profiling and qPCR recovered a nonuniform distribution of microRNAs (**Fig. 2a**); we observed up to four orders of magnitude difference between the most and least frequently detected microRNAs. Only 61% (SREK-SOLiD) and 52% (modban-Solexa) of the microRNAs varied within a single order of magnitude (**Fig. 2a**). These results showed the inherent quantification biases of both DGE profiling and qPCR based on microRNA sequence, complicating comparison of microRNA amounts in a sample.

Correction of the biological dataset with the frequency bias obtained using the synthetic RNA pool did not improve the correlation between the library-preparation methods (data not shown). We therefore used the synthetic small RNA dataset to explore the potential basis of systematic biases. Although we found clear effects of certain terminal mono- and dinucleotides (**Supplementary Fig. 4**), we could not identify a satisfactory correction model based on primary (RNA sequence) and secondary (for example, folding characteristics) parameters (**Supplementary Fig. 5** and **Supplementary Note**). This might be explained by our observation that even single nucleotide differences influenced the read frequencies (**Supplementary Fig. 6**). RNA ligase preferences[7] may contribute to the observed different terminal nucleotides over the read frequency spectrum. In addition, the reverse-transcriptase reaction as well as the PCR could be a contributor to the bias[8].

To determine whether DGE profiling allows for differential expression analysis, we sequenced small RNA libraries from rat spleen and liver (SREK-SOLiD). In parallel, we analyzed the input RNA by qPCR. Similar to our previous results, qPCR data differed substantially from the read frequencies within a sample (**Fig. 2b**). However, differential expression results between samples obtained by qPCR and DGE profiling were strongly correlated (**Fig. 2b**), showing that the systematic biases do not prohibit the comparison of relative microRNA amounts between samples.

Despite the limitations described here, small RNA profiling by DGE is the method of choice for studying small RNA expression. In contrast to most other existing methods, DGE profiling is hybridization-independent, accurate in discriminating microRNA family members that differ by only a single nucleotide, capable of detecting 5′ and 3′ end variability (for example, isoMirs), and as the approach does not require a priori information, it can be used to simultaneously detect known and discover new biomolecules.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

This work was supported by the Cancer Genomics Center and Netherlands Bioinformatic Center through Netherlands Genomics Initiative (to E.C.), Fred Hutchinson Cancer Research Center and Canary Foundation funding (New Development funds to M.T.), Core Center of Excellence in Hematology Pilot Grant P30 DK56465 (to M.T.), Pacific Northwest Prostate Cancer Specialized Program of Research Excellence Grant P50 CA97186 (to M.T.) and a Rosetta Inpharmatics Fellowship in Molecular Profiling (to S.K.W.). We thank P. Toonen for animal care.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturemethods/.

Sam E V Linsen[1,7], Elzo de Wit[1,7], Georges Janssens[1], Sheila Heater[2], Laura Chapman[2], Rachael K Parkin[3], Brian Fritz[3,6], Stacia K Wyman[3], Ewart de Bruijn[1], Emile E Voest[4], Scott Kuersten[2], Muneesh Tewari[3,5] & Edwin Cuppen[1]

[1]Hubrecht Institute and University Medical Center Utrecht, Cancer Genomics Center, Utrecht, The Netherlands. [2]Life Technologies, Austin, Texas, USA. [3]Human Biology Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [4]Department of Medical Oncology, University Medical Center Utrecht, Utrecht, The Netherlands. [5]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [6]Present address: Illumina, Inc., San Diego, California, USA. [7]These authors contributed equally to this work.
e-mail: e.cuppen@niob.knaw.nl

1. Berezikov, E., *et al. Genome Res.* **16**, 1289–1298 (2006).
2. Ruby, J.G. *et al. Cell* **127**, 1193–1207 (2006).
3. Kuchenbauer, F. *et al. Genome Res.* **18**, 1787–1797 (2008).
4. Berezikov, E. *et al. Nat. Genet.* **38**, 1375–1377 (2006).
5. Lau, N.C., Lim, L.P., Weinstein, E.G. & Bartel, D.P. *Science* **294**, 858–862 (2001).
6. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. & Enright, A.J. *Nucleic Acids Res.* **34**, D140–D144 (2006).
7. Romaniuk, E., McLaughlin, L.W., Neilson, T. & Romaniuk, P.J. *Eur. J. Biochem.* **125**, 639–643 (1982).
8. Taube, R., Loya, S., Avidan, O., Perach, M. & Hizi, A. *Biochem. J.* **329**, 579–587 (1998).

RNAiCut: automated detection of significant genes from functional genomic screens

To the Editor: RNA interference (RNAi) is a popular functional genomic technology for identifying genes involved in a biological process. Although higher scores for genes in an RNAi screen suggest more central roles in the pathway, estimating the score threshold separating pathway- or process-relevant hits from noise remains difficult (**Supplementary Table 1**) and is typically done manually.

To overcome this subjective approach, we built a fully automated system, RNAiCut, that objectively and robustly identifies score thresholds from functional genomic data by introducing the use of the connectivity of subgraphs of protein-protein interaction (PPI) networks[1,2]. Unlike some previous work[3], our method does not overlap RNAi and PPI data to find interacting regulators. Instead, its guiding hypothesis is that true positive hits in an RNAi experiment are densely interconnected in the PPI network. For the k highest-scoring genes ($k = 1, 2, 3\ldots$), RNAiCut computes the edge count of the induced subgraph and estimates the P-value of finding a PPI subgraph of at least this size that is induced by k randomly chosen nodes that have the same degrees as these genes (**Supplementary Methods** and **Supplementary Results**). The plot of these P-values as a function of k is typically V-shaped, and we take the global minimum as the score threshold (**Fig. 1**). We used RNAiCut to compute thresholds for several *Drosophila melanogaster* RNAi screens[4] (**Supplementary Figs. 1–10** and **Supplementary Tables 2–3**).

RNAiCut chose successful thresholds, as measured by Gene Ontology (GO)[5] enrichment: the gene lists with above-threshold scores were enriched for functions relevant to the screen, compared to the rest (**Supplementary Table 4**). When the manual screener's threshold was later in the ranked list of hits than the RNAiCut threshold, choosing RNAiCut's threshold may reduce the potentially high number of false positives. When RNAiCut's threshold was later, the GO enrichment for RNAiCut's cutoff was at least as good as for the manually determined cutoff, revealing additional pathway-relevant genes (**Supplementary Results**). Although some of the additional hits identified by RNAiCut may be false positives, analyzing them may be useful given their apparent connectivity to

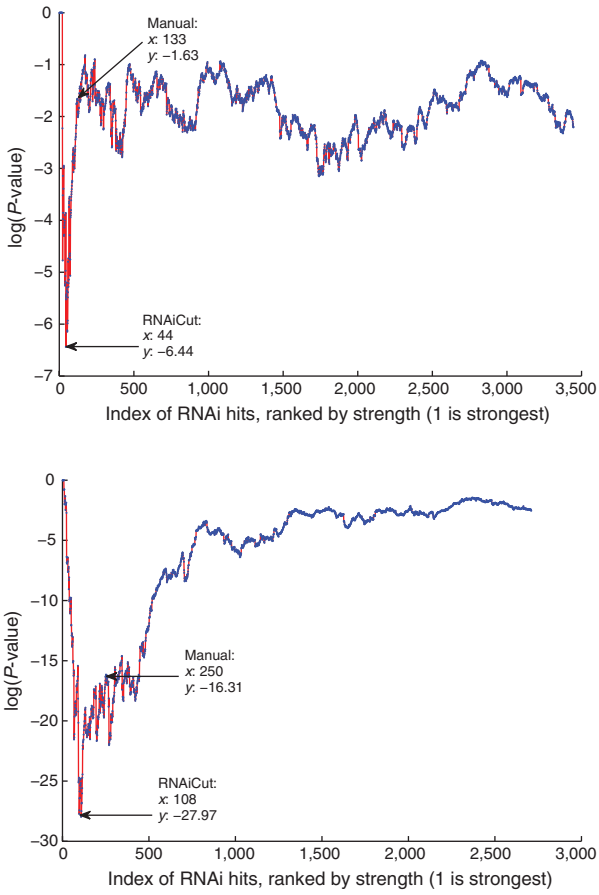

Figure 1 | RNAiCut results for insulin-triggered MAPK pathway screen in *D. melanogaster*[4]. Genes with positive (top) and negative (bottom) *Z* scores in the screen are ordered on the *x* axes from left to right based on the decreasing magnitude of *Z* scores. The *y* axis denotes the *P*-value, as a function of *k*, of finding a random PPI subnetwork as well connected as the one containing the *k* highest-scoring genes from the RNAi screen.

the core signaling pathway. RNAiCut was robust to *Z*-score noise generated by randomly scrambling close *Z* scores (**Supplementary Fig. 11** and **Supplementary Table 5**).

We offer an online server (http://rnaicut.csail.mit.edu) for interpreting functional genomic experiments. Although we developed RNAiCut using a fly PPI network, RNAiCut can also be run on non-fly and non-PPI networks (**Supplementary Fig. 12**). This tool will help functional genomics research by enabling hit-list gene selection using orthogonal datasets.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank J. Bienkowska, M. Booker and K.G. Lu. I.M.K. was partially supported by the John Reed Fund.

Irene M Kaplow[1,6], Rohit Singh[2,6], Adam Friedman[3–5], Chris Bakal[2–4], Norbert Perrimon[3,4] & Bonnie Berger[1,2,5]

[1]Department of Mathematics and [2]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [3]Department of Genetics and [4]Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts, USA. [5]Harvard–Massachusetts Institute of Technology Division of Health Sciences and Technology, Boston, Massachusetts, USA. [6]These authors contributed equally to this work.
e-mail: perrimon@receptor.med.harvard.edu or bab@mit.edu

1. Salwinski, L. *et al. Nucleic Acids Res.* **32**, D449–D451 (2004).
2. Stark, C. *et al. Nucleic Acids Res.* **34**, D535–D539 (2006).
3. König, R. *et al. Cell* **135**, 49–60 (2008).
4. Friedman, A. & Perrimon, N. *Nature* **444**, 230–234 (2006).
5. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).

Enabling IMAC purification of low abundance recombinant proteins from *E. coli* lysates

To the Editor: Currently, the most widely used method for purifying recombinant proteins for biochemical and especially structural studies is immobilized metal affinity chromatography (IMAC), in which a metal-binding polyhistidine tag (His tag) serves as a small purification handle on the target protein. IMAC is a powerful and generic purification method, with high recovery yields and low costs. Additionally, the His tag is compatible with most downstream applications because it is small and relatively inert[1,2]. *Escherichia coli* is by far the most popular expression host owing to its supremacy regarding cost, biomass production and technical simplicity[3,4]. However, a serious drawback of IMAC is the often-experienced failure to purify low-abundance His-tagged proteins from *E. coli* lysates; increasing the culture size and thereby increasing the amount of available His-tagged protein does not result in increased yield. We examined this issue and propose that it is tightly linked to metal-ion leakage from the columns induced by the *E. coli* lysate.

We used His-tagged GFP (His_6-GFP) to examine the effect of *E. coli* lysate on the protein binding capacity of IMAC columns. Application of the soluble fraction of *E. coli* lysate lacking recombinant protein expression to a 1 ml HiTrap Chelating HP column (GE Healthcare) partly loaded with His_6-GFP, caused extensive migration of His_6-GFP whereas application of wash buffer did not (**Supplementary Fig. 1a**).We confirmed this using different column materials and concluded that *E. coli* lysate severely reduces the binding capacity of the column (data not shown). By separating a lysate into high- and low-molecular-weight components we found that the reduced binding capacity was brought about by low-molecular-weight components, and not high-molecular-weight components (**Supplementary Fig. 1b**), implying that the underlying cause for the reduced target protein binding is not the result of native *E. coli* proteins competing with the His-tagged protein for the immobilized nickel-ion binding sites. We determined the amount of nickel present on the different columns before and after sample load and found that the decrease in binding capacity correlated with loss of immobilized nickel ions from the column (**Supplementary Fig. 1c**).

IMAC is very sensitive to the presence of metal chelators[1], and the *E. coli* lysate contains many unspecific weak chelators such as dicarboxylic acids from the citric acid cycle. Under stress conditions, *E. coli* can also produce highly specific metal chelators, metallophores[5]. We speculated that such metallophores, if produced, would be mainly associated with the periplasmic space of *E. coli* but not with the cytosol. We therefore hypothesized that removing the periplasmic material before cell lysis could improve His-tagged recombinant protein purification yields. We subjected *E. coli* cells to osmotic shock to remove the periplasmic material before cell lysis (**Supplementary Methods**). His_6-GFP did not migrate substantially on IMAC columns treated with lysate devoid of periplasmic