistic features are undoubtedly less expensive than the conservation features. They can therefore be seen as an alternative to an evolutionary approach, especially for organisms with long 3′-UTRs [between *Drosophila* and human their accuracy significantly drops in *Drosophila* (18)].

While we observed generally consistent results among the transcriptomics, polysome fractionation and proteomics experimental methods, they were distinguishable from IP experiments. The experimental methods measuring the repression, i.e. the effect of the miRNA, are more accurate to measure the repression strength than methods measuring only miRNA binding. Chi *et al.* (9) observed that 86% of conserved miR-124 seeds were

Downloaded from http://nar.oxfordjournals.org/ by Charles Vejnar on October 15, 2012

present within the Ago footprint region, i.e. the HITS-CLIP method accurately identifies miRNA target sites but does not provide a quantitative measure of miRNA repression. We also noticed that, although polysome fractionation is not commonly used to test miRNA targets, the ribosome number measure performs as well as most other methods.

According to our model, a large part of the variance of the miRNA repression observed from experimental measurements remains to be explained. Indeed, the overall variance includes miRNA indirect effects, such as regulation feedback loops. The proportions of variance explained by our model or TargetScan are therefore underestimates of the explainable variance by miRNA direct repression. An improved understanding of the molecular mechanisms of repression, beyond the currently considered thermodynamic, evolutionary, probabilistic or sequence-based aspects will undoubtedly lead to better predictions. Nevertheless, our model shows that capturing more information with complementary features already significantly improves the predictive power. Additional considerations may extend these improvements. For example, our 'PhyloP' feature is based on a 'per base' model, i.e. positions in the alignment are considered independently. However, for RNAs in general, stacking energies are important, so a context-dependent model, when integrated in PHAST (33), should increase performance and would also quantify the importance of stacking energies. Other considerations are, for the moment, less tractable, e.g. taking the kinetics of the repression into account. The availability of the different components, such as enzymes, miRNAs and mRNAs is ignored in the existing models. However, this system approach requires substantially more information, notably the concentration of different components.

The miRmap library implements 11 features from 4 categories, making it currently the most comprehensive miRNA target prediction resource. All the features and the model evaluated in this study are available as an open-source Python library on a public revision control service, allowing tracking of all contributions. As such, miRmap establishes a solid foundation for the future development of approaches to miRNA target prediction, facilitating meaningful comparisons between existing and new features, and providing the community with direct access to state-of-the-art analytical tools.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2 and Supplementary Figures 1–8.

ACKNOWLEDGEMENTS

We thank the authors of the experimental studies used in this article for making their data and detailed annotation available. We also thank Melissa J. Hubisz for her help to create a Python interface for PhyloP and Grégory Nuel for his help with Spatt. We are grateful to Robert M. Waterhouse, Thomas J. Petty and other member of the Zdobnov group for comments and corrections on the article.

FUNDING

Swiss National Science Foundation [PDFMA3-118375 and 31003A-125350]. Funding for open access charge: Swiss Institute of Bioinformatics.

Conflict of interest statement. None declared.

REFERENCES

1. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
2. Muljo,S.A., Kanellopoulou,C. and Aravind,L. (2010) MicroRNA targeting in mammalian genomes: genes and mechanisms. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, **2**, 148–161.
3. Axtell,M.J., Westholm,J.O. and Lai,E.C. (2011) Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol.*, **12**, 221.
4. Brennecke,J., Stark,A., Russell,R.B. and Cohen,S.M. (2005) Principles of microRNA-target recognition. *PLoS Biol.*, **3**, e85.
5. Didiano,D. and Hobert,O. (2006) Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat. Struct. Mol. Biol.*, **13**, 849–851.
6. Grimson,A., Farh,K.K., Johnston,W.K., Garrett-Engele,P., Lim,L.P. and Bartel,D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, **27**, 91–105.
7. Selbach,M., Schwanhäusser,B., Thierfelder,N., Fang,Z., Khanin,R. and Rajewsky,N. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**, 58–63.
8. Baek,D., Villén,J., Shin,C., Camargo,F.D., Gygi,S.P. and Bartel,D.P. (2008) The impact of microRNAs on protein output. *Nature*, **455**, 64–71.
9. Chi,S.W., Zang,J.B., Mele,A. and Darnell,R.B. (2009) Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps. *Nature*, **460**, 479–486.
10. Ellwanger,D.C., Büttner,F.A., Mewes,H. and Stümpflen,V. (2011) The sufficient minimal set of miRNA seed types. *Bioinformatics (Oxford, England)*, **27**, 1346–1350.
11. Yue,D., Liu,H. and Huang,Y. (2009) Survey of computational algorithms for microRNA target prediction. *Curr. Genomics*, **10**, 478–492.
12. Kertesz,M., Iovino,N., Unnerstall,U., Gaul,U. and Segal,E. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.
13. Krek,A., Grün,D., Poy,M.N., Wolf,R., Rosenberg,L., Epstein,E.J., MacMenamin,P., da Piedade,I., Gunsalus,K.C., Stoffel,M. *et al.* (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.
14. John,B., Enright,A.J., Aravin,A., Tuschl,T., Sander,C. and Marks,D.S. (2004) Human microRNA targets. *PLoS Biol.*, **2**, e363.
15. Rehmsmeier,M., Steffen,P., Hochsmann,M. and Giegerich,R. (2004) Fast and effective prediction of microRNA/target duplexes. *RNA*, **10**, 1507–1517.
16. Kiriakidou,M., Nelson,P.T., Kouranov,A., Fitziev,P., Bouyioukos,C., Mourelatos,Z. and Hatzigeorgiou,A. (2004) A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.*, **18**, 1165–1178.
17. Gaidatzis,D., van Nimwegen,E., Hausser,J. and Zavolan,M. (2007) Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics*, **8**, 69.
18. Marín,R.M. and Vanícek,J. (2011) Efficient use of accessibility in microRNA target prediction. *Nucleic Acids Res.*, **39**, 19–29.
19. Friedman,R.C., Farh,K.K., Burge,C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.

Downloaded from http://nar.oxfordjournals.org/ by Charles Vejnar on October 15, 2012

20. Rajewsky,N. (2006) MicroRNA target predictions in animals. *Nat. Genet.*, **38(Suppl.)**, S8–S13.
21. Hausser,J., Landthaler,M., Jaskiewicz,L., Gaidatzis,D. and Zavolan,M. (2009) Relative contribution of sequence and structure features to the mRNA binding of Argonaute/EIF2C-miRNA complexes and the degradation of miRNA targets. *Genome Res.*, **19**, 2009–2020.
22. Betel,D., Koppal,A., Agius,P., Sander,C. and Leslie,C. (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.*, **11**, R90.
23. Wen,J., Parker,B.J., Jacobsen,A. and Krogh,A. (2011) MicroRNA transfection and AGO-bound CLIP-seq data sets reveal distinct determinants of miRNA action. *RNA*, **17**, 820–834.
24. Linsley,P.S., Schelter,J., Burchard,J., Kibukawa,M., Martin,M.M., Bartz,S.R., Johnson,J.M., Cummins,J.M., Raymond,C.K., Dai,H. *et al.* (2007) Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol. Cell. Biol.*, **27**, 2240–2252.
25. Hendrickson,D.G., Hogan,D.J., McCullough,H.L., Myers,J.W., Herschlag,D., Ferrell,J.E. and Brown,P.O. (2009) Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS Biol.*, **7**, e1000238.
26. Pruitt,K.D., Tatusova,T., Klimke,W. and Maglott,D.R. (2009) NCBI reference sequences: current status, policy and new initiatives. *Nucleic Acids Res.*, **37**, D32–D36.
27. Fujita,P.A., Rhead,B., Zweig,A.S., Hinrichs,A.S., Karolchik,D., Cline,M.S., Goldman,M., Barber,G.P., Clawson,H., Coelho,A. *et al.* (2011) The UCSC genome browser database: update 2011. *Nucleic Acids Res.*, **39**, D876–D882.
28. Kozomara,A. and Griffiths-Jones,S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, **39**, D152–D157.
29. Hofacker,I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429–3431.
30. Parker,J.S., Roe,S.M. and Barford,D. (2005) Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature*, **434**, 663–666.
31. Nuel,G., Regad,L., Martin,J. and Camproux,A. (2010) Exact distribution of a pattern in a set of random sequences generated by a markov source: applications to biological data. *Algorithms Mol. Biol.*, **5**, 15.
32. Stark,A., Lin,M.F., Kheradpour,P., Pedersen,J.S., Parts,L., Carlson,J.W., Crosby,M.A., Rasmussen,M.D., Roy,S., Deoras,A.N. *et al.* (2007) Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature*, **450**, 219–232.
33. Pollard,K.S., Hubisz,M.J., Rosenbloom,K.R. and Siepel,A. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.*, **20**, 110–121.
34. Sukumaran,J. and Holder,M.T. (2010) DendroPy: a python library for phylogenetic computing. *Bioinformatics*, **26**, 1569–1571.
35. Papaioannou,M.D., Lagarrigue,M., Vejnar,C.E., Rolland,A.D., Kühne,F., Aubry,F., Schaad,O., Fort,A., Descombes,P., Neerman-Arbez,M. *et al.* (2011) Loss of Dicer in sertoli cells has a major impact on the testicular proteome of mice. *Mol. Cell. Proteomics*, **10**, M900587MCP200.
36. Zuber,V. and Strimmer,K. (2011) High-dimensional regression and variable selection using CAR scores. *Stat. Appl. Genet. Mol. Biol.*, **10**, 1–27.
37. Lindblad-Toh,K., Garber,M., Zuk,O., Lin,M.F., Parker,B.J., Washietl,S., Kheradpour,P., Ernst,J., Jordan,G., Mauceli,E. *et al.* (2011) A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, **478**, 476–482.
38. Guo,H., Ingolia,N.T., Weissman,J.S. and Bartel,D.P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, **466**, 835–840.
39. Huntzinger,E. and Izaurralde,E. (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat. Rev. Genet.*, **12**, 99–110.
40. Baek,D., Villén,J., Shin,C., Camargo,F.D., Gygi,S.P. and Bartel,D.P. (2008) The impact of microRNAs on protein output. *Nature*, **455**, 64–71.
41. Shin,C., Nam,J., Farh,K.K., Chiang,H.R., Shkumatava,A. and Bartel,D.P. (2010) Expanding the microRNA targeting code: functional sites with centered pairing. *Mol. Cell*, **38**, 789–802.
42. Lu,J., Shen,Y., Wu,Q., Kumar,S., He,B., Shi,S., Carthew,R.W., Wang,S.M. and Wu,C. (2008) The birth and death of microRNA genes in *Drosophila*. *Nat. Genet.*, **40**, 351–355.
43. Chen,K., Maaskola,J., Siegal,M.L. and Rajewsky,N. (2009) Reexamining microRNA site accessibility in *Drosophila*: a population genomics study. *PLoS One*, **4**, e5681.

Downloaded from http://nar.oxfordjournals.org/ by Charles Vejnar on October 15, 2012

miRmap: Comprehensive prediction of microRNA target repression strength

Supplementary information

The source code and documentation of the miRmap Python library is available online at:

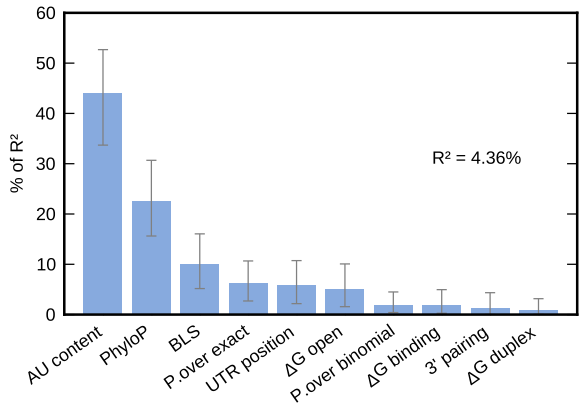
http://cegg.unige.ch/mirmap

		ΔG duplex	ΔG binding	ΔG open	ΔG total	P.over binomial	P.over exact	BLS	PhyloP	AU content	UTR position	3′ pairing
ΔG duplex	r	1.000	0.962	-0.456	0.222	0.101	0.097	0.069	-0.118	0.507	0.029	-0.263
	p	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00
ΔG binding	r	0.962	1.000	-0.416	0.241	0.102	0.098	0.065	-0.110	0.464	0.026	-0.260
	p	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00
ΔG open	r	-0.456	-0.416	1.000	0.725	-0.024	-0.037	-0.058	0.142	-0.635	0.019	0.041
	p	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00
ΔG total	r	0.222	0.241	0.725	1.000	0.058	0.037	-0.005	0.070	-0.317	0.034	-0.157
	p	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	2.78e-27	0.00e+00	0.00e+00	0.00e+00	0.00e+00
P.over binomial	r	0.101	0.102	-0.024	0.058	1.000	0.806	0.405	-0.002	0.044	0.451	-0.003
	p	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	1.05e-06	0.00e+00	0.00e+00	4.73e-11
P.over exact	r	0.097	0.098	-0.037	0.037	0.806	1.000	0.151	-0.060	0.051	0.256	-0.005
	p	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	9.41e-27
BLS	r	0.069	0.065	-0.058	-0.005	0.405	0.151	1.000	-0.386	0.079	0.232	-0.001
	p	0.00e+00	0.00e+00	0.00e+00	2.78e-27	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	1.99e-02
PhyloP	r	-0.118	-0.110	0.142	0.070	-0.002	-0.060	-0.386	1.000	-0.186	0.138	0.011
	p	0.00e+00	0.00e+00	0.00e+00	0.00e+00	1.05e-06	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	7.86e-114
AU content	r	0.507	0.464	-0.635	-0.317	0.044	0.051	0.079	-0.186	1.000	0.044	-0.062
	p	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00
UTR position	r	0.029	0.026	0.019	0.034	0.451	0.256	0.232	0.138	0.044	1.000	-0.000
	p	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00	7.35e-01
3′ pairing	r	-0.263	-0.260	0.041	-0.157	-0.003	-0.005	-0.001	0.011	-0.062	-0.000	1.000
	p	0.00e+00	0.00e+00	0.00e+00	0.00e+00	4.73e-11	9.41e-27	1.99e-02	7.86e-114	0.00e+00	7.35e-01	0.00e+00

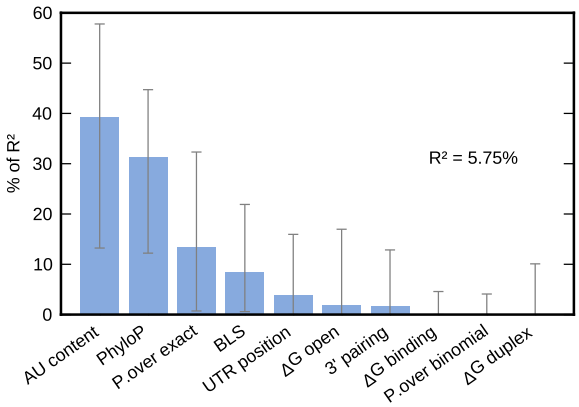
Supplementary Table 1 Correlation among features for human miRNAs and mRNAs. Each cell contains the Spearman rank correlation coefficient (r) and the corresponding p-value (p).

		Trans.Grimson	Trans.Linsley	Prot.Selbach	IPcross.Chi	IP.Hendrickson	RibN.Hendrickson	Trans.Hendrickson
AU content	r	-0.229	-0.166	-0.163	0.160	0.249	-0.228	-0.293
AU content	p	1.17e-48	1.92e-57	1.70e-06	2.51e-05	8.02e-10	1.86e-08	3.47e-13
AU content	r′	-0.967	-0.868	-0.922	0.944	0.994	-0.982	-0.986
AU content	p′	4.68e-09	2.72e-05	8.97e-03	4.68e-03	5.13e-05	4.79e-04	2.87e-04
UTR position	r	0.117	0.060	0.064	-0.022	-0.278	0.125	0.206
UTR position	p	1.39e-13	8.33e-09	6.26e-02	5.63e-01	5.18e-12	2.20e-03	3.92e-07
UTR position	r′	0.880	0.643	0.612	-0.178	-0.921	0.774	0.804
UTR position	p′	1.51e-05	9.79e-03	1.96e-01	7.36e-01	9.02e-03	7.09e-02	5.36e-02
3′ pairing	r	-0.051	-0.041	-0.058	0.030	0.099	-0.043	-0.028
3′ pairing	p	1.16e-03	7.49e-05	9.13e-02	4.26e-01	1.62e-02	2.95e-01	4.90e-01
3′ pairing	r′	-0.493	-0.764	-0.757	0.167	0.607	-0.441	-0.269
3′ pairing	p′	6.20e-02	9.05e-04	8.17e-02	7.52e-01	2.02e-01	3.81e-01	6.06e-01
ΔG duplex	r	0.000	-0.062	-0.017	0.017	-0.002	-0.059	-0.055
ΔG duplex	p	9.87e-01	3.34e-09	6.13e-01	6.57e-01	9.53e-01	1.49e-01	1.79e-01
ΔG duplex	r′	0.039	-0.657	-0.167	-0.009	0.086	-0.521	-0.430
ΔG duplex	p′	8.89e-01	7.75e-03	7.52e-01	9.86e-01	8.71e-01	2.89e-01	3.94e-01
ΔG binding	r	0.023	-0.050	0.001	0.020	-0.007	-0.042	-0.027
ΔG binding	p	1.43e-01	1.70e-06	9.69e-01	5.99e-01	8.63e-01	3.11e-01	5.05e-01
ΔG binding	r′	0.320	-0.598	0.198	0.184	0.054	-0.563	-0.272
ΔG binding	p′	2.45e-01	1.85e-02	7.06e-01	7.27e-01	9.20e-01	2.44e-01	6.02e-01
ΔG open	r	0.196	0.107	0.085	-0.108	-0.234	0.194	0.257
ΔG open	p	4.42e-36	1.61e-24	1.31e-02	4.47e-03	7.40e-09	1.82e-06	1.95e-10
ΔG open	r′	0.950	0.788	0.939	-0.914	-0.875	0.949	0.947
ΔG open	p′	6.23e-08	4.79e-04	5.56e-03	1.09e-02	2.25e-02	3.84e-03	4.12e-03
ΔG total	r	0.204	0.086	0.078	-0.096	-0.249	0.183	0.251
ΔG total	p	8.31e-39	1.95e-16	2.28e-02	1.18e-02	7.02e-10	7.08e-06	5.09e-10
ΔG total	r′	0.939	0.833	0.735	-0.802	-0.824	0.865	0.940
ΔG total	p′	2.27e-07	1.17e-04	9.57e-02	5.51e-02	4.39e-02	2.60e-02	5.21e-03
P.over exact	r	0.170	0.065	0.073	0.015	-0.222	0.076	0.221
P.over exact	p	2.93e-27	4.12e-10	3.32e-02	7.02e-01	4.75e-08	6.45e-02	5.27e-08
P.over exact	r′	0.920	0.745	0.687	0.069	-0.875	0.914	0.837
P.over exact	p′	1.25e-06	1.45e-03	1.31e-01	8.97e-01	2.25e-02	1.07e-02	3.76e-02
P.over binomial	r	0.147	0.067	0.069	0.029	-0.207	0.085	0.188
P.over binomial	p	1.24e-20	1.08e-10	4.30e-02	4.40e-01	3.74e-07	3.84e-02	3.91e-06
P.over binomial	r′	0.867	0.634	0.702	0.329	-0.826	0.677	0.860
P.over binomial	p′	2.84e-05	1.12e-02	1.20e-01	5.24e-01	4.29e-02	1.40e-01	2.80e-02
BLS	r	-0.158	-0.108	-0.124	0.112	0.148	-0.160	-0.226
BLS	p	9.93e-24	2.99e-25	2.78e-04	3.36e-03	2.96e-04	8.95e-05	2.66e-08
BLS	r′	-0.954	-0.815	-0.931	0.893	0.661	-0.817	-0.883
BLS	p′	3.52e-08	2.14e-04	6.97e-03	1.66e-02	1.53e-01	4.70e-02	1.97e-02
PhyloP	r	-0.205	-0.143	-0.172	0.196	0.116	-0.155	-0.223
PhyloP	p	3.00e-39	2.97e-43	4.49e-07	2.22e-07	4.74e-03	1.49e-04	4.12e-08
PhyloP	r′	-0.949	-0.857	-0.946	0.856	0.719	-0.918	-0.929
PhyloP	p′	6.82e-08	4.46e-05	4.31e-03	2.96e-02	1.07e-01	9.83e-03	7.34e-03
PhastCons	r	-0.205	-0.128	-0.143	0.223	0.020	-0.093	-0.124
PhastCons	p	3.26e-39	6.51e-35	2.69e-05	3.22e-09	6.20e-01	2.35e-02	2.38e-03
PhastCons	r′	-0.920	-0.823	-0.975	0.905	0.157	-0.704	-0.690
PhastCons	p′	1.24e-06	1.67e-04	9.59e-04	1.31e-02	7.67e-01	1.18e-01	1.30e-01

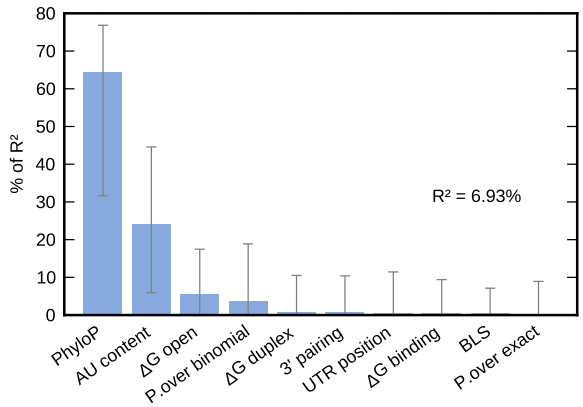
Supplementary Table 2 Correlation between each feature and the seven experimental miRNA repression measures. Data points were binned in 15 equally-sized bins. r is the correlation on the full dataset (p: corresponding p-value); r' is the correlation on the binned dataset (p': corresponding p-value.



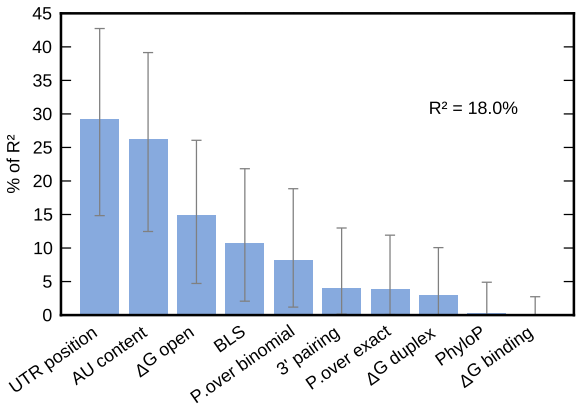
Linsley



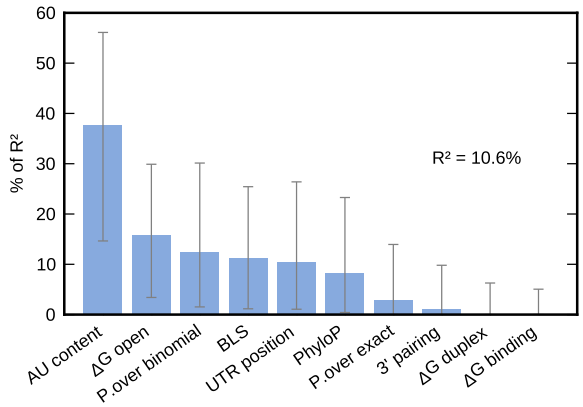
Selbach



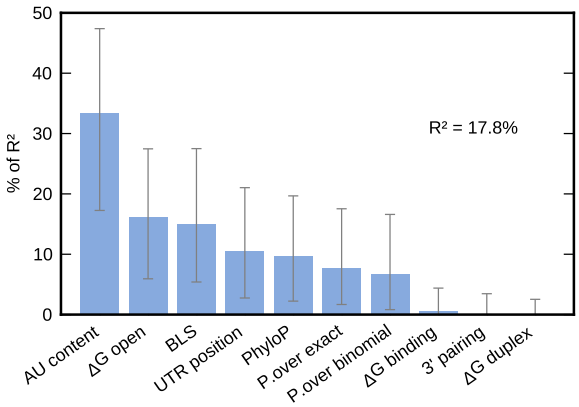
Chi



Hendrickson IP

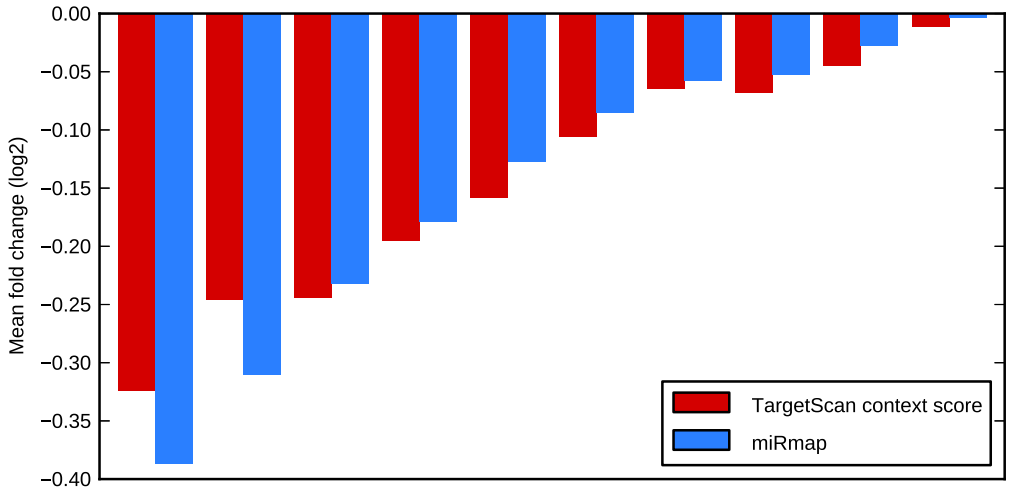


Hendrickson Ribosome-number

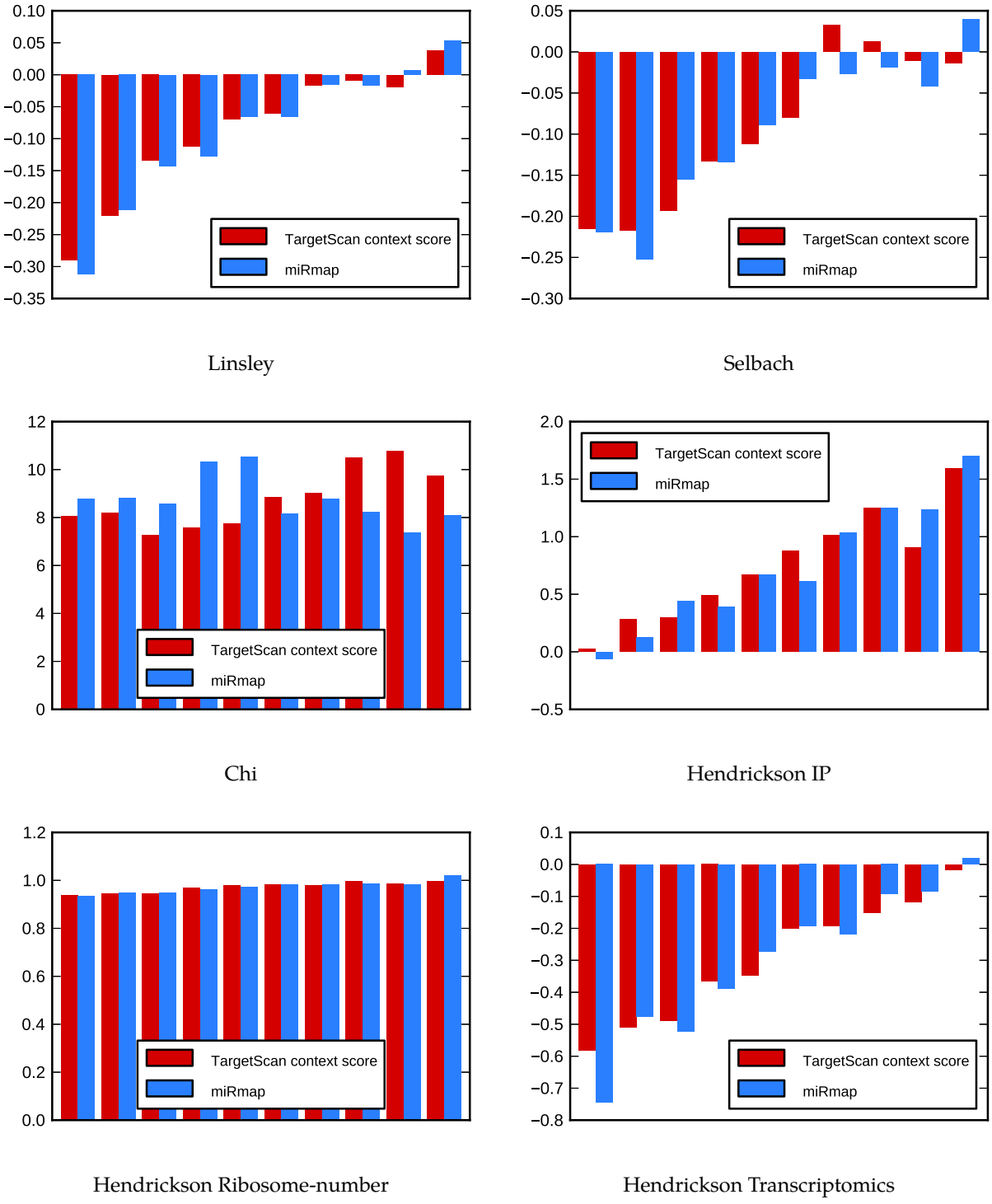


Hendrickson Transcriptomics

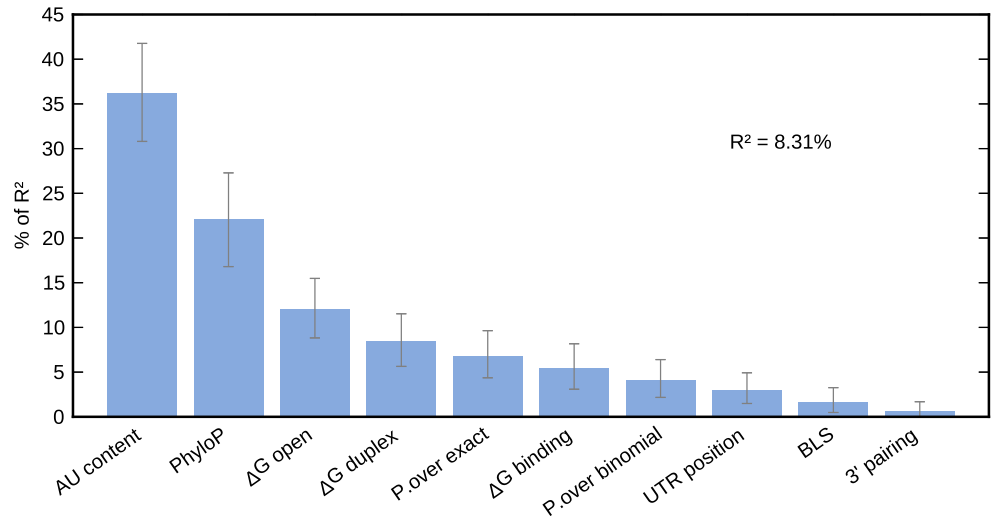
Supplementary Figure 1 Feature relative importance in the miRmap multiple linear regression models predicting miRNA repression strength for all datasets used in this study considering 7-mer seeds. R^2 is the proportion of variance explained by the model.



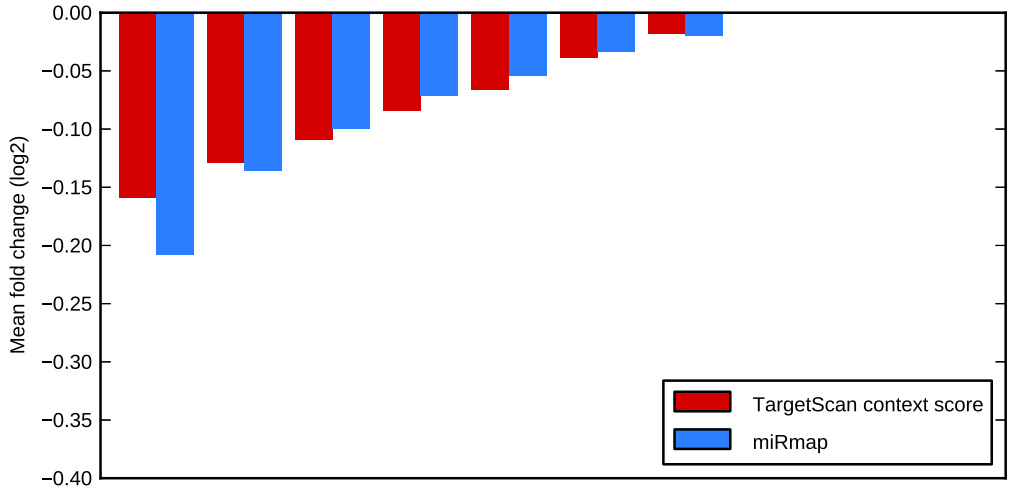
Supplementary Figure 2 Predictive performance of miRmap and TargetScan context score models for 7-mer seeds. All predicted miRNA-mRNA pairs were ranked according to their predicted strength and binned into 10 equally-sized bins. For each bin, the average experimental fold-change was computed. The same procedure was repeated with the miRmap and TargetScan context score models. (Trans.Grimson dataset)



Supplementary Figure 3 Predictive performance of miRmap and TargetScan context score models for 7-mer seeds. All predicted miRNA-mRNA pairs were ranked according to their predicted strength and binned into 10 equally-sized bins. For each bin, the average experimental fold-change was computed. The same procedure was repeated with the miRmap and TargetScan context score models.

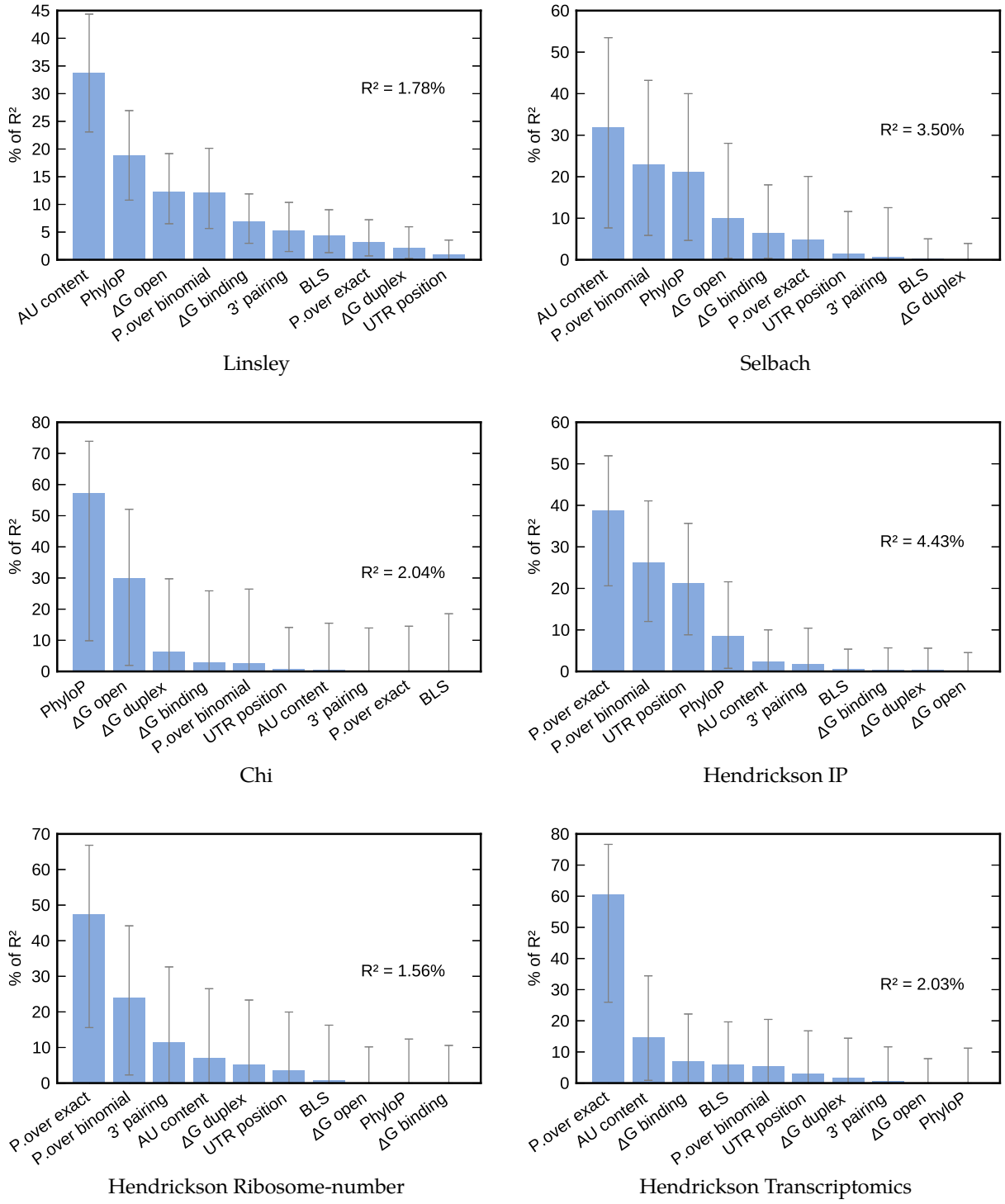


A

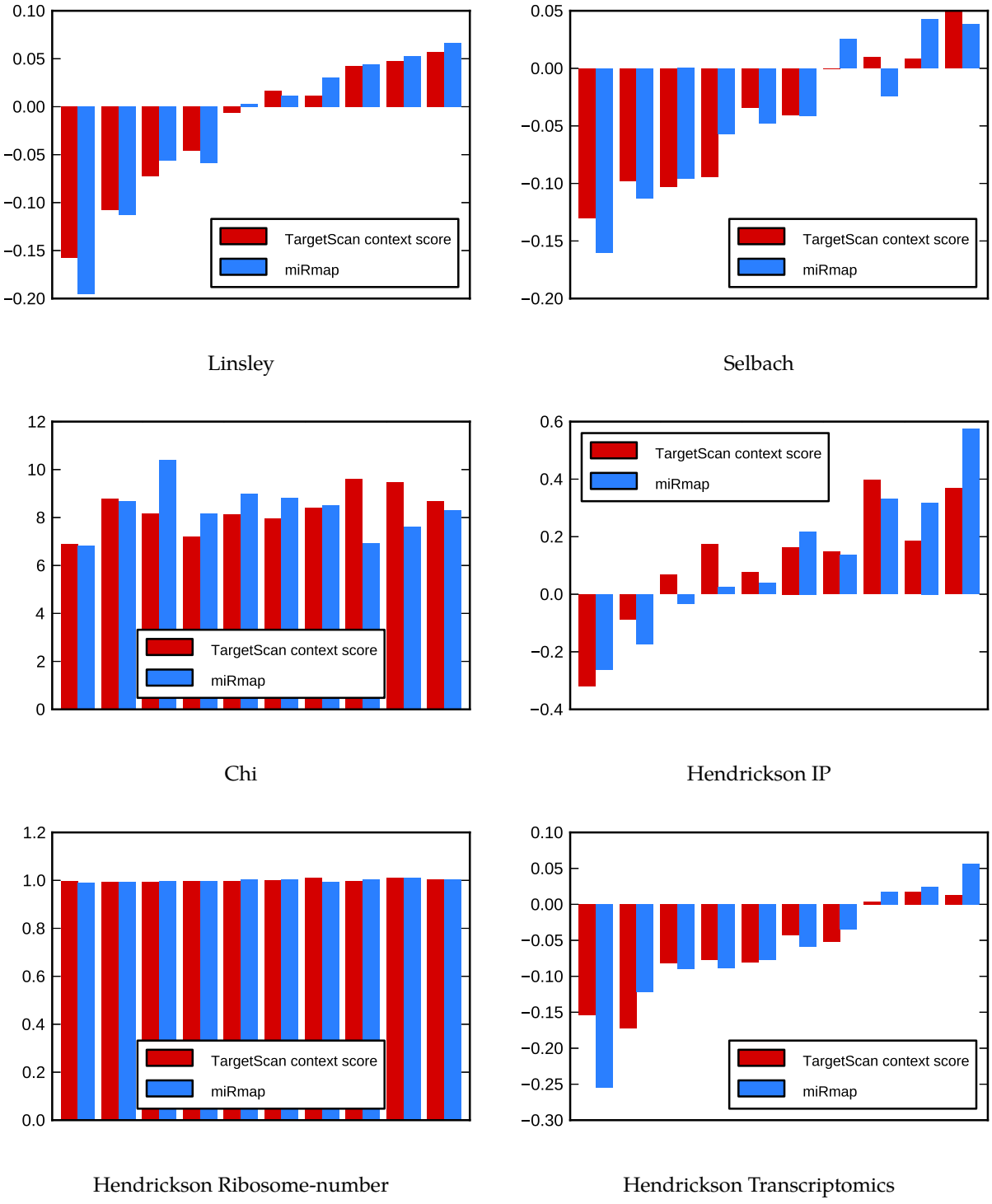


B

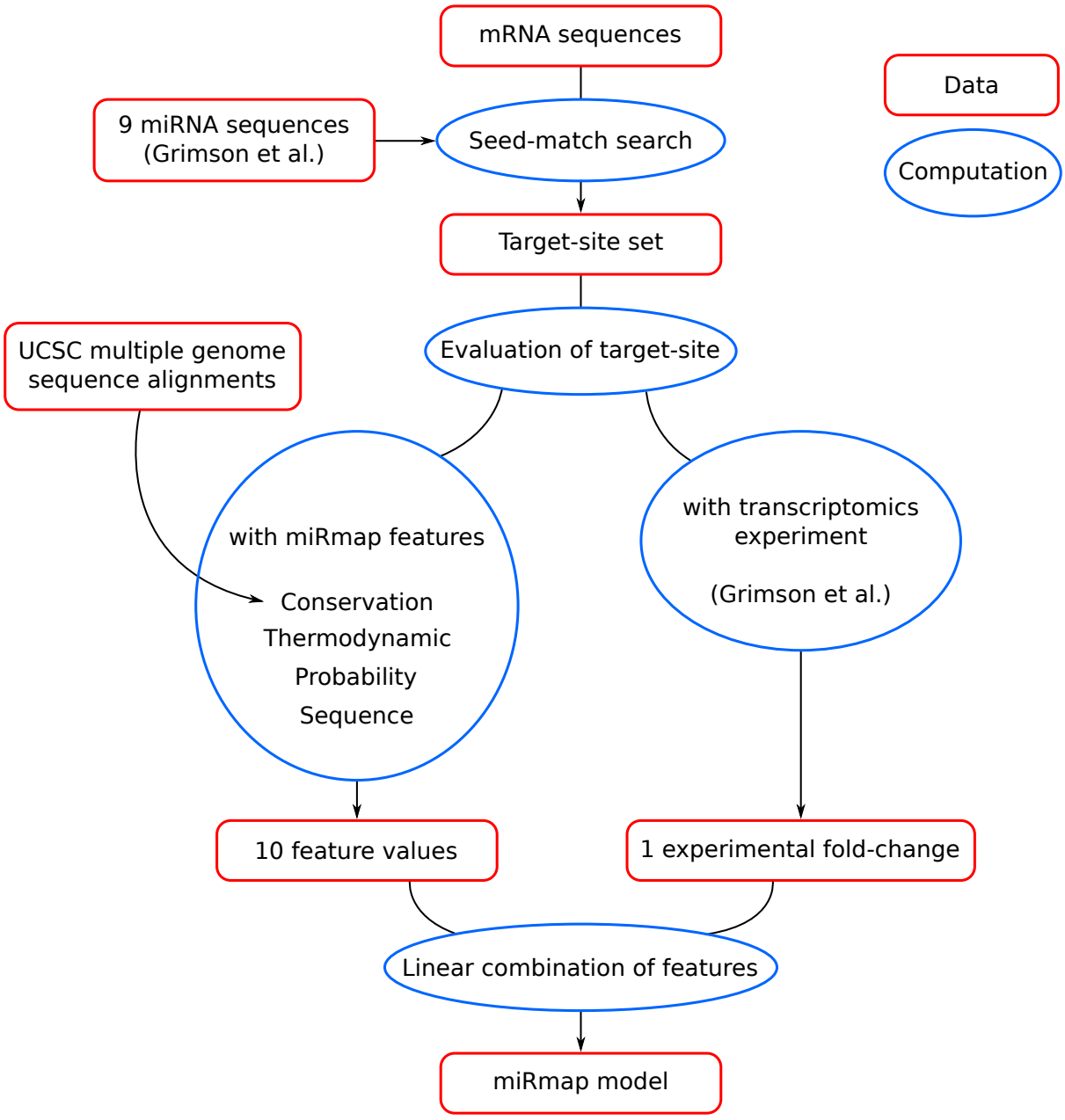
Supplementary Figure 4 Performance and feature relative importance for 6-mer seeds. The same procedures as in Figure 5B (A) and in Supplementary Figure 3 (B) were applied to the set of potential target sites defined by the presence of a single 6-mer seed-match. (Trans.Grimson dataset)



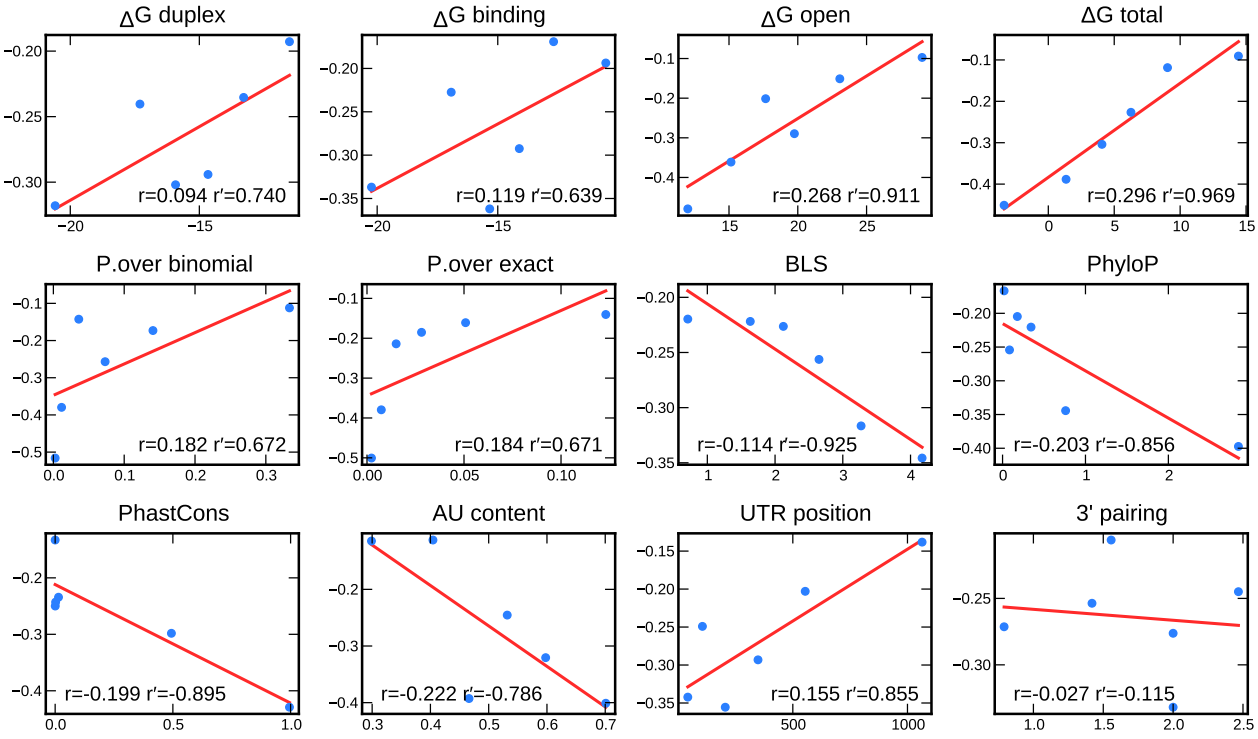
Supplementary Figure 5 Feature relative importance in the miRmap multiple linear regression models predicting miRNA repression strength for all datasets used in this study considering 6-mer seeds. R^2 is the proportion of variance explained by the model.



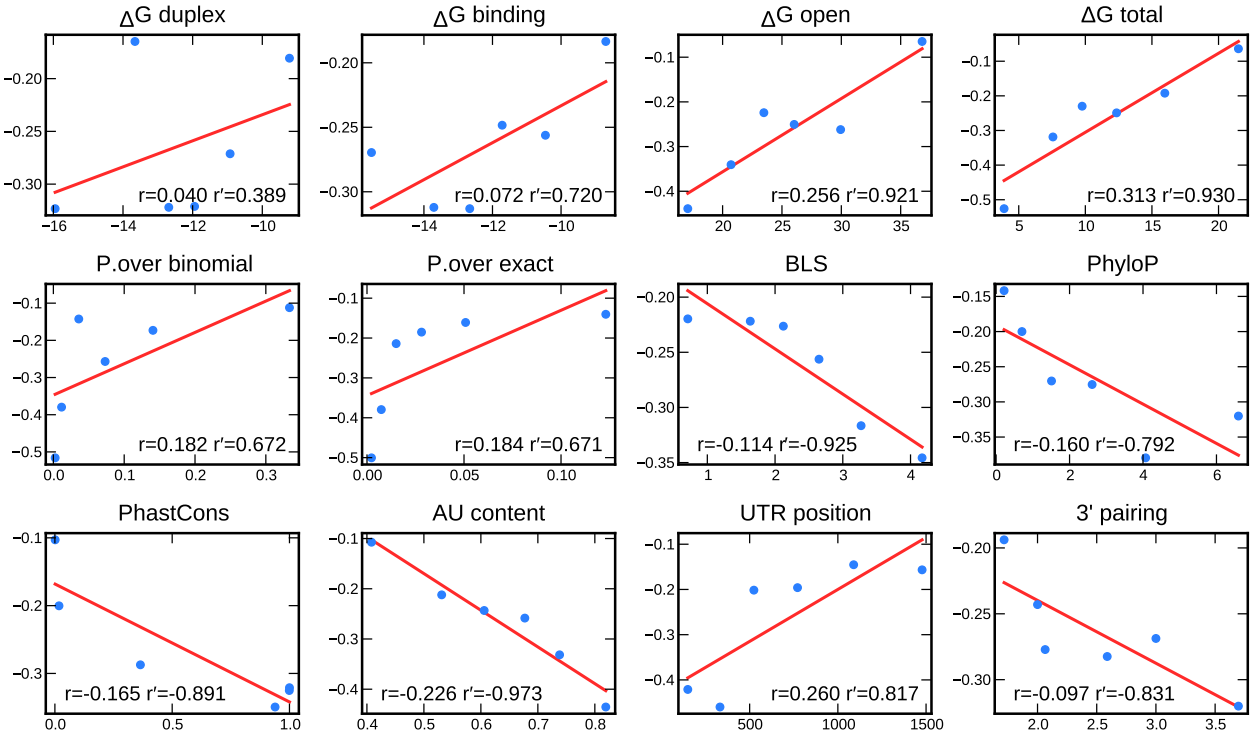
Supplementary Figure 6 Predictive performance of miRmap and TargetScan context score models for 6-mer seeds. All predicted miRNA-mRNA pairs were ranked according to their predicted strength and binned into 10 equally-sized bins. For each bin, the average experimental fold-change was computed. The same procedure was repeated with the miRmap and TargetScan context score models.



Supplementary Figure 7 Description of the miRmap model building. After the target site identification with seed-matches, their repression strength is evaluated computationally and experimentally. The linear combination of the features based on the experimental output is the miRmap model.

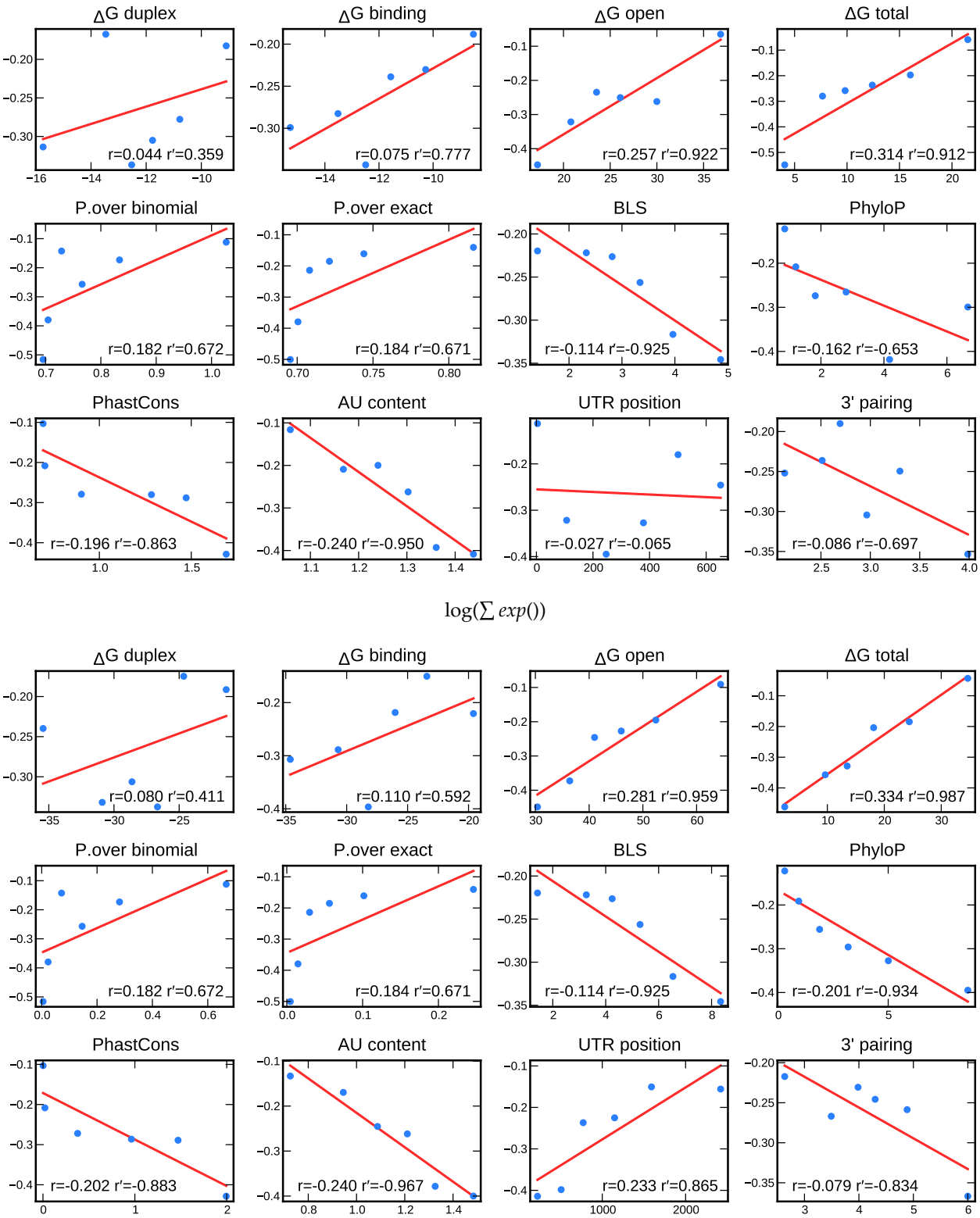


min()



max()

Supplementary Figure 8 (see next page for legend)



Supplementary Figure 8 Correlation similar to Figure 3 for mRNAs with 2 target sites (*Trans.Grimson* dataset) combined with 4 different functions, shown at the bottom of each picture block.

5 DISCUSSION

The miRmap library (Vejnar and Zdobnov [16]) predicts the strength of miRNA repression on targeted mRNA with a linear model composed of eleven features that comprehensively cover thermodynamic, probabilistic, evolutionary, and sequence-based approaches. The strength of miRNA repression is a biologically meaningful criterion to rank potential miRNA-mRNA pairs. Three features included in the library were novel methods. The individual features and the model combining them were evaluated using high-throughput experimental data addressing both miRNA-mediated regulation, *i.e.* mRNA destabilization and translational repression. These publicly available datasets include immunopurification, transcriptomics, proteomics, and polysome fractionation experiments. While the large size of these datasets is beneficial for my analysis, most of them are based on artificially over-expressed miRNAs in cell lines. During the development of the miRmap library, several features were evaluated on more realistic systems, in particular two knock-down experiments using naturally highly expressed miRNAs in mouse tissues. Target site enrichments in the up-regulated gene fraction due to the miRNA knock-down were indeed significant for miR-122 in hepatocytes and miR-155 in dendritic cells (see Results). Controls confirming the enrichment were also performed via enrichment in the opposite fraction (down-regulated), and via enrichment for all miRNAs. The overall predictive power of the miRmap model appears to almost double that of the most renowned TargetScan software, and outperform PITA and PACMIT that are single and double features tools respectively (see **Section 3.2**).

5.1 Enhancing the quality of bioinformatics tools

The miRmap library is, in the first place, a software library written in the Python language. The computing efficiency and quick prototyping oriented the choice towards Python, but also the ease (i) to interface with pre-existing C libraries, and (ii) to scale to multiple genome-wide prediction runs.

Indeed, to compute the thermodynamic features, evolutionary features and one probabilistic feature, C libraries were used (Hofacker [76], and Nuel et al. [77]). Their size and complexity prohibited their reimplementation in Python. I implemented interfaces using the *ctypes* module written in pure Python. Writing such interfaces has many advantages. First, not only they are by themselves of general interest to the bioinformatics community, but they also provided me with a deeper understanding of the libraries I was writing an interface for. Second, while these libraries also have command line interfaces, some options, and even computation, were not available through the command-line interface (the components of ΔG *binding* for instance). Third, computing times are shorter when linking directly to a library compared to a command-line program. The main reason for this difference in computing time is the cost of forking new processes and the consequent inputs/outputs. This cost is highly dependent on the library, and the number of times the library is called in the global process. For example, it is about 10% for a 100000 sequence sets folded with RNAfold (Hofacker [76]). If the overall processing requires a single call to the C library, the cost of a single fork would not be significant to the overall processing time. While the performance difference is not always big, the CPU usage profile is very different. Indeed, during the test described above, for each sequence a new process was created with inputs and outputs resulting in multiple small tasks distributed on multiple CPUs (but still one at a time) instead

of one task using 100% of a single CPU. Fourth, related to the previous point, all the necessary computation encapsulated in a single process makes program parallelelization easier. With the *multiprocessing* Python module, I implemented a parallelelized version of miRmap, making the computing time of all target site strengths for a single miRNA on a 48 CPU-core machine a matter of a few minutes instead of hours.

Contrary to most miRNA target prediction tools (see introduction), miRmap was made available as an open source software. It is also designed to ease the extension, making an effort towards more integration of the different miRNA target prediction tools. Moreover, the source code is distributed on a cooperative development platform. These platforms, such as Bitbucket or Github, promote a novel process for open source software allowing anyone to modify the code, in turn making any modifications easily accessible on the platform, and to request integration of the changes in the main code by a so-called "pulling request". While widely adopted by the computer science community, this development model is still rarely used for bioinformatics projects. I hope to promote the benefits of using these platforms with my miRNA target prediction software.

The miRmap library provides an Application Programming Interface (API) usable inside other programs that predict miRNA targets. While an example of such a program is provided in the miRmap distribution, it requires the installation of the library. This requirement, despite being small, is too large for a few queries or even small projects. To address this type of user need, I am developing a public API, making miRmap computation accessible through a simple HTTP query. The computation will be done on our server. Based on this public API, I am also developing a web interface usable to browse already computed predictions for known miRNAs and genes, and also to predict targets on user submitted sequences.

5.2 Enhancing the quality of miRNA target predictions

As about 85% of the variance of the transcriptomics data used to parameterize the miRmap model is still unexplained (Vejnar and Zdobnov [16]), the performance of miRmap and miRNA target prediction software in general can be largely improved. Improvements are possible for the target recognition model but also at a higher level, for instance the mRNA level. It is worth noting that indirect effects are included in the global variance: miRNA direct effects will not be able to explain 100% of the variance.

The target recognition relies in particular on the thermodynamic energies used to compute RNA folding structures. The principles and parameters of these techniques are based on a small molecule, a few base-pairs, which are then extrapolated to longer RNAs. New software integrating high-throughput methods to fold RNAs are starting to emerge (Quarrier et al. [78]) and can help in obtaining better models of RNA structures (Reviewed in Wan et al. [79]). However, proteins are ignored by these methods. First, the RISC can change the stability of the miRNA-mRNA duplex. The energies for RNA duplex in the context of a RISC are unknown. Second, in the extreme case, an RNA-binding protein (RBP) can mask the miRNA binding site completely, for example PUM1 for miR-221 and miR-222 (Kedde and Agami [80] and Kedde et al. [81]).

Improvement in the assessment of miRNA target site conservation is also possible. Indeed, PhyloP (Pollard et al. [82]) uses "per base" evolutionary models, *i.e.* positions in the alignment are considered independently. However, for RNAs in general, stacking energies are important: a context dependent model that considers positions non-independently, when integrated in PHAST (Pollard et al. [82]), should increase the performance of the *PhyloP* feature of miRmap. Interestingly, it would also quantify the importance of stacking energies.

Conversely, the detection of naturally selected RNA structure would be more accurate, opening large scale studies to detect these structures with more confidence.

The variance in expression fold-changes is large in miRNA knock-down transcriptomics experiments. My model explains part of this variance, but a large proportion remains to be explained. Increasing that proportion could be achieved with improved features, in particular with the few leads described above. Nevertheless, experimental noise and miRNA indirect effects undoubtedly remain explanatory factors. More specific experiments are therefore required to increase miRNA prediction tools performance. As luciferase assays are the reference experiment used in experimental biology, attempts were made to use them in prediction tools (see introduction). However, a limited number of such experiments are available in general, but in addition, there are almost as many experiments as experimentalists, inducing a high experimental noise over the full dataset. Moreover the stability and turn-over of the luciferase proteins is not constant enough between different experiments to consistently predict the strength of the repression, parameter I am aiming to predict. Recently, immunopurification experiments, preceded with cross-linking Ago to the neighbor mRNA followed by RNA-Seq (Chi et al. [42] and Hafner et al. [43]), have shown high sensitivity. Chi et al. [42] observed that 86% of conserved miR-124 seeds were present within the Ago footprint region, *i.e.* the HITS-CLIP method used by the authors accurately identifies miRNA target sites. However, I have shown this method is not as reliable to predict the repression strength (Vejnar and Zdobnov [16]). While the two experiments published are relevant, HITS-CLIP experiments should be performed either on cells with a predominantly expressed miRNA (hepatocyte for instance) or on cell lines with induced miRNA expression to be able to predict miRNA repression strength, as such an experimental setup would reduce indirect regulation (At least 20 miRNAs are expressed in relevant level in Chi et al. [42]).

The target recognition step is only the first level determining miRNA repression. Indeed, mRNA can have multiple target sites for one or more miRNA(s). With the available experimental data, the effects of multiple target sites have been shown to be multiplicative, as the sum of multiple sites log-fold-changes is the most appropriate way to combine the repression strength of individual sites (Grimson et al. [33], Selbach et al. [29] and Vejnar and Zdobnov [16]). However, this multiplicative effect has been shown with mRNAs bearing target sites for the same miRNA, and with experiments involving in most cases overexpression of a miRNA. Indeed, on a local scale, cooperativity has been described when the distance between target sites is appropriate (Saetrom et al. [83]). This simple rule to predict global repression, *i.e.* the sum, can still be enhanced while using the same available experimental data. Saito and Sætrom [34] used an SVM to combine target sites. The main feature of the SVM is the predicted target probability distribution transformed to a set of bins. The bin limits were unequal and chosen to maximize the SVM performance. This method is particularly efficient for integrating the more abundant non-canonical seed-matches: the authors were able to distinguish targeting differences caused by Single Nucleotide Polymorphisms (SNPs) (Thomas et al. [84]).

While the effort towards better prediction of variation in mRNA levels due to miRNA-mediated repression is relevant, kinetic models have not yet been used to solve this problem. Indeed, *in vivo*, multiple target sites must compete for RISC binding. During this competition, both the concentrations of RISC and mRNA molecules *etc* are key factors in determining repression strength. In support of using a kinetic model, preliminary investigations show that an increase in available target sites (Target-site Abundance, TA) (Garcia et al. [47]), is correlated with less efficient repression. However, kinetic models and their related parameters, such as the different concentrations mentioned above, are missing from current prediction algorithms. Moreover, experimental data are not yet available to deduce these parameters.

While being uncertain and challenging, a systems approach to miRNA repression prediction could increase prediction performance, and may enhance our knowledge of miRNA regulation by pointing out missing parameters in the current model. A deeper understanding of miRNA regulation will lead to greater precision in the prediction of protein concentrations, which are influenced by miRNA-mediated repression. As tissue identity, and more generally health, are tightly linked to the maintenance and regulation of protein concentrations in the cell, miRmap participates in solving this challenge by providing the community with direct access to state-of-the-art predictions of miRNA repression.

REFERENCES 6

1. Crick F (1970) Central dogma of molecular biology Nature 227, no. 5258, 561–3
2. Lee RC, Feinbaum RL, Ambros V (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14 Cell 75, no. 5, 843–54
3. Brennecke J, Stark A, Russell RB, Cohen SM (2005) Principles of microRNA-target recognition PLoS Biol 3, no. 3, e85 10.1371/journal.pbio.0030085
4. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A (2004) Identification of mammalian microRNA host genes and transcription units Genome Res 14, no. 10A, 1902–10 10.1101/gr.2722704
5. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function Cell 116, no. 2, 281–97
6. Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song J-J, Hammond SM, Joshua-Tor L, Hannon GJ (2004) Argonaute2 is the catalytic engine of mammalian RNAi Science 305, no. 5689, 1437–41 10.1126/science.1102513
7. Kozomara A and Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data Nucleic Acids Res 39, no. Database issue, D152–7 10.1093/nar/gkq1027
8. Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP (2008) Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals Nature 455, no. 7217, 1193–7 10.1038/nature07415
9. Axtell MJ, Westholm JO, Lai EC (2011) Vive la différence: biogenesis and evolution of microRNAs in plants and animals Genome Biol 12, no. 4, 221 10.1186/gb-2011-12-4-221
10. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E et al (2005) Identification of hundreds of conserved and nonconserved human microRNAs Nat Genet 37, no. 7, 766–70 10.1038/ng1590
11. Berezikov E, Robine N, Samsonova A, Westholm JO, Naqvi A, Hung J-H, Okamura K, Dai Q, Bortolamiol-Becet D, Martin R et al (2011) Deep annotation of Drosophila melanogaster microRNAs yields insights into their processing, modification, and emergence Genome Res 21, no. 2, 203–15 10.1101/gr.116657.110
12. Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE et al (2010) Mammalian microRNAs: experimental evaluation of novel and previously annotated genes Genes Dev 24, no. 10, 992–1009 10.1101/gad.1884710
13. Muljo SA, Kanellopoulou C, Aravind L (2010) MicroRNA targeting in mammalian genomes: genes and mechanisms Wiley Interdiscip Rev Syst Biol Med 2, no. 2, 148–61 10.1002/wsbm.53
14. Drinnenberg IA, Fink GR, Bartel DP (2011) Compatibility with killer explains the rise of RNAi-deficient fungi Science 333, no. 6049, 1592 10.1126/science.1209575
15. Chen K and Rajewsky N (2007) The evolution of gene regulation by transcription factors and microRNAs Nat Rev Genet 8, no. 2, 93–103 10.1038/nrg1990
16. Vejnar CE and Zdobnov EM (2012) miRmap: Comprehensive prediction of microRNA target repression strength In submission
17. Didiano D and Hobert O (2006) Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions Nat Struct Mol Biol 13, no. 9, 849–51 10.1038/nsmb1138

18. Shin C, Nam J-W, Farh KK-H, Chiang HR, Shkumatava A, Bartel DP (2010) Expanding the microRNA targeting code: functional sites with centered pairing Mol Cell 38, no. 6, 789–802 10.1016/j.molcel.2010.06.005
19. Wang Y, Li Y, Ma Z, Yang W, Ai C (2010) Mechanism of microRNA-target interaction: molecular dynamics simulations and thermodynamics analysis PLoS Comput Biol 6, no. 7, e1000866 10.1371/journal.pcbi.1000866
20. Wang Y, Juranek S, Li H, Sheng G, Tuschl T, Patel DJ (2008) Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex Nature 456, no. 7224, 921–6 10.1038/nature07666
21. Gu S, Jin L, Zhang F, Sarnow P, Kay MA (2009) Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs Nat Struct Mol Biol 16, no. 2, 144–50 10.1038/nsmb.1552
22. Eulalio A, Huntzinger E, Nishihara T, Rehwinkel J, Fauser M, Izaurralde E (2009) Deadenylation is a widespread effect of miRNA regulation RNA 15, no. 1, 21–32 10.1261/rna.1399509
23. Filipowicz W, Bhattacharyya SN, Sonenberg N (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet 9, no. 2, 102–14 10.1038/nrg2290
24. Huntzinger E and Izaurralde E (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay Nat Rev Genet 12, no. 2, 99–110 10.1038/nrg2936
25. Pillai RS, Bhattacharyya SN, Artus CG, Zoller T, Cougot N, Basyuk E, Bertrand E, Filipowicz W (2005) Inhibition of translational initiation by Let-7 MicroRNA in human cells Science 309, no. 5740, 1573–6 10.1126/science.1115079
26. Hendrickson DG, Hogan DJ, McCullough HL, Myers JW, Herschlag D, Ferrell JE, Brown PO (2009) Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA PLoS Biol 7, no. 11, e1000238 10.1371/journal.pbio.1000238
27. Guo H, Ingolia NT, Weissman JS, Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels Nature 466, no. 7308, 835–40 10.1038/nature09267
28. Baek D, Villén J, Shin C, Camargo FD, Gygi SP, Bartel DP (2008) The impact of microRNAs on protein output Nature 455, no. 7209, 64–71 10.1038/nature07242
29. Selbach M, Schwanhäusser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N (2008) Widespread changes in protein synthesis induced by microRNAs Nature 455, no. 7209, 58–63 10.1038/nature07228
30. Didiano D, Cochella L, Tursun B, Hobert O (2010) Neuron-type specific regulation of a 3'UTR through redundant and combinatorially acting cis-regulatory elements RNA 16, no. 2, 349–63 10.1261/rna.1931510
31. Vergoulis T, Vlachos IS, Alexiou P, Georgakilas G, Maragkakis M, Reczko M, Gerangelos S, Koziris N, Dalamagas T, Hatzigeorgiou AG (2012) TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support Nucleic Acids Res 40, no. Database issue, D222–9 10.1093/nar/gkr1161
32. Hsu S-D, Lin F-M, Wu W-Y, Liang C, Huang W-C, Chan W-L, Tsai W-T, Chen G-Z, Lee C-J, Chiu C-M et al (2011) miRTarBase: a database curates experimentally validated microRNA-target interactions Nucleic Acids Res 39, no. Database issue, D163–9 10.1093/nar/gkq1107
33. Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing Mol Cell 27, no. 1, 91–105 10.1016/j.molcel.2007.06.017

34. Saito T and Sætrom P (2010) A two-step site and mRNA-level model for predicting microRNA targets BMC Bioinformatics 11, 612 10.1186/1471-2105-11-612
35. Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure J Mol Biol 288, no. 5, 911–40 10.1006/jmbi.1999.2700
36. Nussinov R and Jacobson AB (1980) Fast algorithm for predicting the secondary structure of single-stranded RNA Proc Natl Acad Sci U S A 77, no. 11, 6309–13
37. Zuker M (1989) On finding all suboptimal foldings of an RNA molecule Science 244, no. 4900, 48–52
38. McCaskill JS (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure Biopolymers 29, no. 6-7, 1105–19 10.1002/bip.360290621
39. Mathews DH (2004) Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization RNA 10, no. 8, 1178–90 10.1261/rna.7650904
40. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E et al (2011) A high-resolution map of human evolutionary constraint using 29 mammals Nature 478, no. 7370, 476–82 10.1038/nature10530
41. Ellwanger DC, Büttner FA, Mewes H-W, Stümpflen V (2011) The sufficient minimal set of miRNA seed types Bioinformatics 27, no. 10, 1346–50 10.1093/bioinformatics/btr149
42. Chi SW, Zang JB, Mele A, Darnell RB (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps Nature 460, no. 7254, 479–86 10.1038/nature08170
43. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jungkamp A-C, Munschauer M et al (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP Cell 141, no. 1, 129–41 10.1016/j.cell.2010.03.009
44. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R (2004) Fast and effective prediction of microRNA/target duplexes RNA 10, no. 10, 1507–17 10.1261/rna.5248604
45. Friedman RC, Farh KK-H, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs Genome Res 19, no. 1, 92–105 10.1101/gr.082701.108
46. Stark A, Lin MF, Kheradpour P, Pedersen JS, Parts L, Carlson JW, Crosby MA, Rasmussen MD, Roy S, Deoras AN et al (2007) Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures Nature 450, no. 7167, 219–32 10.1038/nature06340
47. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP (2011) Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs Nat Struct Mol Biol 18, no. 10, 1139–46 10.1038/nsmb.2115
48. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS (2004) Human MicroRNA targets PLoS Biol 2, no. 11, e363 10.1371/journal.pbio.0020363
49. Betel D, Wilson M, Gabow A, Marks DS, Sander C (2008) The microRNA.org resource: targets and expression Nucleic Acids Res 36, no. Database issue, D149–53 10.1093/nar/gkm995
50. Betel D, Koppal A, Agius P, Sander C, Leslie C (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites Genome Biol 11, no. 8, R90 10.1186/gb-2010-11-8-r90
51. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E (2007) The role of site accessibility in microRNA target recognition Nat Genet 39, no. 10, 1278–84 10.1038/ng2135
52. Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M et al (2005) Combinatorial microRNA target predictions Nat Genet 37, no. 5, 495–500 10.1038/ng1536

53. Rigoutsos I and Floratos A (1998) Combinatorial pattern discovery in biological sequences: The TEIRESIAS algorithm Bioinformatics 14, no. 1, 55–67
54. Miranda KC, Huynh T, Tay Y, Ang Y-S, Tam W-L, Thomson AM, Lim B, Rigoutsos I (2006) A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes Cell 126, no. 6, 1203–17 10.1016/j.cell.2006.07.031
55. Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M (2007) Inference of miRNA targets using evolutionary conservation and pathway analysis BMC Bioinformatics 8, 69 10.1186/1471-2105-8-69
56. Marín RM and Vanícek J (2011) Efficient use of accessibility in microRNA target prediction Nucleic Acids Res 39, no. 1, 19–29 10.1093/nar/gkq768
57. Kiriakidou M, Nelson PT, Kouranov A, Fitziev P, Bouyioukos C, Mourelatos Z, Hatzigeorgiou A (2004) A combined computational-experimental approach predicts human microRNA targets Genes Dev 18, no. 10, 1165–78 10.1101/gad.1184704
58. Maragkakis M, Alexiou P, Papadopoulos GL, Reczko M, Dalamagas T, Giannopoulos G, Goumas G, Koukis E, Kourtis K, Simossis VA et al (2009) Accurate microRNA target prediction correlates with protein repression levels BMC Bioinformatics 10, 295 10.1186/1471-2105-10-295
59. Maragkakis M, Vergoulis T, Alexiou P, Reczko M, Plomaritou K, Gousis M, Kourtis K, Koziris N, Dalamagas T, Hatzigeorgiou AG (2011) DIANA-microT Web server upgrade supports Fly and Worm miRNA target prediction and bibliographic miRNA to disease association Nucleic Acids Res 39, no. Web Server issue, W145–8 10.1093/nar/gkr294
60. Liu H, Yue D, Chen Y, Gao S-J, Huang Y (2010) Improving performance of mammalian microRNA target prediction BMC Bioinformatics 11, 476 10.1186/1471-2105-11-476
61. Saito T and Saetrom P (2010) MicroRNAs–targeting and target prediction N Biotechnol 27, no. 3, 243–9 10.1016/j.nbt.2010.02.016
62. Rajewsky N (2006) microRNA target predictions in animals Nat Genet 38 Suppl, S8–13 10.1038/ng1798
63. Hendrickson DG, Hogan DJ, Herschlag D, Ferrell JE, Brown PO (2008) Systematic identification of mRNAs recruited to argonaute 2 by specific microRNAs and corresponding changes in transcript abundance PLoS One 3, no. 5, e2126 10.1371/journal.pone.0002126
64. Linsley PS, Schelter J, Burchard J, Kibukawa M, Martin MM, Bartz SR, Johnson JM, Cummins JM, Raymond CK, Dai H et al (2007) Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression Mol Cell Biol 27, no. 6, 2240–52 10.1128/MCB.02005-06
65. Wen J, Parker BJ, Jacobsen A, Krogh A (2011) MicroRNA transfection and AGO-bound CLIP-seq data sets reveal distinct determinants of miRNA action RNA 17, no. 5, 820–34 10.1261/rna.2387911
66. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T (2002) Identification of tissue-specific microRNAs from mouse Curr Biol 12, no. 9, 735–9
67. Brennan J and Capel B (2004) One tissue, two fates: molecular genetic events that underlie testis versus ovary development Nat Rev Genet 5, no. 7, 509–21 10.1038/nrg1381
68. Papaioannou MD, Pitetti J-L, Ro S, Park C, Aubry F, Schaad O, Vejnar CE, Kühne F, Descombes P, Zdobnov EM et al (2009) Sertoli cell Dicer is essential for spermatogenesis in mice Dev Biol 326, no. 1, 250–9 10.1016/j.ydbio.2008.11.011
69. Vasudevan S, Tong Y, Steitz JA (2007) Switching from repression to activation: microRNAs can up-regulate translation Science 318, no. 5858, 1931–4 10.1126/science.1149460
70. Gachon F, Nagoshi E, Brown SA, Ripperger J, Schibler U (2004) The mammalian circadian timing system: from gene expression to physiology Chromosoma 113, no. 3, 103–12 10.1007/s00412-004-0296-2

71. Storch K-F, Lipan O, Leykin I, Viswanathan N, Davis FC, Wong WH, Weitz CJ (2002) Extensive and divergent circadian gene expression in liver and heart Nature 417, no. 6884, 78–83 10.1038/nature744
72. Xiao C and Rajewsky K (2009) MicroRNA control in the immune system: basic principles Cell 136, no. 1, 26–36 10.1016/j.cell.2008.12.027
73. Rodriguez A, Vigorito E, Clare S, Warren MV, Couttet P, Soond DR, van Dongen S, Grocock RJ, Das PP, Miska EA et al (2007) Requirement of bic/microRNA-155 for normal immune function Science 316, no. 5824, 608–11 10.1126/science.1139253
74. Gatfield D, Le Martelot G, Vejnar CE, Gerlach D, Schaad O, Fleury-Olela F, Ruskeepää A-L, Oresic M, Esau CC, Zdobnov EM et al (2009) Integration of microRNA miR-122 in hepatic circadian gene expression Genes Dev 23, no. 11, 1313–26 10.1101/gad.1781009
75. Dunand-Sauthier I, Santiago-Raber M-L, Capponi L, Vejnar CE, Schaad O, Irla M, Seguín-Estévez Q, Descombes P, Zdobnov EM, Acha-Orbea H et al (2011) Silencing of c-Fos expression by microRNA-155 is critical for dendritic cell maturation and function Blood 117, no. 17, 4490–500 10.1182/blood-2010-09-308064
76. Hofacker IL (2003) Vienna RNA secondary structure server Nucleic Acids Res 31, no. 13, 3429–31
77. Nuel G, Regad L, Martin J, Camproux A-C (2010) Exact distribution of a pattern in a set of random sequences generated by a Markov source: applications to biological data Algorithms Mol Biol 5, 15 10.1186/1748-7188-5-15
78. Quarrier S, Martin JS, Davis-Neulander L, Beauregard A, Laederach A (2010) Evaluation of the information content of RNA structure mapping data for secondary structure prediction RNA 16, no. 6, 1108–17 10.1261/rna.1988510
79. Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY (2011) Understanding the transcriptome through RNA structure Nat Rev Genet 12, no. 9, 641–55 10.1038/nrg3049
80. Kedde M and Agami R (2008) Interplay between microRNAs and RNA-binding proteins determines developmental processes Cell Cycle 7, no. 7, 899–903
81. Kedde M, van Kouwenhove M, Zwart W, Oude Vrielink JAF, Elkon R, Agami R (2010) A Pumilio-induced RNA structure switch in p27-3′ UTR controls miR-221 and miR-222 accessibility Nat Cell Biol 12, no. 10, 1014–20 10.1038/ncb2105
82. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A (2010) Detection of nonneutral substitution rates on mammalian phylogenies Genome Res 20, no. 1, 110–21 10.1101/gr.097857.109
83. Saetrom P, Heale BSE, Snøve O, Aagaard L, Alluin J, Rossi JJ (2007) Distance constraints between microRNA target sites dictate efficacy and cooperativity Nucleic Acids Res 35, no. 7, 2333–42 10.1093/nar/gkm133
84. Thomas LF, Saito T, Sætrom P (2011) Inferring causative variants in microRNA target sites Nucleic Acids Res 39, no. 16, e109 10.1093/nar/gkr414

APPENDIX 7

The following pages contain:

- the co-authored manuscript with Dr Daniel Gerlach entitled *miROrtho: computational survey of microRNA genes* published in 2009,
- the co-authored manuscript with Dr Christelle Borel entitled *Identification of cis- and trans-regulatory variation modulating microRNA expression levels in human fibroblasts* published in 2011,
- the documentation of the miRmap library.

Published online 15 October 2008 *Nucleic Acids Research, 2009, Vol. 37, Database issue* ***D111–D117***
doi:10.1093/nar/gkn707

miROrtho: computational survey of microRNA genes

Daniel Gerlach[1,2], Evgenia V. Kriventseva[1], Nazim Rahman[1], Charles E. Vejnar[1,2] and Evgeny M. Zdobnov[1,2,3,*]

[1]Department of Genetic Medicine and Development, University of Geneva Medical School, [2]Swiss Institute of Bioinformatics, 1 Rue Michel-Servet, 1211 Geneva, Switzerland and [3]Imperial College London, South Kensington Campus, SW7 2AZ London, UK

Received August 19, 2008; Revised September 26, 2008; Accepted September 29, 2008

ABSTRACT

MicroRNAs (miRNAs) are short, non-protein coding RNAs that direct the widespread phenomenon of post-transcriptional regulation of metazoan genes. The mature ~22-nt long RNA molecules are processed from genome-encoded stem-loop structured precursor genes. Hundreds of such genes have been experimentally validated in vertebrate genomes, yet their discovery remains challenging, and substantially higher numbers have been estimated. The miROrtho database (http://cegg.unige.ch/mirortho) presents the results of a comprehensive computational survey of miRNA gene candidates across the majority of sequenced metazoan genomes. We designed and applied a three-tier analysis pipeline: (i) an SVM-based *ab initio* screen for potent hairpins, plus homologs of known miRNAs, (ii) an orthology delineation procedure and (iii) an SVM-based classifier of the ortholog multiple sequence alignments. The web interface provides direct access to putative miRNA annotations, ortholog multiple alignments, RNA secondary structure conservation, and sequence data. The miROrtho data are conceptually complementary to the miRBase catalog of experimentally verified miRNA sequences, providing a consistent comparative genomics perspective as well as identifying many novel miRNA genes with strong evolutionary support.

INTRODUCTION

MicroRNAs (miRNAs) represent an abundant class of short non-protein coding RNAs that direct post-transcriptional regulation of metazoan genes through repression of mRNA translation or transcript degradation. Since their initial discovery in *Caenorhabditis elegans*, the roles of miRNAs have been recognized as a widespread phenomenon, implicated in processes such as cell differentiation and cancer (1–6). Intensive studies have begun to unravel the mechanisms and characteristics of these single-stranded, ~22-nt long RNA molecules that are processed from genome-encoded precursor genes with a defining stem-loop RNA structure. Nevertheless, the discovery and characterization of novel miRNA genes have proved to be challenging both experimentally and computationally, and the miRNA gene repertoire therefore remains largely unexplored. The human genome tops the fast growing number of miRNA genes, with several hundreds now cataloged in the miRBase database of published miRNA sequences (7) and many more estimated (8,9).

The high-throughput experimental approaches usually identify only the short mature segments of the miRNA genes along with other types of endogenous small RNAs (10,11) and degradation products of mRNAs or structural RNAs. Robust computational post-processing of the experimentally derived sequences is therefore essential to identify the underlying miRNA genes. The widely applied discriminatory requirement of a characteristic stem-loop structure for the putative precursor is, however, insufficient as hairpin structures are common in eukaryotic genomes and are not a unique feature of miRNAs (12). Nonetheless, the rapid accumulation of genome-wide sequencing data provides another line of evolutionary evidence from comparative sequence analyses.

Computational screening methods that rely heavily on sequence conservation criteria, such as MirScan (13), were among the first to appear. These characteristically exhibit high specificity [e.g. predicting 35 new miRNA candidates in *C. elegans* (13) and 107 in human (14), many of which were experimentally confirmed], but their sensitivity, the ability to predict novel or divergent homologs in other organisms, is low. Methods that relax sequence conservation requirements in favor of conservation patterns specific to miRNAs (such as a more diverged loop sequence and a more conserved hairpin stem) gained

*To whom correspondence should be addressed. Tel: +41 22 379 59 73; Fax: +41 22 379 57 06; Email: evgeny.zdobnov@unige.ch

© 2008 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from http://nar.oxfordjournals.org/ at University of Geneva on January 17, 2012

D112 Nucleic Acids Research, 2009, Vol. 37, Database issue

substantially higher sensitivity, e.g. Snarloop has been used to predict 214 candidate miRNAs in *C. elegans* (15) and miRSeeker (16) to predict 48 candidate miRNAs in *Drosophila melanogaster*. A similar approach was proposed that takes into account the shapes of conservation patterns of known miRNAs, e.g. phylogenetic shadowing (17,18). The first 7 nt from the second position of the 5′-end of the mature miRNA, termed the seed sequence, are presumed to be critical for the interaction between the miRNA and its targets (19–22). The intra-species abundance or inter-species conservation of such potential seeds have also been proposed as alternative starting points for miRNA gene hunting (23,24).

Secondary structure thermodynamic stability is another important characteristic that can be used to distinguish miRNAs from other hairpins (25). The recently developed software RNAz combines thermodynamic stability and conservation of secondary structure to predict non-coding RNAs (26) from multiple alignments of orthologous regions. Methods relying on phylogenetic conservation of miRNA structure and sequence are by definition restricted in terms of their predictive power. To overcome this limitation, several groups have developed *ab initio* approaches (12,27–32) to predict novel, non-conserved genes. However, these approaches often suffer from high rates of false positives.

Aiming to fuel further studies of microRNA'omes, we present here the database of computationally derived miRNA gene candidates using a novel comparative genomics approach coupled with machine-learning techniques that we consistently applied to a comprehensive set of available metazoan genomes. The three-tier pipeline consists of: (i) a custom designed SVM-based *ab initio* predictor, plus screening for known miRNA homologs, (ii) an orthology delineation procedure and (iii) an SVM-based classifier of the multiple sequence alignments of the putative orthologs. These data are conceptually complementary to the miRBase catalog of experimentally verified miRNA sequences (7). High-throughput experimental exploration of small RNAs requires rigorous follow-up bioinformatic analyses to claim evidence of microRNA genes. Decoupling experimental and bioinformatics approaches, the miROrtho data effectively provide independent supporting evidence for the numerous ongoing experimental interrogations of microRNA'omes.

MATERIAL AND METHODS

Ab initio predictors

The first tier of our analysis pipeline is a novel *ab initio* miRNA prediction procedure. We scanned the genomic sequences using RNALfold (33) for locally stable hairpins characteristic of miRNA precursors, requiring a length of 60–120 nt, a minimum free-folding energy less than −15 kcal/mol, a stem of 20–60 base pairs, a maximal interior loop size of 8 nt, and a maximum bulge loop size of 5 nt. The loop, however, was allowed to include short stem-loops e.g. hsa-let-7b. Those properties accommodate the vast majority of experimentally validated miRNAs (although there are exceptions, e.g. dme-mir-31b and dme-mir-1017). As stem-loop structures are abundant and not exclusive to miRNA genes, this step yields hundreds of millions of candidates: 1.3 million for the ∼170 Mb genome of fruitfly *Drosophila melanogaster*. The availability of many experimentally validated miRNAs revealed that although there are biases in biophysical properties of miRNA stem-loops in comparison to non-miRNA sequences, such as higher thermodynamic stability (25), no clear discriminatory features have yet been identified. We investigated a number of the most discriminating features, such as the minimum free-energy index (34) or the mean base pair distance in the ensemble of structures, and trained an SVM (support vector machine) classifier using LIBSVM (http://www.csie.ntu.edu.tw/∼cjlin/libsvm). The total number of features used for this first SVM was 253. The radial basis function kernel (RBF) was used on 1000 experimentally verified animal pre-miRNAs from miRBase (7) and a negative set of 3000 potent stem-loops from other confirmed ncRNAs [Rfam (35)]. Optimal parameters for the RBF kernel (C-SVC c = 2.0, gamma = 0.03 125) were estimated using a heuristic approach implemented in grid.py, which is a part of the LIBSVM package. A non-redundant training dataset was compiled using CD-Hit-EST (36) at a cutoff of 90% sequence identity. We tested the performance of the SVM on a test set of 237 miRNA sequences and 568 non-miRNA stem-loops which where not used for training the SVM model. Using the SVM posterior probability cutoff at 0.5, the accuracy was estimated to be 95.03%, the area under the ROC curve (receiver operating characteristic) was 0.984, corresponding to a sensitivity and specificity of 0.84 and 0.97, respectively. Using a 10-fold cross-validation procedure on the training data, we received an average AUC (area under the ROC curve) of 0.982. If the potent hairpins had >70% sequence overlap at the same locus, the one with the lower SVM score was discarded.

This single sequence SVM filter allows the space of likely candidates to be reduced by about 95%, yet still yields rather high numbers of gene candidates: 42 000 for *D. melanogaster*. The miRNA structure itself is likely to contribute to these elevated numbers: miRNAs have complementary arms in their stem-loop structure and the reverse complement of a precursor often also folds into a stable RNA hairpin. Nevertheless, we did not explicitly require a choice between the sense and the anti-sense candidates (if both of them passed the other filters) as there is evidence of miRNA loci with both strands yielding a functional miRNA, e.g. dme-mir-iab-4 and dme-mir-iab-4as.

Homology-based predictor

Screening for homologs of currently known miRNAs (miRBase 11.0) captures putative miRNAs that either did not pass the stem-loop screen, e.g. 13 (8%) of known *D. melanogaster* miRNAs, or failed the *ab initio* SVM classification, another 19 (13%). Our procedure initially performs a WU-BLAST (http://blast.wustl.edu) search using the default parameters, plus the DUST

Downloaded from http://nar.oxfordjournals.org/ at University of Geneva on January 17, 2012

filter and the hspsepSmax = 30 option, which defines the maximal separating distance between two high score pairs to allow for a varying loop while still matching the better conserved 5′ and 3′ arms. Next, blast hits longer than 20 nt are extended at both ends to match the length of the query sequence. These hits are further filtered using a minimum free energy filter (≤ -15 kcal/mol) and a RANDFOLD (25) filter ($P \leq 0.05$ on 100 sequence randomizations). We investigated the RNAshapes (37) filter, which predicts the probability of a sequence to fold into a simple stem-loop like structure, but it was not employed as several known miRNAs, e.g. hsa-let-7a-1, would not pass the filter. The candidate miRNAs were then aligned to the query sequence using MAFFT (38) and the conservation of the seed region was calculated by mapping the known mature miRNA region on the query miRNA to the alignment. The hits were then tested for the following criteria: a 100% conserved seed region, >90% conservation of the putative mature part, and a total hairpin identity >65%. As close paralogs (like hsa-let-7, mmu-let-7, etc) can map to the same locus when searched again one genome (e.g. the chimp), the matches were then clustered using GALAXY (http://main.g2.bx.psu.edu) and choosing one representative with the lowest *e*-value of all queries.

Orthology delineation

Groups of likely orthologous genes were automatically identified using a strategy employed previously for protein-coding genes (39) based on all-against-all sequence comparisons using the ParAlign algorithm (40) with NT2 substitution matrix; followed by clustering of best reciprocal hits (BRHs) from highest scoring ones to 10^{-6} *e*-value cutoff for triangulating BRHs or 10^{-10} cutoff for unsupported BRHs, and requiring a sequence alignment overlap of at least 20 nt across all members of a group. Furthermore, the orthologous groups were expanded by genes that are more similar to each other within a genome than to any gene in any of the other species, and by very similar copies that share over 97% sequence identity, which were identified initially using CD-Hit (36). The orthology filter allowed us to reduce the space of the miRNA candidates by a further 92%. Passing the orthology filter provides evolutionary support for the predicted miRNAs; however, detailed inspection highlighted the need for further rigorous sequence classification to remove questionable predictions.

Multi-species conservation classifier

We further analyzed the R-COFFEE (41) multiple sequence alignments of orthologous groups of putative miRNA sequences. From the alignments we gathered the 13 most descriptive features for conservation properties of sequence, energy and structures such as: GC content, number of taxa, mean pairwise sequence identity, number of consistent mutations, conservation of the mature part, etc. Those descriptors were chosen among a larger set of features, in order to optimally describe the typical conservation profile of a miRNA gene family and to reduce false positive predictions. Alignments that mapped to at least one known miRNA from miRBase 11.0 were used as the positive training and testing sets (344 and 100 alignments, respectively). Among those alignments which did not map to any known miRNA family, we randomly selected (with manual checking) the negative training and testing sets (344 and 100 alignments, respectively). The GIST SVM software package (http://www.cs.columbia.edu/compbio) was used for training, testing and classification using the default parameter. The final set of newly predicted miRNAs based on the alignment SVM was selected from all alignments which had SVM score ≥ 0.5, a 100% conserved seed, a mature part >90% conserved and having representatives in at least four taxa. Performance estimation of the alignment SVM on the independent test set showed an accuracy of 91%, with the area under the ROC curve (AUC) of 0.97, and sensitivity and specificity of 0.9 and 0.92, respectively. The AUC for the 10-fold cross validation using the training data averaged to 0.998. The alignment SVM filter allowed us to reduce the space of the miRNA candidates by a further 98%, followed limited manual curation of novel miRNA candidates. We further analyzed the multiple alignments of novel miRNAs (without known homologs) to predict the mature part using a sliding 23-nt long sliding window and scanning for the region with the highest information content in the 5′ or the 3′ arms. The predictions, however, should be taken with caution without further experimental support.

DATABASE CONTENT

The miROrtho database (http://cegg.unige.ch/mirortho) presents computationally predicted putative miRNA genes for a comprehensive set of sequenced animal genomes (selection of genomes in Table 1), employing an in-house developed pipeline combining SVM-based classifiers and orthology delineation procedure adapted from OrthoDB (39). The alignments shown on the website were calculated using R-COFFEE (41), which combines MUSCLE (42), Probcons4RNA (43), MAFFT (38) and the secondary structures predicted by RNAplfold (33). Based on these alignments consensus secondary structures color-coded according to consistent/compensatory mutation were calculated using RNAalifold (44) which incorporates a ribosome scoring matrix suited for aligned RNA sequences. The database aims to provide a comprehensive comparative perspective on the animal repertoire of miRNA genes with direct reference to the putative ortholog multiple alignments, RNA secondary structure conservation, etc. As there seem to be numerous lineage specific miRNAs and miRNA-like sequences that are difficult to differentiate without experimental evidence, we see miROrtho as complementary to miRBase, the repository of experimentally verified miRNA sequences. Overall, miROrtho contains 7887 putative miRNA genes that are homologous to known miRNAs in miRBase 11.0, and 1437 confident predictions that are as yet without experimental support or homology to known miRNAs. Most experimental surveys provide support for mature miRNA sequences, while the identities of the underlying miRNA precursor genes remain somewhat uncertain.

Downloaded from http://nar.oxfordjournals.org/ at University of Geneva on January 17, 2012

D114 Nucleic Acids Research, 2009, Vol. 37, Database issue

Table 1. Analyzed genomes

Species name	Abbreviation	Size (Mb)	Number of miRNA genes			Source
			Homologs[a]	New[b]	miRBase 11.0	
Aedes aegypti	Aaeg	1384	58	1	0	AaegL1
Anopheles gambiae	Agam	273	55	1	45	AgamP3
Apis mellifera	Amel	235	60	1	54	Amel_4.0
Bombyx mori	Bmor	397	33	0	21	SW_scaffold_ge2k
Caenorhabditis elegans	Cele	100	149	0	154	WB170
Canis familiaris	Cfam	2532	383	138	203	CanFam 2.0
Ciona intestinalis	Cint	173	25	0	34	JGI2
Danio rerio	Drer	1626	324	22	337	ZFISH6
Drosophila ananassae	Dana	230	108	12	0	CAF1
Drosophila erecta	Dere	152	136	16	0	CAF1
Drosophila grimshawi	Dgri	200	110	13	0	CAF1
Drosophila melanogaster	Dmel	129	153	15	152	CAF1
Drosophila mojavensis	Dmoj	194	98	14	0	CAF1
Drosophila persimilis	Dper	188	108	16	0	CAF1
Drosophila pseudoobscura	Dpse	153	106	15	76	CAF1
Drosophila sechellia	Dsec	167	139	16	0	CAF1
Drosophila simulans	Dsim	142	131	15	0	CAF1
Drosophila virilis	Dvir	206	101	14	0	CAF1
Drosophila willistoni	Dwil	237	112	12	0	CAF1
Drosophila yakuba	Dyak	169	135	16	0	CAF1
Gallus gallus	Ggal	1100	168	49	149	WASHUC2
Gasterosteus aculeatus	Gacu	462	320	12	0	BROAD S1
Homo sapiens	Hsap	3665	626	151	678	NCBI36
Macaca mulatta	Mmul	3097	530	145	464	MMUL_1
Monodelphis domestica	Mdom	3606	205	82	119	monDom5
Mus musculus	Mmus	2661	505	117	472	NCBIM36
Ornithorhynchus anatinus	Oana	2073	207	57	0	Oana-5.0
Pan troglodytes	Ptro	3524	546	147	100	PanTro 2.1
Rattus norvegicus	Rnor	2719	440	110	287	RGSC 3.4
Strongylocentrotus purpuratus	Surc	907	13	0	0	Spur_v2.1
Takifugu rubripes	Trub	393	250	13	131	FUGU4
Tetraodon nigroviridis	Tnig	402	282	14	132	TETRAODON7
Tribolium castaneum	Tcas	200	37	1	0	Tcas_2.0
Xenopus tropicalis	Xtro	1511	351	24	184	JGI4.1

[a]Homologs to miRBase 11.0 miRNAs.
[b]New predictions that do not show any homology to any annotated miRNA.

In contrast, computational procedures rely on recognizing characteristic sequence and structural properties of the precursors, where even approximate prediction of mature miRNAs is rarely possible. This complementarity extends further, where computational predictions at different stringencies can either be used to prioritize experimental verification, or as direct independent support of miRNAs identified through high throughput experimental screens. Although miRBase accepts annotation of very close homologs of experimentally supported miRNAs, the comparative perspective is heavily biased towards favorite experimental model species. Such a bias is avoided in miROrtho through the consistent application of the same procedures across all the available genomes, delineating groups of orthologous miRNAs over distantly related organisms. The miROrtho methodology has also been applied to the task of miRNA gene annotation in a number of ongoing initial genome analyses, and this database will provide the supporting information for these predictions.

It should be noted that there is still no defining feature that clearly discriminates between *bona fide* miRNA precursors and other abundant genomic sequences capable of similar hairpin folding. Classification filters will therefore inevitably suffer from false negatives and false positives (see Materials and Methods section for estimates), leading to errors at each step along the pipeline. Even the most inclusive initial screen for locally stable stem-loop structures misses some miRNAs reported in miRBase as experimentally validated (e.g. dme-mir-1017). Despite the strict 97% specificity of our *ab initio* SVM, the abundance of false positives is clear and overloads the orthology filter. Computational methods developed for miRNA gene discovery are constantly improving, and will continue to do so as our knowledge of experimentally validated miRNAs grows.

WEB INTERFACE

The miROrtho database presents all predicted miRNA genes within the context of family groups of orthologous miRNAs. For each such family, we provide (Figure 1): (i) a table of annotated miRNA names and genomic coordinates, (ii) a multiple alignment of the miRNA sequences displaying RNA structure conservation, (iii) the minimum

Downloaded from http://nar.oxfordjournals.org/ at University of Geneva on January 17, 2012

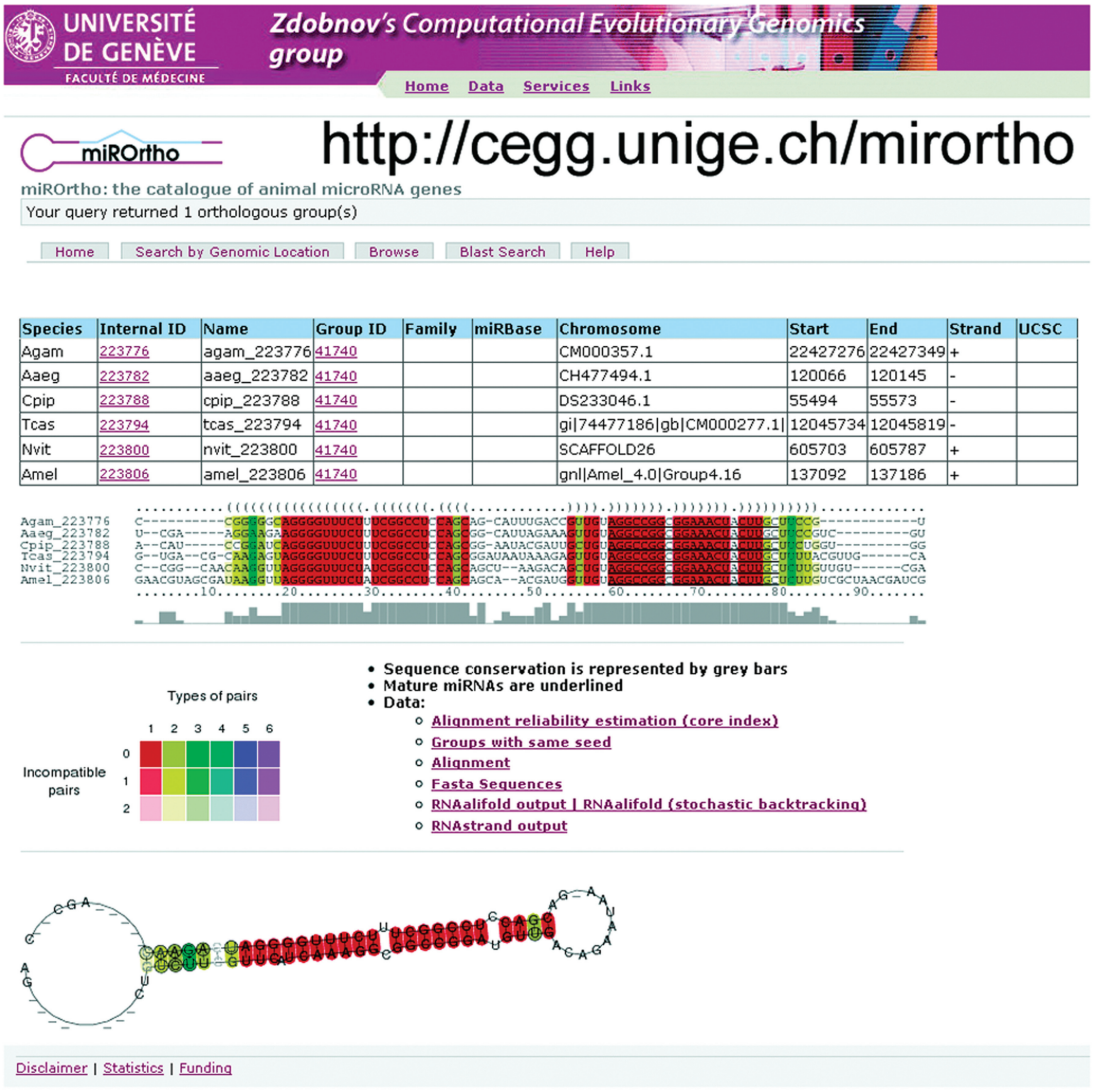


Figure 1. miROrtho screenshot showing a novel miRNA gene family. The results page consists of three parts: (i) a table with detailed information about the individual miRNAs; (ii) a multiple sequence alignment with the consensus secondary structure displayed above in dot-bracket format and conservation profile bars displayed below, with the sequence of the mature miRNAs underlined; (iii) the consensus secondary structure of the orthologous sequences. Both alignment and consensus secondary structure are color-coded according to consistent and compensatory base changes.

energy consensus miRNA hairpin fold, (iv) FASTA sequences and multiple alignment files. Color coding of the alignments and the depicted folds enables clear visualization of compensatory and consistent mutations within a given miRNA family. The mature miRNA sequences are underlined: as annotated in miRBase for known miRNAs or as predicted for novel families. Furthermore, we provide detailed folding information of individual pre-miRNAs including minimum free energy folding, the partition function folding and the centroid structure of the stem-loop. Three images show the secondary structure of a single pre-miRNA with the mature part annotated in red, color-coded according to base pairing probabilities and positional entropy per position. The data can be browsed by the species tree, or can be queried by annotation such as known families (e.g. let-7), identifiers or chromosomes. The predictions can be also searched by sequence homology using WU-BLAST (http://blast.wustl.edu).

ACKNOWLEDGEMENTS

We thank R.M. Waterhouse for help with the article, and the Vital-IT facility (http://www.vital-it.ch/vitalit-intro.htm). We would also like to acknowledge the sequencing centers that made the genome sequences that were used for this study, available before publication: The Baylor College of Medicine (www.hgsc.bcm.tmc.edu), the Washington University School of Medicine

Downloaded from http://nar.oxfordjournals.org/ at University of Geneva on January 17, 2012

(genome.wustl.edu), the Broad Institute (www.broad.mit.edu), the J. Craig Venter Institute (www.jcvi.org), the DOE Joint Genome Institute (www.jgi.doe.gov), the Sanger Center (www.sanger.ac.uk), the Institute for Genomic Research (www.tigr.org), Celera Genomics (www.celera.com), and Genoscope (www.genoscope.cns.fr).

FUNDING

Swiss National Science Foundation (SNF PDFMA3-118375 and 3100A0-112588). Funding for open access charges: Swiss National Science Foundation (SNF 3100A0-112588).

Conflict of interest statement. None declared.

REFERENCES

1. Ambros,V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.
2. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
3. Du,T. and Zamore,P.D. (2005) microPrimer: the biogenesis and function of microRNA. *Development*, **132**, 4645–4652.
4. Calin,G.A. and Croce,C.M. (2006) MicroRNA signatures in human cancers. *Nat. Rev. Cancer*, **6**, 857–866.
5. Zhang,B., Pan,X., Cobb,G.P. and Anderson,T.A. (2007) microRNAs as oncogenes and tumor suppressors. *Dev. Biol.*, **302**, 1–12.
6. Barbarotto,E., Schmittgen,T.D. and Calin,G.A. (2008) MicroRNAs and cancer: profile, profile, profile. *Int. J. Cancer*, **122**, 969–977.
7. Griffiths-Jones,S., Saini,H.K., van Dongen,S. and Enright,A.J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.
8. Miranda,K.C., Huynh,T., Tay,Y., Ang,Y.S., Tam,W.L., Thomson,A.M., Lim,B. and Rigoutsos,I. (2006) A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*, **126**, 1203–1217.
9. Berezikov,E., van Tetering,G., Verheul,M., van de Belt,J., van Laake,L., Vos,J., Verloop,R., van de Wetering,M., Guryev,V., Takada,S. *et al.* (2006) Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis. *Genome Res.*, **16**, 1289–1298.
10. Kim,V.N. and Nam,J.W. (2006) Genomics of microRNA. *Trends Genet.*, **22**, 165–173.
11. Aravin,A. and Tuschl,T. (2005) Identification and characterization of small RNAs involved in RNA silencing. *FEBS Lett.*, **579**, 5830–5840.
12. Bentwich,I., Avniel,A., Karov,Y., Aharonov,R., Gilad,S., Barad,O., Barzilai,A., Einat,P., Einav,U., Meiri,E. *et al.* (2005) Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.*, **37**, 766–770.
13. Lim,L.P., Lau,N.C., Weinstein,E.G., Abdelhakim,A., Yekta,S., Rhoades,M.W., Burge,C.B. and Bartel,D.P. (2003) The microRNAs of Caenorhabditis elegans. *Genes Dev.*, **17**, 991–1008.
14. Lim,L.P., Glasner,M.E., Yekta,S., Burge,C.B. and Bartel,D.P. (2003) Vertebrate microRNA genes. *Science*, **299**, 1540.
15. Grad,Y., Aach,J., Hayes,G.D., Reinhart,B.J., Church,G.M., Ruvkun,G. and Kim,J. (2003) Computational and experimental identification of C. elegans microRNAs. *Mol. Cell*, **11**, 1253–1263.
16. Lai,E.C., Tomancak,P., Williams,R.W. and Rubin,G.M. (2003) Computational identification of Drosophila microRNA genes. *Genome Biol.*, **4**, R42.
17. Boffelli,D., McAuliffe,J., Ovcharenko,D., Lewis,K.D., Ovcharenko,I., Pachter,L. and Rubin,E.M. (2003) Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science*, **299**, 1391–1394.
18. Berezikov,E., Guryev,V., van de Belt,J., Wienholds,E., Plasterk,R.H. and Cuppen,E. (2005) Phylogenetic shadowing and computational identification of human microRNA genes. *Cell*, **120**, 21–24.
19. Lewis,B.P., Shih,I.H., Jones-Rhoades,M.W., Bartel,D.P. and Burge,C.B. (2003) Prediction of mammalian microRNA targets. *Cell*, **115**, 787–798.
20. Doench,J.G. and Sharp,P.A. (2004) Specificity of microRNA target selection in translational repression. *Genes Dev.*, **18**, 504–511.
21. Brennecke,J., Stark,A., Russell,R.B. and Cohen,S.M. (2005) Principles of microRNA-target recognition. *PLoS Biol.*, **3**, e85.
22. Stark,A., Brennecke,J., Russell,R.B. and Cohen,S.M. (2003) Identification of Drosophila MicroRNA targets. *PLoS Biol.*, **1**, E60.
23. Xie,X., Lu,J., Kulbokas,E.J., Golub,T.R., Mootha,V., Lindblad-Toh,K., Lander,E.S. and Kellis,M. (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature*, **434**, 338–345.
24. Weaver,D.B., Anzola,J.M., Evans,J.D., Reid,J.G., Reese,J.T., Childs,K.L., Zdobnov,E.M., Samanta,M.P., Miller,J. and Elsik,C.G. (2007) Computational and transcriptional evidence for microRNAs in the honey bee genome. *Genome Biol.*, **8**, R97.
25. Bonnet,E., Wuyts,J., Rouze,P. and Van de Peer,Y. (2004) Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics*, **20**, 2911–2917.
26. Washietl,S., Hofacker,I.L. and Stadler,P.F. (2005) Fast and reliable prediction of noncoding RNAs. *Proc. Natl Acad. Sci. USA*, **102**, 2454–2459.
27. Sewer,A., Paul,N., Landgraf,P., Aravin,A., Pfeffer,S., Brownstein,M.J., Tuschl,T., van Nimwegen,E. and Zavolan,M. (2005) Identification of clustered microRNAs using an ab initio prediction method. *BMC Bioinformatics*, **6**, 267.
28. Xue,C., Li,F., He,T., Liu,G.P., Li,Y. and Zhang,X. (2005) Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. *BMC Bioinformatics*, **6**, 310.
29. Nam,J.W., Kim,J., Kim,S.K. and Zhang,B.T. (2006) ProMiR II: a web server for the probabilistic prediction of clustered, nonclustered, conserved and nonconserved microRNAs. *Nucleic Acids Res.*, **34**, W455–W458.
30. Helvik,S.A., Snove,O. Jr. and Saetrom,P. (2007) Reliable prediction of Drosha processing sites improves microRNA gene prediction. *Bioinformatics*, **23**, 142–149.
31. Ng,K.L. and Mishra,S.K. (2007) De novo SVM classification of precursor microRNAs from genomic pseudo hairpins using global and intrinsic folding measures. *Bioinformatics*, **23**, 1321–1330.
32. Jiang,P., Wu,H., Wang,W., Ma,W., Sun,X. and Lu,Z. (2007) MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. *Nucleic Acids Res.*, **35**, W339–W344.
33. Hofacker,I.L., Priwitzer,B. and Stadler,P.F. (2004) Prediction of locally stable RNA secondary structures for genome-wide surveys. *Bioinformatics*, **20**, 186–190.
34. Zhang,B.H., Pan,X.P., Cox,S.B., Cobb,G.P. and Anderson,T.A. (2006) Evidence that miRNAs are different from other RNAs. *Cell Mol. Life Sci.*, **63**, 246–254.
35. Griffiths-Jones,S., Moxon,S., Marshall,M., Khanna,A., Eddy,S.R. and Bateman,A. (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.*, **33**, D121–D124.
36. Li,W. and Godzik,A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.
37. Steffen,P., Voss,B., Rehmsmeier,M., Reeder,J. and Giegerich,R. (2006) RNAshapes: an integrated RNA analysis package based on abstract shapes. *Bioinformatics*, **22**, 500–503.
38. Katoh,K. and Toh,H. (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform.*, **9**, 286–298.
39. Kriventseva,E.V., Rahman,N., Espinosa,O. and Zdobnov,E.M. (2008) OrthoDB: the hierarchical catalog of eukaryotic orthologs. *Nucleic Acids Res.*, **36**, D271–D275.
40. Saebo,P.E., Andersen,S.M., Myrseth,J., Laerdahl,J.K. and Rognes,T. (2005) PARALIGN: rapid and sensitive sequence similarity searches powered by parallel computing technology. *Nucleic Acids Res.*, **33**, W535–W539.

Downloaded from http://nar.oxfordjournals.org/ at University of Geneva on January 17, 2012

41. Wilm,A., Higgins,D.G. and Notredame,C. (2008) R-Coffee: a method for multiple alignment of non-coding RNA. *Nucleic Acids Res.*, **36**, e52.
42. Edgar,R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.
43. Do,C.B., Mahabhashyam,M.S., Brudno,M. and Batzoglou,S. (2005) ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Res.*, **15**, 330–340.
44. Hofacker,I.L., Fekete,M. and Stadler,P.F. (2002) Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.*, **319**, 1059–1066.

Downloaded from http://nar.oxfordjournals.org/ at University of Geneva on January 17, 2012

Downloaded from genome.cshlp.org on December 16, 2010 - Published by Cold Spring Harbor Laboratory Press

Research

Identification of *cis*- and *trans*-regulatory variation modulating microRNA expression levels in human fibroblasts

Christelle Borel,[1] Samuel Deutsch,[1] Audrey Letourneau, Eugenia Migliavacca, Stephen B. Montgomery, Antigone S. Dimas,[2] Charles E. Vejnar, Homa Attar, Maryline Gagnebin, Corinne Gehrig, Emilie Falconnet, Yann Dupré, Emmanouil T. Dermitzakis, and Stylianos E. Antonarakis[3]

Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva 1211, Switzerland

MicroRNAs (miRNAs) are regulatory noncoding RNAs that affect the production of a significant fraction of human mRNAs via post-transcriptional regulation. Interindividual variation of the miRNA expression levels is likely to influence the expression of miRNA target genes and may therefore contribute to phenotypic differences in humans, including susceptibility to common disorders. The extent to which miRNA levels are genetically controlled is largely unknown. In this report, we assayed the expression levels of miRNAs in primary fibroblasts from 180 European newborns of the GenCord project and performed association analysis to identify eQTLs (expression quantitative traits loci). We detected robust expression for 121 miRNAs out of 365 interrogated. We have identified significant *cis*- (10%) and *trans*- (11%) eQTLs. Furthermore, we detected one genomic locus (rs1522653) that influences the expression levels of five miRNAs, thus unraveling a novel mechanism for coregulation of miRNA expression.

[Supplemental material is available online at http://www.genome.org. The miRNA expression data from this study has been submitted to the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under accession no. GSE24610. The genotyping data from this study have been submitted to the EMBL-EBI European Genome-phenome Archive (http://www.ebi.ac.uk/ega/) under accession no. EGAS00000000056.]

The discovery of microRNAs (miRNAs) (19- to 25-nt-long single-stranded RNA molecules) has revealed a new mechanism for the regulation of protein-coding gene expression (Ambros 2004; Bartel 2004; Baek et al. 2008; Selbach et al. 2008). Dosage alterations of miRNA levels are thought to be involved in human disease pathogenesis (Bartel 2004; Kloosterman and Plasterk 2006; Bushati and Cohen 2007; Bartel 2009; Xiao and Rajewsky 2009). One of the least understood aspects of miRNA biogenesis concerns the regulation of its expression levels. Approximately half of the miRNAs identified to date are located in intergenic regions and are therefore likely to possess their own promoter and enhancer elements. The remaining miRNAs map to introns of protein-coding genes and are transcribed from the same strand (Saini et al. 2008). However, it is not yet clear whether these miRNAs are the by-products of protein-coding gene transcription or whether their transcription is controlled by independent regulatory elements. Since miRNA genes are transcribed by RNA polymerase II, it is likely that they share a similar mode of regulation with protein-coding mRNAs.

The goal of this study was to identify genetic variation associated with miRNA levels, as a way to dissect the elements and mechanisms governing miRNA expression.

[1]These authors contributed equally to this work.
[2]Present address: Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, UK.
[3]Corresponding author.
E-mail stylianos.antonarakis@unige.ch.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.109371.110.

Recent genetic analyses have demonstrated that transcription levels of protein-coding genes behave as heritable quantitative traits and display significant associations with genetic variants, including single nucleotide polymorphisms (SNPs) and copy number variants (CNVs) (Morley et al. 2004; Cheung et al. 2005; Deutsch et al. 2005; Stranger et al. 2007; Dermitzakis 2008).

In this study, we conducted an association analysis using mature miRNA expression levels as the primary phenotype, with the aim of identifying regulatory polymorphic variants (expression quantitative traits loci [eQTLs]) significantly associated with miRNA expression levels in human primary fibroblasts.

Results

Primary fibroblasts were derived from the umbilical cord of 180 newborns of western European origin recruited for the GenCord project (see Methods). All samples were genotyped using the Illumina Hap550 SNP array. Mature miRNA expression phenotypes were generated using the micro-fluidics-based TaqMan Human MiRNA Array v1.0 (Applied Biosystems). For each sample, the expression levels for 365 known human mature miRNAs were assayed (Supplemental Table S1). We detected expression above the background for 57% (n = 208) of the miRNAs in cultured primary fibroblasts. These were further filtered to include miRNAs with expression above the background in at least 50% of the samples (n = 90). One hundred twenty-one miRNAs were retained for association analysis.

We identified *cis*-eQTLs, by testing for association between expression levels and SNP genotypes, within 1 Mb 5′ and 3′ of each miRNA. SNPs were considered to be significantly associated with

Downloaded from genome.cshlp.org on December 16, 2010 - Published by Cold Spring Harbor Laboratory Press

miRNA expression levels (i.e., eQTLs) if they passed the 0.05 permutation level threshold for 10,000 permutations (see Methods).

Twelve (i.e., 10%) of the 121 miRNAs tested showed significant evidence for *cis*-regulatory variation (permutated *P*-value < 0.05) (Table 1A; Fig. 1; Supplemental Fig. S1). Given that we tested 121 miRNAs and we expect 5% of them to have significant *cis*-associations by chance at the permutation level 0.05, we estimate that our false-discovery rate (FDR) is about 50% of the 12 miRNA signals.

Examples of these *cis*-eQTLs are shown in Figure 1. The most highly significant *cis*-eQTL detected was rs10750218, intronic to *UBASH3B*, which associates with levels of miR-100 533 kb away (Fig. 1). The distance between the *cis*-eQTLs and their respective miRNA was variable and ranged from 13.6 kb to 886 kb (Table 1A). In one case (miR-218-1), *cis*-eQTLs mapped within the protein-coding sequences of *SLIT2* that also contain the miRNA sequence. This raises the interesting question of whether both the miR-218-1 and the *SLIT2* mRNA share regulatory sequences (Table 1). To address this, we investigated whether the specific *cis*-eQTL for the miRNA was also associated with *SLIT2* mRNA levels. Transcription levels of protein-coding genes were assayed using Illumina's WG-6 v3 Expression BeadChip array (Dimas et al. 2009). We found no evidence of shared regulatory variation between mRNA and miRNA, and no correlation between the miR-218-1 and *SLIT2* mRNA levels was observed (Pearson correlation = −0.023, $n = 55$), implying absence of coregulation of these two transcripts in fibroblasts.

We then aimed to identify *trans*-eQTLs by performing a genome-wide association study (GWAS) for the 121 miRNA expression phenotypes. We observed 18 significant *trans*-eQTLs for 13 miRNAs (10.7%) after Bonferroni correction for multiple testing at the 95% significance level (Table 1B; Supplemental Fig. S2). Since under the null hypotheses we would expect on average six associations, we can estimate our FDR at about 30% for 18 reported miRNAs.

The most significant *trans*-eQTL was detected for miR-140 (chromosome 16) with SNP rs6039847 located on chromosome 20 (unadjusted $P = 1.5 \times 10^{-9}$). The majority of *trans*-eQTLs (72%) mapped to intergenic regions. We detected cases where multiple *trans*-eQTLs, located in different chromosomes, associate with the expression levels of single miRNAs (Table 1B), suggesting that multiple loci may act together to regulate miRNA expression. For example, two significant *trans*-eQTLs were detected for miR-134, the first on chromosome 21 (rs2824791, unadjusted $P = 1 \times 10^{-8}$) and the second on chromosome 3 (rs17533447, unadjusted $P = 3.6 \times 10^{-8}$) (Fig. 2). Similar observations were made for miR-103, miR-130b, miR-29a, and miR-410 (Table 1B; Supplemental Fig. S2). We also observed two cases in which a single SNP was associated with the expression of multiple, unrelated miRNAs: rs1522653 is significantly associated with the expression of miR-103 and miR-29a; rs6039847, with miR-140 and miR-130b (Table 1B).

These observations prompted us to analyze in-depth for the presence of statistically significant miRNA "master regulators," defined as *trans*-eQTLs involved in the regulation of multiple miRNA genes.

To this end, we ascertained for each SNP the number of miRNA associations detected using a reduced stringency (unadjusted *P*-value < 10^{-6}) (Supplemental Table S2). This analysis identified one *trans*-eQTL, rs1522653 on chromosome 11 that was associated with the expression of five miRNAs (miR-15b, miR-26a, miR-29a, miR-30c, and miR-103) (Fig. 3). To determine the significance of this finding, we permuted 1,000 times the expression levels of all miRNAs (preserving the miRNA expression matrix per individual) and performed GWAS for each permuted data set. From this, we estimated the empirical significance of our master regulator to be equal to 0.005 (Fig. 3; see Methods).

Remarkably, rs1522653 is an intergenic SNP, located in a large gene desert (3.29 Mb with no annotated protein-coding or non-coding RNAs); the nearest gene, *FAM181B*, maps 1.59 Mb away (Supplemental Fig. S2). The identification of regulatory variants associated with the expression levels of multiple miRNAs may point to potential "master regulatory" properties and suggests that the expression levels of groups of miRNAs may be coordinated through the use of common regulatory elements. This hypothesis predicts that the five miRNAs associated with rs1522653 should display related expression profiles. To test this hypothesis, we compared the average of the correlation values of the five miRNAs associated with rs1522653 to 10,000 sets of five randomly selected miRNAs. We found that the observed average correlation of 0.44 is higher than that expected by chance (permutated *P*-value of 0.0012) (Supplemental Fig. S3). We also examined whether the predicted target transcripts of the five miRNAs associated with a master regulator share molecular functions. We investigated Gene Ontology (GO) terms from computational target predictions of the five coregulated miRNAs (miRanda [John et al. 2004] from the miRBase-Targets database [Griffiths-Jones et al. 2008]). This analysis revealed that the mRNA targets for these five miRNAs are significantly enriched for "protein-binding process" ($P = 4.4 \times 10^{-8}$, Fisher's exact test), "transcription regulator activity" ($P = 7.8 \times 10^{-8}$), and "transcription factor activity" ($P = 1.2 \times 10^{-6}$) (Supplemental Table S3).

We therefore propose a model in which certain eQTLs act as master regulators by comodulating the expression of multiple miRNAs, thus revealing a novel mechanism for coregulation of miRNA expression.

Discussion

This study provides an initial assessment of the expression level variation of mature human miRNAs and explores how these levels are regulated by common genetic variants in fibroblasts from European individuals. Since we only studied one cell type, the eQTLs identified here are likely to represent a small subset of regulatory variation affecting miRNA levels. Indeed, many miRNAs are expressed in a tissue-restricted manner (Landgraf et al. 2007) and are thus likely to have tissue-specific regulators, as reported recently for protein coding genes (Dimas et al. 2009).

Earlier studies have shown that common genetic variants contribute significantly to the individual differences in protein-coding gene expression variation (Cheung et al. 2003, 2005; Morley et al. 2004; Deutsch et al. 2005; Stranger et al. 2005, 2007; Spielman et al. 2007; Storey et al. 2007) and transcript isoform variation (Hull et al. 2007; Kwan et al. 2007, 2008; Zhang et al. 2009). Our study adds a level of complexity to cellular gene expression regulation by revealing that *cis*- and *trans*-eQTLs can affect the expression of miRNAs that are themselves regulatory molecules. eQTLs identified in this study are potential candidates for the involvement in human phenotypes. Differences in the quantity of mature miRNAs have a clear impact on the level of targeted proteins and result in phenotypic differences (Sethupathy et al. 2007; Baek et al. 2008; Selbach et al. 2008; Bartel 2009). The subsequent identification of the functional variation related to each eQTL type may provide important genomic targets for dissecting the molecular basis of susceptibility to genetic disorders.

Downloaded from genome.cshlp.org on December 16, 2010 - Published by Cold Spring Harbor Laboratory Press

Table 1. Summary of significant *cis*-eQTLs and *trans*-eQTLs detected for miRNA expression variation

	microRNA		Associated SNP							
eQTLs	ID	Location	ID	Distance SNP-miR midpoint (bp)	Location	Unadjusted *P*-value LR[a]	Unadjusted *P*-value SRC (adjusted *P*-value SRC)[b]	Bonferroni adjusted *P*-value	FDR-BH adjusted *P*-value[c]
A. *Cis*-eQTLs									
Cis-2Mb	miR-218_1	Chr 4 (intronic, *SLIT2*)	rs675465	13,614	Chr 4 (intronic, *SLIT2*)	5.4×10^{-6}	2.6×10^{-5} (0.0087)	2.1×10^{-3}	2.1×10^{-3}
Cis-2Mb	miR-594	Chr 7 (annotated tRNA gene)	rs6467784	711,316	Chr 7 (intronic, *SVOPL*)	1.6×10^{-5}	9.3×10^{-5} (0.0267)	5.4×10^{-3}	5.4×10^{-3}
Cis-2Mb	miR-100	Chr 11 (intronic, LOC399959)	rs10750218	533,601	Chr 11 (intronic, *UBASH3B*)	1.7×10^{-5}	1.2×10^{-5} (0.0044)	6.6×10^{-3}	2.2×10^{-3}
Cis-2Mb	miR-125b_1	Chr 11 (intronic, LOC399959)	rs11218891	824,929	Chr 11 (intronic, *c11orf63*)	7.4×10^{-5}	1.2×10^{-4} (0.0355)	2.9×10^{-2}	1.7×10^{-2}
Cis-2Mb	miR-16_2	Chr 3 (intronic, *SMC4*)	rs692890	423,250	Chr 3 (intergenic)	9.1×10^{-5}	1.3×10^{-4} (0.0209)	2.3×10^{-2}	2.3×10^{-2}
Cis-2Mb	miR-654	Chr 14 (intergenic)	rs17099976	146,512	Chr 14 (intergenic)	1.2×10^{-4}	8.9×10^{-5} (0.0223)	4.1×10^{-2}	4.1×10^{-2}
Cis-2Mb	miR-29c	Chr 1 (intergenic)	rs2000059	838,152	Chr 1 (intronic, *FCAMR*)	1.5×10^{-4}	2.2×10^{-5} (0.0053)	5.1×10^{-2}	5.1×10^{-2}
Cis-2Mb	miR-550	Chr 7 (intergenic)	rs12670233	227,301	Chr 7 (intronic, *AVL9*)	2.4×10^{-4}	1.3×10^{-4} (0.0315)	9.6×10^{-2}	4.8×10^{-2}
Cis-2Mb	miR-224	Chr X (intronic, *GABRE*)	rs547043	790,095	Chr X (intergenic)	4.1×10^{-4}	0.2×10^{-3} (0.0494)	1.1×10^{-1}	1.1×10^{-1}
Cis-2Mb	miR-320	Chr 8 (intergenic)	rs7014043	886,085	Chr 8 (intergenic)	1.1×10^{-3}	4.1×10^{-5} (0.0145)	4.7×10^{-1}	1.6×10^{-1}
Cis-2Mb	miR-660	Chr X (intronic, *CLCN5*)	rs4554617	555,255	Chr X (intronic,*DGKK*)	3.5×10^{-3}	1.1×10^{-3} (0.0458)	2.6×10^{-1}	1.0×10^{-1}
Cis-2Mb	miR-24	Chr 9 (intronic, *C9orf3*)	rs600130	509,590	Chr 9 (intronic,*FBP2*)	4.3×10^{-3}	1.8×10^{-4} (0.033)	1	4.0×10^{-1}
B. *Trans*-eQTLs									
Trans	miR-140	Chr 16 (intronic, *WWP2*)	rs6039847	Interchromosomal	Chr 20 (intergenic)	1.5×10^{-9}	3.0×10^{-2}	7.4×10^{-4}	7.4×10^{-4}
Trans	miR-98	Chr X (intronic, *HUWE1*)	rs11150154	Interchromosomal	Chr 16 (intergenic)	1.5×10^{-9}	3.1×10^{-3}	7.4×10^{-4}	7.4×10^{-4}
Trans	miR-92a	Chr 13 (intronic, *C13orf25*) or chr X (intergenic)	rs12324904	Interchromosomal	Chr 15 (intronic, *ZNF710*)	2.9×10^{-9}	9.2×10^{-3}	1.3×10^{-3}	1.3×10^{-3}
Trans	miR-130b	Chr 22 (intergenic)	rs6039847	Interchromosomal	Chr 20 (intergenic)	4.2×10^{-9}	7.3×10^{-2}	1.9×10^{-3}	1.9×10^{-3}
Trans	miR-134	Chr 14 (intergenic)	rs2824791	Interchromosomal	Chr 21 (intronic, *PRSS7*)	1.0×10^{-8}	2.4×10^{-7}	4.8×10^{-3}	4.8×10^{-3}
Trans	miR-24	Chr 9 (intronic, *C9orf3*) or chr 19 (intergenic)	rs16931830	Interchromosomal	Chr 10 (intergenic)	1.2×10^{-8}	7.0×10^{-4}	6.0×10^{-3}	6.0×10^{-3}
Trans	miR-130b	Chr 22 (intergenic)	rs13334253	Interchromosomal	Chr 16 (intergenic)	1.3×10^{-8}	5.9×10^{-4}	6.2×10^{-3}	3.1×10^{-3}
Trans	miR-221	Chr X (intergenic)	rs10275283	Interchromosomal	Chr 7 (intergenic)	2.0×10^{-8}	1.7×10^{-6}	9.3×10^{-3}	9.3×10^{-3}
Trans	miR-425	Chr 3 (intronic, *DALRD3*)	rs7859900	Interchromosomal	Chr 9 (intergenic)	2.0×10^{-8}	1.2×10^{-7}	9.3×10^{-3}	9.3×10^{-3}
Trans	miR-29a	Chr 7 (intergenic)	rs1522653	Interchromosomal	Chr 11 (intergenic)	2.1×10^{-8}	8.6×10^{-8}	1.0×10^{-2}	1.0×10^{-2}
Trans	miR-134	Chr 14 (intergenic)	rs17533447	Interchromosomal	Chr 3 (intronic, *NAALADL2*)	3.6×10^{-8}	7.9×10^{-6}	1.7×10^{-2}	8.6×10^{-3}
Trans	miR-99a	Chr 21 (intronic, *C21orf34*)	rs4829489	Interchromosomal	Chr X (intronic, *LHFPL1*)	3.7×10^{-8}	7.6×10^{-3}	1.8×10^{-2}	1.7×10^{-2}
Trans	miR-410	Chr 14 (intergenic)	rs4751986	Interchromosomal	Chr 10 (intergenic)	4.3×10^{-8}	2.1×10^{-4}	2.1×10^{-2}	2.0×10^{-2}
Trans	miR-191	Chr 3 (intronic, *DALRD3*)	rs12479616	Interchromosomal	Chr 20 (intergenic)	5.3×10^{-8}	6.6×10^{-7}	2.5×10^{-2}	2.5×10^{-2}
Trans	miR-103	Chr 5(intronic, *PANK3*) or chr 20 (intronic, *PANK2*)	rs1522653	Interchromosomal	Chr 11 (intergenic)	5.6×10^{-8}	2.3×10^{-6}	2.7×10^{-2}	2.1×10^{-2}
Trans	miR-29a	Chr 7 (intergenic)	rs396146	Interchromosomal	Chr 2 (intergenic)	6.3×10^{-8}	1.4×10^{-5}	3.0×10^{-2}	1.5×10^{-2}
Trans	miR-410	Chr 14 (intergenic)	rs278977	Interchromosomal	Chr 4 (intergenic)	9.0×10^{-8}	1.2×10^{-3}	4.3×10^{-2}	2.1×10^{-2}
Trans	miR-103	Chr 5(intronic, *PANK3*) or chr 20 (intronic, *PANK2*)	rs8063973	Interchromosomal	Chr 16 (intronic, *ATP2C2*)	9.1×10^{-8}	8.6×10^{-7}	4.4×10^{-2}	2.1×10^{-2}

[a]Unadjusted *P*-value using linear regression (LR).
[b]Unadjusted *P*-value using Spearman's rank correlation (SRC) and adjusted *P*-value based on 10,000 permutations (in parentheses).
[c]FDR-BH adjusted *P*-value indicates Benjamini-Hochberg false discovery rate.

Downloaded from genome.cshlp.org on December 16, 2010 - Published by Cold Spring Harbor Laboratory Press

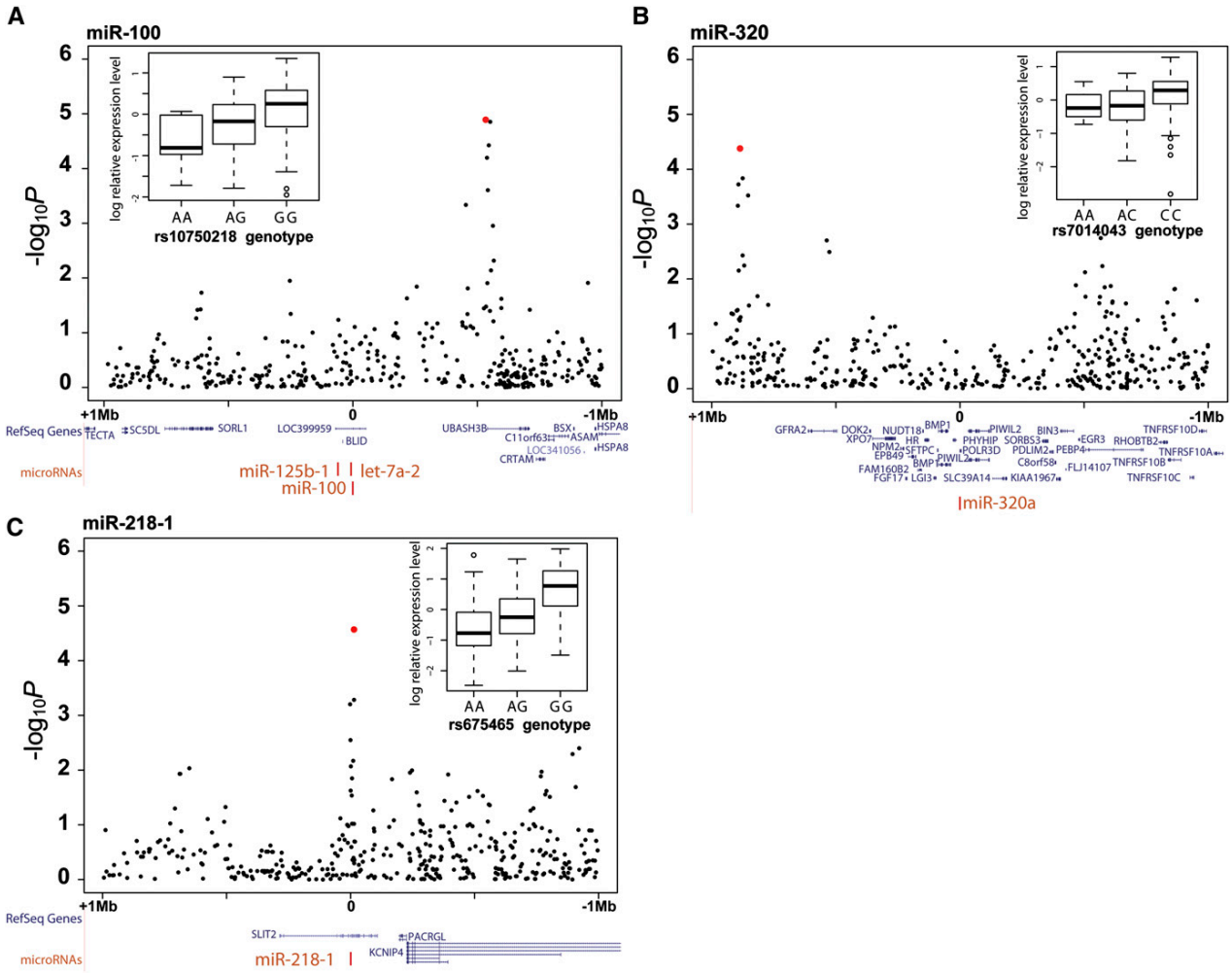


Figure 1. Examples of *cis*-eQTLs for miR-100 (*A*), miR-320 (*B*), and miR-218-1 (*C*). The panels show the distribution of $-\log_{10}$ *P*-values for SNPs across a 1-Mb region surrounding the miRNA ("0" position). The highest significant $-\log_{10}$ *P*-values are shown as red dots. Also shown are the mapping of RefSeq genes in blue and miRNAs in red. The boxplots depict the relationship between miRNA relative expression levels ($\log_2$) and genotypes for the most significant SNPs. Boxplots are divided by median values.

Methods

Cell culture and RNA preparation

We obtained primary fibroblasts from 180 individuals of the GenCord project. This collection was established from umbilical cords of newborns of western European origin (following appropriate informed consent and approval by the Geneva University Hospital's ethics committee). All cell lines were grown in DMEM with Glutamax I (Invitrogen) supplemented with 10% fetal calf serum (Invitrogen) and 1% penicillin/streptomycin/fungizone mix (Amimed, BioConcept) at 37°C and 5% CO_2. Confluent cell lines were trypsinized and diluted at a density of 7×10^5 cells/mL (40% of confluence) and harvested the following day. Total RNA was isolated using TRIzol (Invitrogen) according to the manufacturer's instructions. RNA quality was assessed using RNA 6000 NanoChips with the Agilent 2100 Bioanalyzer (Agilent), and RNA was quantified with a NanoDrop spectrophotometer (NanoDrop Technologies).

miRNA expression measurement and data normalization

Expression of 365 known human miRNAs was analyzed using the TaqMan Human MiRNA Array v1.0 early access (Applied Biosystems), according to the manufacturer's instructions. Briefly, 800 ng of total RNA samples was used as template for eight multiplex reverse transcriptions containing up to 48 specific primers, using the Multiplex RT for TaqMan miRNA Assays Kit (Applied Biosystems) under conditions defined by the supplier. Each cDNA generated was amplified by quantitative PCR using 365 sequence-specific primers from the TaqMan miRNA Assays Human Panel on an Applied Biosystems 7900 Fast Real Time PCR system. Absolute threshold cycle values (Ct) were determined with the SDS 2.2 software (Applied Biosystems). A threshold value was determined for each miRNA and used for all the 180 samples. All signals with a Ct value of ≥34 (background threshold) were manually set to undetermined. Indeed, we considered miRNA with a Ct value of <34 as an "expressed miRNA." Values were normalized across individuals using median normalization and were reported as an

Downloaded from genome.cshlp.org on December 16, 2010 - Published by Cold Spring Harbor Laboratory Press

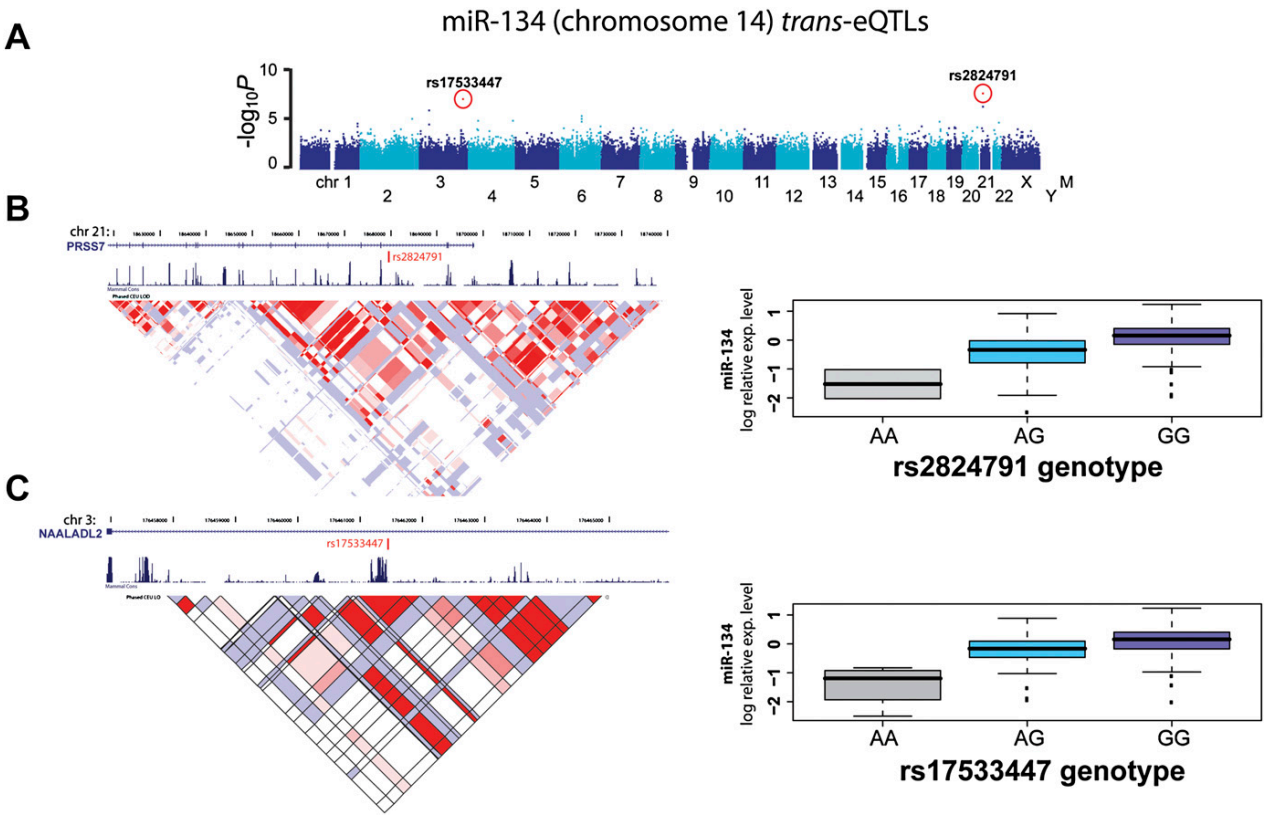


Figure 2. Example of *trans*- eQTLs for miR-134. (*A*) Manhattan plot displays $-\log_{10}$ *P*-values of a GWAS for miR-134 expression variation. Each chromosome is depicted as different shades of blue. Chromosome Y and mitochondrial genotypes have not been included in this study. Red dots indicate SNPs with the highest significant *P*-values after Bonferroni correction for multiple testing (see Table 1). Those two eQTLs, rs2824791 on chromosome 21 and rs7533447 on chromosome 3, are detailed in panels *B* and *C*, respectively. (*B*,*C*, *left* panels) The location of associated SNPs, as well as RefSeq transcripts, conservation, and LD information (LOD scores for CEU population). (*B*,*C*, *right* panels) Boxplots for miRNA expression for different genotypic groups.

expression relative to the population mean for each miRNA as described (Deutsch et al. 2005; Prandini et al. 2007). Log_2 values were used for the association analysis. TaqMan miRNA data sets have been submitted to the NCBI Gene Expression Omnibus (GEO) database under accession number GSE24610.

Genotyping

Genotyping was performed using the Illumina Hap550 or Hap550-duo arrays. Genotype calling was performed using the BeadStudio 3.1 software. SNPs were filtered in a stepwise fashion using the following criteria: (1) a SNP call frequency of at least 99%, (2) cluster separation greater than 0.3, (3) SNPs with Het Excess values between [−1.0 to −0.1] and [0.1 to 1.0] were removed, (4) SNPs that violate Hardy-Weinberg equilibrium (HWE = $P < 0.05$) were removed, (5) SNPs with a minimum allele frequency (MAF) < 0.02 were removed (at least seven heterozygous in our sample). After filtering, 479,314 SNPs were retained for statistical analyses. Genotyping data sets have been submitted to the European Genome-phenome Archive (EGA) database under accession number EGAS00000000056.

Genome-wide and *cis*-association analysis

eQTLs were detected using linear regression as implemented in the PLINK package (Purcell et al. 2007). For the *cis*-analysis, the association of genotype with expression levels was calculated for each miRNA within a 2-Mb window around its transcription start site (1 Mb either side). Association was also calculated using Spearman's rank correlation and was compared to the extreme

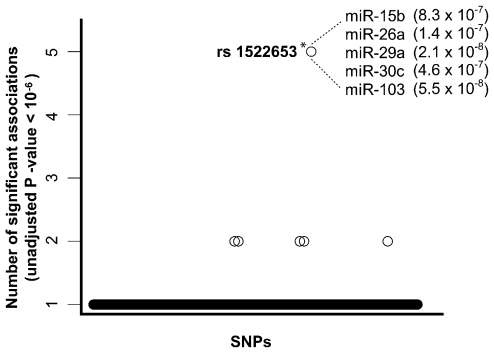


Figure 3. Master miRNA *trans*-eQTLs. Plot shows SNPs associated with the expression variation of multiple miRNAs (using a threshold of an unadjusted *P*-value < 10^{-6} per association) (see Supplemental Table S2). Each circle represents a single SNP. Only SNPs with at least one association below the $P < 10^{-6}$ threshold are shown. One SNP (rs1522653) is significantly associated with the expression of five miRNAs (*, permutated *P*-value of 0.005). The identities and unadjusted *P*-values for these miRNAs are shown.

Downloaded from genome.cshlp.org on December 16, 2010 - Published by Cold Spring Harbor Laboratory Press

P-value distribution of similar associations calculated for 10,000 permutations of the expression phenotype for each miRNA (permutation threshold) as previously reported (Stranger et al. 2007; Dimas et al. 2009). We applied a permutation threshold of 0.05 per gene, and we subsequently estimated the FDR on our number of discoveries based on the fact that we expected 5% of the miRNA genes to have a significant signal under the null. This design, which we have extensively applied in the past (Stranger et al. 2005, 2007; Bartel 2009; Dimas et al. 2009; Montgomery et al. 2010), allows for simultaneous assessment of the multiple testing effect of all markers tested within a 2-Mb window as well as across all phenotypes tested. For visualization and graphical displays, we used WGAviewer (Ge et al. 2008).

Gene Ontology annotation analysis

Analysis were conducted using Bioconductor GO stats version 2.8.0 and annotation Ms.eg.db version 2.2.6 packages (FDR adjusted *P*-value < 0.05) (Falcon and Gentleman 2007).

Expression clustering analysis

Hierarchical clustering was performed using Pearson correlation as a similarity measure and average linkage as an agglomerative hierarchical clustering algorithm.

Statistical analysis for master regulator identification

We tested for each SNP how many miRs were associated using an unadjusted *P*-value < 10^{-6}. To estimate the significance for our findings, we permuted 1000 times the miR expression phenotypes (preserving the miR expression matrix per individual) and performed GWAS for each permuted data set.

Acknowledgments

We thank P. Descombes and the members of the genomics platform of the University of Geneva for their assistance, A.J. Sharp for comments on the manuscript, and Vital-IT for computational support. This study was funded by the Swiss National Science Foundation, the National Center for Competence in Research (NCCR) "Frontiers in Genetics," the European Union FP6 "AnEUploidy" integrated project, and the Infectigen, ChildCare, and J. Lejeune foundations.

References

Ambros V. 2004. The functions of animal microRNAs. *Nature* **431:** 350–355.
Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. 2008. The impact of microRNAs on protein output. *Nature* **455:** 64–71.
Bartel DP. 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116:** 281–297.
Bartel DP. 2009. MicroRNAs: Target recognition and regulatory functions. *Cell* **136:** 215–233.
Bushati N, Cohen SM. 2007. microRNA functions. *Annu Rev Cell Dev Biol* **23:** 175–205.
Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen KY, Morley M, Spielman RS. 2003. Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat Genet* **33:** 422–425.
Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT. 2005. Mapping determinants of human gene expression by regional and genome-wide association. *Nature* **437:** 1365–1369.
Dermitzakis ET. 2008. From gene expression to disease risk. *Nat Genet* **40:** 492–493.
Deutsch S, Lyle R, Dermitzakis ET, Attar H, Subrahmanyan L, Gehrig C, Parand L, Gagnebin M, Rougemont J, Jongeneel CV, et al. 2005. Gene expression variation and expression quantitative trait mapping of human chromosome 21 genes. *Hum Mol Genet* **14:** 3741–3749.
Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M, et al. 2009. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325:** 1246–1250.
Falcon S, Gentleman R. 2007. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23:** 257–258.
Ge D, Zhang K, Need AC, Martin O, Fellay J, Urban TJ, Telenti A, Goldstein DB. 2008. WGAViewer: Software for genomic annotation of whole genome association studies. *Genome Res* **18:** 640–643.
Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. 2008. miRBase: Tools for microRNA genomics. *Nucleic Acids Res* **36:** D154–D158.
Hull J, Campino S, Rowlands K, Chan MS, Copley RR, Taylor MS, Rockett K, Elvidge G, Keating B, Knight J, et al. 2007. Identification of common genetic variation that modulates alternative splicing. *PLoS Genet* **3:** e99. doi: 10.1371/journal.pgen.0030099.
John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. 2004. Human microRNA targets. *PLoS Biol* **2:** e363. doi: 10.1371/journal.pbio.0020363.
Kloosterman WP, Plasterk RH. 2006. The diverse functions of microRNAs in animal development and disease. *Dev Cell* **11:** 441–450.
Kwan T, Benovoy D, Dias C, Gurd S, Serre D, Zuzan H, Clark TA, Schweitzer A, Staples MK, Wang H, et al. 2007. Heritability of alternative splicing in the human genome. *Genome Res* **17:** 1210–1218.
Kwan T, Benovoy D, Dias C, Gurd S, Provencher C, Beaulieu P, Hudson TJ, Sladek R, Majewski J. 2008. Genome-wide analysis of transcript isoform variation in humans. *Nat Genet* **40:** 225–231.
Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, et al. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129:** 1401–1414.
Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET. 2010. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464:** 773–777.
Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. 2004. Genetic analysis of genome-wide variation in human gene expression. *Nature* **430:** 743–747.
Prandini P, Deutsch S, Lyle R, Gagnebin M, Delucinge Vivier C, Delorenzi M, Gehrig C, Descombes P, Sherman S, Dagna Bricarelli F, et al. 2007. Natural gene-expression variation in Down syndrome modulates the outcome of gene-dosage imbalance. *Am J Hum Genet* **81:** 252–263.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81:** 559–575.
Saini HK, Enright AJ, Griffiths-Jones S. 2008. Annotation of mammalian primary microRNAs. *BMC Genomics* **9:** 564. doi: 10.1186/1471-2164-9-564.
Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455:** 58–63.
Sethupathy P, Borel C, Gagnebin M, Grant GR, Deutsch S, Elton TS, Hatzigeorgiou AG, Antonarakis SE. 2007. Human microRNA-155 on chromosome 21 differentially interacts with its polymorphic target in the AGTR1 3′ untranslated region: A mechanism for functional single-nucleotide polymorphisms related to phenotypes. *Am J Hum Genet* **81:** 405–413.
Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG. 2007. Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet* **39:** 226–231.
Storey JD, Madeoy J, Strout JL, Wurfel M, Ronald J, Akey JM. 2007. Gene-expression variation within and among human populations. *Am J Hum Genet* **80:** 502–509.
Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis SE, Tavare S, et al. 2005. Genome-wide associations of gene expression variation in humans. *PLoS Genet* **1:** e78. doi: 10.1371/journal.pgen.0010078.
Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, et al. 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315:** 848–853.
Xiao C, Rajewsky K. 2009. MicroRNA control in the immune system: Basic principles. *Cell* **136:** 26–36.
Zhang W, Duan S, Bleibel WK, Wisel SA, Huang RS, Wu X, He L, Clark TA, Chen TX, Schweitzer AC, et al. 2009. Identification of common genetic variants that account for transcript isoform variation between human populations. *Hum Genet* **125:** 81–93.

Received April 20, 2010; accepted in revised form November 1, 2010.

miRmap Documentation

Release 1.0

Charles E. Vejnar

April 18, 2012

CONTENTS

ii

`miRmap` library is Python library organized with ...

The `miRmap` library is a Python library predicting the repression strength of microRNA (miRNA) targets. The model combines:

- **thermodynamic** features: ΔG *duplex*, ΔG *binding*, ΔG *open* and ΔG *total*,
- **evolutionary** features: *BLS* and *PhyloP*,
- **probabilistic** features: *P.over binomial* and *P.over exact*, and
- **sequence-based** features: *AU content*, *UTR position* and *3' pairing*.

CHAPTER

ONE

DOWNLOAD

`miRmap` distribution is available http://cegg.unige.ch/mirmap/miRmap-1.0.tar.gz.

Note: To the reviewers.

This section is temporary.

After acceptance of the article, we will open the public repository hosted on BitBucket http://dev.vejnar.org/mirmap. The source code will be separated from the binaries. Features of BitBuket, including bugs and issues trackings, wiki, etc... will be available at the same time.

CHAPTER

TWO

INSTALLATION

2.1 Requirements

The `miRmap` library has the following requirements:

1. `miRmap` requires Python 2.7 but it can be used with Python 2.6 if the `collections` module is installed (A version compatible with Python 2.4-2.6 is available as the ordereddict module.).
2. For the evolutionary features, the Python library `DendroPy` is needed for tree manipulation. You can install DendroPy directly from the Python Package Index.
3. C librairies. A compiled version of the 3 libraries (*.so) is included in the `miRmap` distribution. If you want/have to compile them, please follow these intructions:
   - For the thermodynamic features, the Vienna RNA library is required.

     Download the latest Vienna RNA tarball (Versions 1.8.x were successfully tested), then do:

     ```
     cd ViennaRNA-<version>
     ./configure --without-kinfold --without-forester --without-svm --without-perl
     make
     gcc -shared -Wl,-O2 -o lib/libRNAvienna.so `find lib/ -name "*.o"` -lm
     ```

   - For the evolutionary features, the PHAST library is required (The CLAPACK has to be compiled first, please follow the instructions in Phast package).

     ```
     svn co http://compgen.bscb.cornell.edu/svnrepo/phast/trunk phast
     cd phast/src
     ```

     In the file *make-include.mk*, add the `-DUSE_PHAST_MEMORY_HANDLER` parameter to the line starting with `CFLAGS += -I${INC} -DPHAST_VERSION=${PHAST_VERSION}`. Then replace the path to the CLAPACK and compile with:

     ```
     make CLAPACKPATH=../CLAPACK-3.2.1 sharedlib
     ```

   - For the *P.over exact* feature, the Spatt library is required (You will need a working copy of CMake on your system).

     Download the latest Spatt tarball (Version 2.0 was successfully tested), then do:

     ```
     cd spatt-<version>
     mkdir build
     cd build
     cmake -DWITH_SHARED_LIB=ON ..
     make
     ```

From the directory you compiled the C libraries:

```
mv spatt-<version>/libspatt2/libspatt2.so mirmap/libs/default
mv ViennaRNA-<version>/lib/libRNAvienna.so mirmap/libs/default
mv phast/lib/sharedlib/libphast.so mirmap/libs/default
```

CHAPTER

THREE

USAGE

Example with the pure Python features.

```
>>> import mirmap
>>> seq_target = 'GCUACAGUUUUUAUUUAGCAUGGGGAUUGCAGAGUGACCAGCACACUGGACUCCGAGGUGGUUCAGACAAGACAGAGGGGAGC
... UCCCGCCAGGAGCUUCUUCGUUCCUGCGCAUAUAGACUGUACAUUAUGAAGAAUACCCAGGAAGACUUUGUGACUGUCACUUGCUGCUUUUUCUGCC
... GUUGGCAAACGAGACUUUCUCCUGGCCCCUGCCUGCUGGAGAUCAGCAUGCCUGUCCUUUCAGUCUGAUCCAUCCAUCUCUCUCUUGCCUGAGGGGA
... AGGCAGAGAACAGAACUGGAGGCAGUCCAUCUA'
>>> seq_mirna = 'UAGCAGCACGUAAAUAUUGGCG'
>>> mim = mirmap.mm(seq_target, seq_mirna)
>>> mim.find_potential_targets_with_seed(allowed_lengths=[6,7], allowed_gu_wobbles={6:0,7:0},\
... allowed_mismatches={6:0,7:0}, take_best=True)
>>> mim.end_sites                                          # Coordinate(s) (3' end) of the target site on t
[186]
>>> mim.eval_tgs_au(with_correction=False)                 # TargetScan features manually evaluated with
>>> mim.eval_tgs_pairing3p(with_correction=False)          # a non-default parameter.
>>> mim.eval_tgs_position(with_correction=False)
>>> mim.prob_binomial                                      # mim's attribute: the feature is automatically
0.03311825751646191
>>> print mim.report()
155                            186
|                              |
CAGGAAGACUUUGUGACUGUCACUUGCUGCUUUUUCUGCGCU
                        |||||||.
          GCGGUUAUAAAUGCACGACGAU
  AU content                     0.64942
  UTR position                   166.00000
  3' pairing                     1.00000
  Probability (Binomial)         0.03312
```

With the C libraries installed:

```
>>> import mirmap.library_link
>>> mim.libs = mirmap.library_link.LibraryLink('libs/compiled') # Change to the path where you unzipp
>>> mim.dg_duplex
-13.5
>>> mim.dg_open
12.180591583251953
>>> mim.prob_exact
0.06798900807193115
>>> print mim.report()
155                            186
|                              |
CAGGAAGACUUUGUGACUGUCACUUGCUGCUUUUUCUGCGCU
                        |||||||.
```

```
        GCGGUUAUAAAUGCACGACGAU
ΔG duplex (kcal/mol)          -13.50000
ΔG binding (kcal/mol)         -11.91708
ΔG open (kcal/mol)             12.18059
AU content                     0.64942
UTR position                   166.00000
3′ pairing                     1.00000
Probability (Exact)            0.06799
Probability (Binomial)         0.03312
```

CHAPTER

FOUR

CLASSES

mm and mmPP base classes of miRmap that inherit their methods from all the modules. Each module define the methods for one category.

class mirmap.**mm**(*target_seq*, *mirna_seq*, *min_target_length=None*)

Bases: mirmap.evolution.mmEvolution, mirmap.model.mmModel, mirmap.prob_binomial.mmProbBinomial, mirmap.prob_exact.mmProbExact, mirmap.report.mmReport, mirmap.thermo.mmThermo, mirmap.targetscan.mmTargetScan

miRNA and mRNA containing class.

Parameters

- **target_seq** (*str*) – Target sequence (mRNA).
- **mirna_seq** (*str*) – miRNA sequence.
- **min_target_length** (*int*) – Target site length, base-pairing independent.

eval_cons_bls(*aln_fname=None*, *aln=None*, *aln_format=None*, *aln_alphabet=None*, *subst_model=None*, *tree=None*, *fitting_tree=None*, *use_em=None*, *libphast=None*, *motif_def=None*, *motif_upstream_extension=None*, *motif_downstream_extension=None*)

Computes the Branch Length Score (*BLS*).

Parameters

- **aln_fname** (*str*) – Alignment filename.
- **aln** (*str*) – Alignment it-self.
- **aln_format** (*str*) – Alignment format. Currently supported is FASTA.
- **aln_alphabet** (*list*) – List of nucleotides to consider in the aligned sequences (others get filtered).
- **subst_model** (*str*) – PhyloFit substitution model (REV...).
- **tree** (*str*) – Tree in the Newick format.
- **fitting_tree** (*bool*) – Fitting or not the tree on the alignment.
- **use_em** (*bool*) – Fitting or not the tree with Expectation-Maximization algorithm.
- **libphast** (LibraryLink) – Link to the Phast library.
- **motif_def** (*str*) – ‘seed’ or ‘seed_extended’ or ‘site’.
- **motif_upstream_extension** (*int*) – Upstream extension length.
- **motif_downstream_extension** (*int*) – Downstream extension length.

eval_dg_duplex(*librna=None*, *mirna_start_pairing=None*)
Computes the ΔG *duplex* and ΔG *binding* scores.

Parameters

- **librna** (`LibraryLink`) – Link to the Vienna RNA library.
- **mirna_start_pairing** (*int*) – Starting position of the seed in the miRNA (from the 5').

eval_dg_open(*librna=None*, *upstream_rest=None*, *downstream_rest=None*, *dg_binding_area=None*)
Computes the ΔG *open* score.

Parameters

- **librna** (`LibraryLink`) – Link to the Vienna RNA library.
- **upstream_rest** (*int*) – Upstream unfolding length.
- **downstream_rest** (*int*) – Downstream unfolding length.
- **dg_binding_area** (*int*) – Supplementary sequence length to fold (applied twice: upstream and downstream).

eval_dg_total()
Computes the ΔG *total* score combining ΔG *duplex* and ΔG *open* scores.

eval_prob_binomial(*markov_order=None*, *alphabet=None*, *transitions=None*, *motif_def=None*, *motif_upstream_extension=None*, *motif_downstream_extension=None*)
Computes the *P.over binomial* score.

Parameters

- **markov_order** (*int*) – Markov Chain order
- **alphabet** (*list*) – List of nucleotides to consider in the sequences (others get filtered).
- **transitions** (*list*) – Transition matrix of the Markov Chain model
- **motif_def** (*str*) – 'seed' or 'seed_extended' or 'site'.
- **motif_upstream_extension** (*int*) – Upstream extension length.
- **motif_downstream_extension** (*int*) – Downstream extension length.

eval_prob_exact(*libspatt=None*, *markov_order=None*, *alphabet=None*, *transitions=None*, *motif_def=None*, *motif_upstream_extension=None*, *motif_downstream_extension=None*)
Computes the *P.over binomial* score.

Parameters

- **libspatt** (`LibraryLink`) – Link to the Spatt library.
- **markov_order** (*int*) – Markov Chain order
- **alphabet** (*list*) – List of nucleotides to consider in the sequences (others get filtered).
- **transitions** (*list*) – Transition matrix of the Markov Chain model
- **motif_def** (*str*) – 'seed' or 'seed_extended' or 'site'.
- **motif_upstream_extension** (*int*) – Upstream extension length.
- **motif_downstream_extension** (*int*) – Downstream extension length.

eval_selec_phylop(*aln_fname=None*, *aln=None*, *aln_format=None*, *aln_alphabet=None*, *mod_fname=None*, *libphast=None*, *method=None*, *mode=None*, *motif_def=None*, *motif_upstream_extension=None*, *motif_downstream_extension=None*)

Computes the *PhyloP* score.

Parameters

- **aln_fname** (*str*) – Alignment filename.
- **aln** (*str*) – Alignment it-self.
- **aln_format** (*str*) – Alignment format. Currently supported is FASTA.
- **aln_alphabet** (*list*) – List of nucleotides to consider in the aligned sequences (others get filtered).
- **mod_fname** (*str*) – Model filename.
- **libphast** (`LibraryLink`) – Link to the Phast library.
- **method** (*str*) – Test name performed by PhyloP (SPH...).
- **mode** (*str*) – Testing for conservation (CON), acceleration (ACC) or both (CONACC).
- **motif_def** (*str*) – 'seed' or 'seed_extended' or 'site'.
- **motif_upstream_extension** (*int*) – Upstream extension length.
- **motif_downstream_extension** (*int*) – Downstream extension length.

eval_tgs_au(*ts_types=None*, *ca_window_length=None*, *with_correction=None*)

Computes the *AU content* score.

Parameters

- **ts_types** (*object*) – Parameters by seed-type.
- **ca_window_length** (*int*) – Sequence length to compute the score with.
- **with_correction** (*bool*) – Apply the linear regression correction or not.

eval_tgs_pairing3p(*ts_types=None*, *with_correction=None*)

Computes the *3' pairing* score.

Parameters

- **ts_types** (*object*) – Parameters by seed-type.
- **with_correction** (*bool*) – Apply the linear regression correction or not.

eval_tgs_position(*ts_types=None*, *with_correction=None*)

Computes the *UTR position* score.

Parameters

- **ts_types** (*object*) – Parameters by seed-type.
- **with_correction** (*bool*) – Apply the linear regression correction or not.

eval_tgs_score(*ts_types=None*, *with_correction=None*)

Computes the *TargetScan* score combining *AU content*, *UTR position* and *3' pairing* scores.

Parameters

- **ts_types** (*object*) – Parameters by seed-type.
- **with_correction** (*bool*) – Apply the linear regression correction or not.

find_potential_targets_with_seed(*mirna_start_pairing=None*, *allowed_lengths=None*, *allowed_gu_wobbles=None*, *allowed_mismatches=None*, *take_best=None*)

Searches for seed(s) in the target sequence.

Parameters

- **mirna_start_pairing** (*int*) – Starting position of the seed in the miRNA (from the 5').
- **allowed_lengths** (*list*) – List of seed length(s).
- **allowed_gu_wobbles** (*dict*) – For each seed length (key), how many GU wobbles are allowed (value).
- **allowed_mismatches** (*dict*) – For each seed length (key), how many mismatches are allowed (value).
- **take_best** (*bool*) – If seed matches are overlapping, taking or not the longest.

report()

Returns a formatted report of already computed features for all target site(s).

cons_bls

Branch Length Score (*BLS*) with default parameters.

dg_duplex

ΔG *duplex* score with default parameters.

dg_open

ΔG *open* score with default parameters.

dg_total

ΔG *total* score with default parameters.

prob_binomial

P.over binomial score with default parameters.

prob_exact

P.over exact score with default parameters.

selec_phylop

PhyloP score with default parameters.

tgs_score

TargetScan score with default parameters.

class mirmap.**mmPP**(*target_seq*, *mirna_seq*, *min_target_length=None*)

Bases: mirmap.model.mmModel, mirmap.prob_binomial.mmProbBinomial, mirmap.report.mmReport, mirmap.targetscan.mmTargetScan

miRNA and mRNA containing class with pure Python methods only.

Parameters

- **target_seq** (*str*) – Target sequence (mRNA).
- **mirna_seq** (*str*) – miRNA sequence.
- **min_target_length** (*int*) – Target site length, base-pairing independent.

eval_prob_binomial(*markov_order=None*, *alphabet=None*, *transitions=None*, *motif_def=None*, *motif_upstream_extension=None*, *motif_downstream_extension=None*)

Computes the *P.over binomial* score.

Parameters

- **markov_order** (*int*) – Markov Chain order
- **alphabet** (*list*) – List of nucleotides to consider in the sequences (others get filtered).
- **transitions** (*list*) – Transition matrix of the Markov Chain model
- **motif_def** (*str*) – 'seed' or 'seed_extended' or 'site'.
- **motif_upstream_extension** (*int*) – Upstream extension length.
- **motif_downstream_extension** (*int*) – Downstream extension length.

eval_tgs_au(*ts_types=None*, *ca_window_length=None*, *with_correction=None*)
Computes the *AU content* score.

Parameters

- **ts_types** (*object*) – Parameters by seed-type.
- **ca_window_length** (*int*) – Sequence length to compute the score with.
- **with_correction** (*bool*) – Apply the linear regression correction or not.

eval_tgs_pairing3p(*ts_types=None*, *with_correction=None*)
Computes the *3' pairing* score.

Parameters

- **ts_types** (*object*) – Parameters by seed-type.
- **with_correction** (*bool*) – Apply the linear regression correction or not.

eval_tgs_position(*ts_types=None*, *with_correction=None*)
Computes the *UTR position* score.

Parameters

- **ts_types** (*object*) – Parameters by seed-type.
- **with_correction** (*bool*) – Apply the linear regression correction or not.

eval_tgs_score(*ts_types=None*, *with_correction=None*)
Computes the *TargetScan* score combining *AU content*, *UTR position* and *3' pairing* scores.

Parameters

- **ts_types** (*object*) – Parameters by seed-type.
- **with_correction** (*bool*) – Apply the linear regression correction or not.

find_potential_targets_with_seed(*mirna_start_pairing=None*, *allowed_lengths=None*, *allowed_gu_wobbles=None*, *allowed_mismatches=None*, *take_best=None*)
Searches for seed(s) in the target sequence.

Parameters

- **mirna_start_pairing** (*int*) – Starting position of the seed in the miRNA (from the 5').
- **allowed_lengths** (*list*) – List of seed length(s).
- **allowed_gu_wobbles** (*dict*) – For each seed length (key), how many GU wobbles are allowed (value).
- **allowed_mismatches** (*dict*) – For each seed length (key), how many mismatches are allowed (value).
- **take_best** (*bool*) – If seed matches are overlapping, taking or not the longest.

report()
Returns a formatted report of already computed features for all target site(s).

prob_binomial
P.over binomial score with default parameters.

tgs_score
TargetScan score with default parameters.

class mirmap.library_link.**LibraryLink**(*library_path=None*)

Parameters **library_path** (*str*) – Path to the C dynamic libraries.

CHAPTER

FIVE

COPYRIGHT, LICENSE AND WARRANTY

The miRmap library is:

Copyright 2011 Charles E. Vejnar

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License version 3 as published by the Free Software Foundation.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program (in /LICENCE). If not, see GNU Licenses.

PYTHON MODULE INDEX

INDEX