

Technical Note

Open Access

Fisher: a program for the detection of H/ACA snoRNAs using MFE secondary structure prediction and comparative genomics – assessment and update

Eva Freyhult^{1,6}, Sverker Edvardsson², Ivica Tamas^{3,7}, Vincent Moulton⁴ and Anthony M Poole^{*3,5}

Address: ¹The Linnaeus Centre for Bioinformatics, Uppsala University, Box 598, S-751 24 Uppsala, Sweden, ²Department of Natural Sciences, Mid Sweden University, S-851 70, Sundsvall, Sweden, ³Department of Molecular Biology & Functional Genomics, Arrhenius Laboratories for Natural Sciences, Stockholm University, SE-106 91 Stockholm, Sweden, ⁴School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ, UK, ⁵School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch, New Zealand, ⁶Department of Clinical Microbiology, Clinical Bacteriology, Umeå University, 901 85 Umeå, Sweden and ⁷Department of Biochemistry & Molecular Biology, Dalhousie University, Halifax, NS, Canada

Email: Eva Freyhult - Eva.Freyhult@climi.umu.se; Sverker Edvardsson - Sverker.Edvardsson@miun.se; Ivica Tamas - i.tamas@dal.ca; Vincent Moulton - vincent.moulton@cmp.uea.ac.uk; Anthony M Poole* - anthony.poole@molbio.su.se

* Corresponding author

Published: 21 July 2008

Received: 3 April 2008

BMC Research Notes 2008, 1:49 doi:10.1186/1756-0500-1-49

Accepted: 21 July 2008

This article is available from: <http://www.biomedcentral.com/1756-0500/1/49>

© 2008 Freyhult et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The H/ACA family of small nucleolar RNAs (snoRNAs) plays a central role in guiding the pseudouridylation of ribosomal RNA (rRNA). In an effort to systematically identify the complete set of rRNA-modifying H/ACA snoRNAs from the genome sequence of the budding yeast, *Saccharomyces cerevisiae*, we developed a program – Fisher – and previously presented several candidate snoRNAs based on our analysis [1].

Findings: In this report, we provide a brief update of this work, which was aborted after the publication of experimentally-identified snoRNAs [2] identical to candidates we had identified bioinformatically using Fisher. Our motivation for revisiting this work is to report on the status of the candidate snoRNAs described in [1], and secondly, to report that a modified version of Fisher together with the available multiple yeast genome sequences was able to correctly identify several H/ACA snoRNAs for modification sites not identified by the snoGPS program [3]. While we are no longer developing Fisher, we briefly consider the merits of the Fisher algorithm relative to snoGPS, which may be of use for workers considering pursuing a similar search strategy for the identification of small RNAs. The modified source code for Fisher is made available as supplementary material.

Conclusion: Our results confirm the validity of using minimum free energy (MFE) secondary structure prediction to guide comparative genomic screening for RNA families with few sequence constraints.

Findings

Small nucleolar RNAs (snoRNAs) guide nucleotide modi-

fications of ribosomal RNAs (rRNAs), as well as an expanding repertoire of cellular RNAs [4,5]. SnoRNAs can

be divided into two broad families: the C/D-box snoRNAs that guide site-specific 2'-O-methylation of ribose [6,7], and the H/ACA snoRNAs [8,9] that guide specific conversions of uridine (U) to pseudouridine (ψ).

Both H/ACA and C/D-box snoRNAs target the exact nucleotide for modification via base-complementarity to their target molecule, thereby permitting sequence-specific binding to the region of the target to be modified. Based on the guide regions as well as other known sequence motifs and structural elements, search tools have been developed both to identify C/D-box [10,11] and H/ACA snoRNAs [1,3,11], and, more recently, snoRNAs with no known target sequence [12,11]. Tools are also available for screening archaeal genomes for snoRNA-like sRNAs [13-15].

We previously developed the program Fisher as a tool to screen for H/ACA snoRNAs in the genome of the budding yeast, *Saccharomyces cerevisiae* [1]. Since the conserved sequence motifs of the H/ACA box snoRNAs are short (and in the case of regions of intermolecular interaction with rRNA target molecules, discontinuous) (Figure 1) the H/ACA box snoRNAs are harder to identify based on sequence motifs only, and Fisher therefore relies on both primary and secondary structure searches. However, Fisher suffers from a high frequency of false positive pre-

dictions (a large number of potential candidates are predicted, but many of these will not be true snoRNAs), and so additional screening of the Fisher candidates is necessary. This is likewise true for the snoGPS program [3].

In this report we describe some minor modifications to Fisher, a brief update on the candidate snoRNAs we reported in Edvardsson et al. [1], and a comparative genome analysis using 14 available yeast genome sequences which heavily reduced the number of false positives and enabled us to identify three yeast snoRNAs not identified by Schattner and colleagues' snoGPS program. We therefore provide a brief comparison of the two search strategies.

The core bioinformatic analyses reported forthwith were completed in 2004, but while we were working to experimentally characterize three candidates arising from our screen, Torchet et al. [2] published a characterization of those same candidates, which they found independently using a lab-based experimental approach. Hence we elected not to pursue the project further. We present the bioinformatics part of our previous study on account of the fact there are ongoing requests for the source code of Fisher, and we therefore wish to provide additional information and results to those originally presented in Edvardsson et al. [1] concerning the potential utility of

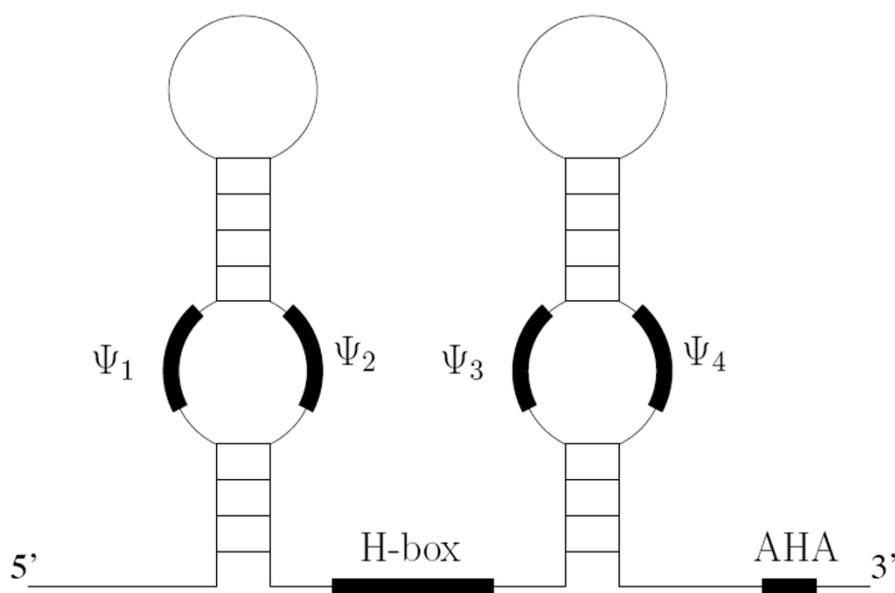


Figure 1

Idealised secondary structural representation of an H/ACA family snoRNA. H/ACA snoRNAs generally form a double hairpin structure. Each hairpin carries an internal bulge, called a pseudouridylation pocket, which may possess regions of base complementarity to a target RNA (not all snoRNAs carry two functional pseudouridylation pockets). We have designated these regions ψ_{1-4} for descriptive purposes. Other key motifs are the H-box, located in the hinge region between the two hairpins, and the ACA box at the 3' end of the molecule.

this program, including confirmation that such a strategy can in principle be used on other datasets. A brief description of the algorithm follows. For a full description, see [1]. For the source code of Fisher, see Additional file 1.

The Fisher algorithm in brief

The essence of the algorithm is as follows. First the search is made for an H-box of the form $AN_1AN_2N_3N_4N_5$. This is then scored using a probabilistic model. If the score is acceptable the H-box is accepted. The algorithm then searches downstream for ψ_3 and ψ_4 motifs (see Figure 1, Table S1 in Additional file 2) in appropriate locations. Then the algorithm continues to search for the AHA sequence. The complete H-AHA region together with a variable upstream sequence from the H-box is then passed to the secondary structure filters where acceptable folds are investigated. It is also possible to require a hairpin to the left of the H-box containing ψ_1 and ψ_2 motifs (see Figure 1, Table S1 in Additional file 2) in appropriate locations. The candidates are scored depending on both primary and secondary structure.

Previously reported candidate snoRNAs

In Edvardsson et al. [1], we presented three possible candidate *S. cerevisiae* snoRNAs, two of which were located in the introns of ribosomal protein genes (coding for RPL43A and RPS11A). A third candidate was overprinted on the coding sequence of the gene coding for snoRNP U3 protein MPP10. Neither these, nor any other of the 50 high-scoring candidates from our initial analysis could be detected using either RT-PCR or Northern Blots, whereas control snoRNAs were readily identified (AMP, P. A. McLenachan, A. R. Gore, A. M. Idicula, unpublished observations). We therefore concluded that the real false positive rate was prohibitively high, despite favourable preliminary results reported in Edvardsson et al. [1].

Combining Fisher with comparative genome data

Next, we made use of genome sequences from 14 additional yeasts, in order to establish whether sequence conservation could be employed as a filter to reduce the high false positive rate. The genomes used were as follows: *S. bayanus*, *S. mikatae* [16,17], *S. paradoxus*, *K. waltii* [17], *S. castellii*, *S. kluyveri*, *S. kudriavzevii* [16], *C. glabrata*, *D. hansenii*, *K. lactis*, *Y. lipolytica* [18], *C. albicans* [19], *S. pombe* [20] & *A. gossypii* [21] (see Figure 2 for overview of their phylogenetic relationships).

As in Edvardsson et al. [1], the *S. cerevisiae* genome was downloaded from the Saccharomyces Genome Database (SGD) [22], and only a reduced version of the genome was used for the snoRNA screen: i.e. all regions corresponding to open reading frames were removed, with introns added back into the dataset. We modified Fisher to accept ψ_4 to ACA distances of between 13 and 16

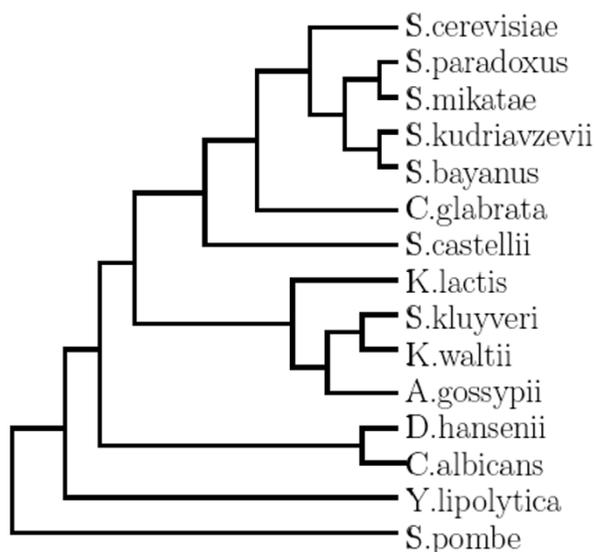


Figure 2
Evolutionary relationships of the yeast genomes used in the comparative genomics screen. Schematic based on a tree computed for the 25S rRNA using maximum parsimony, as presented in the supplementary material of Dujon et al. [18] Branch lengths are not preserved from the original maximum parsimony analysis, and only those taxa used in the current analysis are shown.

nucleotides, and allowed any base at the middle position of the ACA-box, which increased the number of potential candidates. We then used blastn [23] to screen the other yeast genomes for sequences homologous to these candidates, using a penalty of nucleotide mismatch of -4, a reward for nucleotide match of +5, and a gap creation and extension penalty of -10. Blast hits shorter than 40 nucleotides, with an identity below 40% or an E-value above 0.01 were not considered as potential homologues. Blast alignments were extended to cover the full length of the *S. cerevisiae* candidate sequence and realigned using CLUSTAL W [24]. This was done iteratively by adding nucleotides to the homologous sequences and realigning until the alignment covered the full candidate sequence. This gave a list of pairwise alignments between each *S. cerevisiae* candidate and potentially homologous sequences in other yeast genomes, for which pairwise identities could be established.

We compared these results with the results of blasting known *S. cerevisiae* H/ACA snoRNAs against the comparative genome dataset. We found that all snoRNAs known at the time of analysis had homologs (as identified by blastn) in at least three of the four genomes that are most closely related to *S. cerevisiae* (*S. paradoxus*, *S. mikatae*, *S. bayanus* and *S. kudriavzevii*), with high sequence identities

(see Table S2 in Additional file 2). The sequence identities were all above 89, 82, 78 and 82%, for *S. paradoxus*, *S. mikatae*, *S. bayanus* and *S. kudriavzevii*, respectively. We therefore further examined those candidates where potential homologues were identified in at least three of these four genomes, and with a pairwise sequence identity not below 87, 80, 76 and 80%, respectively.

Sequences in other genomes with high sequence identity to Fisher candidates were examined for the presence of the following features: H-box, ACA-box, and a 3' pseudouridylation pocket (consisting of ψ_3 and ψ_4 and elements – See Figure 1, Table S1 in Additional file 2) downstream of the H-box. These features were required to be completely conserved or to at least fulfil the same requirements as specified by Fisher (see [1]). We examined rRNA regions around the pseudouridylated nucleotides in *S. cerevisiae* and confirmed that these were conserved across the yeast species investigated. Hence the ψ_3 and ψ_4 and elements making up the pseudouridylation pockets of potentially homologous sequences were all required to be complementary to the same sequence (see Table S1 in Additional file 2). We likewise required the ACA-box to be located at a distance of between 13 and 16 nucleotides from the 5' end of the ψ_4 element. Finally, we made sure that all homologous sequences had an H-box upstream of the ψ_3 element at a position corresponding to the H-box position in the aligned candidate sequence, or not more than ten nucleotides upstream or downstream from this position.

We next examined the secondary structure of all candidates that fulfilled the above sequence criteria. H/ACA snoRNAs form a hairpin in the region between H-box and the ACA-box, with the ψ_3 and ψ_4 and elements forming an interior bulge (the 3' pseudouridylation pocket). Observation of known snoRNA structures indicated that the region just downstream of ψ_3 and just upstream of ψ_4 are always base paired. This observation was investigated in more detail for all known snoRNAs by folding the regions just downstream/upstream of ψ_3 and ψ_4 using RNAfold [25,26]. Table S3 (Additional file 2) shows the number of base pairs predicted for the stem-loop structure formed between the nucleotides immediately 3' and 5' of the ψ_3 and ψ_4 elements, and the number of unpaired bases between the pseudouridylation pockets and the stem region immediately above. The investigation showed that every known H/ACA snoRNA with a 3' pseudouridylation pocket have, at a distance of at most one nucleotide from the ψ_3 and ψ_4 boxes, a region of at least three complementary bases.

We therefore used these criteria for assessing candidate snoRNA sequences with RNAfold, and discarded all sequences that did not conform to these folding criteria.

Finally a manual check of folding for the complete sequences of all remaining snoRNA candidates served to filter out candidates that did not conform to the expected snoRNA structure.

Identification of candidate snoRNAs and conservation across yeast genomes

To be able to compare our results with the results reported by Schattner et al. [3], we performed a search for snoRNAs predicted to guide pseudouridylation at nine sites on *S. cerevisiae* rRNA, for which snoRNAs were identified for six targets (Schattner et al. [3], and three of which remained unidentified at the time of our analyses (maps 10, 13, 15, 22, 32, 35 and 7, 14, 33, respectively – see Table S1 in Additional file 2)). For the six targets studied by Schattner et al., Fisher reported 1590 unique snoRNA candidates. The comparative filters reduced this number to eight candidates (see Table 1). Among these candidates were snR80, snR81 and snR82, i.e. three out of the six novel snoRNAs identified by snoGPS [3]. Two of the novel snoRNAs, snR84 and snR85, cannot be found by Fisher since they only possess a 5' pseudouridylation pocket and Fisher can only identify snoRNAs with a single 3' pocket and snoRNAs carrying two (i.e. both a 5' and a 3') pocket. snR83 possesses a 3' pseudouridylation pocket, but was not among the 1590 candidates. It is rejected by Fisher because it does not fold into a secondary structure with two hairpins. It has a canonical 3' hairpin, but lacks the 5' hairpin and is therefore rejected.

For the remaining three target sites (maps 7, 14 and 33 in Table S1, Additional file 2), Fisher reported 2256 candidate snoRNAs for a reduced genome dataset (described in [1]). The comparative filter reduced this number down to two candidates for each target site (Table 1).

We subsequently performed Northern blots to probe for expression of these candidates, and found that candidates 2 and 11 were expressed in *S. cerevisiae* (IT & AMP, unpublished observations). Candidate 2 corresponds to snR80, which was determined by Torchet et al. [2] to guide pseudouridylation of SSU₇₅₉ and LSU_{776'}, and candidate 11 is snR86, which Torchet et al. demonstrated was required for pseudouridylation of LSU₂₃₁₄. None of the candidates for position 14 (SSU₁₄₁₅) were detected on Northern blots, and we concluded that these are false positives. Torchet et al. demonstrated that this pseudouridylation was guided by snR83, which Fisher fails to detect. As is evident from Table 1, snR80, snR82 and snR86 are conserved in diverse yeast species. Both snR80 and snR86 both form hairpin structures (3' hairpins shown in Figure 3). Alignment of snR80 from *S. cerevisiae* with snR80 sequences from other yeasts, shows that the H-box, ACA-box and the ψ_3 and ψ_4 boxes forming the 3' pseudouridylation pocket are almost perfectly conserved, with overall sequence similarity

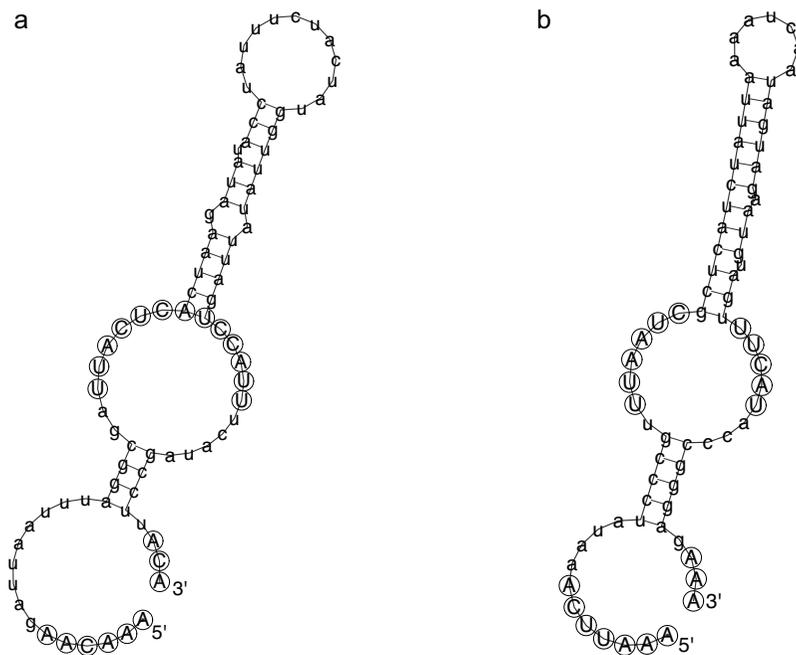


Figure 3
Secondary structure predictions showing 3' hairpin structures of the two top candidates identified by Fisher. 3' hairpin structures from H- to ACA-box for **(a)** candidate 7 from Table 1, corresponding to snR80, and **(b)** Candidate 11 from table 1, corresponding to snR86. Residues forming the H-box, ψ -pocket and ACA-box are capitalised and circled.

being high among sensu stricto species (*S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, and *S. bayanus*) (see Figure S1, Additional file 3). In the case of snR86, only the sequence corresponding to the 3' hairpin is conserved (i.e. from just upstream from the H-box to the 3' end) (see Figure S2, Additional file 3). Based on a comparison between snR86 sequences from *S. cerevisiae* and *C. glabrata*, Torchet et al. proposed that the large 5' region of snR86 was structurally conserved. We examined snR86 sequences from 10 yeast species, and, given the low sequence similarity, uncertainty on the genomic position of the 5'-end for these sequences, and the consequent difficulty in reliably aligning these, we are only confident of the conservation of structure of the 3' region (Figure 3, Figure S2 in Additional file 3).

Comparison of our results to those for snoGPS

snoGPS [3] uses a combination of deterministic tests and a probabilistic model to search for snoRNA genes. The program has the option to search for only the 5' stem-loop or only the 3' stem-loop, but as stated in [3] the two-stem alternative is usually used, where a two-hairpin structure is required. snoGPS starts by searching for the guide sequences and the downstream ACA- or H-box. The hairpin structure is investigated by, for example, measuring

distances between boxes, and complementarity in the stem regions. For detecting two-stem snoRNAs snoGPS then proceeds to look for a second stem-loop either upstream or downstream of the first. Candidates are then scored using a probabilistic model trained on known snoRNAs. The deterministic tests in snoGPS are generally looser than the corresponding tests in Fisher, but the probabilistic model helps cut down the search space. However, the snoGPS search of the *S. cerevisiae* genome reported in Schattner et al. [3] was complemented with comparative genome analysis and free-energy calculations, to further reduce the number of false candidates. This search resulted in the identification of three new snoRNAs (snR80, snR81, snR85), and confirmation of three additional candidate H/ACA snoRNAs identified previously using QRNA (snR84, snR82, snR83) [27].

The comparative genomics filter presented here significantly reduced the number of snoRNA candidates output by Fisher. As described above, Fisher detected three of the six snoRNAs identified with snoGPS (snR80, snR81, snR82), and snR86, not identified by snoGPS. In addition, Fisher correctly identified snR80 as the candidate responsible for modification at SSU₇₅₉. Given that the deterministic filters used in snoGPS are more generous

than the corresponding filters in Fisher, we would expect that the candidates we identify would probably have been excluded due to the probabilistic filters implemented in snoGPS. Another possibility is that the candidates did not pass the comparative genome analysis in [3], which contained a manual evaluation stage.

Comparing the results from our method with the results reported in [3] we conclude that our algorithm is more restrictive, but still at the same level of predictive power. A clear drawback with the Fisher algorithm is that it cannot detect snoRNAs with only a 5' pseudouridylation pocket, providing an obvious explanation as to why we could not identify snR84 and snR85. In the case of snR83, which possesses a 3' pseudouridylation pocket but not a 5' pocket, *in silico* folding reveals it possesses only a single hairpin structure. As Fisher requires snoRNA candidates to conform to a two-hairpin secondary structure, any snoRNAs with the structural characteristics of snR83 should not be identified. Ironically, snR86 is a very large snoRNA (~1000 nt), and the unusual size and predicted secondary structure reported therein [2] is such that we would expect Fisher to not detect it.

While Fisher reports a double hairpin structure for snR86, a multispecies alignment and secondary structure predictions do not strongly support this. Upstream of the H-box, there is limited conservation across diverse yeast, in contrast to the well-conserved downstream region (see Figure S2 in Additional file 3). It would certainly be possible to modify Fisher to allow single-stem structures. However, a caveat of such a modification would be an increased number of false positives. In that the success of multiple searches depends heavily upon the availability of multiple genome sequences from related species, an increase in false positive results may not necessarily represent a major obstacle for such datasets.

Conclusion

Our analyses demonstrate that, in spite of the high false positive rate apparent when screening a single genome, the Fisher algorithm is effective if used in combination with a comparative genomics analysis. The combined approach reported here resulted in a small number of candidates, thereby permitting subsequent experimental screening for bona fide snoRNAs. This report confirms the utility of minimum free energy (MFE) secondary structure prediction as a method for screening for families of structural RNA with few sequence constraints, and demonstrates the effectiveness of combining secondary structure with comparative genome data.

Availability and requirements

- Project name: Fisher (snoRNA search)

- Project home page: N/A
- Operating system(s): Platform independent
- Programming language: C
- Other requirements: Vienna RNA Package, see
- <http://www.tbi.univie.ac.at/~ivo/RNA/>
- License: Freeware
- Any restrictions to use by non-academics: None

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

EF performed the majority of the computational analyses described herein; AMP performed additional analyses to update results. SE made modifications to Fisher, ran the program, and assisted in analysis of results. IT performed Northern hybridizations for candidates in Table 1. VM & AMP initiated the project, which was developed with input from all authors. EF and AMP analysed results and wrote the manuscript. All authors read and approved the final version.

Additional material

Additional file 1

The source code of Fisher.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1756-0500-1-49-S1.zip>]

Additional file 2

Three target sites, maps 7, 14 and 33.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1756-0500-1-49-S2.doc>]

Additional file 3

Figure S1.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1756-0500-1-49-S3.pdf>]

Acknowledgements

We gratefully acknowledge the efforts of A. R. Gore, A. M. Idicula & P. A. McLenachan who contributed to the experimental screen of the candidates presented in [1]. We also thank Paul Gardner for assistance and advice on RNA structural analyses. AMP acknowledges the support of the Knut and Alice Wallenberg Foundation, and the New Zealand Marsden Fund.

References

- Edvardsson S, Gardner PP, Poole AM, Hendy MD, Penny D, Moulton V: **A search for H/ACA snoRNAs in yeast using MFE secondary structure prediction.** *Bioinformatics* 2003, **19**:865-873.
- Torchet C, Badis G, Devaux F, Costanzo G, Werner M, Jacquier A: **The complete set of H/ACA snoRNAs that guide rRNA pseudouridylations in *Saccharomyces cerevisiae*.** *RNA* 2005, **11**:928-938.
- Schattner P, Decatur WA, Davis CA, Ares M Jr, Fournier MJ, Lowe TM: **Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the *Saccharomyces cerevisiae* genome.** *Nucleic Acids Res* 2004, **32**:4281-4296.
- Bachelier JP, Cavaille J, Huttenhofer A: **The expanding snoRNA world.** *Biochimie* 2002, **84**:775-790.
- Kishore S, Stamm S: **The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C.** *Science* 2006, **311**:230-232.
- Kiss-Laszlo Z, Henry Y, Bachelier JP, Caizergues-Ferrer M, Kiss T: **Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs.** *Cell* 1996, **85**:1077-1088.
- Nicoloso M, Qu LH, Michot B, Bachelier JP: **Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points to their direct role as guides for the 2'-O-ribose methylation of rRNAs.** *J Mol Biol* 1996, **260**:178-195.
- Ganot P, Bortolin ML, Kiss T: **Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs.** *Cell* 1997, **89**:799-809.
- Ni J, Tien AL, Fournier MJ: **Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA.** *Cell* 1997, **89**:565-573.
- Lowe TM, Eddy SR: **A computational screen for methylation guide snoRNAs in yeast.** *Science* 1999, **283**:1168-1171.
- Yang JH, Zhang XC, Huang ZP, Zhou H, Huang MB, Zhang S, Chen YQ, Qu LH: **snoSeeker: an advanced computational package for screening of guide and orphan snoRNA genes in the human genome.** *Nucleic Acids Res* 2006, **34**:5112-5123.
- Hertel J, Hofacker IL, Stadler PF: **SnoReport: computational identification of snoRNAs with unknown targets.** *Bioinformatics* 2008, **24**:158-164.
- Gaspin C, Cavaille J, Erauso G, Bachelier JP: **Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes.** *J Mol Biol* 2000, **297**:895-906.
- Muller S, Leclerc F, Behm-Ansmant I, Fourmann JB, Charpentier B, Branlant C: **Combined in silico and experimental identification of the *Pyrococcus abyssi* H/ACA sRNAs and their target sites in ribosomal RNAs.** *Nucleic Acids Res* 2008, **36**:2459-2475.
- Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP: **Homologs of small nucleolar RNAs in Archaea.** *Science* 2000, **288**:517-522.
- Clifften P, Sudarsanam P, Desikan A, Fulton L, Fulton B, Majors J, Waterston R, Cohen BA, Johnston M: **Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting.** *Science* 2003, **301**:71-76.
- Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES: **Sequencing and comparison of yeast species to identify genes and regulatory elements.** *Nature* 2003, **423**:241-254.
- Dujon B, Sherman D, Fischer G, Durrrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuvéglise C, Talla E, et al.: **Genome evolution in yeasts.** *Nature* 2004, **430**:35-44.
- Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, Magee BB, Newport G, Thorstenson YR, Agabian N, Magee PT, et al.: **The diploid genome sequence of *Candida albicans*.** *Proc Natl Acad Sci USA* 2004, **101**:7329-7334.
- Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, et al.: **The genome sequence of *Schizosaccharomyces pombe*.** *Nature* 2002, **415**:871-880.
- Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, Mohr C, Pohlmann R, Luedi P, Choi S, et al.: **The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome.** *Science* 2004, **304**:304-307.
- Weng S, Dong Q, Balakrishnan R, Christie K, Costanzo M, Dolinski K, Dwight SS, Engel S, Fisk DG, Hong E, et al.: ***Saccharomyces Genome Database (SGD)* provides biochemical and structural information for budding yeast proteins.** *Nucleic Acids Res* 2003, **31**:216-218.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
- Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680.
- Hofacker IL: **Vienna RNA secondary structure server.** *Nucleic Acids Res* 2003, **31**:3429-3431.
- Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P: **Fast folding and comparison of RNA secondary structures.** *Monatshefte für Chemie* 1994, **125**:167-188.
- McCutcheon JP, Eddy SR: **Computational identification of non-coding RNAs in *Saccharomyces cerevisiae* by comparative genomics.** *Nucleic Acids Res* 2003, **31**:4119-4128.
- Bon E, Neuvéglise C, Lepingle A, Wincker P, Artiguenave F, Gaillardin C, Casaregola S: **Genomic exploration of the hemiascomycetous yeasts: 6. *Saccharomyces exiguus*.** *FEBS Lett* 2000, **487**:42-46.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

