doi:10.1093/nar/gkp926

Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates

Samuel Hiard[1], Carole Charlier[2], Wouter Coppieters[2], Michel Georges[2,]* and Denis Baurain[2]

[1]Systems and Modeling, Montefiore Institute and [2]Unit of Animal Genomics, GIGA-R and Faculty of Veterinary Medicine, University of Liège, Belgium

Received July 14, 2009; Revised September 18, 2009; Accepted October 9, 2009

ABSTRACT

The Patrocles database (http://www.patrocles.org/) compiles DNA sequence polymorphisms (DSPs) that are predicted to perturb miRNA-mediated gene regulation. Distinctive features include: (i) the coverage of seven vertebrate species in its present release, aiming for more when information becomes available, (ii) the coverage of the three compartments involved in the silencing process (i.e. targets, miRNA precursors and silencing machinery), (iii) contextual information that enables users to prioritize candidate 'Patrocles DSPs', including graphical information on miRNA-target coexpression and eQTL effect of genotype on target expression levels, (iv) the inclusion of Copy Number Variants and eQTL information that affect miRNA precursors as well as genes encoding components of the silencing machinery and (v) a tool (Patrocles finder) that allows the user to determine whether her favorite DSP may perturb miRNA-mediated gene regulation of custom target sequences. To support the biological relevance of Patrocles' content, we searched for signatures of selection acting on 'Patrocles single nucleotide polymorphisms (pSNPs)' in human and mice. As expected, we found a strong signature of purifying selection against not only SNPs that destroy conserved target sites but also against SNPs that create novel, illegitimate target sites, which is reminiscent of the Texel mutation in sheep.

INTRODUCTION

The expression level of at least one-third of mammalian genes is fine-tuned by one or more of a total set of ~1000 miRNAs. This posttranscriptional regulation requires a functional silencing pathway with many components involved in nuclear and cytoplasmic miRNA processing, loading of the miRNP, recognition of the target and actual silencing. The corresponding sequence space, i.e. target sites, miRNA precursors and silencing machinery, is bound to suffer its toll of DNA sequence polymorphisms (DSPs) of which some will be functional and possibly affect phenotype. That this is indeed the case that has been demonstrated by (i) the identification of a mutation in the 3′-UTR of the ovine *MSTN* gene that causes increased muscle mass by creating an illegitimate target site for coexpressed miR-1 and miR-206 (1), and the report of >10 associations of polymorphisms in miRNA target sites (poly-miRTS) with human disease [reviewed in (2)], (ii) the identification of mutations in the seed region of human miR-96 responsible for nonsyndromic progressive hearing loss (3,4) and (iii) the identification of *DICER1* mutations in familial pleuropulmonary blastoma (5). To assist in the identification of DSPs that affect miRNA-mediated regulation, we have searched the public domain databases for single nucleotide polymorphisms (SNPs) and other polymorphisms in the three sequence compartments involved in miRNA control (targets, miRNA precursors and silencing machinery). The outcome of this search is browsable via the Patrocles website (http://www.patrocles.org/).

METHODS

Patrocles contents

Patrocles is built using data from public databases and from the primary literature (i.e. Supplementary data). The lingua franca used to merge all sources of genomic data is Ensembl annotations (6). This means that any gene/probe identifier or genome coordinate is mapped to one or more Ensembl genes (using cross-reference tables) prior to further processing. For miRNA catalogs, Patrocles relies on miRBase (7), which implies ignoring genuine homologs not yet annotated in miRBase. To maximize consistency, Patrocles performs all its mapping tasks internally. Thus, only miRNA names, coordinates

*To whom correspondence should be addressed. Tel: +32 4 366 41 51; Fax: +32 4 366 41 98; Email: michel.georges@ulg.ac.be and sequences (both precursors and matures) are fetched from miRBase, whereas other annotations (e.g. host genes) are computed on the fly. Our software architecture ensures that all input is mapped and all output is built using the same versions of Ensembl and miRBase throughout a given Patrocles release. However, in ancestrality and conservation assessments, some species may be represented by an older genome build than the one normally used in the corresponding Ensembl release. This is due to the Galaxy server (8) offering uneven access to the various genome-wide multiple species alignments stored in the University of California Santa Cruz (UCSC) genome database (9).

Patrocles has three species-templated pipelines written as a mixture of Perl and SQL queries. Each pipeline handles one of the sequence compartments involved in miRNA control, i.e. polymorphic targets, polymorphic miRNA precursors and polymorphic silencing machinery. As the target pipeline is relatively complex, a flowchart of its major steps is provided in Supplementary Figure S3. The two other pipelines start from miRNA precursors available in miRBase and from silencing machinery components manually selected among Ensembl genes, respectively. In both cases, DSPs are processed as for targets, except that neither ancestrality nor conservation is assessed. Using genomic intervals, miRNA precursors and machinery genes are tested for their inclusion in Copy Number Variants (CNVs) [human (Database of Genomic Variants) (10), mouse (11), rat (12)]. Similarly, machinery and protein genes hosting miRNA precursors are searched for identity with known human eQTL (13,14,15–19) or with genes subject to allelic imbalance (20,21). miRNA secondary structures are first computed with RNAfold (22), then constrained to textual stem-loops by 'unrolling' additional arm loops. Unrolled regions (if any) are shown in lowercase in the output.

Patrocles website

The Patrocles website is written in PHP and based on denormalized SQL tables for fast access. Though Patrocles finder relies on the same species-specific octamer lists as the static version, its algorithms are slightly cruder and directly implemented in PHP. This is likely to change in the future. Patrocles builds upon Ensembl 49 and miRBase 11.

Expression plots

Patrocles plots were generated with gnuplot (http://www.gnuplot.info/) and ImageMagick (http://www.imagemagick.org/). Coexpression plots comparing the expression of a given miRNA with its target gene were computed for all miRNA–target pairs affected by at least one Patrocles DSPs (pDSPs). For target gene expression, MAS5-condensed fluorescence intensities from SymAtlas (23) were reduced to one replicate-averaged value per tissue and per Ensembl gene. When several probes were available for a single gene, we selected the probe yielding the highest replicate-averaged expression summed across tissues. For miRNAs, we used either mature counts directly extracted from Landgraf *et al.*'s (24) atlas of miRNA expression or the expression level of the host gene (if any) as computed from SymAtlas. Since expression data for target and host genes derive both from SymAtlas, establishing tissular correspondence was straightforward, comprising only one-to-one relationships. However, as miRNA read counts were obtained from a distinct set of libraries, matching was slightly more complicated, including one-to-many, many-to-one or many-to-many miRNA–target links depending on mapping onto larger systems (e.g. central nervous system, hematopoietic system). In the latter case, a summary score corresponding to the mean of the 'many' was generated as well. Counts were affiliated to miRBase precursors based on mature sequences (allowing 3p extensions) and only matures reaching either ≥10 copies or ≥1% in a single library were considered. For all libraries, Supplementary Table S5 lists abbreviations and colors used in plots, along with mapping to tissues and larger systems.

eQTL plots relating target gene expression in lymphoblastoid cell lines to individual genotypes were computed for all pSNPs found in HapMap (25). pSNPs affecting several octamers have >1 eQTL plot. Normalized expression data were taken from Stranger *et al.* (14). Multiple probes per Ensembl gene were allowed but only probes for which at least one individual had an expression ≥8.0 were considered. In eQTL plots, the genotype leading to the functional octamer (either destroyed or created) is always shown on the left, while the ancestral genotype is denoted by a star. Mean expressions broken by genotype and HapMap population are shown as black dots with error bars for standard error.

RESULTS

Polymorphic targets

To identify DSPs in protein-coding genes that might influence miRNA-mediated regulation we downloaded aligned 3′-UTR sequences from the UCSC genome browser (9) using Ensembl annotation (6) for gene structures. DSPs mapping to the corresponding genome coordinates were then retrieved from Ensembl. Table 1 and Supplementary Table S1 show the number of genes with 3′-UTR sequences and corresponding DSPs obtained for the species studied so far: human, mouse, chimpanzee, rat, dog, cow and chicken. Ancestral and derived DSP alleles were determined from the alignment with the orthologous sequence of sister species when available (human ↔ chimpanzee; mouse ↔ rat). When no sibling sequence was available, an allele was considered ancestral if shared by at least one primate, one rodent and one nonprimate/nonrodent mammal. Table 1 and Supplementary Table S1 report the percentage of DSPs for which the ancestral allele could be determined. Human DSPs reported in the 1000 genomes project (26) were labelled as validated.

We defined two sets of miRNA target site motifs. The first (X-motifs) corresponds to 540 octamers identified by Xie *et al.* (27) on the basis of their unusually high motif conservation score in 3′-UTRs. The second corresponds to the 8-mer, 7-mer-A1 and 7-mer-m8 sites as defined by

Table 1. Patrocles DSPs in target genes for human and mouse

	Human		Mouse	
3′-UTRs				
No. of genes	24 319		21 911	
Sequence space	26 261 732		21 634 548	
DSPs in 3′-UTRs				
Total	136 159		126 589	
Known ancestral allele	114 305 (83.9%)		111 178 (87.8%)	
Validated	56 807 (41.7%)		62 150 (49.1%)	
Target site motifs				
X-octamers	540		540	
miRNAs	676		484	
miRNAs*	170		117	
L-octamers	683		466	
X- OR L-octamers	1164		948	
X- AND L-octamers	59		58	
L-heptamers	1265		882	
Target sites in 3′-UTRs				
X-targets	323 833		267 644	
L-targets	375 054		219 392	
X- OR L-targets	661 187		455 620	
X- AND L-targets	37 700		31 416	
Conserved X- AND L-targets	10 425 (27.7%)		9436	
Conserved X- NOT L-targets	64 010 (22.4%)		57 154	
Conserved L- NOT X-targets	30 290 (9.0%)		19 595	
Conserved 7-mer L-targets[a]	183 320		111 759	
Sequence space	4 072 176 (15.5%)		2 674 395 (12.4%)	
DSPs affecting target sites	X	L	X	L
3′-UTR pDSPs—total	20 679	26 719	19 657	17 505
3′-UTR pDSPs—DC + CC	1546 + 50	959 + 58	951 + 102	496 + 65
3′-UTR pDSPs—DNC	7392	10 328	7732	7250
3′-UTR pDSPs—CNC	9006	11 244	8545	7573
3′-UTR pDSPs—P	1944	3295	2290	2065
3′-UTR pDSPs—S	741	837	37	56
3′-UTR pDSPs—DC + CC (7-mers)[a]	–	4310	–	2664

[a] Only considering 7-mer L-targets not included in 8-mer X- or L-targets.

Lewis *et al.* (28) (L-motifs). '8-mer sites' correspond to the Watson–Crick (WC) reverse complement of nucleotides 2–8 of known miRNAs followed by an 'A anchor' at its 3′-end, '7-mer-A1 sites' to the WC reverse complement of nucleotides 2–7 of known miRNAs plus the 'A anchor' and '7-mer-m8 sites' to the WC reverse complement of nucleotides 2–8. Species-specific sets of miRNAs were downloaded from miRBase (7). Both mature miRNAs and passenger miRNAs* were considered, as abundant miRNAs* may reach higher tissular concentrations than rare miRNAs [e.g. (24)]. Table 1 and Supplementary Table S1 show the number of miRNAs (including 5p- and 3p-forms) and miRNAs* identified in the different species, as well as the corresponding numbers of L-8- and L-7-mers.

In human, X- and L-8-mers jointly define 1164 unique octamers of which only 59 (5%) are common. Unexpectedly, it thus appears at first glance that X and L-targets explore very distinct sequence domains. To further characterize the concordance between the two sets, we examined the degree of overlap between 7- and 6-mers embedded within the human X- and L-octamers. The 540 X-8-mers encompass 577 7-mers and 554 6-mers. The corresponding figures for the 683 human L-8-mers are 1265 and 1448, respectively. One hundred and eight (16%) human L-8-mers share at least one 7-mer with 335 X-8-mers (62%), while 277 L-8-mers (40%) share at least one 6-mer with 491 X-8-mers (91%) (Figure 1). Assuming that 6-mer sharing indeed reflects functional overlap, ~40% of the L-motifs thus capture most of the biology (91%) related to X-8-mers. At any rate, X-8-mers are also bound to include functional elements not related to miRNA-mediated regulation.

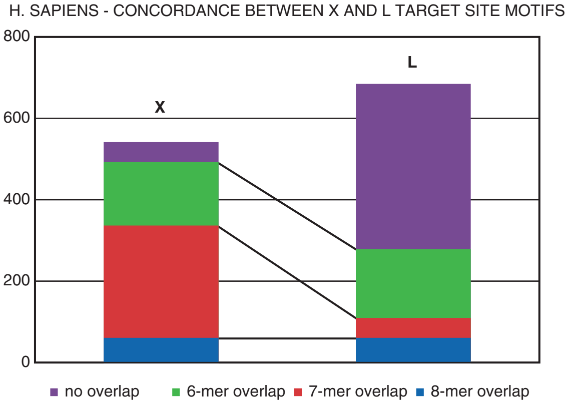


Figure 1. Number of X-motifs (27) (left column) and human L-motifs (28) (right column) with overlapping 8-mer (blue), 7-mer (orange), 6-mer (green) or without overlap (purple).

We then identified putative miRNA target sites in the selected 3′-UTR sequences, considering all matches to the defined X- and L-8-mers as well as matches to 'conserved' L-7-mers (A1 and m8). A conservation criterion was applied to L-7-mers to control the number of false positive predictions. Target sites were considered conserved if they were shared by at least one primate, one rodent and one nonprimate/nonrodent mammal. Table 1 and Supplementary Table S1 report the number of X- and L-targets identified with this procedure in the different species.

In human, for instance, 28% of the target sites that match both an X- and an L-octamer are conserved, versus 22% for those that only match an X-octamer, and 9% for those that only match an L-octamer. Thus, considering that conservation is indicative of functionality, matching an X- and an L-motif increases the probability to be a true miRNA target site. The higher proportion of conserved X-matching target sites versus L-matching target sites is as expected given the strategy underlying X-motif identification (27). Amongst L-target, the proportion of conserved target sites is higher for octamer motifs corresponding to mature miRNAs than to passenger miRNAs* (Figure 2A and B).

We then searched for DSPs that were altering the X- or L-target site content of the 3′-UTRs. We refer to these DSPs as pDPSs. pDSP for which the ancestral allele is known can modify target site content in the following ways: (i) destruction of a conserved target site (DC), (ii) destruction of a nonconserved target site (DNC) and (iii) creation of a nonconserved target site (CNC). pDSPs for which the ancestral allele is unknown (the general situation in species other than primates and rodents) were assigned to a fourth category of polymorphic target sites (P). Finally, pDSPs shifting the position of a target site were assigned to a fifth class (S). Note that the same DSP may cause multiple such events by affecting several overlapping target site motifs. Table 1 and Supplementary Table S1 show the number of events of each category observed for the two sets of target site motifs in the studied species. Figures for 8- and 7-mer L-targets are provided separately. It is worthwhile noting that, for 8-mer target sites, the number of target site destructions (DC + DNC) is virtually identical to the number of creations (CNC) for all species.

Exceptionally, we found primate (respectively rodent) pDSPs for which the derived allele corresponded to a target-site motif conserved across nonprimate (respectively nonrodent) mammals. In these cases, we assumed that it was more likely that the allele initially labelled as derived was in fact ancestral, and that the DSP actually appeared prior to the divergence of the two sibling species used to infer the ancestral state. Thus, these creations of a conserved site (CC) were parsimoniously added to the DC class, yet identified as such (the created target site is shown in the column corresponding to the derived allele). Occasionally, DSPs destroy a conserved or nonconserved 8-mer target site yet maintain a conserved 7-mer L-target site. Such events are identified using a weakening (W) label. Likewise, the CNC class include events converting a conserved 7-mer in an 8-mer target site. Such events are identified with a strengthening (S) label.

Taken together, our results indicate that there are thousands of common DSPs that alter the content of 3′-UTRs in putative miRNA target sites. This is not unexpected given the fact that >10% of the 3′-UTR sequence space is occupied by putative target sites (Table 1). What is the evidence that any of those are truly affecting gene function? We addressed this by looking for signatures of purifying selection on pSNPs in human and mice. To that end, we simulated sets of pSNPs matching the true human and mouse sets as follows. We first selected SNPs (i.e. excluding DSPs affecting >1 nucleotide residue and those corresponding to indels) in human (114 641 SNPs) and mice (125 693 SNPs). We then determined the ancestral allele or, when not possible, arbitrarily assigned ancestral status to the nucleotide in the reference sequence. Finally, we randomly shifted the position of the SNPs in the 3′-UTR space, yet respecting their trinucleotide context. For instance, a cAt (ancestral) → cGt (derived) transition was moved to a randomly selected cAt trinucleotide within the 3′-UTR space. Substitution rates are indeed known to depend on immediately surrounding nucleotides (29–31), while the trinucleotide composition of the miRNA target motifs differs from the general trinucleotide composition of 3′-UTRs (Supplementary Figure S1). To see this, assume that CpG dinucleotides (known to be C→T mutational hot spots) are enriched in miRNA target sites, shifting the mutated T residues to any C in the 3′-UTRs would reduce the proportion of pSNPs in the simulated data sets. The number of DC, DNC and CNC events was then compiled for this *in silico* generated SNP set. This operation was repeated 100 times. The number of 'Patrocles events' obtained with the true set was then compared with the distribution of number of events across simulations. Functional sites under purifying selection are expected to be more often affected by *in silico* SNPs than by real SNPs. On the contrary, nonfunctional, neutral sites are expected to be more often affected by real than by *in silico* SNPs (Supplementary Figure S2).

As expected, there is a strong signature of purifying selection against SNPs that destroy conserved target sites, whether X-targets, L-targets corresponding to mature miRNAs or L-targets corresponding to passenger miRNAs* (Figure 3 and Table 2). SNP avoidance is more pronounced in mice than in human, which could be due to a more effective selection against mildly deleterious mutations in the larger effective population of wild mice (prior to domestication) when compared with human, combined with a strong selection against deleterious recessive mutations as a result of inbreeding (after domestication). Purifying selection may have eliminated of the order of 22–35% of SNPs affecting conserved target sites in human versus 53–67% in mice. This observation corroborates the findings of Chen and Rajewsky (32) who noticed a depletion of SNPs in conserved miRNA target sites when compared to other conserved 3′-UTR sequences.

Interestingly, we also obtained evidence of purifying selection against SNPs that create novel, illegitimate

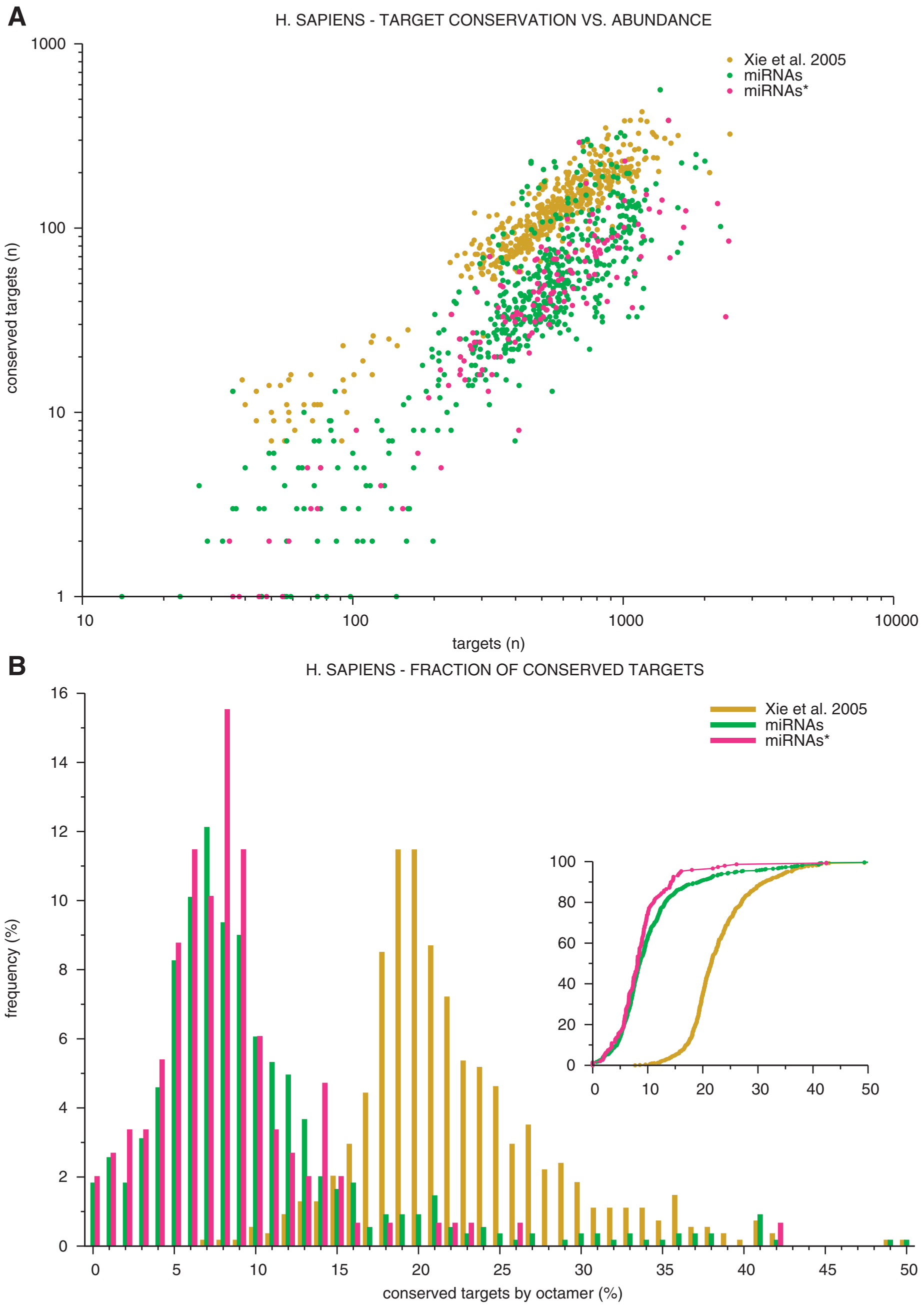


Figure 2. (**A**) Conserved versus total numbers of putative target sites in human 3′-UTRs for X-octamers (yellow), L-octamers corresponding to mature miRNAs (green) and L-octamers corresponding to passenger miRNAs* (red). (**B**) Frequency distribution of the proportion of conserved target sites in human 3′-UTRs for X-octamers (yellow), L-octamers corresponding to mature miRNAs (green) and L-octamers corresponding to passenger miRNAs* (red). Inset: corresponding cumulative frequency distributions.

target sites in human and mice (Table 2). Such events are reminiscent of the Texel mutation in sheep (1). The effect was most pronounced for X-targets, but clearly noticeable for L-targets corresponding to mature miRNAs as well. The observed ratios between real and simulated events suggest that as much as 10% of illegitimate target sites might be functional. A more modest signal of purifying selection against SNPs destroying nonconserved target sites (X-target sites and L-target corresponding to mature miRNAs) was also observed in human but not in mice (Table 2).

pDSPs that are most likely to affect gene function include (i) those destroying conserved target sites (DC) and (ii) those creating illegitimate target sites (CNC) in ‘anti-target’ genes. pDSP causing DC events can be selected as such in the Patrocles database. As mentioned before, target sites are considered conserved only when shared by at least one primate, one rodent and one other mammal. As a matter of fact, the DNC set must include a number of DSPs destroying target sites whose function and hence conservation is restricted to specific lineages. The alignment of the 3′-UTRs across a larger

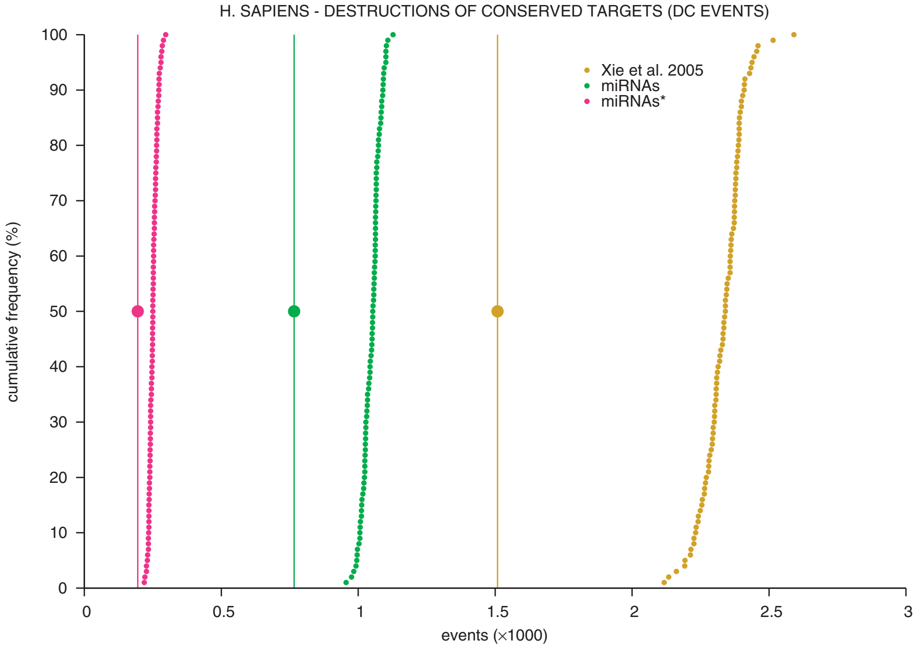


Figure 3. Large dots: number of destructions of X-8-mer targets (yellow), L-8-mer targets corresponding to mature miRNAs (green) or L-8-mer targets corresponding to passenger miRNAs* observed with the collection of real human SNPs. Small dots: cumulative frequency distribution of the corresponding number of DC events obtained with 100 matched collections of SNPs generated *in silico* as described in the text. Corresponding figures for DC, DNC and CNC events in human and mice are summarized in Table 2.

Table 2. Signatures of purifying selection on pSNPs in human and mice

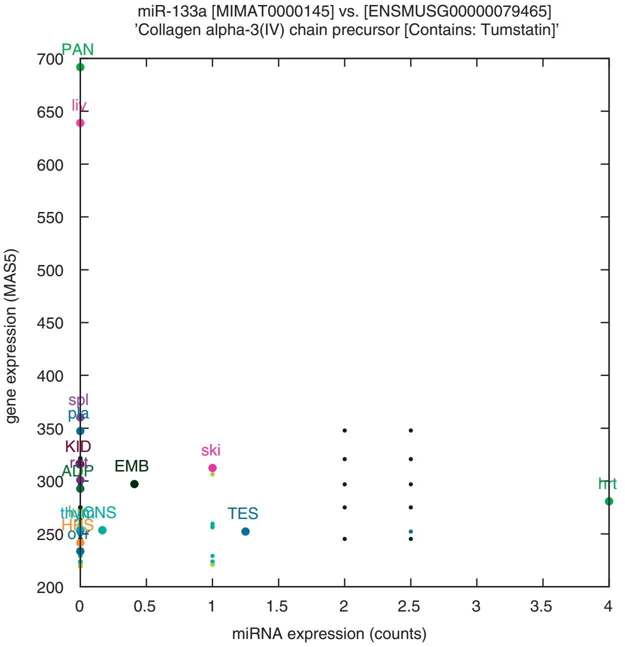
	DC			DNC			CNC			
	X	L	L*	X	L	L*	X	L	L*	
Human	1509	766	195	7592	8160	2530	9202	9005	2618	OBS
	0.647	0.730	0.775	0.968	0.968	1.009	0.916	0.953	0.987	[OBS/SIM]
	−10.832	−8.843	−3.766	−2.249	−3.302	0.503	−5.743	−4.489	−0.559	[OBS-SIM]/SD_SIM
Mouse	951	410	94	8850	6604	1759	9598	6711	1933	OBS
	0.324	0.390	0.467	0.987	0.997	1.070	0.890	0.959	1.014	[OBS/SIM]
	−23.933	−18.267	−8.320	−0.814	−0.240	2.706	−8.061	−2.949	0.627	[OBS-SIM]/SD_SIM

X, octamer motifs identified by Xie *et al.* (27); L, octamer motifs corresponding to 8-mer sites defined as in Lewis *et al.* (28) based on mature human miRNAs compiled in miRBase (7); L*, octamer motifs corresponding to 8-mer sites defined as in Lewis *et al.* (28) based on passenger human miRNAs* compiled in miRBase (7); OBS, numbers of corresponding events observed with real SNPs in 3′-UTRs; [OBS/SIM], ratio of the number of events observed with real SNPs divided by the mean number of corresponding events observed with *in silico* generated SNPs; SD_SIM, standard deviation of the number of corresponding events observed across 100 sets of simulated SNPs.

number of mammalian species may identify such lineage-specific target sites.

'Anti-targets' are genes that are under selective pressure to avoid target sites (33,34). The *G + 6723G-A* mutation in the ovine *MSTN* 3′-UTR is a good example of a pSNP that creates an illegitimate target site in an miR-1/206 anti-target and hence an hypomorphic *MSTN* allele (1). To assist in the identification of other such pDSPs, we provide graphical information about the coexpression of the polymorphic target gene and the cognate miRNA (when known) across tissues. This is achieved in the form of 2D plots in which each point reports the expression level of the corresponding target–miRNA pair in a given tissue. For target genes, expression levels correspond to fluorescence intensities obtained from SymAtlas (23), while for the miRNAs, expression levels correspond either (i) to the number of sequence reads as reported by Landgraf *et al.* (24) and/or (ii) to the expression level of the host gene in SymAtlas. Indeed, several reports indicate that the expression levels of host genes and intronic miRNAs expressed from the same strand are, in general, positively correlated [e.g. (35–38)]. It is noteworthy that an estimated ~1/3 of intronic miRNAs are predicted to be under control of an independent promoter (39). For those, host gene expression level may not be an appropriate surrogate of miRNA expression level.

Figure 4 shows an example of a coexpression plot based on the number of sequence reads for a murine pSNP recently shown to affect the binding of miR-133a to a collagen precursor (*Col4a3*) (40). At present, coexpression plots are available in Patrocles for human and mice.



target gene expression taken from SymAtlas 2.0
miRNA expression taken from Landgraf et al. (2007)

Figure 4. Example of coexpression plot for a miRNA–target pair in mice. Relative expression levels of mir-133a (based on the number of sequence reads reported in Landgraf *et al.* (24) versus *Col4a3* (collagen alpha-3(IV) chain precursor) (based on SymAtlas). The graph shows (i) that mir-133a is muscle specific (hrt) and (ii) that *Col4a3* mRNA levels in the heart are lower than in most other tissues. Interestingly, the murine *Col4a3* 3′-UTR encompasses an experimentally confirmed mir-133a target site corresponding to the ancestral allele (T), this allele being present in all sequenced mice lines (30) except in m.m. Castaneus, which harbours the derived allele (C) (rs30240795). By analyzing allelic imbalance in F1 mice heterozygous for this pSNP, Kim and Bartel (40) convincingly showed a higher mRNA steady state level for the derived allele (C) compared with the ancestral allele (T).

Assuming that a pDSP is truly functional, the steady state mRNA level of the targeted allele is predicted to be lower than that of the untargeted one in tissues where target and miRNA are coexpressed. Genome-wide expression data for cohorts of individuals genotyped for large numbers of SNPs are becoming available in human, albeit today only for lymphoblastoid cell lines. These are being used to examine associations between gene expression levels and SNP genotype across the genome, i.e. eQTL studies [e.g. (13,14)]. We have exploited the corresponding HapMap resource (25) to look for associations between pSNPs and the expression level of the corresponding gene (i.e. *cis* eQTL effects). Patrocles provides eQTL plots for pSNPs genotyped in the HapMap population. For a functional pSNP, the resulting *cis* eQTL effect is expected to be consistent across populations (Yerubans, Asians and Europeans). Figure 5 illustrates such a consistent eQTL effect associated with pSNP rs9874 causing the creation of a nonconserved target site for miR-181 in the selenoprotein S (*SELS*) gene, thereby supporting its functionality.

The Patrocles database also lists reported associations between pDSPs and phenotypes. In addition to the Texel mutation in sheep (1), 13 associations between pDSPs and human phenotypes had been reported at the time of writing: (i) miR-189-*SLITRK1* and Tourette's syndrome (41–46), (ii) miR-140-*REEP1* and hereditary spastic paraplegia (47,48), (iii) miR-206-*ERα* and breast cancer (49), (iv) miR-155-*AGTR1* and hypertension (50), (v) miR-24-*DHFR* and drug resistance (51) (note that the corresponding rs34764978 SNP lies outside of the target site, yet seems to have a strong effect on miR-24 dependent regulation), (vi) miR148a-*HLA-G* and childhood asthma (52), (vii) miR-96-*HTR1B* and aggressive behaviour (53), (viii) five miRNAs-*CD86* and colorectal cancer (54), (ix) miR-433-*FGF20* and Parkinson disease (55), (x) miR-510-*HTR3E* and irritable bowel syndrome (56), (xi) let-7-*KRAS* and nonsmall cell lung cancer (57), (xii) miR-34a-*ITGB4* and breast cancer (58) and (xiii) miR-657-*IGF2R* and Type II diabetes (59). It was recently commented by Sethupathy and Collins (2) that the evidence supporting the majority of associations reported in human should be considered tentative, requiring further confirmation as well as functional and mechanistic support. The Patrocles interface invites the community to assist in updating the list of published associations between pDSPs and phenotypes.

The 'Polymorphic target' section of the Patrocles database allows interrogation for pDSPs including filtering by species (presently human, chimpanzee, mouse, rat, dog, cow and chicken), by type of target site (X- and/or L-targets, miRNA identifier or octamer motif), by target gene (gene identifier or chromosomal interval) and by DSP category (effect on target site content, DSP validation status, DSP identifier). The output can be visualized on screen or downloaded as a text file.

Polymorphic miRNAs

To identify DSPs that might affect either the biogenesis or the sequence of miRNAs, we downloaded sequences annotated as pre-miRNAs from miRBase. We then downloaded from Ensembl all DSPs mapping to the corresponding genome coordinates. Table 3 and Supplementary Table S2 summarize the number of pre-miRNAs available in the species analyzed to date, as well as the number of DSPs in them. DSPs are sorted depending on whether they affect the seed sequence (residues 2–8), the mature miRNA outside the seed or other parts of the pre-miRNA. In human for instance, we identified 184 DSPs affecting 136 out of 676 pre-miRNAs. Twelve of these mapped to the miRNA seed and 26 to the mature miRNA outside the seed. It is noteworthy that the 12 human miRNAs with a DSP in their seed sequence are either members of a seed-sharing miRNA family or more recently discovered miRNAs that are likely to be expressed at lower levels [e.g. (60)]. The effect of the DSPs on pre-miRNA structure was evaluated using RNAfold (22) and the predicted secondary structures are viewable in Patrocles. Patrocles includes the DSPs in pre-miRNAs that were previously identified by Iwai and Naraba (61) as well as by Chen and Rajewsky (60).

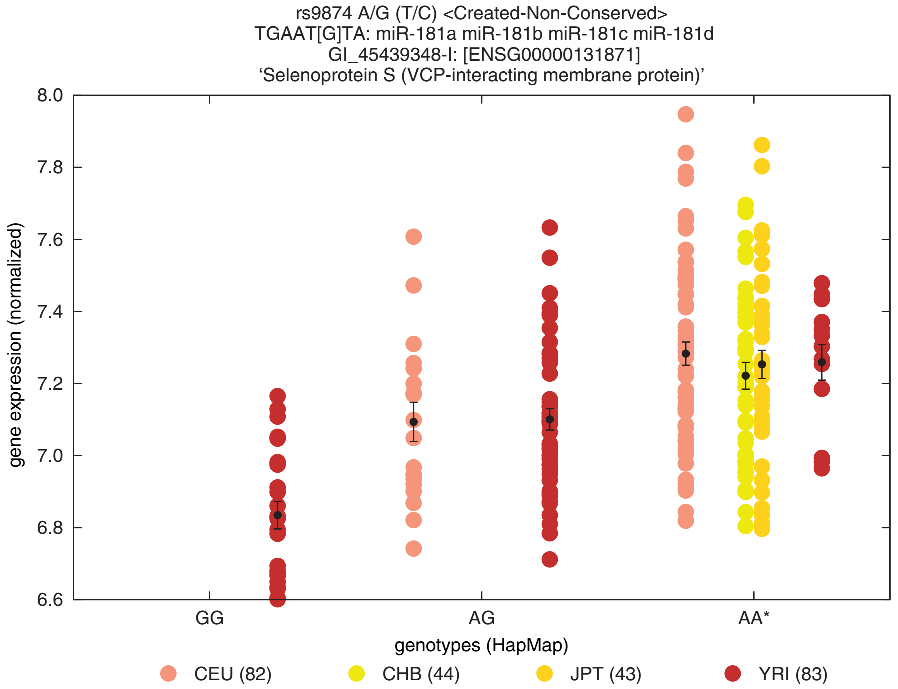


Figure 5. Comparison of the expression levels of *SELS* in lymphoblastoid cell lines of Yerubans (brown), Han Chinese (yellow), Japanese (orange) and Caucasians (pink), sorted by genotype for pSNP *rs9874* predicted to create a nonconserved miR-181 target site. Homozygous genotypic class for the ancestral allele (AA) is marked by an asterisk. Error bars correspond to standard error.

For human (10), mouse (11) and rat (12), Patrocles also lists CNVs [e.g. (62,63)] encompassing known miRNAs (Table 3 and Supplementary Table S2). These dosage differences may cause differences in miRNA concentration and hence influence miRNA-mediated gene regulation. In human, 158 reported CNVs jointly encompass 256 miRNA genes. The corresponding figures may have to be reevaluated in light of recently redefined CNV boundaries (64).

Finally, Patrocles lists host genes with evidence for inherited variation in expression level obtained either from eQTL experiments (13–19) or from genome-wide scans for allelic imbalance (20,21). Inherited variations in host gene expression levels, whether caused by sequence variants in *cis*- or *trans*-acting regulators, may affect the concentration of embedded miRNAs, and hence influence miRNA-mediated gene regulation. This information is presently only compiled for human, for which we found 78 eQTL potentially affecting the expression level of 85 miRNAs (Table 3).

Reports of DSPs in miRNAs that have been associated with altered miRNA expression or with a phenotype are listed as such in Patrocles as well. At the time of writing, this includes DSPs modulating miRNA expression level (65) or processing (66), as well as DSPs in miRNAs associated with different cancers (67–74) or with schizophrenia (75). As in the previous section, Patrocles invites the community to contribute in updating the list of published associations between miRNA polymorphisms and phenotypes.

The 'Polymorphic miRNA' tables of the Patrocles database can be queried using a variety of filters including species, miRNA (identifier and map position), type of variation (DSP, CNV or eQTL), DSP category (position with respect to miRNA structure, DSP validation status, DSP identifier), CNV identifier and host gene affected by a reported eQTL.

Table 3. Patrocles DSPs in miRNA precursors for human and mouse

	Human	Mouse
No. of pre-miRNAs	676	466
DSPs in pre-miRNAs		
No. of affected miRNAs	136	71
Total	184	89
Seed	12	4
Mature non-seed	26	6
Other	146	79
miRNAs in CNVs		
No. of CNVs	158	0
No. of affected miRNAs	256	0
miRNAs hosted in eQTL genes		
No. of eQTL	78	ND
No. of affected miRNAs	85	ND

ND, non-determined.

Polymorphic silencing machinery

DSPs in core components of the silencing machinery may affect the efficacy of specific steps in the silencing process. Not all biochemical pathways will be equally sensitive to such perturbations. DSPs affecting the silencing machinery may thus contribute to the genetic variation observed for specific phenotypes.

To aid in the identification of such variants, we established a manually curated list of gene products

Table 4. Patrocles DSPs in components of the silencing machinery for human and mouse

	Human	Mouse
No. of genes	52	51
DSPs in machinery genes		
No. of affected genes	49	35
Total	237	127
Non-synonymous	151	73
Stops/frameshifts	45	2
Splicing sites	42	52
Machinery genes in CNVs		
No. of CNVs	17	0
No. of affected genes	17	0
Machinery genes identified as eQTL		
No. of eQTL/affected genes	21	ND

ND, non-determined.

(Supplementary Table S4) participating in the silencing process. We then searched for (i) DSPs altering the corresponding coding sequences, (ii) CNVs encompassing the corresponding genes and (iii) evidence for eQTL or allelic imbalance for the corresponding genes. In human for instance, we observed 237 DSPs in 49 genes, 17 CNVs affecting 17 genes and 21 genes with evidence for variation in expression level. Table 4 and Supplementary Table S3 reports the corresponding numbers for the other species.

All these events are listed in the Patrocles database which can be interrogated by species, gene identifier, DSP identifier and chromosomal location.

Patrocles finder

Resequencing efforts will reveal novel, undocumented DSPs in candidate genes of interest. We have generated a tool (Patrocles finder) that allows convenient examination of the miRNA target site content of a sequence of interest and examination of the effect of DSPs in that sequence on target-site content. Target sites are defined as described above, i.e. either as one of the octamer motifs discovered by Xie *et al.* (27) or as species-specific 8- and 7-mer sites as defined by Lewis *et al.* (28). Patrocles finder analyzes both isolated sequences as well as alignments of orthologous sequences in FASTA format. When selecting the latter option, Patrocles finder provides direct information about the conservation or not of the identified miRNA target sites.

DISCUSSION

The Patrocles database aims at providing the community with a bioinformatic tool to assist in the identification of DSPs that may affect miRNA-mediated gene regulation and possibly phenotype. Patrocles may be particularly useful in the final stages of a positional cloning effort when a chromosomal region corresponding to a phenotype of interest has been identified either by linkage analysis or association studies.

At present, the majority of the information provided by Patrocles concerns DSPs in putative miRNA target sites. Patrocles reports tens of thousands of pDSPs for human alone.

Evidently, information provided by Patrocles should be considered with appropriate caution. It is worthwhile remembering in this regard that as much as 67% of coexpressed target genes predicted with the most effective packages on the basis of conserved target sites are likely to be false positives given the absence of a detectable response at the mRNA or protein level (76). This figure rises to >85% for target predictions based on nonconserved target sites, despite ample experimental evidence supporting the existence of functional yet nonconserved target sites [e.g. (33,76)]. Predictions are bound to be even less specific when ignoring coexpression information as most packages do.

While cautious interpretation is of the order, population genetic data strongly suggest that Patrocles and related databases contain biologically relevant information. Indeed, in addition to a strong signature of purifying selection against SNPs destroying conserved target sites, we present evidence for purifying selection against a significant proportion of SNPs creating nonconserved target sites in human and mice, and against SNPs destroying nonconserved target sites in human. One could rightfully argue that such signatures are indicative of past selection against SNPs that have since been eliminated from the population. That extant pSNPs affecting nonconserved target sites might affect gene function was supported by the finding of Chen and Rajewsky (32) of a shift towards lower frequencies of the derived allele for pSNPs altering nonconserved target sites for coexpressed miRNAs.

To assist in the prioritization of pDSPs, Patrocles provides convenient access to contextual information, including miRNA-target coexpression and eQTL data. Target site 'context score' as defined by Grimson *et al.* (77) could be considered in future versions of Patrocles to improve pDSP ranking.

The search for pDSPs has been restricted to 3′-UTRs despite accumulating evidence for functional miRNA target sites in coding segments as well [e.g. (76,78)]. However, as the specificity of target site predictions is considerably lower in open reading frames, including coding sequences (representing a much larger sequence space than 3′-UTRs) in our search would have greatly inflated the proportion of false positive pDSP predictions.

Bioinformatic evidence supporting polymorphic miRNA-mediated gene regulation should be testable experimentally. Most approaches described so far rely on the transfection of cultured cells with reporter vectors carrying alternate allelic forms of the predicted target sites in the 3′-UTR, as well as miRNA expression vectors. Whether differential regulation observed in such artificial conditions can be trusted as evidence for what happens *in vivo* is a matter of debate. To overcome these limitations, we have successfully developed an allelic imbalance test following coimminupreciptation of RNA-induced silencing complex-bound mRNAs from tissue samples of individuals heterozygous for the candidate pDSPs (H. Takeda *et al.*, unpublished data). Combined with hybridization on genome-wide, high-density SNP arrays, this and related approaches such as high-throughput sequencing of RNAs isolated by cross-linking immunoprecipitation [e.g. (79)] could be used to systematically scan the genome for polymorphic miRNA-mediated gene regulation. Recently, Kim and Bartel (40) tested the effect of 67 pDSPs altering target sites for miR-1/206, miR-133 and miR-122 in the 3′-UTR of coexpressed genes using allelic imbalance sequencing. They estimated that ~15% of their pDSPs were indeed functional, resulting in a ~2-fold difference in expression level between the mRNA allelomorphs. Thus, between any pair of mouse strains, >100 genes might be differentially regulated as a result of pDSPs. These experimental data emphasize the importance of polymorphic miRNA-mediated gene regulation and the utility of the Patrocles database.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Xavier Tordoir for his help with early versions of Patrocles.

FUNDING

European Union's Framework 6 (Callimir STREP, Epigenome NoE, Eadgene NoE); Belgian Science Policy organisation (SSTC Genefunc, BioMAGNet PAI); Belgian Fonds National de la Recherche Scientifique; Communauté Française de Belgique (Game, BIOMOD ARC); University of Liège. C.C. is Chercheur Qualifié of the Fonds National de la Recherche Scientifique.

Conflict of interest statement. None declared.

REFERENCES

1. Clop,A., Marcq,F., Takeda,H., Pirottin,D., Tordoir,X., Bibe,B., Bouix,J., Caiment,F., Elsen,J.M., Eychenne,F. *et al.* (2006) A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat. Genet.*, **38**, 813–818.
2. Sethupathy,P. and Collins,F.S. (2008) MicroRNA target site polymorphisms and human disease. *Trends Genet.*, **24**, 489–497.
3. Georges,M., Coppieters,W. and Charlier,C. (2007) Polymorphic miRNA-mediated gene regulation: contribution to phenotypic variation and disease. *Curr. Opin. Genet. Dev.*, **17**, 166–176.
4. Lewis,M.A., Quint,E., Glazier,A.M., Fuchs,H., De Angelis,M.H., Langford,C., van Dongen,S., Abreu-Goodger,C., Piipari,M., Redshaw,N. *et al.* (2009) An ENU-induced mutation of miR-96 associated with progressive hearing loss in mice. *Nat. Genet.*, **41**, 614–618.
5. Hill,D.A., Ivanovich,J., Priest,J.R., Gurnett,C.A., Dehner,L.P., Desruisseau,D., Jarzembowski,J.A., Wikenheiser-Brokamp,K.A., Suarez,B.K., Whelan,A.J. *et al.* (2009) DICER1 mutations in familial pleuropulmonary blastoma. *Science*, **325**, 965.
6. Hubbard,T.J., Aken,B.L., Ayling,S., Ballester,B., Beal,K., Bragin,E., Brent,S., Chen,Y., Clapham,P., Clarke,L. *et al.* (2009) Ensembl 2009. *Nucleic Acids Res.*, **37**, D690–D697.
7. Griffiths-Jones,S., Saini,H.K., van Dongen,S. and Enright,A.J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.
8. Giardine,B., Riemer,C., Hardison,R.C., Burhans,R., Elnitski,L., Shah,P., Zhang,Y., Blankenberg,D., Albert,I., Taylor,J. *et al.* (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.*, **15**, 1451–1455.
9. Kuhn,R.M., Karolchik,D., Zweig,A.S., Wang,T., Smith,K.E., Rosenbloom,K.R., Rhead,B., Raney,B.J., Pohl,A., Pheasant,M. *et al.* (2009) The UCSC Genome Browser Database: update 2009. *Nucleic Acids Res.*, **37**, D755–D761.
10. Zhang,J., Feuk,L., Duggan,G.E., Khaja,R. and Scherer,S.W. (2006) Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. *Cytogenet. Genome Res.*, **115**, 205–214.
11. She,X., Cheng,Z., Zollner,S., Church,D.M. and Eichler,E.E. (2008) Mouse segmental duplication and copy number variation. *Nat. Genet.*, **40**, 909–914.
12. Guryev,V., Saar,K., Adamovic,T., Verheul,M., van Heesch,S.A., Cook,S., Pravenec,M., Aitman,T., Jacob,H., Shull,J.D. *et al.* (2008) Distribution and functional impact of DNA copy number variation in the rat. *Nat. Genet.*, **40**, 538–545.
13. Dixon,A.L., Liang,L., Moffatt,M.F., Chen,W., Heath,S., Wong,K.C., Taylor,J., Burnett,E., Gut,I., Farrall,M. *et al.* (2007) A genome-wide association study of global gene expression. *Nat. Genet.*, **39**, 1202–1207.
14. Stranger,B.E., Nica,A.C., Forrest,M.S., Dimas,A., Bird,C.P., Beazley,C., Ingle,C.E., Dunning,M., Flicek,P., Koller,D. *et al.* (2007) Population genomics of human gene expression. *Nat. Genet.*, **39**, 1217–1224.
15. Cheung,V.G., Spielman,R.S., Ewens,K.G., Weber,T.M., Morley,M. and Burdick,J.T. (2005) Mapping determinants of human gene expression by regional and genome-wide association. *Nature*, **437**, 1365–1369.
16. Goring,H.H., Curran,J.E., Johnson,M.P., Dyer,T.D., Charlesworth,J., Cole,S.A., Jowett,J.B., Abraham,L.J., Rainwater,D.L., Comuzzie,A.G. *et al.* (2007) Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat. Genet.*, **39**, 1208–1216.
17. Morley,M., Molony,C.M., Weber,T.M., Devlin,J.L., Ewens,K.G., Spielman,R.S. and Cheung,V.G. (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature*, **430**, 743–747.
18. Spielman,R.S., Bastone,L.A., Burdick,J.T., Morley,M., Ewens,W.J. and Cheung,V.G. (2007) Common genetic variants account for differences in gene expression among ethnic groups. *Nat. Genet.*, **39**, 226–231.
19. Stranger,B.E., Forrest,M.S., Clark,A.G., Minichiello,M.J., Deutsch,S., Lyle,R., Hunt,S., Kahl,B., Antonarakis,S.E., Tavare,S. *et al.* (2005) Genome-wide associations of gene expression variation in humans. *PLoS Genet.*, **1**, e78.
20. Ge,B., Gurd,S., Gaudin,T., Dore,C., Lepage,P., Harmsen,E., Hudson,T.J. and Pastinen,T. (2005) Survey of allelic expression using EST mining. *Genome Res.*, **15**, 1584–1591.
21. Pant,P.V., Tao,H., Beilharz,E.J., Ballinger,D.G., Cox,D.R. and Frazer,K.A. (2006) Analysis of allelic differential expression in human white blood cells. *Genome Res.*, **16**, 331–339.
22. Hofacker,I., Fontana,W., Stadler,P., Bonhoeffer,L., Tacker,M. and Schuster,P. (1994) Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, **125**, 167–188.
23. Su,A.I., Wiltshire,T., Batalov,S., Lapp,H., Ching,K.A., Block,D., Zhang,J., Soden,R., Hayakawa,M., Kreiman,G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.
24. Landgraf,P., Rusu,M., Sheridan,R., Sewer,A., Iovino,N., Aravin,A., Pfeffer,S., Rice,A., Kamphorst,A.O., Landthaler,M. *et al.* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, **129**, 1401–1414.
25. Consortium,T.I.H. (2005) A haplotype map of the human genome. *Nature*, **437**, 1299–1320.
26. Siva,N. (2008) 1000 Genomes project. *Nat. Biotechnol.*, **26**, 256.
27. Xie,X., Lu,J., Kulbokas,E.J., Golub,T.R., Mootha,V., Lindblad-Toh,K., Lander,E.S. and Kellis,M. (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature*, **434**, 338–345.
28. Lewis,B.P., Burge,C.B. and Bartel,D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.

29. Siepel,A. and Haussler,D. (2004) Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol. Biol. Evol.*, **21**, 468–488.
30. Hwang,D.G. and Green,P. (2004) Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proc. Natl Acad. Sci. USA*, **101**, 13994–14001.
31. Arndt,P.F. and Hwa,T. (2005) Identification and measurement of neighbor-dependent nucleotide substitution processes. *Bioinformatics*, **21**, 2322–2328.
32. Chen,K. and Rajewsky,N. (2006) Natural selection on human microRNA binding sites inferred from SNP data. *Nat. Genet.*, **38**, 1452–1456.
33. Farh,K.K., Grimson,A., Jan,C., Lewis,B.P., Johnston,W.K., Lim,L.P., Burge,C.B. and Bartel,D.P. (2005) The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science*, **310**, 1817–1821.
34. Stark,A., Brennecke,J., Bushati,N., Russell,R.B. and Cohen,S.M. (2005) Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution. *Cell*, **123**, 1133–1146.
35. Tsang,J., Zhu,J. and van Oudenaarden,A. (2007) MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. *Mol. Cell*, **26**, 753–767.
36. Baskerville,S. and Bartel,D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, **11**, 241–247.
37. Kim,Y.K. and Kim,V.N. (2007) Processing of intronic microRNAs. *EMBO J.*, **26**, 775–783.
38. Gennarino,V.A., Sardiello,M., Avellino,R., Meola,N., Maselli,V., Anand,S., Cutillo,L., Ballabio,A. and Banfi,S. (2009) MicroRNA target prediction by expression analysis of host genes. *Genome Res.*, **19**, 481–490.
39. Ozsolak,F., Poling,L.L., Wang,Z., Liu,H., Liu,X.S., Roeder,R.G., Zhang,X., Song,J.S. and Fisher,D.E. (2008) Chromatin structure analyses identify miRNA promoters. *Genes Dev.*, **22**, 3172–3183.
40. Kim,J. and Bartel,D.P. (2009) Allelic imbalance sequencing reveals that single-nucleotide polymorphisms frequently alter microRNA-directed repression. *Nat. Biotechnol.*, **27**, 472–477.
41. Abelson,J.F., Kwan,K.Y., O'Roak,B.J., Baek,D.Y., Stillman,A.A., Morgan,T.M., Mathews,C.A., Pauls,D.L., Rasin,M.R., Gunel,M. *et al.* (2005) Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science*, **310**, 317–320.
42. Chou,I.C., Wan,L., Liu,S.C., Tsai,C.H. and Tsai,F.J. (2007) Association of the Slit and Trk-like 1 gene in Taiwanese patients with Tourette syndrome. *Pediatr. Neurol.*, **37**, 404–406.
43. Deng,H., Le,W.D., Xie,W.J. and Jankovic,J. (2006) Examination of the SLITRK1 gene in Caucasian patients with Tourette syndrome. *Acta Neurol. Scand.*, **114**, 400–402.
44. Fabbrini,G., Pasquini,M., Aurilia,C., Berardelli,I., Breedveld,G., Oostra,B.A., Bonifati,V. and Berardelli,A. (2007) A large Italian family with Gilles de la Tourette syndrome: clinical study and analysis of the SLITRK1 gene. *Mov. Disord.*, **22**, 2229–2234.
45. Keen-Kim,D., Mathews,C.A., Reus,V.I., Lowe,T.L., Herrera,L.D., Budman,C.L., Gross-Tsur,V., Pulver,A.E., Bruun,R.D., Erenberg,G. *et al.* (2006) Overrepresentation of rare variants in a specific ethnic group may confuse interpretation of association analyses. *Hum. Mol. Genet.*, **15**, 3324–3328.
46. Scharf,J.M., Moorjani,P., Fagerness,J., Platko,J.V., Illmann,C., Galloway,B., Jenike,E., Stewart,S.E. and Pauls,D.L. (2008) Lack of association between SLITRK1var321 and Tourette syndrome in a large family-based sample. *Neurology*, **70**, 1495–1496.
47. Beetz,C., Schule,R., Deconinck,T., Tran-Viet,K.N., Zhu,H., Kremer,B.P., Frints,S.G., van Zelst-Stams,W.A., Byrne,P., Otto,S. *et al.* (2008) REEP1 mutation spectrum and genotype/phenotype correlation in hereditary spastic paraplegia type 31. *Brain*, **131**, 1078–1086.
48. Zuchner,S., Wang,G., Tran-Viet,K.N., Nance,M.A., Gaskell,P.C., Vance,J.M., Ashley-Koch,A.E. and Pericak-Vance,M.A. (2006) Mutations in the novel mitochondrial protein REEP1 cause hereditary spastic paraplegia type 31. *Am. J. Hum. Genet.*, **79**, 365–369.
49. Adams,B.D., Furneaux,H. and White,B.A. (2007) The micro-ribonucleic acid (miRNA) miR-206 targets the human estrogen receptor-alpha (ERalpha) and represses ERalpha messenger RNA and protein expression in breast cancer cell lines. *Mol. Endocrinol.*, **21**, 1132–1147.
50. Sethupathy,P., Borel,C., Gagnebin,M., Grant,G.R., Deutsch,S., Elton,T.S., Hatzigeorgiou,A.G. and Antonarakis,S.E. (2007) Human microRNA-155 on chromosome 21 differentially interacts with its polymorphic target in the AGTR1 3′ untranslated region: a mechanism for functional single-nucleotide polymorphisms related to phenotypes. *Am. J. Hum. Genet.*, **81**, 405–413.
51. Mishra,P.J., Humeniuk,R., Longo-Sorbello,G.S., Banerjee,D. and Bertino,J.R. (2007) A miR-24 microRNA binding-site polymorphism in dihydrofolate reductase gene leads to methotrexate resistance. *Proc. Natl Acad. Sci. USA*, **104**, 13513–13518.
52. Tan,Z., Randall,G., Fan,J., Camoretti-Mercado,B., Brockman-Schneider,R., Pan,L., Solway,J., Gern,J.E., Lemanske,R.F., Nicolae,D. *et al.* (2007) Allele-specific targeting of microRNAs to HLA-G and risk of asthma. *Am. J. Hum. Genet.*, **81**, 829–834.
53. Jensen,K.P., Covault,J., Conner,T.S., Tennen,H., Kranzler,H.R. and Furneaux,H.M. (2009) A common polymorphism in serotonin receptor 1B mRNA moderates regulation by miR-96 and associates with aggressive human behaviors. *Mol. Psychiatry*, **14**, 381–389.
54. Landi,D., Gemignani,F., Naccarati,A., Pardini,B., Vodicka,P., Vodickova,L., Novotny,J., Forsti,A., Hemminki,K., Canzian,F. *et al.* (2008) Polymorphisms within micro-RNA-binding sites and risk of sporadic colorectal cancer. *Carcinogenesis*, **29**, 579–584.
55. Wang,G., van der Walt,J.M., Mayhew,G., Li,Y.J., Zuchner,S., Scott,W.K., Martin,E.R. and Vance,J.M. (2008) Variation in the miRNA-433 binding site of FGF20 confers risk for Parkinson disease by overexpression of alpha-synuclein. *Am. J. Hum. Genet.*, **82**, 283–289.
56. Kapeller,J., Houghton,L.A., Monnikes,H., Walstab,J., Moller,D., Bonisch,H., Burwinkel,B., Autschbach,F., Funke,B., Lasitschka,F. *et al.* (2008) First evidence for an association of a functional variant in the microRNA-510 target site of the serotonin receptor-type 3E gene with diarrhea predominant irritable bowel syndrome. *Hum. Mol. Genet.*, **17**, 2967–2977.
57. Chin,L.J., Ratner,E., Leng,S., Zhai,R., Nallur,S., Babar,I., Muller,R.U., Straka,E., Su,L., Burki,E.A. *et al.* (2008) A SNP in a let-7 microRNA complementary site in the KRAS 3′ untranslated region increases non-small cell lung cancer risk. *Cancer Res.*, **68**, 8535–8540.
58. Brendle,A., Lei,H., Brandt,A., Johansson,R., Enquist,K., Henriksson,R., Hemminki,K., Lenner,P. and Forsti,A. (2008) Polymorphisms in predicted microRNA-binding sites in integrin genes and breast cancer: ITGB4 as prognostic marker. *Carcinogenesis*, **29**, 1394–1399.
59. Lv,K., Guo,Y., Zhang,Y., Wang,K., Jia,Y. and Sun,S. (2008) Allele-specific targeting of hsa-miR-657 to human IGF2R creates a potential mechanism underlying the association of ACAA-insertion/deletion polymorphism with type 2 diabetes. *Biochem. Biophys. Res. Commun.*, **374**, 101–105.
60. Chen,K. and Rajewsky,N. (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.*, **8**, 93–103.
61. Iwai,N. and Naraba,H. (2005) Polymorphisms in human pre-miRNAs. *Biochem. Biophys. Res. Commun.*, **331**, 1439–1444.
62. Redon,R., Ishikawa,S., Fitch,K.R., Feuk,L., Perry,G.H., Andrews,T.D., Fiegler,H., Shapero,M.H., Carson,A.R., Chen,W. *et al.* (2006) Global variation in copy number in the human genome. *Nature*, **444**, 444–454.
63. Wong,K.K., deLeeuw,R.J., Dosanjh,N.S., Kimm,L.R., Cheng,Z., Horsman,D.E., MacAulay,C., Ng,R.T., Brown,C.J., Eichler,E.E. *et al.* (2007) A comprehensive analysis of common copy-number variations in the human genome. *Am. J. Hum. Genet.*, **80**, 91–104.
64. McCarroll,S.A., Kuruvilla,F.G., Korn,J.M., Cawley,S., Nemesh,J., Wysoker,A., Shapero,M.H., de Bakker,P.I., Maller,J.B., Kirby,A. *et al.* (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.*, **40**, 1166–1174.
65. Calin,G.A., Ferracin,M., Cimmino,A., Di Leva,G., Shimizu,M., Wojcik,S.E., Iorio,M.V., Visone,R., Sever,N.I., Fabbri,M. *et al.* (2005) A MicroRNA signature associated with prognosis and

progression in chronic lymphocytic leukemia. *N. Engl. J. Med.*, **353**, 1793–1801.

66. Duan,R., Pak,C. and Jin,P. (2007) Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. *Hum. Mol. Genet.*, **16**, 1124–1131.
67. Arisawa,T., Tahara,T., Shibata,T., Nagasaka,M., Nakamura,M., Kamiya,Y., Fujita,H., Hasegawa,S., Takagi,T., Wang,F.Y. *et al.* (2007) A polymorphism of microRNA 27a genome region is associated with the development of gastric mucosal atrophy in Japanese male subjects. *Dig. Dis. Sci.*, **52**, 1691–1697.
68. Calin,G.A. and Croce,C.M. (2007) Chromosomal rearrangements and microRNAs: a new cancer link with clinical implications. *J. Clin. Invest.*, **117**, 2059–2066.
69. Diederichs,S. and Haber,D.A. (2006) Sequence variations of microRNAs in human cancer: alterations in predicted secondary structure do not affect processing. *Cancer Res.*, **66**, 6097–6104.
70. Wu,M., Jolicoeur,N., Li,Z., Zhang,L., Fortin,Y., L'Abbe,D., Yu,Z. and Shen,S.H. (2008) Genetic variations of microRNAs in human cancer and their effects on the expression of miRNAs. *Carcinogenesis*, **29**, 1710–1716.
71. Yang,H., Dinney,C.P., Ye,Y., Zhu,Y., Grossman,H.B. and Wu,X. (2008) Evaluation of genetic variants in microRNA-related genes and risk of bladder cancer. *Cancer Res.*, **68**, 2530–2537.
72. Yang,J., Zhou,F., Xu,T., Deng,H., Ge,Y.Y., Zhang,C., Li,J. and Zhuang,S.M. (2008) Analysis of sequence variations in 59 microRNAs in hepatocellular carcinomas. *Mutat. Res.*, **638**, 205–209.
73. Yang,N., Coukos,G. and Zhang,L. (2008) MicroRNA epigenetic alterations in human cancer: one step forward in diagnosis and treatment. *Int. J. Cancer*, **122**, 963–968.
74. Zhang,L., Volinia,S., Bonome,T., Calin,G.A., Greshock,J., Yang,N., Liu,C.G., Giannakakis,A., Alexiou,P., Hasegawa,K. *et al.* (2008) Genomic and epigenetic alterations deregulate microRNA expression in human epithelial ovarian cancer. *Proc. Natl Acad. Sci. USA*, **105**, 7004–7009.
75. Hansen,T., Olsen,L., Lindow,M., Jakobsen,K.D., Ullum,H., Jonsson,E., Andreassen,O.A., Djurovic,S., Melle,I., Agartz,I. *et al.* (2007) Brain expressed microRNAs implicated in schizophrenia etiology. *PLoS ONE*, **2**, e873.
76. Baek,D., Villen,J., Shin,C., Camargo,F.D., Gygi,S.P. and Bartel,D.P. (2008) The impact of microRNAs on protein output. *Nature*, **455**, 64–71.
77. Grimson,A., Farh,K.K., Johnston,W.K., Garrett-Engele,P., Lim,L.P. and Bartel,D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, **27**, 91–105.
78. Selbach,M., Schwanhausser,B., Thierfelder,N., Fang,Z., Khanin,R. and Rajewsky,N. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**, 58–63.
79. Chi,S.W., Zang,J.B., Mele,A. and Darnell,R.B. (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, **460**, 479–486.