ARTICLES

The impact of microRNAs on protein output

Daehyun Baek[1,2]*, Judit Villén[3]*, Chanseok Shin[1,2]*, Fernando D. Camargo[1], Steven P. Gygi[3] & David P. Bartel[1,2]

MicroRNAs are endogenous ~23-nucleotide RNAs that can pair to sites in the messenger RNAs of protein-coding genes to downregulate the expression from these messages. MicroRNAs are known to influence the evolution and stability of many mRNAs, but their global impact on protein output had not been examined. Here we use quantitative mass spectrometry to measure the response of thousands of proteins after introducing microRNAs into cultured cells and after deleting *mir-223* in mouse neutrophils. The identities of the responsive proteins indicate that targeting is primarily through seed-matched sites located within favourable predicted contexts in 3′ untranslated regions. Hundreds of genes were directly repressed, albeit each to a modest degree, by individual microRNAs. Although some targets were repressed without detectable changes in mRNA levels, those translationally repressed by more than a third also displayed detectable mRNA destabilization, and, for the more highly repressed targets, mRNA destabilization usually comprised the major component of repression. The impact of microRNAs on the proteome indicated that for most interactions microRNAs act as rheostats to make fine-scale adjustments to protein output.

Large-scale approaches for studying the regulatory effects of microRNAs (miRNAs) have revealed important insights into target recognition and function. These approaches include computational analysis of the selective maintenance or avoidance of miRNA complementary sites during evolution[1–8] and experimental identification of messages destabilized or those preferentially associated with argonaute proteins in the presence of a miRNA[7–15]. Despite their utility, none of these approaches directly measures the influence of a miRNA on protein output, which is the most relevant readout of its regulatory effects. The influence of miRNAs on protein output has instead been limited to single-protein analyses, primarily immunoblotting and reporter assays, and a medium-size proteomics analysis with detection of 504 proteins[16].

Proteomic consequences of added miRNAs

To acquire data sufficient to investigate the effects of miRNA regulation on the proteome, we applied a quantitative-mass-spectrometry-based approach using SILAC (stable isotope labelling with amino acids in cell culture)[17] to investigate the influence of specific miRNAs on the levels of many proteins (Supplementary Figs 1 and 2). We first measured the effects of introducing miR-124, a brain-specific miRNA, into HeLa cells. To include proteins from a broad expression spectrum, this experiment focused on nuclear-localized proteins. Out of 2,120 proteins detected, the analysis considered 1,544 that mapped to our non-redundant mRNA data set and were each quantified by at least two independent measurements that passed our quality thresholds (Supplementary Data 1 and 5).

Because this and all subsequent SILAC analyses were performed with two technical replicates, and because different peptides from the same protein and different charge states from the same peptide also provided the opportunity for independent measurements, most proteins were quantified by many more than two independent measurements (median of 12 for the 1,544 quantified proteins). The high

reproducibility when comparing technical replicates and when comparing different peptides representing the same protein illustrated the quantification accuracy ($r^2 = 0.72$ and 0.65, respectively, Spearman's correlation; Supplementary Fig. 3).

Messages for proteins that decreased the most relative to the mock-transfection control were compared to the messages of the other quantified proteins (cutoff, 85th percentile), searching for motifs over-represented in their open reading frames (ORFs) or untranslated regions (UTRs). When considering all 16,384 possible 7-nucleotide motifs and the different regions of the mRNA, the only one significantly enriched after Bonferroni correction for multiple hypothesis testing was the GUGCCUU heptanucleotide in the 3′ UTR ($P < 10^{-7}$, Fisher's exact test). This heptanucleotide motif comprised the 6-nucleotide match to the seed of miR-124 (underlined) supplemented by a match to miRNA nucleotide 8, and is named the 7mer-m8 seed-matched site (Fig. 1a). It was the same motif that is most associated with 3′ UTRs of messages destabilized after introduction of miR-124 (ref. 9). The other sites consistently associated both with preferential conservation and with mRNA destabilization after miRNA introduction are named the 6mer, 7mer-A1 and 8mer seed-matched sites[2,7,8] (Fig. 1a). A more directed search for the seed-matched sites revealed that most of the robustly repressed proteins derived from messages with at least one 7–8mer 3′-UTR site (Fig. 1b). For example, 24 out of the 40 proteins repressed by at least 50% had at least one 7–8mer 3′-UTR site, with only 3 of these 24 attributed to chance (Fig. 1b, repression cutoff of 50%). Less stringent repression cutoffs yielded many additional proteins from messages with 7–8mer sites, even after subtracting those expected by chance. The overall enrichment of seed-matched sites in messages of downregulated proteins indicated that miR-124 recognition of mRNAs for repression of protein output used, more than any other type of site, seed-matched sites in 3′ UTRs.

To survey the efficacy of the different seed-matched sites, we plotted the response of proteins from messages with 3′ UTRs possessing

single sites (Fig. 1c). Proteins from messages with single 7–8mer sites had a significant propensity to be downregulated when compared to those from messages without 3′-UTR sites ($P = 0.02$, 0.0008 and 0.02 for 8mer, 7mer-m8 and 7mer-A1, respectively, Kolmogorov–Smirnov test).

We performed analogous SILAC experiments with two additional miRNAs: miR-1 and miR-181, for which 2,312 and 1,774 proteins, respectively, mapped to our non-redundant mRNA data set and passed our quantification quality cutoffs (Supplementary Data 2, 3 and 5). The motifs associated with messages of the most downregulated proteins mirrored those observed for miR-124; for miR-1, the 7mer-m8 match was the most confidently enriched heptanucleotide motif in the 3′ UTRs of downregulated proteins ($P = 0.0004$), and, for miR-181, the 7mer-A1 match was among the top two motifs ($P = 0.007$), slightly less confidently enriched than an unrelated motif, CUGCCCC ($P = 0.006$, Fisher's exact test with Bonferroni correction).

When pooling the data from all three miRNA transfections, thereby combining 5,630 independent protein quantifications, proteins from messages with single 7mer or 8mer sites matching the cognate miRNA had a significant propensity to be downregulated (Fig. 1d, $P < 10^{-14}$ overall, $P < 10^{-4}$ for each site separately, Kolmogorov–Smirnov test). Vertical displacement from the no-site distribution demonstrated that at least 16% of the proteins from messages with single 7–8mer 3′-UTR sites responded to the miRNA (Fig. 1d). The response of proteins from messages with a 6mer site closely tracked that from messages with no site, indicating that in this system 6mer recognition was generally insufficient for detectable protein downregulation (Fig. 1d).

Analysis of site conservation, site depletion, argonaute pull-downs and reporter assays all indicate that targeting can occur in protein-coding regions[2,5,13,18]. Analysis of mRNA destabilization concurs that targeting occurs in coding regions, but indicates that these sites are generally much less effective than those in 3′ UTRs[7]. However, monitoring mRNA destabilization would understate the influence of sites

in coding regions if these sites, by virtue of falling in the path of the ribosome, had a disproportionate effect on translation compared to mRNA destabilization. To address this possibility, we examined our data monitoring protein output and found that sites in coding regions were generally less effective that those in UTRs (Fig. 1e).

Proteomic consequences of disrupting *mir-223*

Measuring the effects of ectopic miRNA addition can provide generic insights into miRNA target recognition, but the responsive proteins are not necessarily the endogenous targets, and the magnitude and kinetics of mRNA and protein changes are not expected to match those of endogenous targeting (Supplementary Discussion). To obtain data relevant to endogenous miRNA–target interactions, with pertinent information on the degree of repression, we examined the effects of the *mir-223* gene knockout in mouse neutrophils. *mir-223* is preferentially expressed in myeloid haematopoietic cells, with high expression in neutrophils and their progenitors[19,20]. To obtain labelled samples suitable for the quantitative proteomics experiment, we isolated bone marrow haematopoietic progenitors from wild-type and *mir-223*-deficient mice[21] and developed a protocol for their proliferation in SILAC media and differentiation into mature neutrophils *in vitro* (Fig. 2a and Supplementary Fig. 4a, b). By day 8, the surviving cells had descended from progenitors that had undergone multiple cell divisions in the presence of SILAC media (Supplementary Fig. 4c), which resulted in >99% heavy isotope incorporation. RNA blots confirmed that both the progenitors and the differentiating neutrophils expressed *mir-223* (Fig. 2b). Array experiments demonstrated that the effect of miR-223 on messages with cognate sites was analogous to that observed for neutrophils isolated directly from mice, although somewhat less robust (Fig. 2c), perhaps in part because the neutrophils differentiated *in vitro* accumulate ~35% less miR-223 (Fig. 2b).

Analysis by mass spectrometry of both nuclear and cytoplasmic fractions provided quantitative information for 5,019 proteins, 3,819

Figure 1 | The impact of transfected miRNAs on protein output.
a, Canonical miRNA seed-matched sites[2]. **b**, The fraction of repressed proteins deriving from messages with miR-124 3′-UTR sites (filled orange bar). At each repression cutoff, the number of repressed proteins from messages without 3′-UTR sites (indicated in the open bar) was used to calculate the additional fraction expected by chance to have a site (dashed line, with the corresponding number of repressed proteins indicated below the dashed line). Above the dashed line is the surplus number of repressed proteins deriving from messages with sites. **c**, Response of proteins from messages with single miR-124 3′-UTR sites. Plotted is the fraction of proteins that change at least to the degree indicated on the *x* axis. Proteins

from messages with multiple 3′-UTR sites were not considered. 6mer sites that were part of larger sites were not included in the 6mer distribution, and 7mers that were part of 8mers were not included in the 7mer distributions. **d**, Efficacy of single 3′-UTR sites when pooling data from miR-124, miR-1 and miR-181 transfections, plotted as in **c**. **e**, ORF and 3′-UTR targeting efficacy. Plotted is the average change ($\pm$ standard error) of protein and corresponding mRNA for quantified proteins from messages with at least one 8mer in the ORF ($n = 83$) or 3′ UTR ($n = 87$) corresponding to the transfected miRNA (excluding messages with sites in both ORF and 3′ UTR).

of which mapped to our mRNA data set and passed our quality cut-offs (Supplementary Data 4 and 5). The effects of removal of endogenous miR-223 on neutrophil protein levels were essentially the reciprocal of those observed when ectopically adding individual miRNAs, except more of the targeting trends were statistically significant, presumably because more proteins were quantified. For instance, derepressed proteins derived from messages with strong enrichment for 6–8mer seed-matched motifs in 3′ UTRs (but not 5′ UTRs or coding regions), with high confidence for all four site types, even after Bonferroni correction (Supplementary Table 1). The fraction of responsive proteins from messages with 3′-UTR sites (Fig. 2d) resembled that observed for ectopic miR-124 delivery (Fig. 1b). Proteins from messages with a single 7–8mer site tended to be derepressed (Fig. 2e, $P < 10^{-5}$, $P < 10^{-6}$ and $P < 10^{-4}$ for 8mer, 7mer-m8 and 7mer-A1, respectively, Kolmogorov–Smirnov test). The apparent hierarchy of site efficacy observed when monitoring protein output (Fig. 2e, 8mer > 7mer-m8 > 7mer-A1 > 6mer) matched that obtained when monitoring mRNA effects[7,8]. Evidence for modest ORF targeting was again observed (Fig. 2f). The 33 quantified proteins from messages with multiple sites tended to be more responsive (Fig. 2g), but the increased output did not exceed that

expected from each site acting independently. This independent, non-cooperative response was in agreement with results monitoring mRNA destabilization and reporter assays, which indicate that cooperative action of sites tends to occur only for those sites falling within 8–40 nucleotides of each other[7]. Taken together, our results demonstrated experimentally that targeting principles elucidated from ectopically added miRNAs apply also to endogenous miRNA targeting, and in particular to endogenous targeting at the level of protein downregulation.

Endogenous response of predicted miRNA targets

The perturbation of endogenous targeting provided the opportunity to test sets of target predictions. When considering current predictions from miRBase Targets[22], miRanda[23,24], PicTar[4,25], PITA[26] and TargetScan[2,7], all of which use site conservation as a prediction criterion, those from TargetScan and PicTar performed the best (Fig. 3a). Predictions from TargetScan and PicTar are primarily those messages with at least one 3′-UTR 7–8mer site conserved among mammals, operationally defined as those sites preserved in orthologous locations of human, mouse, rat and dog UTRs[2,4]. Their enhanced performance over the set of messages with any 3′-UTR
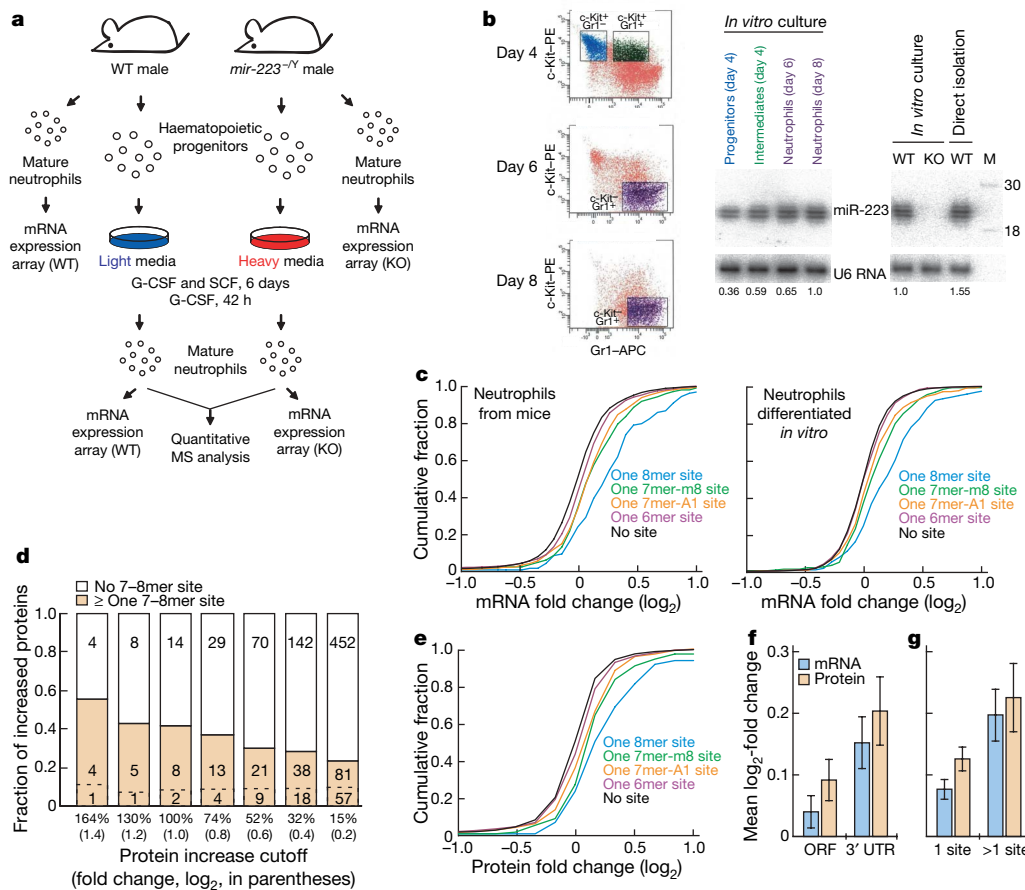
Figure 2 | The proteomic impact of deleting *mir-223* in mouse neutrophils.
a, Schematic of neutrophil labelling and analysis. Haematopoietic progenitors were isolated from wild-type (WT) or *mir-223*$^{-/Y}$ (KO) male mice and cultured in SILAC media containing granulocyte colony-stimulating factor (G-CSF) and stem cell factor (SCF) for six days. To enhance differentiation, SCF was withdrawn over the next 42 h. Mature neutrophils were mixed, and proteins were size-fractionated for quantitative MS analysis. mRNA was also collected from the cultures and directly from mice for expression profiling. **b**, *mir-223* expression detected with RNA blots probing for miR-223. One blot analysed total RNA from sorted subpopulations of cells cultured *in vitro* (left, with sorting profiles shown at the far left). The other blot analysed total RNA from cells cultured *in vitro* for eight days and from neutrophils isolated directly from bone marrow (right).

As a loading control, blots were re-probed for U6 small nuclear RNA. Below each blot are the relative expression levels, normalized using the loading control. Radio-labelled RNA markers (M) are also shown. **c**, Analysis of neutrophils isolated directly from mice (left) and those derived *in vitro* from haematopoietic precursors (right), monitoring the effects of miR-223 loss on messages with single miR-223 sites in their 3′ UTRs. Plotted is the fraction of messages that changed at least to the degree indicated on the x axis, otherwise as in Fig. 1c. **d**, The fraction of upregulated proteins deriving from messages with miR-223 7–8mer 3′-UTR sites, plotted as in Fig. 1b. **e**, The impact of deletion of *mir-223* on neutrophil proteins, considering proteins from messages with single miR-223 sites in their 3′ UTRs, plotted as in Fig. 1c. **f**, Targeting efficacy in ORFs ($n = 69$) and 3′ UTRs ($n = 50$), plotted as in Fig. 1e. **g**, Efficacy of single 7–8mer sites and multiple sites, plotted as in **f**.

7–8mer sites demonstrated that considering site conservation not only enriches for sites with presumed functional roles but also enriches for those that are more effective. All of the other algorithms include many sites with least one mismatch or wobble to the seed, which seems to have compromised their performance. For example, the predictions of miRBase Targets had been generated using the miRanda algorithm[23] with updated parameters, searching for conserved sites with more stringent seed pairing but still allowing one mismatch or wobble to the seed[22]. Analysis of the seed-matched and seed-mismatched predictions separately revealed that any benefit gained in searching for site conservation was offset by the inclusion of many poorly performing predictions with seed mismatches (Supplementary Fig. 5a). Despite the relative success of TargetScan and PicTar, two-thirds of their predicted targets appeared to be non-responsive to miR-223 loss in neutrophils, indicating a false-positive

rate within the range of that inferred from estimates of chance conservation of the target sites of this miRNA (Supplementary Fig. 5b).

The similar performance of PicTar and TargetScan was expected for miR-223, which begins with a U, but might not have been expected for those miRNAs that do not begin with a U. TargetScan rewards an A across from position 1, whereas PicTar (and similar algorithms[3,27]) rewards a Watson–Crick match at this position. Therefore, for miRNAs that begin with A, C or G, only one of the two heptanucleotide matches (the 7mer-m8) is the same for the algorithms[2,4] and thus about half of the predicted targets are expected to differ. To investigate which type of heptanucleotide match is most associated with decreased protein output, we examined the proteomics data from the experiment transfecting miR-181, which does not begin with a U. Plotting the response of proteins from messages with single sites revealed that the 7mer-A1 match was more effective than the Watson–Crick 1–7 match (Fig. 3b, $P = 0.009$, Kolmogorov–Smirnov test). Moreover, the Watson–Crick 1–7 match was no more effective than were 6mer sites with G or C mismatches across from position 1 (Fig. 3b). We conclude that the recognition of an A across from miRNA nucleotide 1 favours miRNA-mediated protein downregulation, which explains the preferential conservation of an A at this position, even when it cannot participate in a Watson–Crick interaction[2].

Target prediction sets are typically ranked, with the assertion that the better scoring predictions are more likely to be authentic or effective. Recent TargetScan predictions (release 4) are ranked by 'total context score', which is based on site type, site number and site context[7]. This ranking correlated with protein downregulation, with the top third significantly more responsive than the bottom third (Fig. 3c). For the other algorithms, the predictions scoring in the top third were not significantly more responsive than those in the bottom third (Fig. 3c, $P > 0.05$, Mann–Whitney U-test). Despite their poor overall performance, the more inclusive algorithms might still have utility when considering only their top few predictions. To investigate this possibility, we considered only the top 29 predictions of each algorithm, choosing 29 because the most restrictive set (that of PicTar) includes this number of predictions. At this stringent cut-off, the performances of the more inclusive algorithms approached that of PicTar (resulting in difference that was no longer statistically significant, $P > 0.05$), but remained lower than that of TargetScan ($P < 0.05$, Fig. 3d). Interestingly, the top 29 quantified proteins ranked only by the total context score of their respective 3′ UTRs, without any regard to site conservation, were at least as responsive as the top 29 TargetScan predictions (Fig. 3d).

Analysis of the evolutionary impact of miRNAs and analysis of messages that are upregulated in miRNA-deficient animals both indicate that many non-conserved sites mediate repression *in vivo*[5,6,10,11]. We also found evidence for widespread non-conserved targeting among natural miR-223–target interactions. In an attempt to predict non-conserved targets, RNA22 (ref. 28) and a more permissive version of PITA[26] do not consider site conservation. When evaluated using our miR-223 data, these algorithms performed no better than did a simple search for messages with 7–8mer seed-matched sites (Fig. 3e). A more effective tool was the total context score, which correlated with derepression when considering only those messages with non-conserved 7–8mer sites (that is, sites missing or mutated in orthologous positions of human, rat or dog 3′ UTRs), with the top third of non-conserved predictions significantly more effective than the bottom third (Fig. 3e). Indeed the top third of non-conserved predictions (Fig. 3e, context score) appeared as effective as the bottom two-thirds of conserved predictions (Fig. 3c, TargetScan), and because proteins from non-conserved predictions outnumbered those from conserved ones by 6 to 1, the non-conserved predictions with favourable context scores were a bountiful source of biological targets.

The success of the total context score in ranking both conserved and non-conserved predictions was due in part to its consideration of
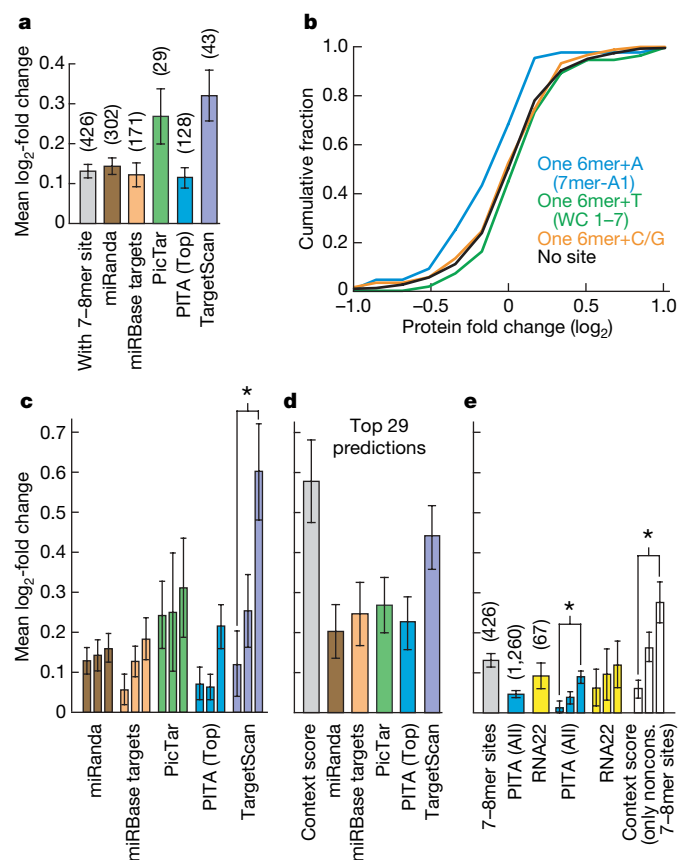

Figure 3 | Correspondence between computational target predictions and observed protein changes. Analogous results were observed using the transfection data sets (Supplementary Fig. 8). **a**, Performance of programs that consider site conservation. Plotted is the average protein derepression (± standard error) of genes with ≥1 conserved or non-conserved 7–8mer 3′-UTR site (grey) and of genes predicted to be miR-223 targets. The number of quantified proteins in each set is in parenthesis. **b**, Recognition of an adenosine (A) opposite the first nucleotide of the miRNA. The cumulative plot of protein changes after miR-181 transfection compares proteins from messages with no seed-matched 3′-UTR site to those from messages with the indicated single 3′-UTR site. **c**, Relationship between the scores of predicted targets and protein derepression. Predictions corresponding to quantified proteins were divided into three equal-size bins according to the scores proposed to indicate the quality of the prediction or degree of repression. Statistically significant differences between the bottom and top third are indicated (asterisk, $P < 0.01$, Mann–Whitney U-test). **d**, Response of the top 29 predictions of each algorithm, plotted as in **a**. **e**, Performance of programs that do not consider site conservation, displayed as in **a** and **c**. Also shown is the response of quantified proteins from messages with only non-conserved 7–8mer 3′-UTR sites, binned by total context score.

site type (Fig. 2e) and the number of sites (Fig. 2g). To isolate its third component (site context) we considered only those quantified proteins deriving from messages with single 7mer-m8 3′-UTR sites and still observed a significant correlation between context score and protein response ($P = 0.001$, Spearman's correlation test). Predicted 3′-UTR structure and other features of site context are reported to influence site accessibility and efficacy[7,8,26,27,29–31]. The context score combines some of these features, including high local AU nucleotide composition (which accounts for effects of predicted 3′-UTR structure on site accessibility), proximity to residues that can pair to miRNA nucleotides 13–16, and positioning away from the centre of long UTRs[7]. As anticipated from analyses of mRNA destabilization data[7], the most influential component was local AU composition, which when examined in isolation significantly correlated with protein response ($P = 0.01$, Spearman's correlation test).

Response of proteins compared to that of mRNAs

Because previously used high-throughput methods were unable to determine the amount of protein repression, the relative contributions of mRNA destabilization and translational repression during miRNA-mediated regulation has been of intense interest. Our miR-223 data was informative for addressing this issue because it examined the response, at both the mRNA and the protein level, of removing an endogenous miRNA, without the confounding influences of exogenous targeting mediated by an ectopically delivered miRNA. The near steady-state nature of our miR-223 system also avoided quantification caveats inherent to transient transfection, such as variable transfection efficiencies and pre-steady-state complexities especially acute when comparing effects on an mRNA to those on its protein because messages and their proteins can have very different intrinsic stabilities. Note that our mRNA quantification used standard array platforms, which include oligo(dT) priming during detection, and thus the

mRNA destabilization we observed encompassed the conversion of the message into a form that was unsuitable for translation because it lacked a poly(A) tail.

To achieve greater quantification accuracy in this analysis of individual proteins, we narrowed our focus to the 2,773 proteins quantified with ≥6 independent measurements. Plotting protein changes as a function of mRNA changes indicated a strong positive correlation for messages with 7mer or 8mer 3′-UTR sites (Fig. 4a; $r^2 = 0.45$ and 0.63, $P < 10^{-33}$ and $P < 10^{-12}$, respectively) and weaker correlation for messages without sites (Fig. 4b; $r^2 = 0.15$, $P < 10^{-11}$, Pearson's correlation test). Proteins in both plots displayed some scatter around the origin; however, when normalizing to those without sites, many more of those from messages with sites increased in response to miR-223 loss (Fig. 4a, b and Supplementary Fig. 6). Immunoblots probing for three of the more responsive proteins confirmed protein derepression in $mir\text{-}223^{-/Y}$ neutrophils differentiated *in vitro* as well as in those isolated directly from mice (Supplementary Fig. 7).

Two of the three most responsive proteins derived from messages with single, non-conserved 7mers (Table 1)—sites that on their own would not be expected to impart such a robust response. Previous work has shown that sites falling within 8–40 nucleotides of sites to co-expressed miRNAs typically act cooperatively, which increases the effect of loosing interactions at particular sites[7]. We performed high-throughput sequencing to identify miRNAs co-expressed in cultured neutrophils (Supplementary Table 2) and found that both of the highly responsive 7mers fell near to sites matching a co-expressed miRNA, with intersite spacing favouring a cooperative response (Table 1). The site in *Ctsl* was near a site for the miR-26 family, one of five families sequenced more frequently than miR-223, whereas the site in *Gns* fell near a site to the miR-103/107 family, sequenced about a third as often as miR-223 (Supplementary Table 2).
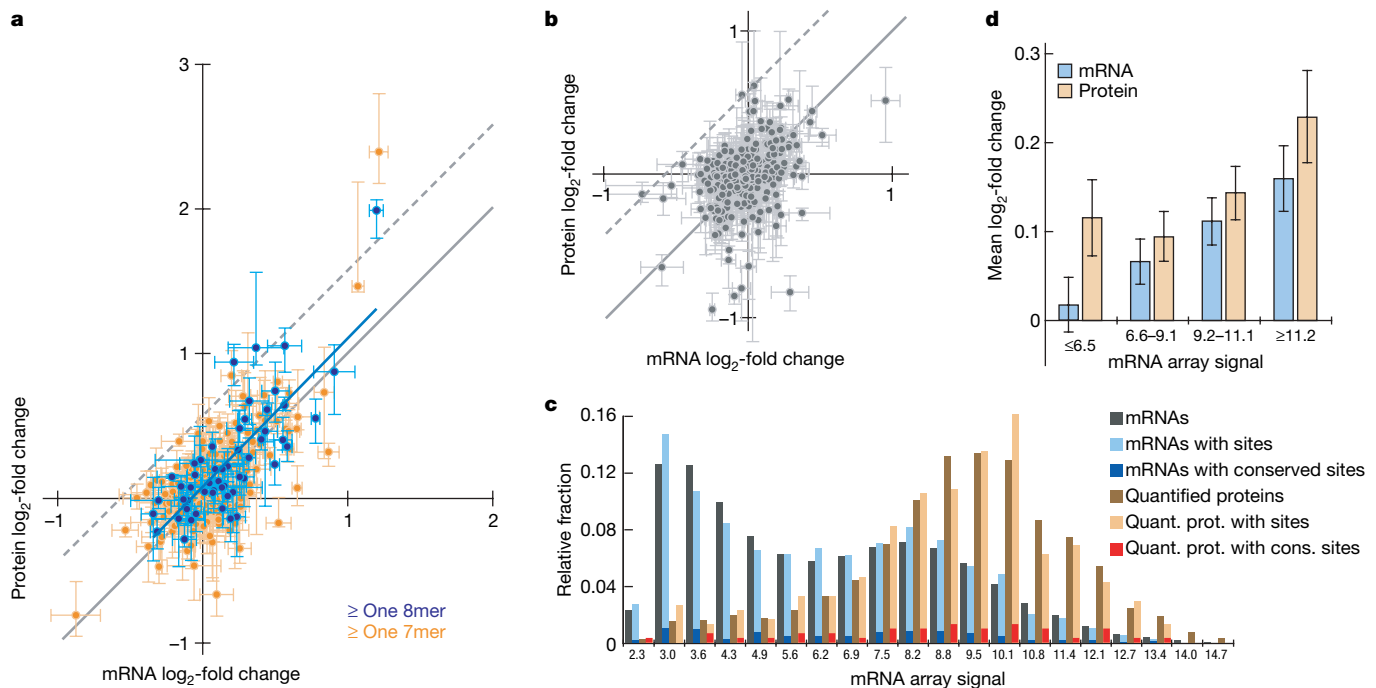

Figure 4 | Comparison of protein and mRNA changes accompanying miR-223 loss. a, Protein and mRNA changes for quantified proteins deriving from messages with at least one 8mer 3′-UTR site (blue, $n = 55$) or at least one 7mer (orange, an additional 250 proteins). The least-squares best fit to the 8mer data are shown (blue line), as are reference lines (grey), which both have slope of 1.0. Vertical error bars indicate 25th and 75th percentiles for independent measurements of protein changes. Horizontal error bars indicate standard errors of mRNA changes from three biological replicates, one of which was also used for the SILAC experiments. **b**, Protein and mRNA changes for quantified proteins deriving from messages without 7–8mer 3′-UTR sites, plotted as in **a**. One of seven random cohorts is plotted here; the other six are in Supplementary Fig. 6. **c**, Distribution of the indicated reference-set mRNAs and quantified proteins with respect to mRNA expression, as indicated by the array signals from cultured neutrophils. **d**, Response of quantified proteins and their respective mRNAs to *mir-223* deletion, considering those messages with 7–8mer 3′-UTR sites, grouped by mRNA expression.

If protein changes merely reflected mRNA changes, with no additional repression at the translational level, then the points would fall on the diagonal (Fig. 4a, grey line). Although many were on the diagonal or very close to it, least-squares linear regression yielded a positive y-intercept ($+0.053$ and $+0.079$ for 7mer and 8mer data, respectively). These modest yet statistically significantly positive y-intercept values ($P = 0.0002$ and $P = 0.042$, t-test) suggested that a cohort of genes were modestly derepressed at the protein level with little or no change at the mRNA level. The messages of such genes were each good candidates for targets affected only at the translational level, although some might have derived from genes undergoing non-miRNA-mediated transcriptional repression as a compensatory feedback response to the loss of miR-223 targeting.

Despite evidence for some translation-only repression, all proteins derepressed by more than 50% ($\log_2 > 0.58$) derived from messages that displayed detectable increases (Fig. 4a and Table 1). Moreover, only five points were more than 0.58 units ($\log_2$) above the diagonal (Fig. 4a, upper dashed line; Table 1, indicated with §). Note that a 33% repression by miR-223 in wild-type neutrophils would correspond to a 50% ($+0.58\ \log_2$) derepression in mutant neutrophils. Thus, in wild-type neutrophils only 5 of the 305 quantified proteins from messages with 7–8mer 3'-UTR sites appeared to undergo translational repression by more than 33%. We conclude that, although in some instances translational repression produces a substantial amount of endogenous miRNA-mediated repression, this occurred for surprisingly few of the many inferred targets. Substantial translational repression appeared so rarely because targets repressed only at the level of translation were repressed quite modestly ($<33\%$); for targets undergoing more robust repression, the major component of the repression was usually mRNA destabilization (Table 1). Further study is required to determine whether those mRNA molecules undergoing miRNA-mediated repression might experience translational repression as a prelude to destabilization, but our results show that mRNA destabilization can explain most of the endogenous miR-223-mediated repression.

Our proteomics data were limited to the confidently quantified proteins, which were expected to be those that were both soluble and more highly expressed in neutrophils. To consider how the expression bias might have influenced our results, we plotted the distributions of mRNAs and quantified proteins as a function of mRNA expression in neutrophils, considering all mRNAs of our non-redundant data set (including those without detectable expression), as well as those with 3'-UTR sites (Fig. 4c). The messages with conserved or non-conserved 3'-UTR sites displayed the full range of expression values, with a distribution matching that of messages more generally. As anticipated, more quantified proteins derived from highly expressed messages (Fig. 4c). However, the distribution of quantified proteins from messages with sites (conserved or non-conserved) closely matched that of those without sites. Moreover, we found no evidence that the greater representation of proteins from more highly expressed messages underrepresented the impact of miRNAs on protein output; if anything, proteins from more highly expressed messages tended to respond more robustly than did those from lowly expressed messages (Fig. 4d). An analysis using Gene Ontology terms[32] came to similar conclusions (data not shown). Therefore, although our experiment monitored the impact on only a portion of the neutrophil proteome and thus missed many miR-223 targets (including some conserved targets, such as Mef2c; refs 2, 21), we found no reason to suspect that undetected targets respond more robustly.

The proteins from the least abundant mRNAs appeared to respond without detectable mRNA changes (Fig. 4d, ≤ 6.5 bin). Apparent dominance of the translational component might have been a consequence of less reliable array signals for these messages, many of which fell within background signals from non-expressed messages. A more intriguing possibility is that very efficient translation of these messages (inferred from the ability to quantify proteins from such lowly expressed messages) makes them more susceptible to greater translational repression.

The regulatory function of miR-223

Some of the most strongly derepressed proteins from messages with miR-223 sites provided potential explanations for the pro-inflammatory phenotype observed in *mir-223*[-/Y] neutrophils[21]. Cathepsin L and cathepsin Z (Ctsl and Ctsz, listed first and fourteenth in Table 1) are cysteine proteases associated with chronic inflammatory conditions, in which they can act as mediators of tissue destruction[33,34]. Another potentially relevant target, the insulin-like growth factor

Table 1 | The most responsive proteins deriving from messages with at least one 7–8mer 3'-UTR site

Protein*	Fold change ($\log_2$) in *mir-223*[-/Y] cells versus wild type				Fold change ($\log_2$) of mRNA during neutrophil differentiation†	3'-UTR sites	Co-expressed miRNA family‡ with cooperatively spaced site
	Neutrophil culture (8 days)		Sorted cultured cells				
	Protein (25th–75th percentiles)	mRNA ($\pm$s.e.m.)	Progenitor mRNA	Neutrophil mRNA			
Ctsl§	2.40 (2.18–2.80)	1.21 ± 0.07	0.79	1.03	1.71	7mer	miR-26, 8mer
Parp9§	1.99 (1.80–2.06)	1.20 ± 0.05	0.43	1.07	1.55	8mer	
Gns	1.47 (1.43–2.19)	1.07 ± 0.04	0.51	1.01	1.30	7mer	miR-103/107, 7mer ||
Rasa1	1.06 (0.87–1.18)	0.56 ± 0.12	0.27	0.62	−0.22	8mer ||	
Acsl3§	1.04 (0.92–1.56)	0.37 ± 0.19	0.44	0.50	−1.49	8mer ¶	
Igf1r§	0.94 (0.78–1.06)	0.22 ± 0.13	0.01	−0.03	0.03	8mer ||, 7mer	
Galnt7	0.87 (0.58–1.06)	0.91 ± 0.14	0.40	0.89	−0.15	8mer, 7mer	miR-103/107, 7mer ||
Myo1c§	0.85 (0.64–1.02)	0.20 ± 0.07	0.40	0.29	0.55	7mer	
Gm885	0.81 (0.76–0.88)	0.52 ± 0.03	0.07	0.27	1.63	7mer	
Smarcd1	0.74 (0.53–0.94)	0.50 ± 0.09	0.24	0.41	−0.11	8mer ||	
Mvp	0.73 (0.63–0.82)	0.60 ± 0.01	0.25	0.48	2.07	7mer	
1110019N10Rik	0.73 (0.28–1.05)	0.84 ± 0.07	0.66	0.88	−0.61	7mer	
Ipo9	0.71 (0.65–0.88)	0.28 ± 0.10	0.19	0.33	−0.72	7mer, 7mer	
Ctsz	0.71 (0.65–0.81)	0.42 ± 0.09	0.41	0.70	−0.26	7mer	miR-27, 7mer
Atp2b1	0.67 (0.51–0.83)	0.32 ± 0.14	−0.01	−0.01	0.71	8mer ||	
Prkcb1	0.66 (0.54–0.76)	0.57 ± 0.09	−0.47	0.37	2.97	7mer, 7mer	
Rrm2	0.64 (0.61–0.70)	0.56 ± 0.01	0.30	0.60	−0.49	8mer	miR-27, 7mer
Ankrd13a	0.64 (0.58–0.87)	0.29 ± 0.08	0.10	0.23	0.05	7mer	
Ywhah	0.61 (0.39–0.66)	0.44 ± 0.03	0.14	0.56	0.33	8mer	miR-142-5p, 7mer ||

*Listed are all proteins quantified using ≥ 6 independent measurements and also upregulated more than 1.5-fold (0.58 $\log_2$) in *mir-223*[-/Y] neutrophils.
†mRNA change when comparing sorted *mir-223*[-/Y] neutrophils (cultured for eight days) with sorted *mir-223*[-/Y] progenitors (cultured for four days; Fig. 2b). For messages of all quantified proteins, median fold change ($\log_2$) was −0.22.
‡Considered were sites for 20 co-expressed miRNA families (Supplementary Table 2).
§Protein upregulated $\geq 50\%$ (0.58 $\log_2$) with miR-223 loss, after accounting for the mRNA change.
||Conserved site.
¶Conserved as the 8mer in human and rat but as a 7mer in dog.

69

receptor 1 (Igf1r, listed sixth), is crucial for the priming and activation of mature neutrophils[35,36].

To examine whether repression begins before neutrophil maturation, we profiled mRNA levels in sorted progenitors and neutrophils (Supplementary Data 4). Messages of most of the highly responsive proteins were derepressed already at the progenitor stage, although usually to a lower degree than in neutrophils, which accumulate more miR-223 (Table 1 and Fig. 2b).

The profiles of miR-223-deficient progenitors and neutrophils provided the opportunity to examine the regulation of putative miR-223 targets in the absence of miR-223 to determine whether miR-223-mediated repression predominantly acts coherently with (that is, in the same direction as) the other gene-regulatory processes acting on these genes. During differentiation from progenitor to neutrophil, putative targets increased and decreased in similar numbers (Table 1 and Supplementary Data 4). This result revealed a proportion of incoherent regulatory relationships larger than that observed for other miRNAs[5,6,11] but nonetheless consistent with the miR-223 loss-of-function phenotype; this phenotype indicates that miR-223 dampens progenitor proliferation and neutrophil differentiation and activation[21]—functions opposite of those expected for coherent regulatory interactions involving a miRNA preferentially expressed in neutrophils.

Because the miR-223 proteomics experiment detected targeting potentially missed by other high-throughput methods, particularly non-conserved targets influenced (albeit modestly) at the level of translation, it provided the clearest picture so far of the scope and magnitude of endogenous miRNA targeting. The vertical displacement from the no-site distribution in Fig. 2e indicated that at least 18.4% of the 426 proteins from messages with 7–8mer 3′-UTR seed-matched sites underwent increased protein output attributable to the sites, thereby implicating messages for at least 78 out of the 3,819 quantified proteins as direct targets. These 78 included ∼33% of those quantified proteins from messages with conserved 3′-UTR sites and ∼16% of those from messages with nonconserved 3′-UTR sites. Assuming that only about one-third of the proteome was quantified, we estimate that miR-223 has >200 targets in neutrophils (3 × 78). These would not include any targets undergoing fail-safe regulation (targeting of messages for proteins not normally expressed at all in neutrophils), which are invisible in derepression experiments. Despite the broad scope of miR-223 targeting, each interaction had only a modest effect, even when observed at the protein level. Many miR-223-responsive targets also have sites for other miRNAs, some of which are also expressed in neutrophils, and thus the aggregate impact of miRNAs on these targets is presumably greater than that observed for miR-223 alone. Nonetheless, the targeting by other miRNAs is not expected to obscure the effect of removing miR-223 because multiple non-overlapping sites to co-expressed miRNAs typically act independently[7,8], and in the rare cases in which they do not act independently, they act cooperatively, which would boost rather than decrease the effect of loosing a single miRNA[7]. The widespread scope but low magnitude of endogenous miR-223-mediated repression indicates that this miRNA often acts as a rheostat to adjust protein output.

METHODS SUMMARY

HeLa cells were grown in media containing either regular (light) Lys and Arg or $^{13}C_6$-labelled (heavy) Lys and Arg. Light cells were transfected with miRNA, and heavy cells were mock-transfected. After 24 h some cells were harvested for mRNA expression profiling. After 48 h the remaining cells were harvested, and equal numbers from both populations were mixed and enriched for soluble nuclear proteins. Neutrophil culture was as outlined in Fig. 2a. Protein mixtures were separated by SDS–PAGE, and fractions were digested with trypsin. Peptides were analysed by liquid chromatography–tandem mass spectrometry (LC–MS/MS), which identified peptides and quantified the relative amounts of isotopic pairs of the same peptide. To prevent double-counting of any targeting interactions, peptides were mapped to a non-redundant complementary DNA data set (Supplementary Data 5), and targeting analyses were as performed previously on

mRNA destabilization data[7]. To compare to target-prediction algorithms, predictions by TargetScan (release 4.1)[2,7], PicTar (human, chimp, mouse, rat, dog)[4,25], miRanda (January 2008 release)[23,24], miRBase Targets (version 5)[22], RNA22 (ref. 28) and PITA[26] were obtained from their respective websites, using the most recent predictions publicly available as of March 2008.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
2. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
3. Brennecke, J., Stark, A., Russell, R. B. & Cohen, S. M. Principles of microRNA-target recognition. *PLoS Biol.* **3**, e85 (2005).
4. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nature Genet.* **37**, 495–500 (2005).
5. Farh, K. K. *et al.* The widespread impact of mammalian microRNAs on mRNA repression and evolution. *Science* **310**, 1817–1821 (2005).
6. Stark, A., Brennecke, J., Bushati, N., Russell, R. B. & Cohen, S. M. Animal microRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution. *Cell* **123**, 1133–1146 (2005).
7. Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27**, 91–105 (2007).
8. Nielsen, C. B. *et al.* Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* **13**, 1894–1910 (2007).
9. Lim, L. P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
10. Krutzfeldt, J. *et al.* Silencing of microRNAs *in vivo* with 'antagomirs'. *Nature* **438**, 685–689 (2005).
11. Giraldez, A. J. *et al.* Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75–79 (2006).
12. Beitzinger, M., Peters, L., Zhu, J. Y., Kremmer, E. & Meister, G. Identification of human microRNA targets from isolated argonaute protein complexes. *RNA Biol.* **4**, 76–84 (2007).
13. Easow, G., Teleman, A. A. & Cohen, S. M. Isolation of microRNA targets by miRNP immunopurification. *RNA* **13**, 1198–1204 (2007).
14. Karginov, F. V. *et al.* A biochemical approach to identifying microRNA targets. *Proc. Natl Acad. Sci. USA* **104**, 19291–19296 (2007).
15. Zhang, L. *et al.* Systematic identification of *C. elegans* miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2. *Mol. Cell* **28**, 598–613 (2007).
16. Vinther, J., Hedegaard, M. M., Gardner, P. P., Andersen, J. S. & Arctander, P. Identification of miRNA targets with stable isotope labeling by amino acids in cell culture. *Nucleic Acids Res.* **34**, e107 (2006).
17. Ong, S. E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
18. Kloosterman, W. P., Wienholds, E., Ketting, R. F. & Plasterk, R. H. Substrate requirements for *let-7* function in the developing zebrafish embryo. *Nucleic Acids Res.* **32**, 6284–6291 (2004).
19. Chen, C. Z., Li, L., Lodish, H. F. & Bartel, D. P. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**, 83–86 (2004).
20. Fazi, F. *et al.* A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBPα regulates human granulopoiesis. *Cell* **123**, 819–831 (2005).
21. Johnnidis, J. B. *et al.* Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. *Nature* **451**, 1125–1129 (2008).
22. Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36**, D154–D158 (2008).
23. John, B. *et al.* Human microRNA targets. *PLoS Biol.* **2**, e363 (2004).
24. Betel, D., Wilson, M., Gabow, A., Marks, D. S. & Sander, C. The microRNA.org resource: targets and expression. *Nucleic Acids Res.* **36**, D149–D153 (2008).
25. Lall, S. *et al.* A genome-wide map of conserved microRNA targets in *C. elegans*. *Curr. Biol.* **16**, 460–471 (2006).
26. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nature Genet.* **39**, 1278–1284 (2007).
27. Gaidatzis, D., van Nimwegen, E., Hausser, J. & Zavolan, M. Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics* **8**, 69 (2007).
28. Miranda, K. C. *et al.* A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell* **126**, 1203–1217 (2006).
29. Robins, H., Li, Y. & Padgett, R. W. Incorporating structure to predict microRNA targets. *Proc. Natl Acad. Sci. USA* **102**, 4006–4009 (2005).
30. Zhao, Y., Samal, E. & Srivastava, D. Serum response factor regulates a muscle-specific microRNA that targets *Hand2* during cardiogenesis. *Nature* **436**, 214–220 (2005).
31. Long, D. *et al.* Potent effect of target structure on microRNA function. *Nature Struct. Mol. Biol.* **14**, 287–294 (2007).

32. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25,** 25–29 (2000).

33. Reddy, V. Y., Zhang, Q. Y. & Weiss, S. J. Pericellular mobilization of the tissue-destructive cysteine proteinases, cathepsins B, L, and S, by human monocyte-derived macrophages. *Proc. Natl Acad. Sci. USA* **92,** 3849–3853 (1995).

34. Menzel, K. *et al.* Cathepsins B, L and D in inflammatory bowel disease macrophages and potential therapeutic effects of cathepsin inhibition *in vivo. Clin. Exp. Immunol.* **146,** 169–180 (2006).

35. Fu, Y. K., Arkins, S., Wang, B. S. & Kelley, K. W. A novel role of growth hormone and insulin-like growth factor-I. Priming neutrophils for superoxide anion secretion. *J. Immunol.* **146,** 1602–1608 (1991).

36. Bjerknes, R. & Aarskog, D. Priming of human polymorphonuclear neutrophilic leukocytes by insulin-like growth factor I: increased phagocytic capacity, complement receptor expression, degranulation, and oxidative burst. *J. Clin. Endocrinol. Metab.* **80,** 1948–1955 (1995).

Author Contributions The order of listing of the first three authors is arbitrary. C.S. and F.C. performed the experimental work with cells and animals. J.V. performed the mass spectrometry and associated computational analysis. D.B. performed the computational analysis of targeting. All authors contributed to the design of the study and preparation of the manuscript.

Author Information Array data and small RNA sequencing data were deposited in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/) under accession number GSE12075. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.P.G. (steven_gygi@hms.harvard.edu) or D.P.B (dbartel@wi.mit.edu).

METHODS

Transfection experiments. HeLa cells (ATCC, CCL-2) were grown in SILAC DMEM media (Invitrogen) supplemented with Pro ($10 \, \text{mg l}^{-1}$) and containing either naturally occurring isotopes of Arg and Lys ($50 \, \text{mg l}^{-1}$ each) or heavy ($^{13}\text{C}_6$)-labelled Arg and Lys ($50 \, \text{mg l}^{-1}$ each, Cambridge Isotope Laboratory). Heavy isotope incorporation in proteins was analysed by mass spectrometry (>99% Arg, >98.5 Lys). Cells grown in heavy amino acids were mock-transfected with lipofectamine 2000 (Invitrogen), whereas those grown in light amino acids were transfected with miRNA duplexes described previously[7,9], using lipofectamine 2000, 25 nM duplex, and supplementing with OPTI-MEM (Invitrogen). After 6 h, media of both mock and miRNA transfections was replaced with SILAC DMEM. Twenty-four hours after transfection, some cells were harvested, and mRNA was purified (RNeasy Plus, Qiagen) for expression profiling (Agilent human 4 × 44K microarray). Forty-eight hours after transfection, the remaining cells were harvested, and equal numbers of miRNA- and mock-tranfected cells were mixed. Soluble nuclear proteins were purified (NE-PER Nuclear and Cytoplasmic Extraction Reagent, Thermo Fisher Scientific) and separated into ten fractions by SDS–PAGE for mass spectrometry analysis.

As an additional control for targeting specificity, analyses of the transfection results were repeated comparing the response of proteins from messages with sites to the cognate miRNA to that of the very same proteins when the non-cognate miRNA was transfected. The overall conclusions from this set of control analyses were the same as for the analyses presented, indicating that the results depended on the identity of the miRNA transfected, rather than on other differences between mock- and miRNA-tranfected cells, such as the mass of the amino acids or the presence of OPTI-MEM in the transfections.

Neutrophil culture and differentiation. All animal experiments were approved by the MIT Committee on Animal Care. Bone marrow was obtained from three 3-month-old wild-type male mice and from three 3-month-old *mir-223*$^{-/\text{Y}}$ mice[21], and bone marrow haematopoietic progenitors were isolated as follows. Bone marrow from the three mice of each genotype was pooled, and suspended cells were depleted of mature cells using a mixture of biotin-conjugated Ter 119, Mac-1, Gr-1, B220 and CD3e antibodies (eBioscience) and anti-biotin microbeads (Miltenyi Biotech, Inc.), followed by magnetic cell sorting (MACS, Miltenyi Biotech, Inc.). The remaining cells were collected and cultured in SILAC IMDM media (Invitrogen) supplemented with Pro ($10 \, \text{mg l}^{-1}$) and containing G-CSF ($100 \, \text{ng ml}^{-1}$, PeproTech) and SCF ($50 \, \text{ng ml}^{-1}$, PeproTech). Media containing light Arg and Lys ($50 \, \text{mg l}^{-1}$ each) was used for cells derived from wild-type mice, and heavy media containing $^{13}\text{C}_6$-Arg and $^{13}\text{C}_6$-Lys ($50 \, \text{mg l}^{-1}$ each) was used for *mir-223* knockout cells. Media was replaced every two days, and after six days SCF was withdrawn to arrest proliferation and induce additional differentiation. Forty-two hours later, cells were harvested, and dead cells were removed (Dead cell removal kit, Miltenyi Biotech). Neutrophil maturity and viability were analysed by flow cytometry (FACSCalibur, BD Biosciences) after staining with PE-conjugated anti-mouse c-Kit antibody (eBioscience), APC-conjugated anti-mouse Gr-1 antibody (eBioscience), and propidium iodide (Supplementary Fig. 4a). The homogeneity of the cell population was also checked by microscopy after Wright–Giemsa stain of cytospun neutrophils (Supplementary Fig. 4b). Mass spectrometry confirmed nearly quantitative (>99%) incorporation of heavy Arg and Lys in cells cultured from the miR-223-deficient mice.

A fraction of each cell population was used to purify mRNA (RNeasy Plus, Qiagen) for expression profiling (Affymetrix mouse 430 2.0 microarray, Fig. 2c). Equal numbers of cells from each population were mixed, and soluble nuclear and cytoplasmic protein preparations were fractionated by SDS–PAGE and analysed independently by LC–MS/MS. Additional biological replicates (each starting with bone marrow pooled from one to four additional mice) were prepared from both wild-type and knockout mice and used for mRNA expression profiling, RNA blotting and immunoblotting. Some of these additional replicates were sorted using FACSAria (BD Biosciences) with PE-conjugated c-Kit antibody and APC-conjugated Gr1 antibody to generate subpopulations for monitoring miR-223 expression (Fig. 2b), for monitoring fates after additional culture (Supplementary Fig. 4c) and for mRNA profiling (Table 1 and Supplementary Data 4). For comparison, neutrophils directly isolated from wild-type and mutant mice, using biotin-conjugated Gr-1 antibody and MACS (each biological replicate pooling cells from three mice), were examined using expression profiling, RNA blotting and immunoblotting.

Mass spectrometry analysis. Protein (50 µg) was reduced (5 mM DTT in 50 mM ammonium bicarbonate, pH 8.2, at 56 °C, 30 min) and alkylated (15 mM iodoacetamide, in 50 mM ammonium bicarbonate in the dark at room temperature, 20–22 °C, 25 min), and then separated into 10 fractions (HeLa samples) or 16 fractions (neutrophils) by SDS–PAGE. Each fraction was in-gel digested with trypsin (5 ng µl^{-1} in 50 mM ammonium bicarbonate, pH 8.2, at 37 °C, 16 h). Peptides were extracted in 50% acetonitrile (ACN) and 5% formic acid (FA), and

then dried down and desalted by reverse-phase (C18 StageTip). Peptide mixtures were resuspended in 5% ACN and 4% FA, and 20% of each mixture was analysed by LC–MS/MS in duplicate. Peptides were separated across a 55-min gradient ranging from 7% to 30% ACN in 0.1% FA in a microcapillary (125 µm × 17 cm) column packed with C18 reverse-phase material (Magic C18AQ, 5 µm particles, 200 Å pore size, Michrom Bioresources) and on-line analysed on a hybrid linear ion trap Orbitrap (LTQOrbitrap, ThermoElectron) mass spectrometer. For each cycle, one full mass spectrometry scan acquired at high mass resolution (60,000 at 400 m/z, AGC target = 1×10^6, maximum ion injection time = 1,000 ms) in the orbitrap analyser was followed by 10 MS/MS spectra on the linear ion trap (AGC target = 5×10^3, maximum ion injection time = 120 ms) from the ten most abundant ions. Fragmented precursor ions were dynamically excluded from further selection for 35 s. Ions were also excluded if their charge was either <2 or unassigned.

Protein database searches and peptide quantification. MS/MS spectra were searched against the IPI protein sequence database using the Sequest algorithm. Peptide matches were filtered to <1% false-discovery rate using a target-decoy database strategy and using as filters mass deviation (in p.p.m.), Sequest Xcorr and dCn scores, and excluding sequences containing simultaneously heavy and light versions of Lys and Arg residues. Peptides were quantified using in-house Vista software[37,38] by peak-area integration, and heavy/light peptide ratios were calculated. Among the set of independent measurements retained for each protein, the median of the heavy/light ratio was defined as the protein fold change (Supplementary Data 1–4). Quality cutoffs were as follows: all measurements were required to have a Vista confidence score ≥75 and a signal-to-noise ratio (S/N) ≥6.0, where the S/N parameter was calculated as the sum of S/N_{Heavy} and S/N_{Light}. Measurements for proteins quantified with only one peptide were required to pass a more stringent S/N cutoff of 10.0. For proteins quantified with multiple peptides, independent measurements from a single peptide were not allowed to exceed half of the total number of independent measurements (by eliminating those measurements with lower S/N); this ensured that measurements for more than one peptide would influence the median.

To link the protein fold change to our reference cDNA set, the genomic coordinates of proteins from the IPI database[39] were used, requiring ≥50 nucleotide overlap between the genomic coordinates of the protein and a reference cDNA. To correct for the overall displacement of heavy and light populations (presumably caused by slightly unequal cell mixtures), we identified the subset of the proteins deriving from messages without 6–8mer seed-matched 3′-UTR sites, computed the difference in the median of heavy and light peaks, and offset all the fold-changes (including those from messages with sites) by this difference. This normalization caused our reported fold-change distribution of the proteins with no seed-matched sites to centre on zero.

cDNA data sets of non-redundant 5′-UTR, ORF or 3′-UTR sequences. We obtained human full-length cDNAs from RefSeq[40] and H-Invitational[41] databases, and aligned them against the human genome[39] using BLAT[42]. Functional cDNAs were enriched as described previously[43], discarding those without introns as well as those with a low alignment quality, multiple high-scoring matches to the human genome, a premature stop codon or an incomplete coding sequence. If cDNAs had overlapping 3′ UTRs, those obtained from the RefSeq database were chosen. If more than one cDNA remained, the cDNA with the longest 3′ UTR was retained. The resulting set of non-redundant cDNAs was designated the 'reference cDNAs' (Supplementary Data 5). Multiple reference cDNAs for a single gene were allowed if the genomic coordinates of their 3′ UTRs did not overlap with each other. However, when performing analysis of sites in ORFs or 5′ UTRs, only a single cDNA was arbitrary chosen (from among the RefSeq cDNA, when present) to represent the gene, to prevent double counting the contribution from a single site. The same criteria were used to choose a unique reference cDNA to match each quantified protein. To search for miRNA seed-matched sites, the genomic sequence of the reference cDNA (with introns removed) was used instead of the cDNA sequence itself. The analogous procedure was repeated for mouse full-length cDNAs, from RefSeq and FANTOM DB[44] databases, aligned against the mouse genome (Supplementary Data 5).

Microarray data processing. The 60-nucleotide probe sequences of Agilent 4 × 44K microarray were aligned against the human genome using BLAT. Any probe that had a less than a perfect match to the human genome or multiple perfect matches was removed. The mRNA fold change and the corresponding error, generated by the Agilent Feature Extraction Software, were linked to our reference cDNA set by a method analogous to that used for the SILAC data described previously (Supplementary Data 1–3). Similarly, a set of probe 'consensus sequences' from the Affymetrix mouse 430 2.0 microarray were aligned against the mouse genome. Any probe consensus sequence that had a BLAT alignment score of <100 or that had multiple high-scoring matches to the genome (that is, whose top two alignments to the genome had <1% difference in percentage identity) was removed. For each probe consensus sequence, the

mRNA fold change between the wild type and *mir-223*$^{-/Y}$ and its standard error were computed after quantile-normalizing the expression data from the multiple chips using the RMAExpress software[45] (Supplementary Data 4). When mapping the Agilent probes and Affymetrix probe consensus sequences to our reference cDNA set, ≥15 nucleotides of genomic coordinates between the probe and a reference cDNA were required to overlap.

Considerations for target site analyses. The minimal fraction of genes responding to the miRNA was calculated from cumulative distributions, determining the maximal cumulative difference between distributions, with correction for distribution bumpiness, as described[7]. To prevent undue impact from a few outliers, fold changes were truncated at ±2.0 before calculating mean log-fold changes. To evaluate sequence conservation of human reference cDNAs, human, mouse, rat and dog alignments were extracted from 28 vertebrate genome alignments (aligned against the human genome) obtained from the UCSC Genome Bioinformatics Site[46]. A site was considered conserved if also found in the orthologous positions of the other three genomes, allowing for horizontal shifts of the site (resulting from presumed artefacts or ambiguities in the alignment), provided that two of the alignment columns (each column being the width of one position in the alignment) overlapped the site in all four species. Similarly, from 30 vertebrate genome alignments (aligned against the mouse genome), the four mammalian sequences were extracted to assess the sequence conservation of mouse reference cDNAs and to identify conserved target sites.

Comparison of target prediction tools. Lists of miRNA targets predicted by TargetScan (release 4.1)[2,7], PicTar (human, chimp, mouse, rat, dog)[4,25], miRanda (January 2008 release)[23,24], miRBase Targets (version 5)[22], RNA22 (ref. 28) and PITA[26] were obtained from their respective websites, using the most recent predictions publicly available as of March 2008. Most of these consisted of gene symbols, sequence identification of full-length cDNAs, and/or scores. To map these predictions to the human or mouse genome, genomic alignments of RefSeq, Ensembl and UCSC genes were obtained from the UCSC Genome Bioinformatics Site[46], and the most informative set of alignments for each prediction tool was used. To prevent double counting, a single prediction was arbitrarily chosen for genes with multiple redundant predictions.

Small-RNA sequencing, RNA blots and protein blots. Small RNAs were sequenced on the Solexa platform using a protocol modified from that used previously[47]. RNA blots analysed 5 μg total RNA per lane and used carbodiimide-mediated cross-linking to the membrane[48]. Protein blots were probed using the following antibody dilutions: anti-Cstl goat monoclonal antibody (R&D Systems), 1:1,600; anti-Igf1r rabbit polyclonal antibody (Santa Cruz Biotechnology), 1:1,000; anti-Cbx5 (HP-1α) mouse monoclonal antibody (Millipore), 1:2,500; anti-actin mouse monoclonal antibody (Abcam), 1:25,000; and anti-actin rabbit polyclonal antibody (Cell Signaling), 1:10,000.

37. Everley, P. A. *et al.* Enhanced analysis of metastatic prostate cancer using stable isotopes and high mass accuracy instrumentation. *J. Proteome Res.* **5**, 1224–1231 (2006).
38. Guo, A. *et al.* Signaling networks assembled by oncogenic EGFR and c-Met. *Proc. Natl Acad. Sci. USA* **105**, 692–697 (2008).
39. Kersey, P. J. *et al.* The International Protein Index: an integrated database for proteomics experiments. *Proteomics* **4**, 1985–1988 (2004).
40. Pruitt, K. D., Katz, K. S., Sicotte, H. & Maglott, D. R. Introducing RefSeq and LocusLink: curated human genome resources at the NCBI. *Trends Genet.* **16**, 44–47 (2000).
41. Imanishi, T. *et al.* Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS Biol.* **2**, e162 (2004).
42. Kent, W. J. BLAT–the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
43. Baek, D., Davis, C., Ewing, B., Gordon, D. & Green, P. Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. *Genome Res.* **17**, 145–155 (2007).
44. Okazaki, Y. *et al.* Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**, 563–573 (2002).
45. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
46. Karolchik, D. *et al.* The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res.* **36**, D773–D779 (2008).
47. Ruby, J. G. *et al.* Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans. Cell* **127**, 1193–1207 (2006).
48. Pall, G. S., Codony-Servat, C., Byrne, J., Ritchie, L. & Hamilton, A. Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic Acids Res.* **35**, e60 (2007).