Supplementary material

	Dataset	Data set	Size [ORFs]	
mRNA expression	Holstege et al. (1998)	gene chip	5455	Reference mRNA expression (6071 ORFs)
	Roth et al. (1998)		6263	
	Jelinsky et al. (1998)		6090	
	Velculescu et al. (1997)	SAGE	3778	
Protein abundance	Gygi et al. (1999)	2-DE gel + mass spectrometry	156	Reference protein abundance (150 ORFs)
	Futcher et al. (1999)		71	

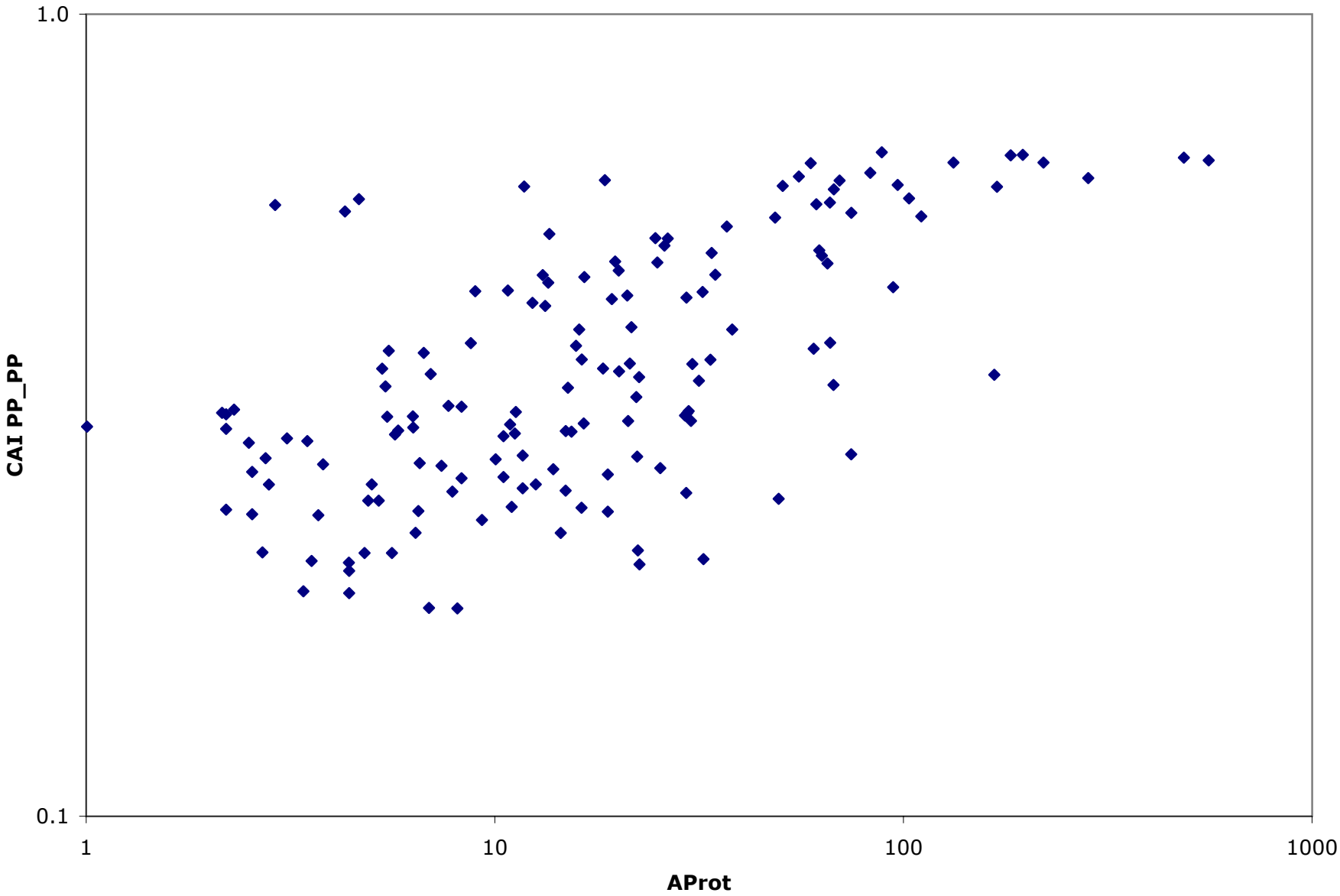
Supplementary table 1

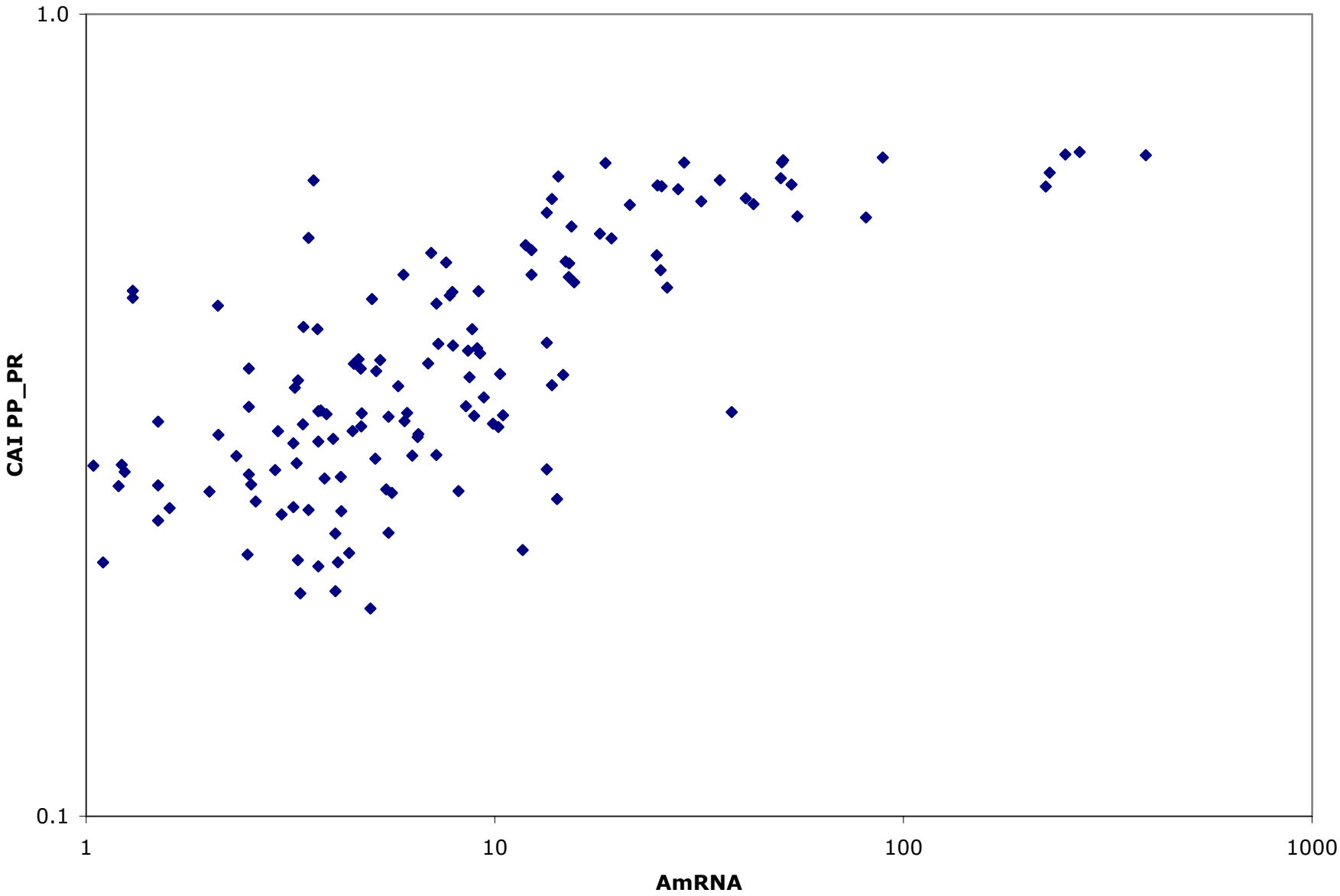
Shows an overview of the expression data sources. We looked at both mRNA and protein abundance data from the sources listed here. In order to reduce error rates and increase coverage of the genome, we combined and merged the four mRNA expression datasets and the two protein abundance datasets into two reference datasets each (the "reference expression" set for mRNA expression and the "reference abundance" set for protein abundance) (15). The reference abundance set is much smaller than the reference expression set (150 versus 6071 ORFs when combined with data about the codon composition) because there is much more data available for available for mRNA expression (from gene chips and SAGE data) than for protein abundance (from 2D-gel experiments) (9-14). In our calculations, we only used the 4270 ORFs in the reference expression set that have an expression level > 0.5 copies/cell since lower expression levels are too close to the detection limit of the gene chips and therefore too noisy.

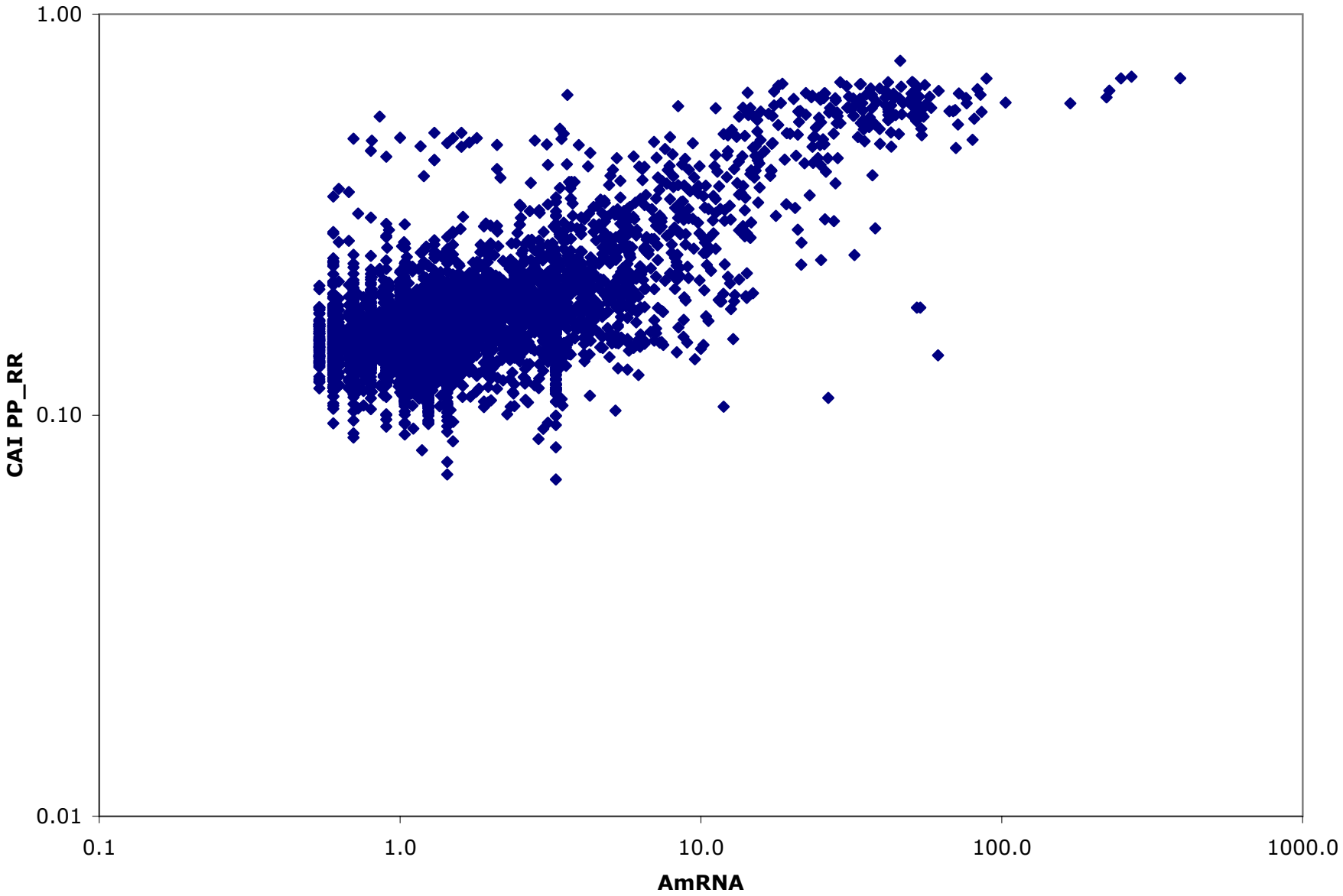
Scatter plots related to table 1

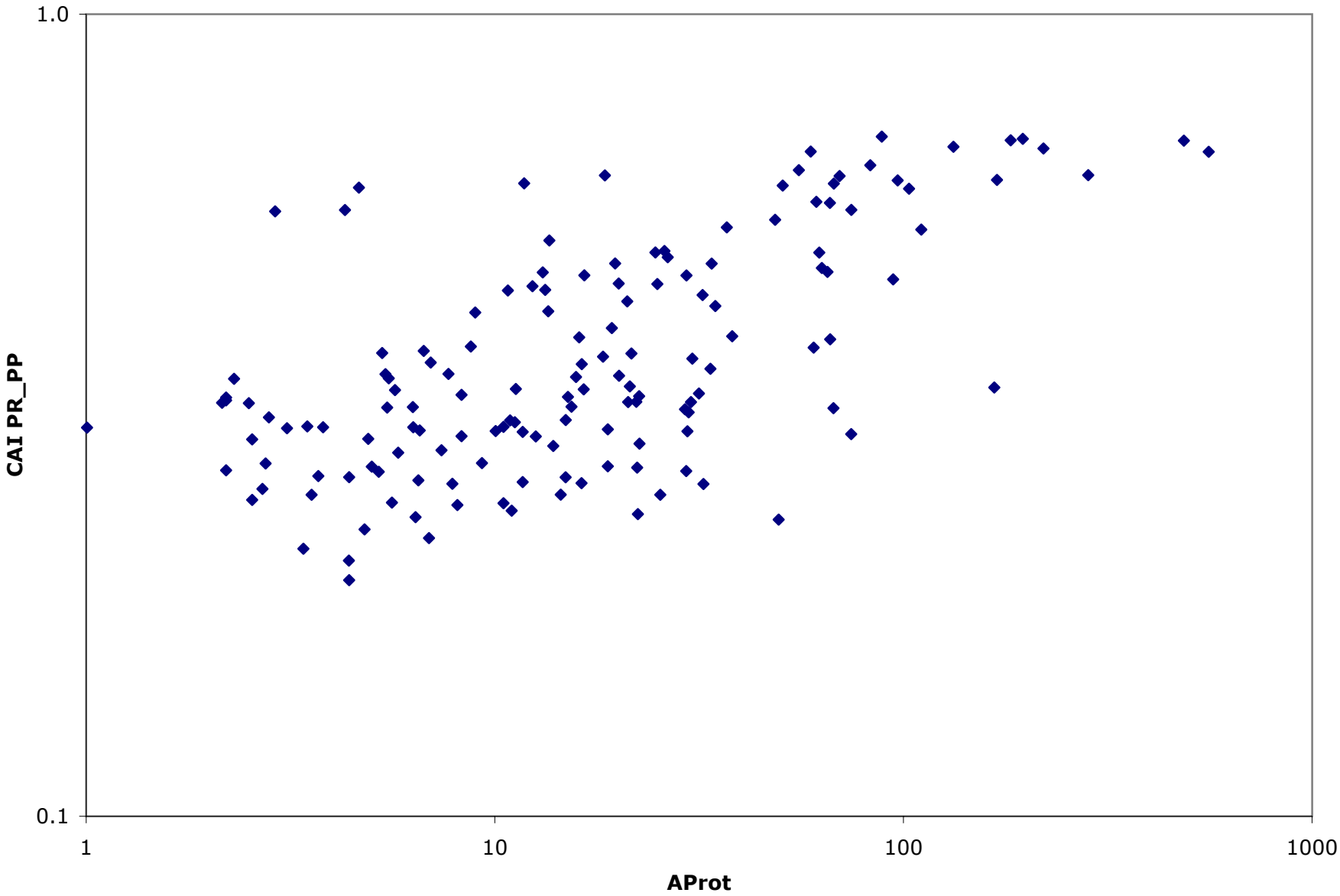
	Parameterization	Evaluation
PP_PP	[GProt, AProt]	[GProt, AProt]
PR_PP	[GProt, AmRNA]	[GProt, AProt]
RR_PP	[GmRNA, AmRNA}	[GProt, AProt]
PP_PR	[GProt, AProt]	[GProt, AmRNA]
PR_PR	[GProt, AmRNA]	[GProt, AmRNA]
RR_PR	[GmRNA, AmRNA}	[GProt, AmRNA]
PP_RR	[GProt, AProt]	[GmRNA, AmRNA}
PR_RR	[GProt, AmRNA]	[GmRNA, AmRNA}
RR_RR	[GmRNA, AmRNA}	[GmRNA, AmRNA}

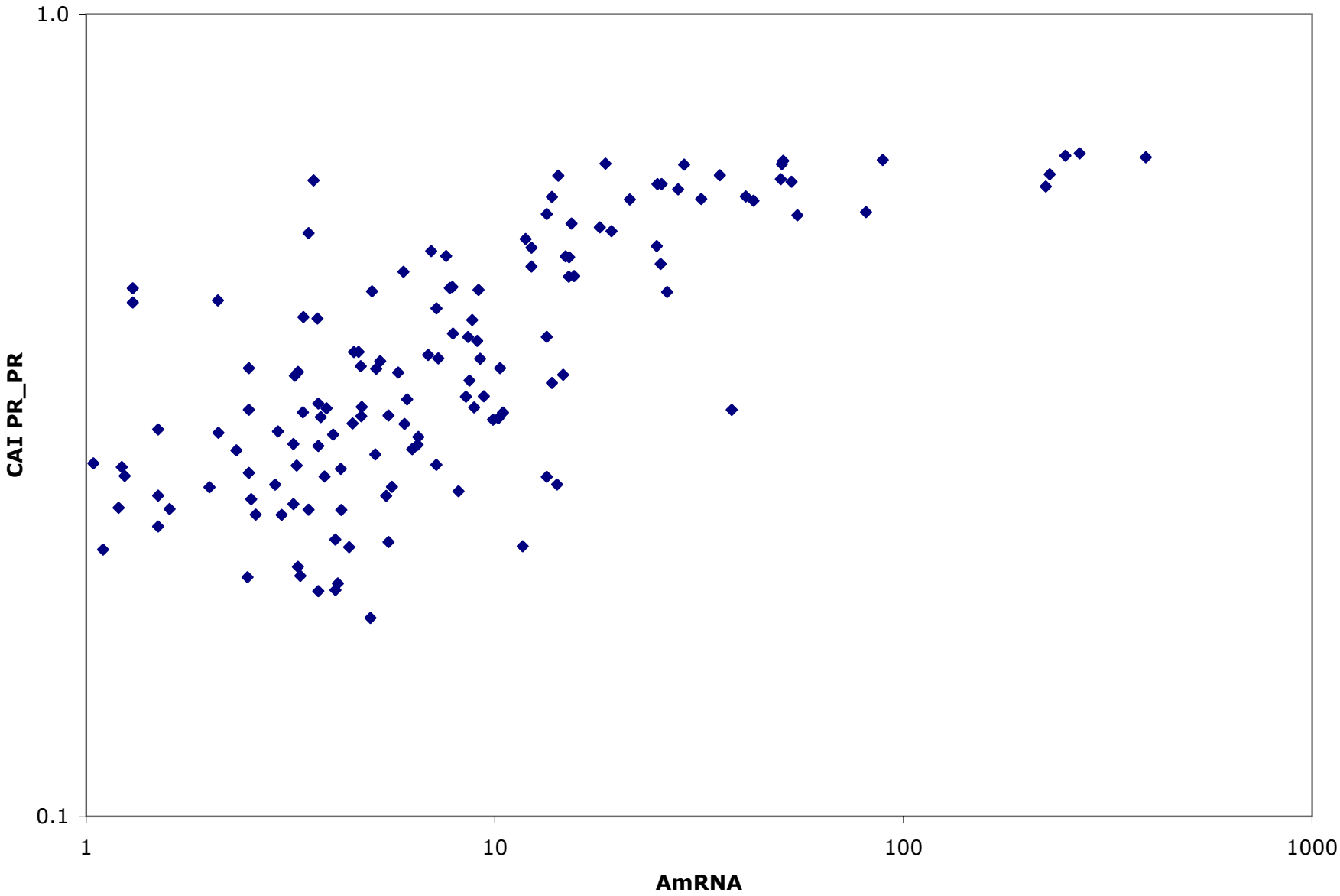
Legend

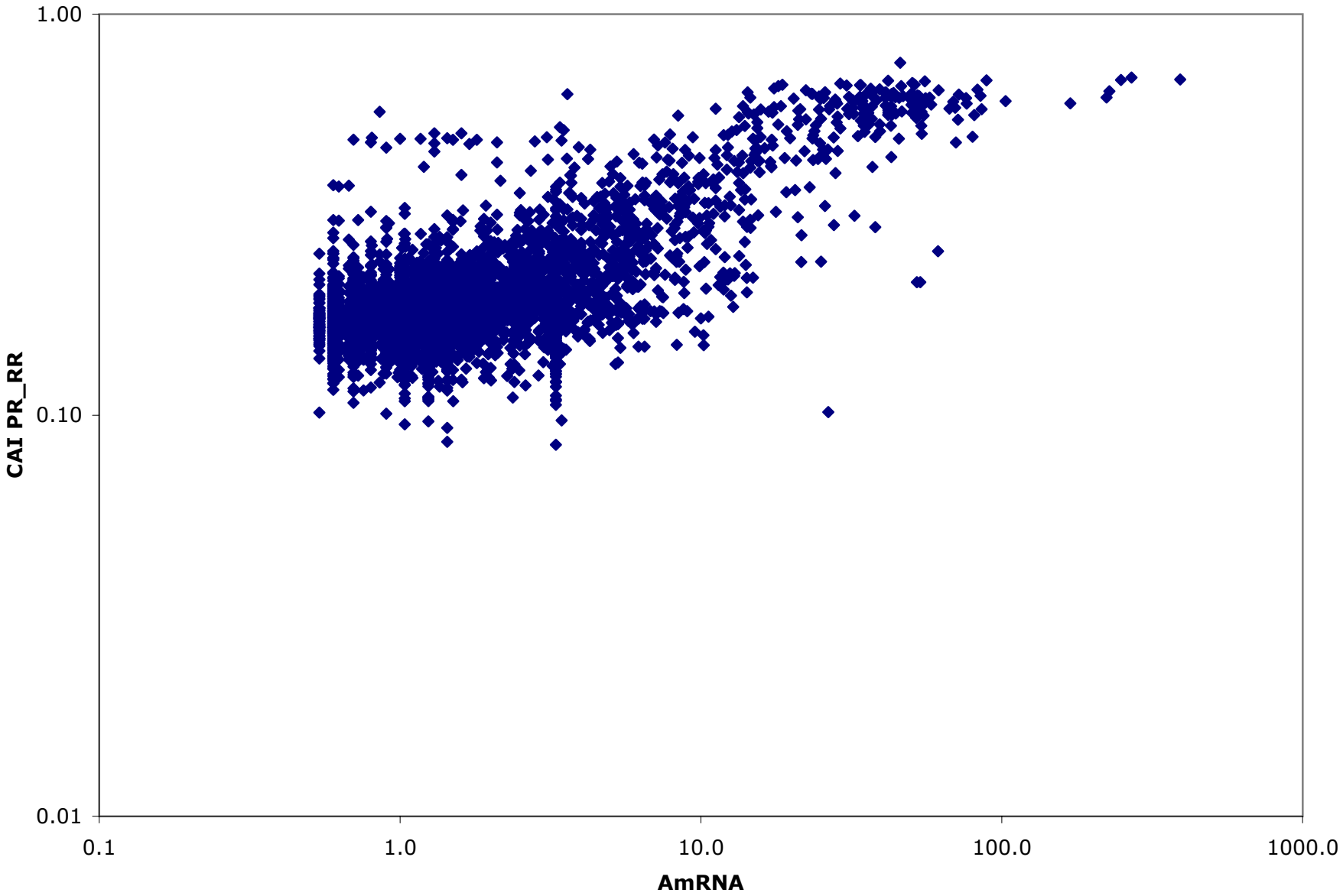

1.0
0.1
CAI PP_PP
1
10
100
1000
AProt

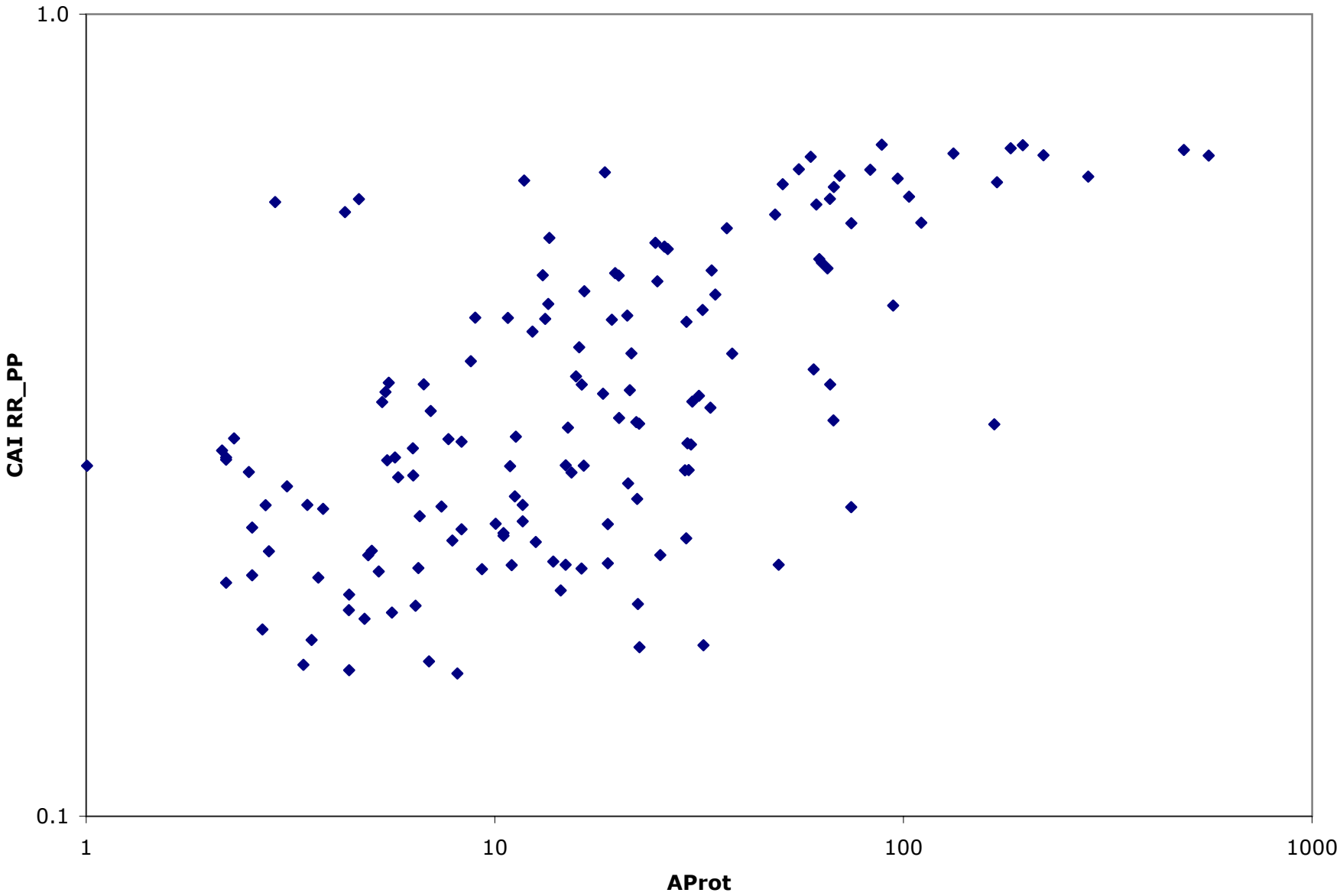

1.0
0.1
CAI PP_PR
1
10
100
1000
AmRNA

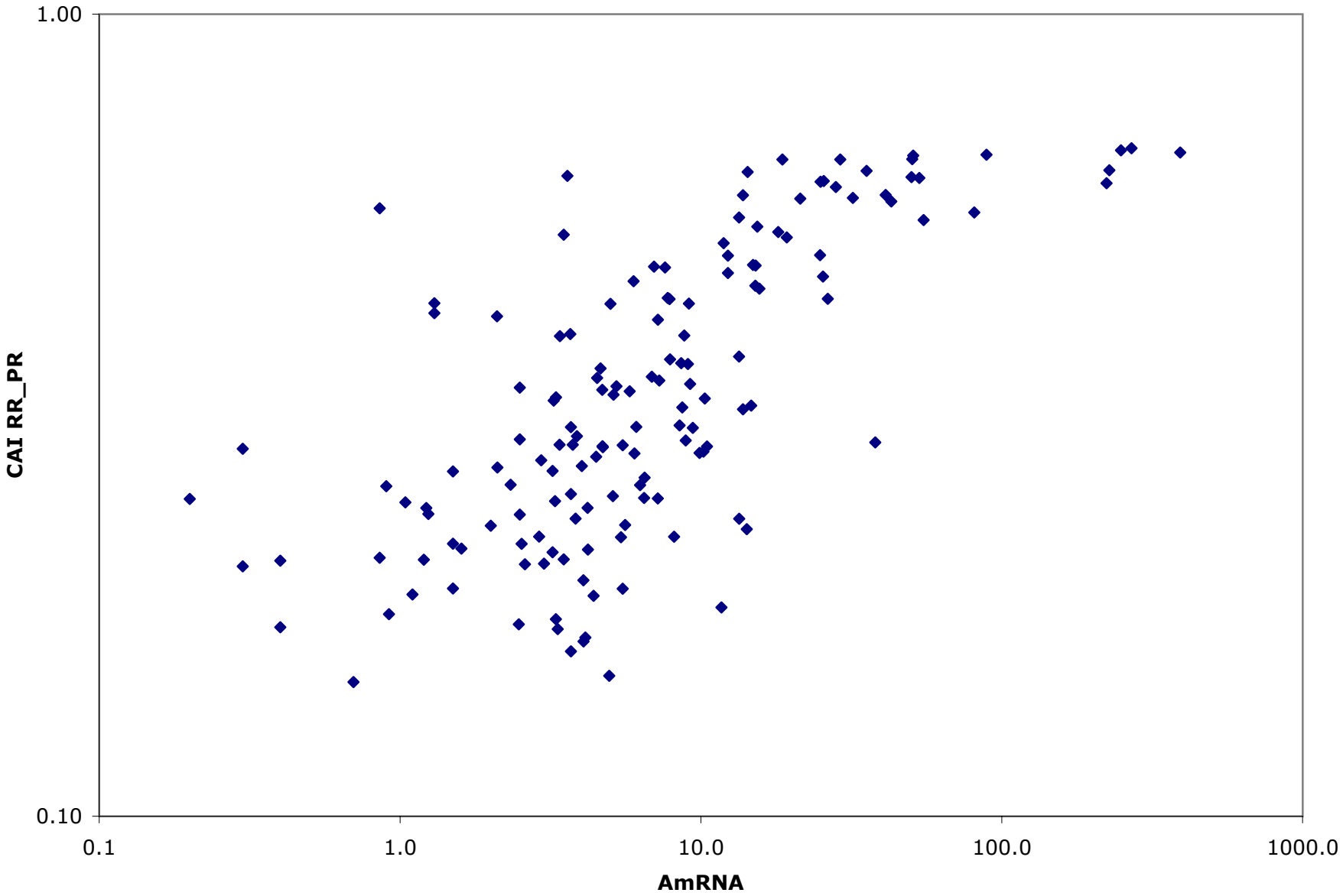

1.00
0.10
0.01
CAI PP_RR
0.1
1.0
10.0
100.0
1000.0
AmRNA

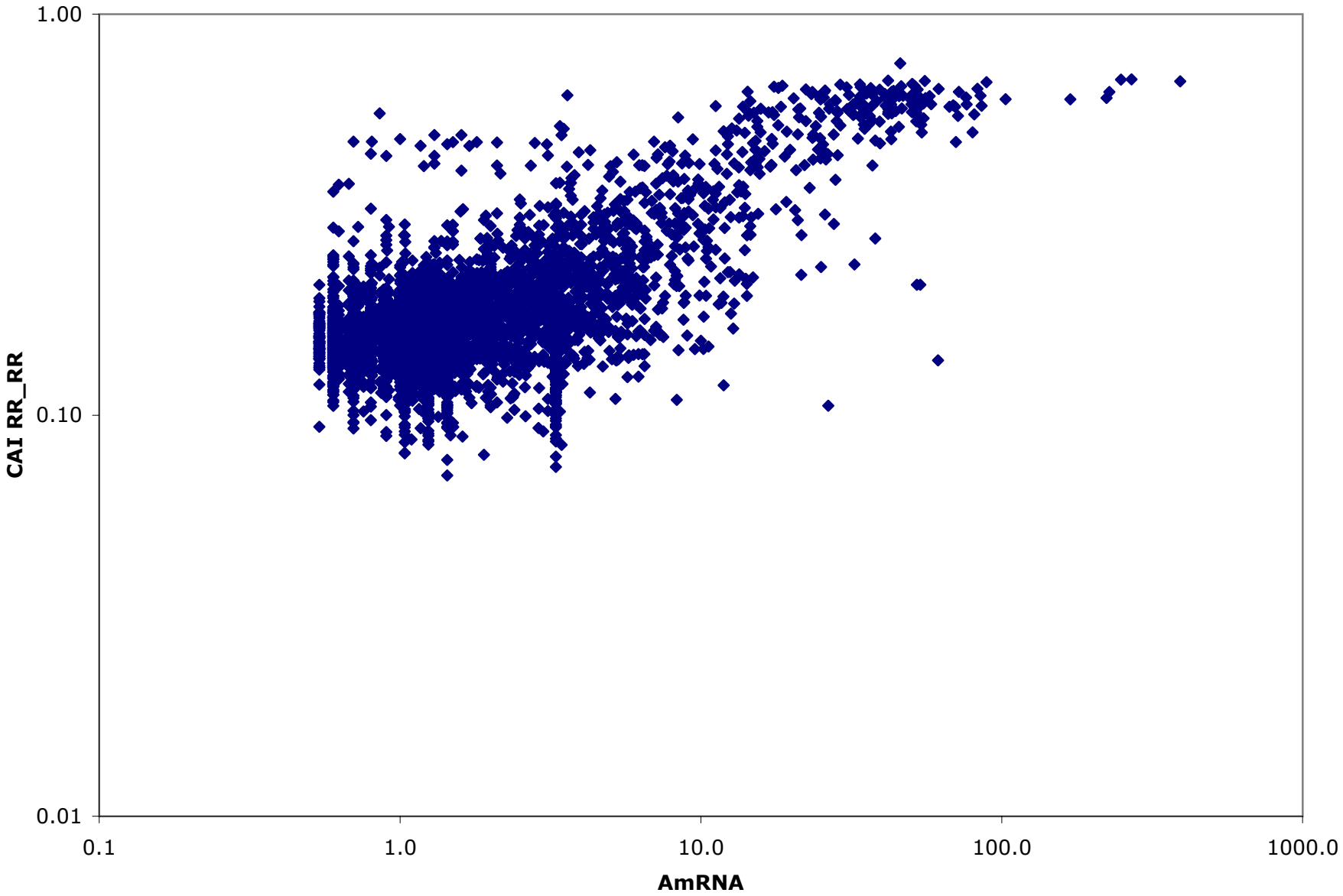

1.0
0.1
CAI PR_PP
1
10
100
1000
AProt

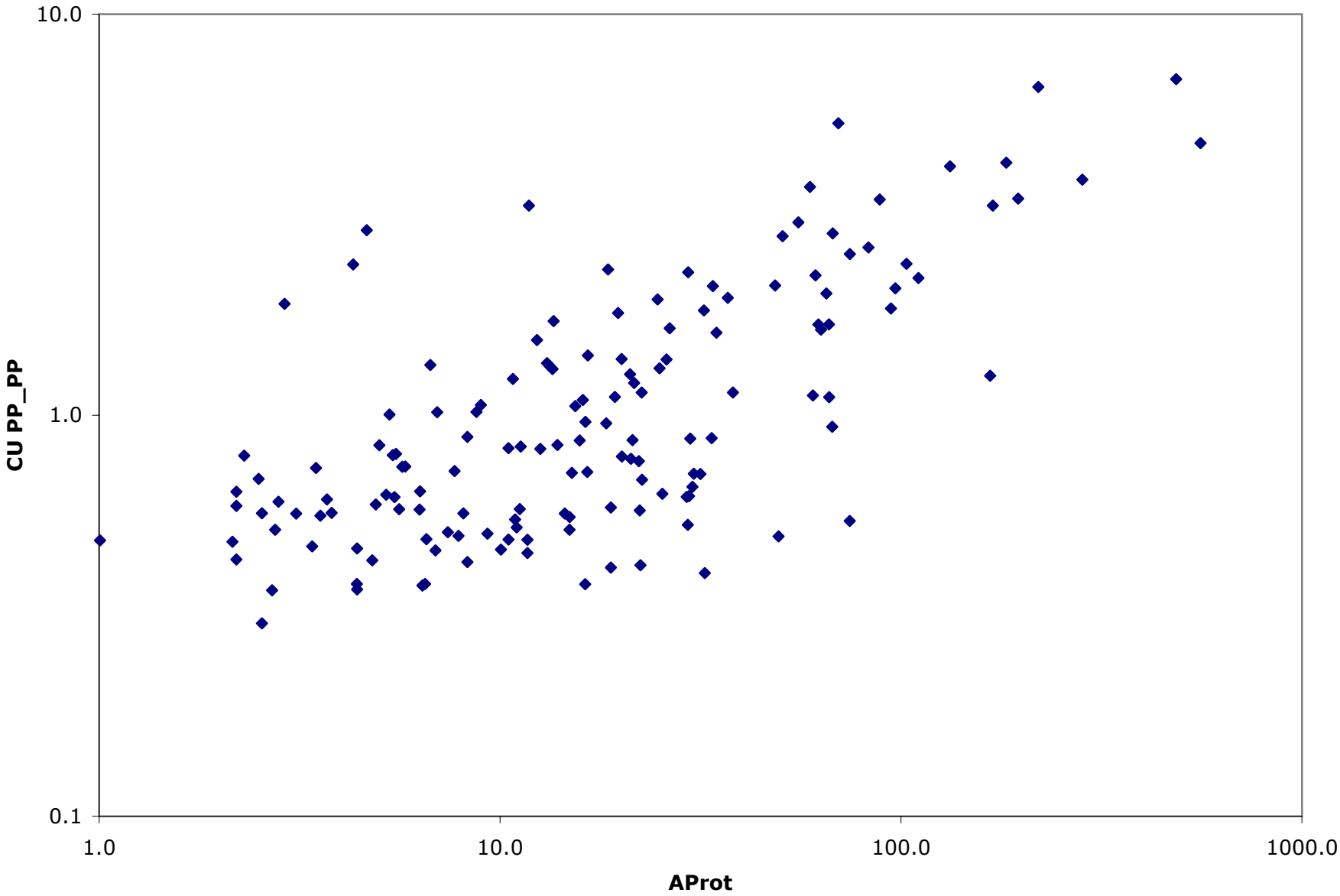

1.0
0.1
CAI PR_PR
1
10
100
1000
AmRNA

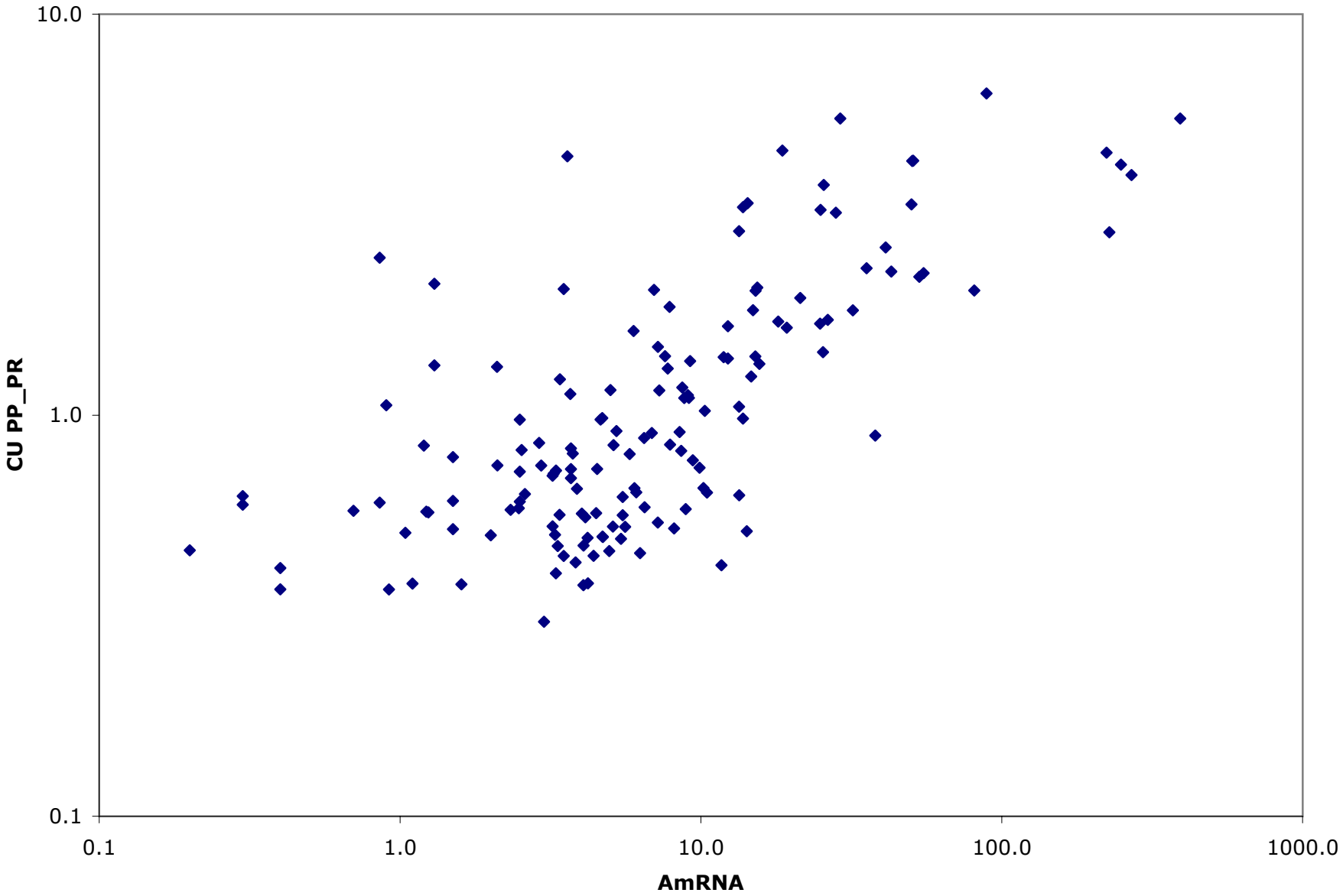

1.00
0.10
0.01
CAI PR_RR
0.1
1.0
10.0
100.0
1000.0
AmRNA

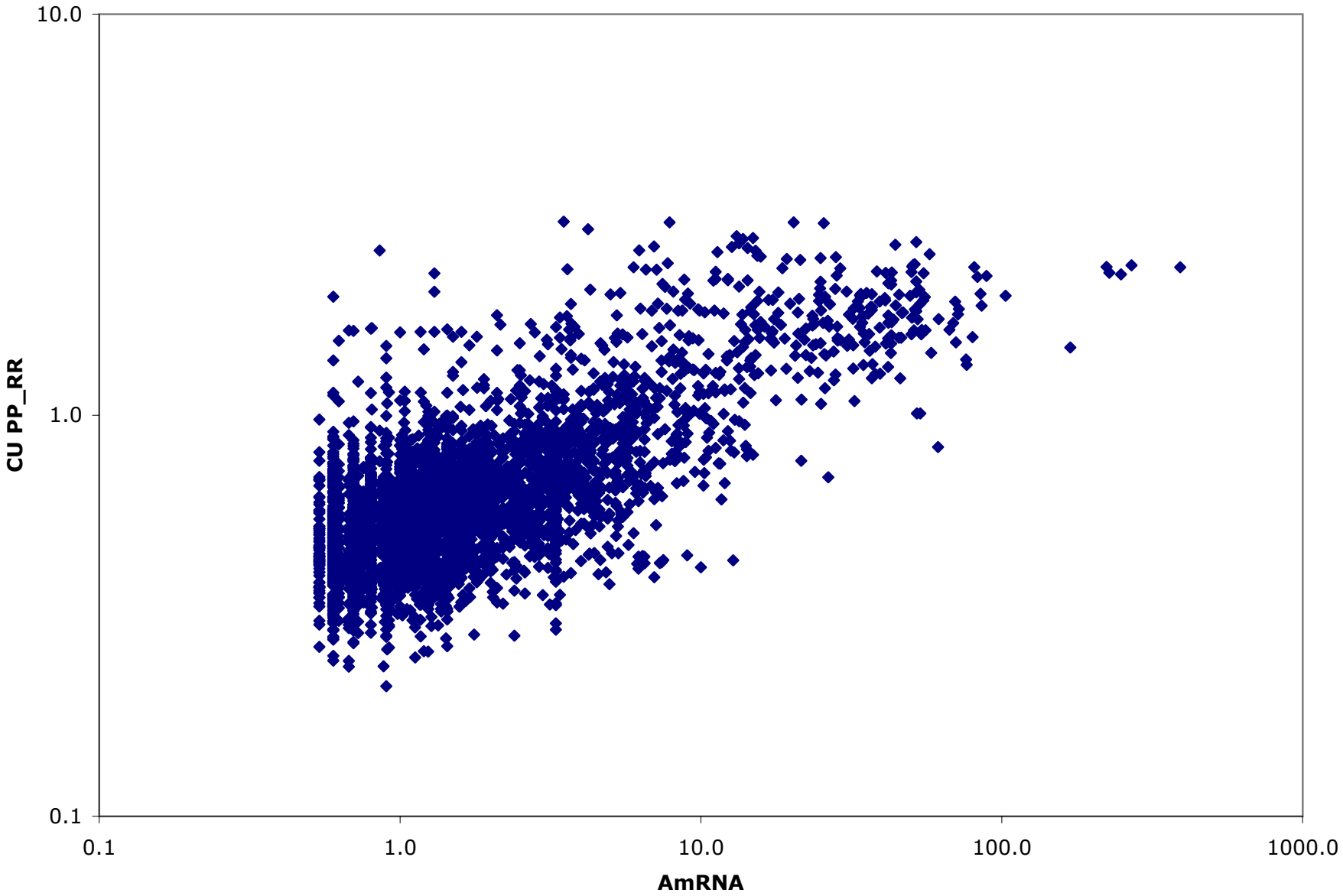

1.0
0.1
CAI RR_PP
1
10
100
1000
AProt

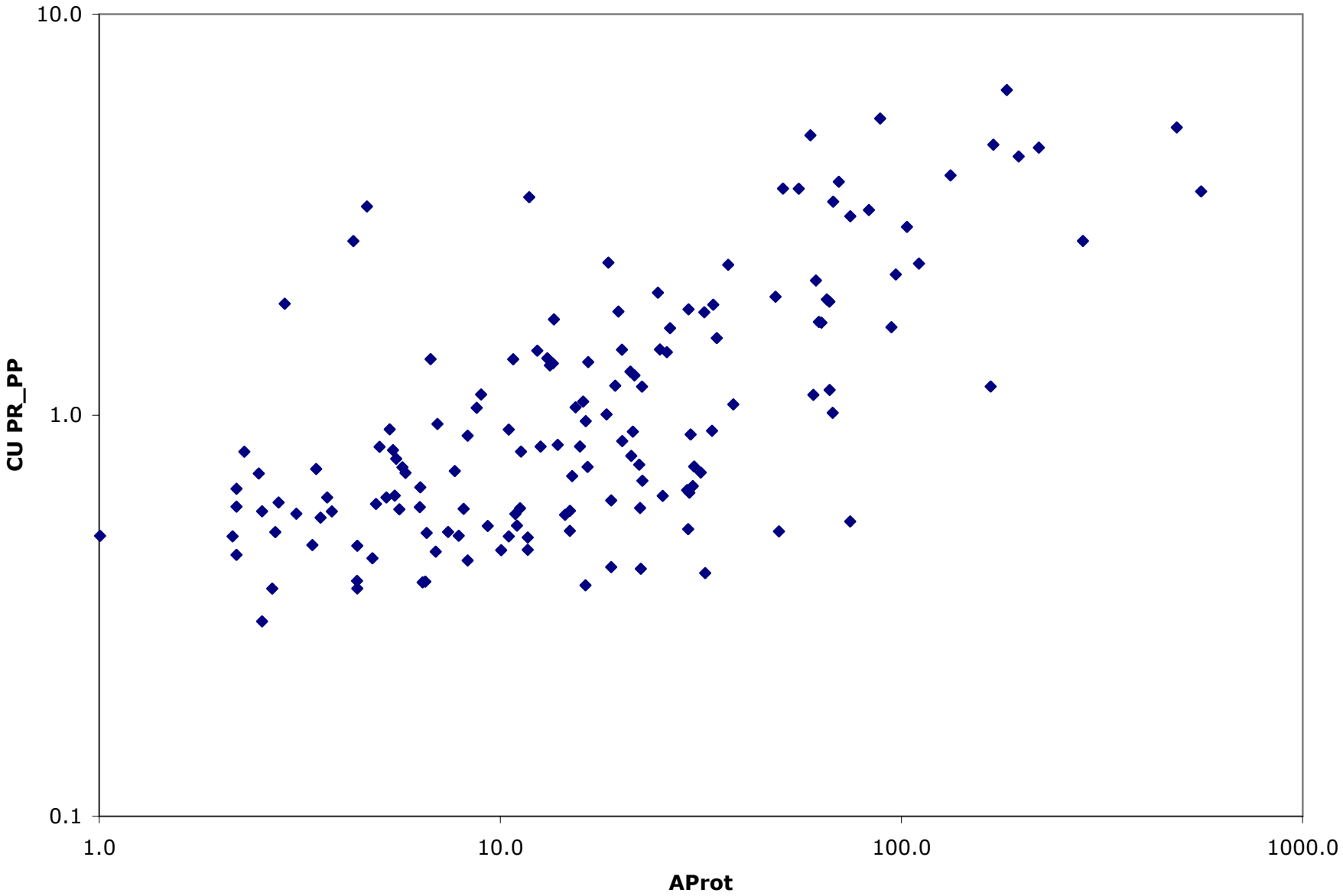

1.00
0.10
CAI RR_PR
0.1
1.0
10.0
100.0
1000.0
AmRNA

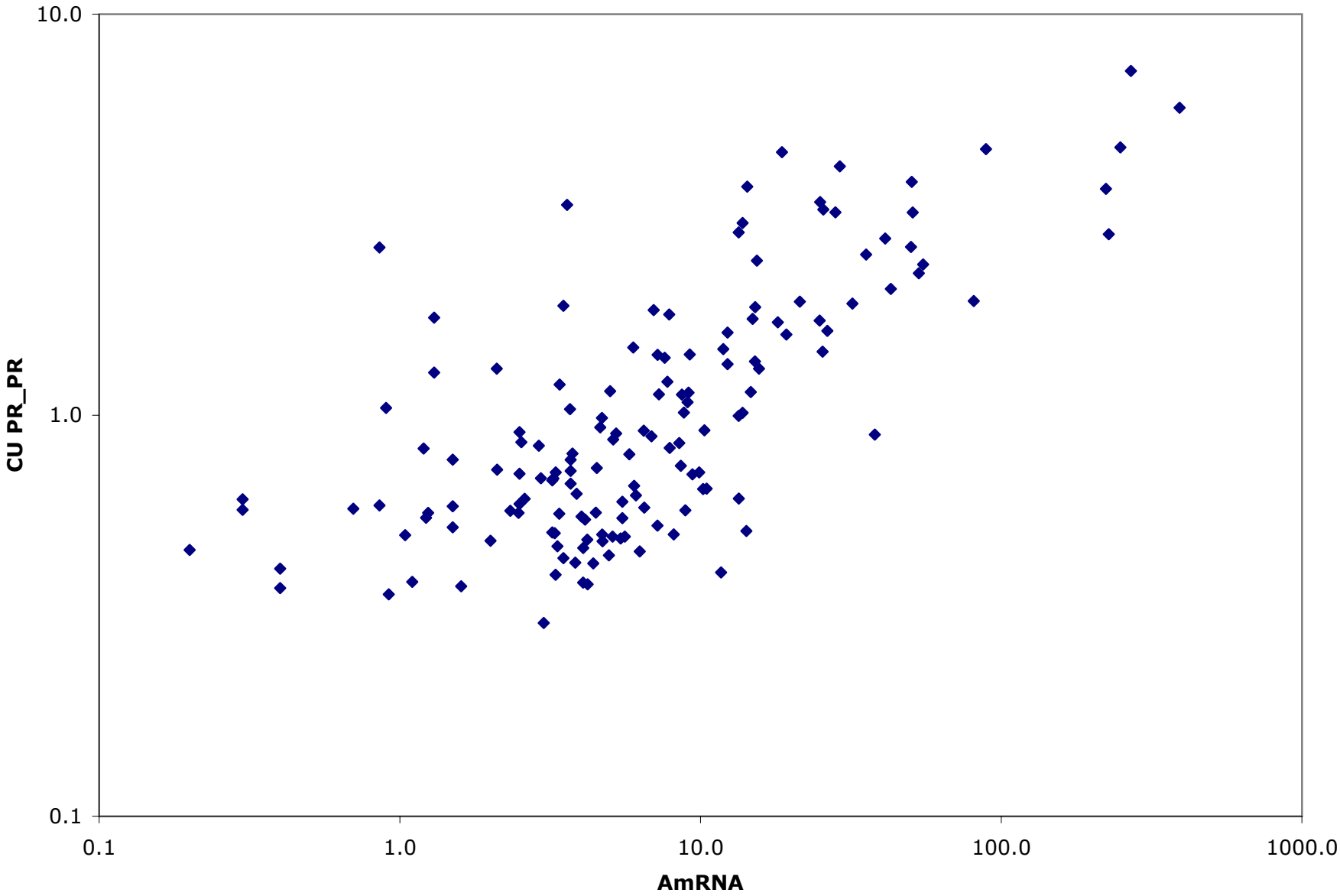

1.00
0.10
0.01
CAI RR_RR
0.1
1.0
10.0
100.0
1000.0
AmRNA

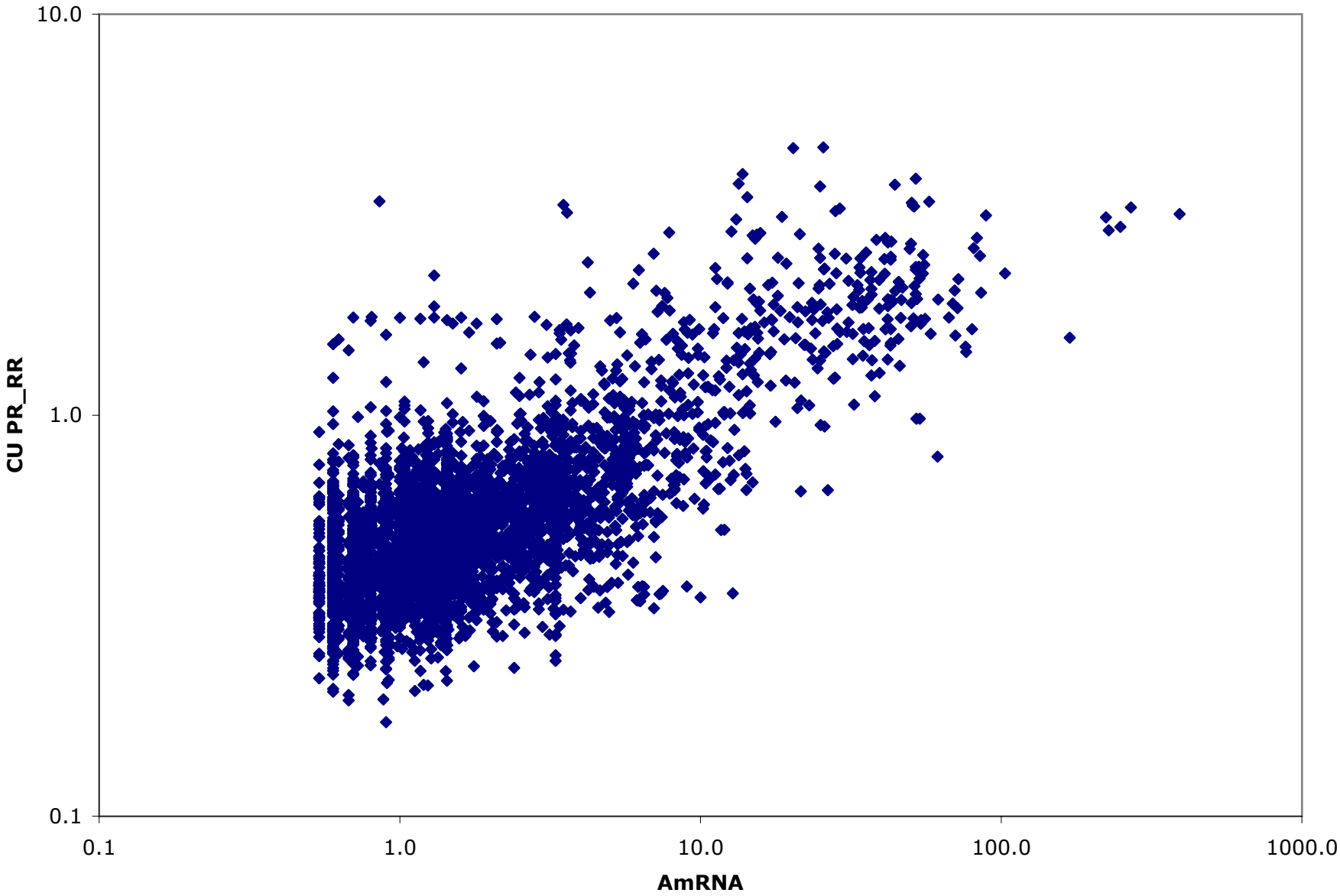

10.0
1.0
0.1
CU PP_PP
1.0
10.0
100.0
1000.0
AProt

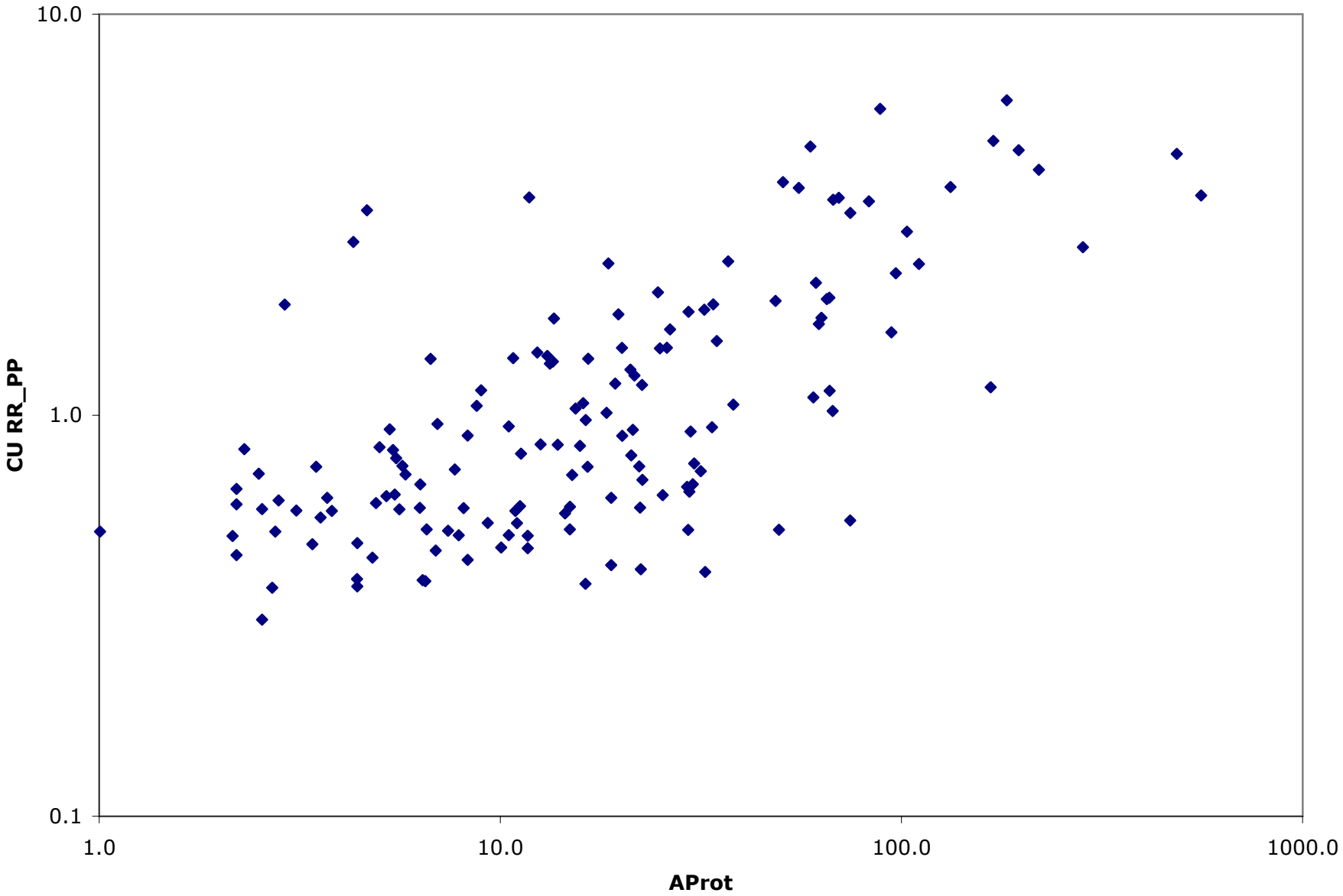

10.0
1.0
0.1
CU PP_PR
0.1
1.0
10.0
100.0
1000.0
AmRNA

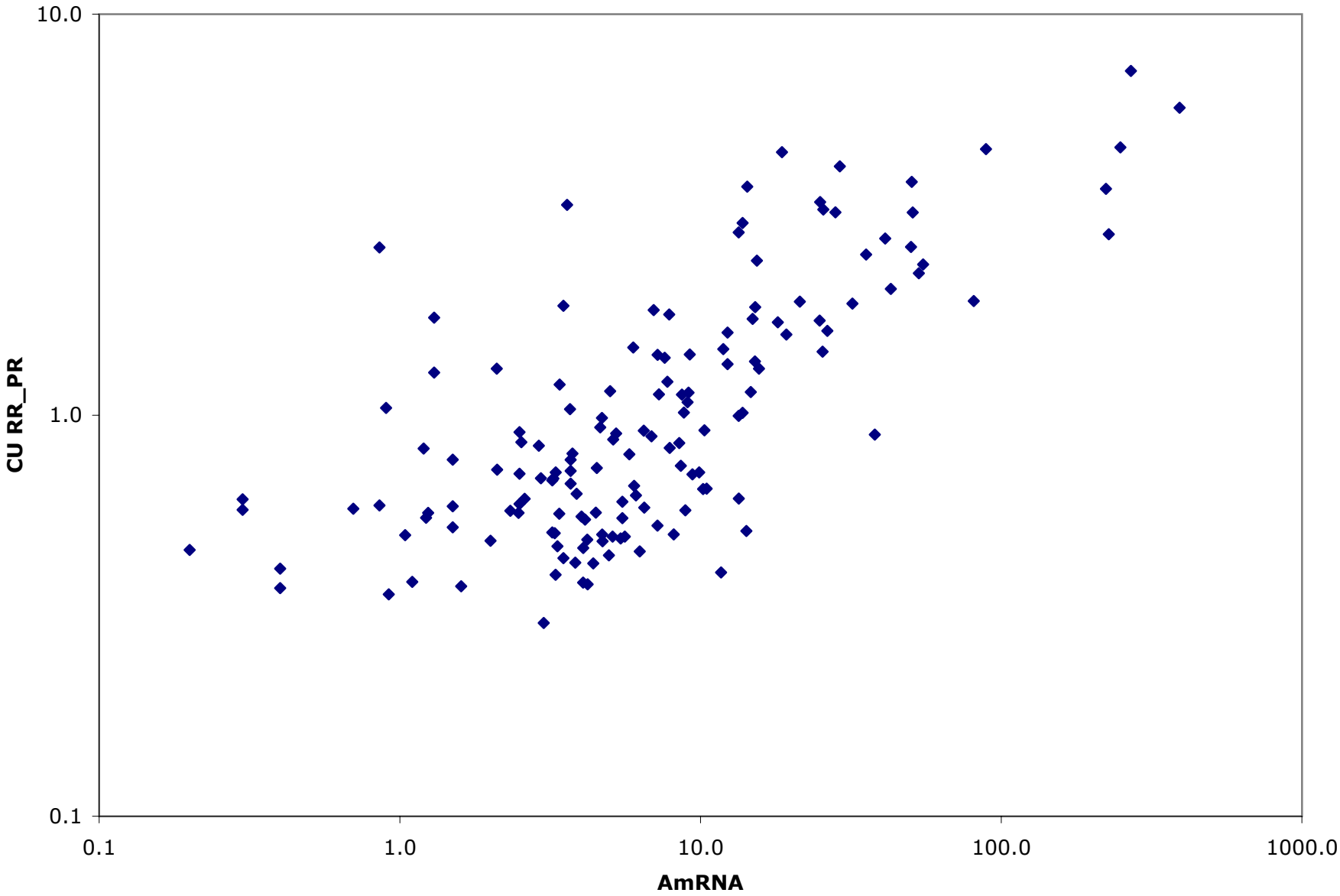

10.0
1.0
0.1
CU PP_RR
0.1
1.0
10.0
100.0
1000.0
AmRNA

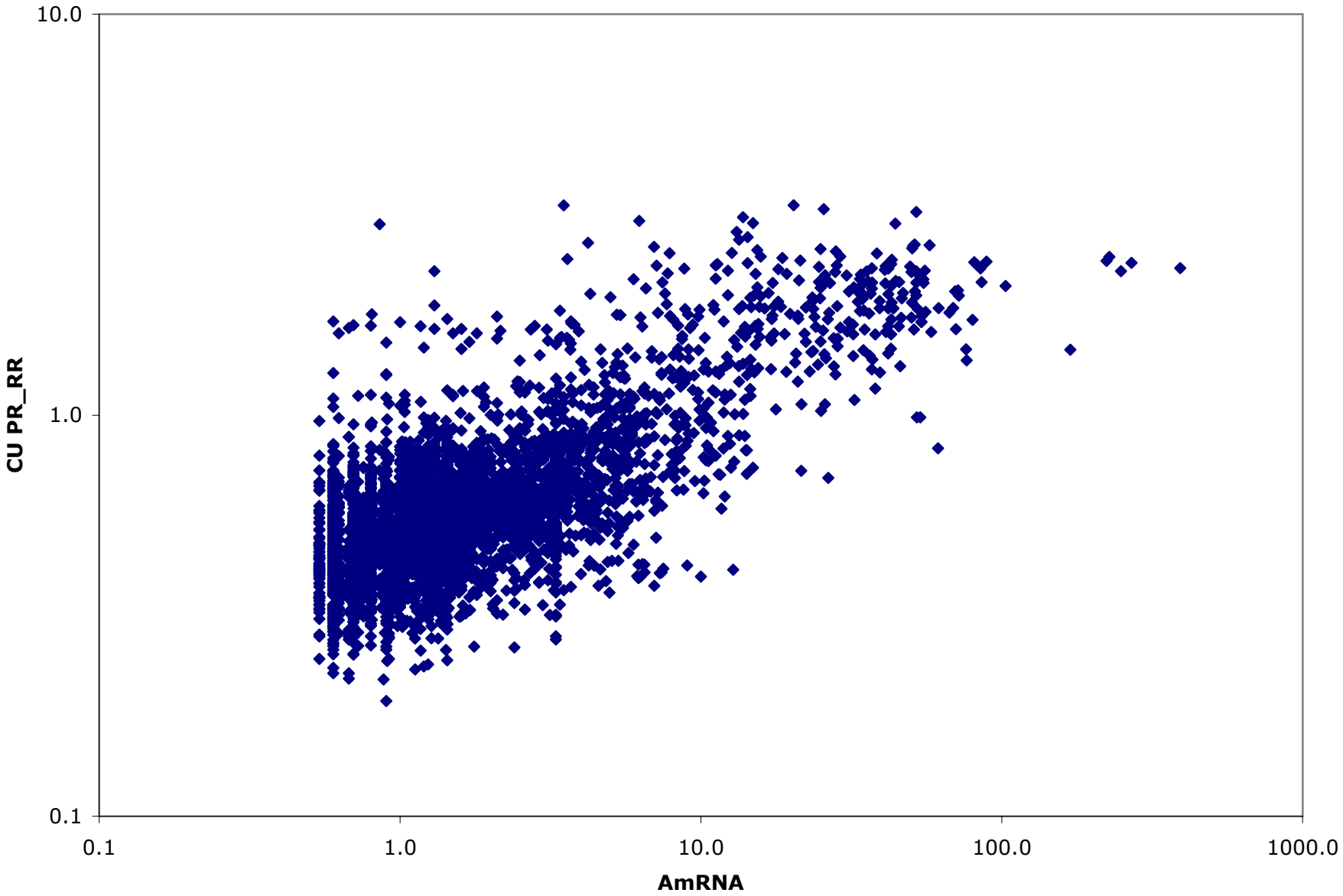

10.0
1.0
0.1
CU PR_PP
1.0
10.0
100.0
1000.0
AProt

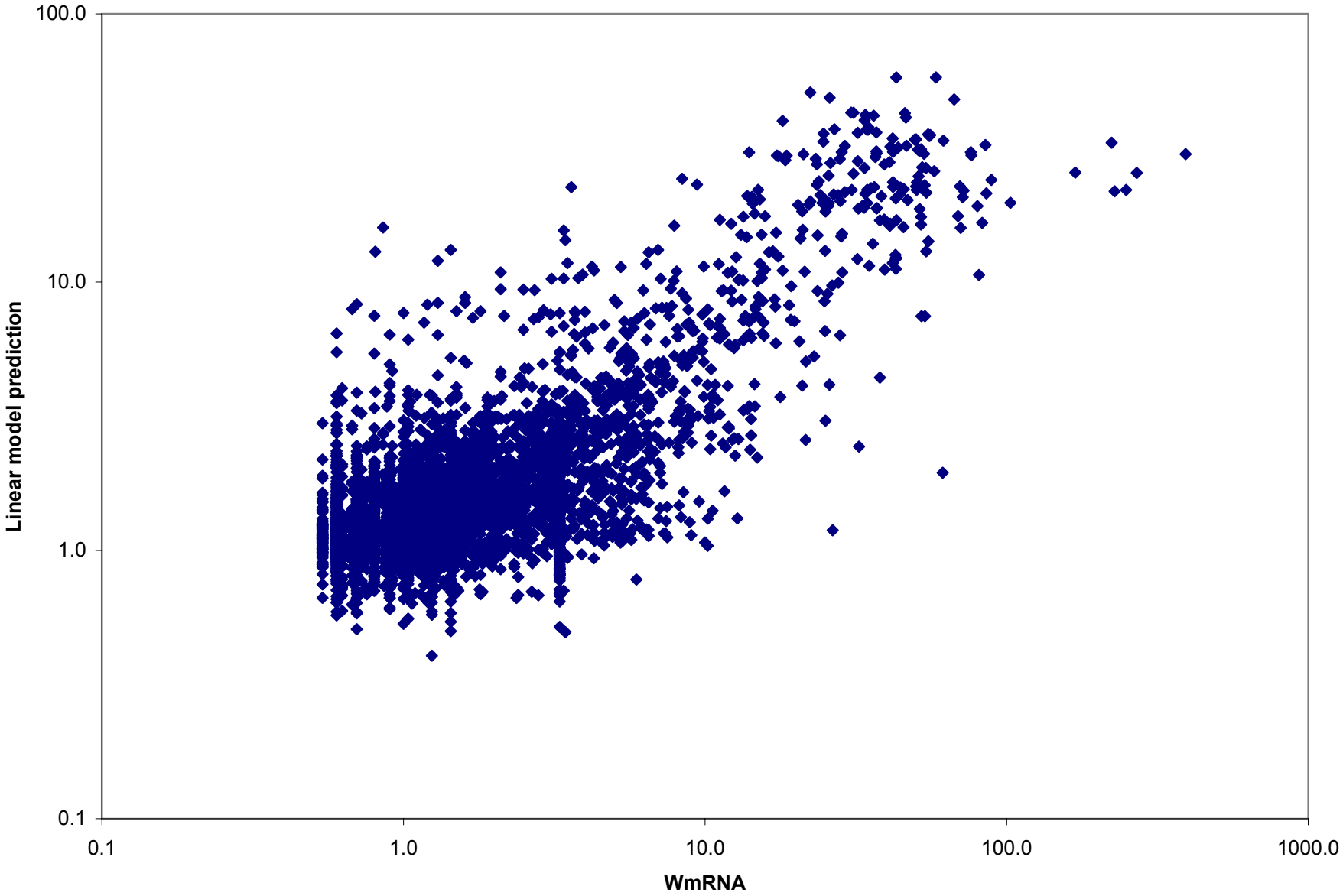

10.0
1.0
0.1
CU PR_PR
0.1
1.0
10.0
100.0
1000.0
AmRNA


10.0
1.0
0.1
CU PR_RR
0.1
1.0
10.0
100.0
1000.0
AmRNA


10.0
1.0
0.1
CU RR_PP
1.0
10.0
100.0
1000.0
AProt


10.0
1.0
0.1
CU RR_PR
0.1
1.0
10.0
100.0
1000.0
AmRNA


10.0
1.0
0.1
CU PR_RR
0.1
1.0
10.0
100.0
1000.0
AmRNA


100.0
10.0
1.0
0.1
Linear model prediction
0.1
1.0
10.0
100.0
1000.0
WmRNA

Linear model parameters

k	**Codon**	*vk*	95% confidence intervall	
0	Intercept	-0.81	-0.97	-0.64
1	GGT	9.00	7.66	10.33
2	AAG	7.72	6.58	8.87
3	GCT	8.90	7.47	10.34
4	AAT	-7.26	-8.74	-5.79
5	GCA	-6.58	-8.81	-4.34
6	TTG	5.86	4.31	7.41
7	CCA	7.49	5.62	9.36
8	CGT	9.67	6.61	12.73
9	AGG	-7.51	-10.56	-4.46
10	AAC	5.46	3.86	7.06
11	TGC	8.47	5.59	11.35
12	CCC	-8.79	-12.24	-5.34
13	GGA	-4.61	-7.20	-2.01
14	AGA	4.74	3.19	6.29
15	GTC	7.50	5.08	9.91
16	TTC	4.65	2.92	6.39
17	CAA	3.60	2.22	4.97
18	GTT	4.90	3.07	6.73
19	TGT	5.17	2.81	7.53
20	TCT	3.21	1.72	4.69

	t value	two-sided P-value	Bonferroni-corrected P-value
GGT	5.76	9.0.E-09	3.8.E-07
AAG	-3.48	5.0.E-04	2.1.E-02
GCT	7.40	1.6.E-13	6.8.E-12
AAT	7.84	5.5.E-15	2.3.E-13
GCA	-5.00	6.1.E-07	2.6.E-05
TTG	5.25	1.6.E-07	6.6.E-06
CCA	13.20	4.9.E-39	2.0.E-37
CGT	5.26	1.5.E-07	6.4.E-06
AGG	6.68	2.8.E-11	1.2.E-09
AAC	6.00	2.2.E-09	9.2.E-08
TGC	-4.82	1.5.E-06	6.2.E-05
CCC	-5.77	8.3.E-09	3.5.E-07
GGA	13.25	2.8.E-39	1.2.E-37
AGA	6.09	1.2.E-09	5.2.E-08
GTC	4.30	1.7.E-05	7.3.E-04
TTC	-9.67	6.7.E-22	2.8.E-20
CAA	5.14	2.9.E-07	1.2.E-05
GTT	12.17	1.7.E-33	7.2.E-32
TGT	6.19	6.4.E-10	2.7.E-08
TCT	4.23	2.4.E-05	9.9.E-04

Bonferroni factor
42