Results are expressed as mean ± SE. Student *t* test or Welch test was used to compare data between 2 groups. P-values less than 0.05 were considered as statistically significant. Individual experiments were performed in triplicate, and each experiment was independently performed three times.

SUPPLEMENTARY TABLE AND FIGURE LEGENDS

Supplementary Table 1. The list represents Ct values of 250 kinds of miRNAs and *RNU6-2* in day 0, day 1 and day 7. This file can be viewed with: Microsoft Excel.

Supplementary Table 2. The list represents 154 genes increased in a similar fashion among "NC" at day 7, "four miRNA" at the day 5 and "miR-338-3p and miR-451" at the day 5 in comparison to "NC" at day 5. This file can be viewed with: Microsoft Excel.

Supplementary Figure 1. Increased expression levels of miR-210, miR-338-3p, miR-33a and miR-451 along with the epithelial cell differentiation of Caco-2 cells. The expression levels of miR-210, miR-338-3p, miR-33a, miR-451 and *RNU6-2* in Caco-2 cells cultured in transwell chambers for the indicated periods were determined by the qRT-PCR. The values are shown as the fold of values obtained from the sample at day 1 (Welch test; *, $p < 0.05$ for cells plated in transwell chamber vs cells at day 1), and are represented as mean ± SE (n=3).

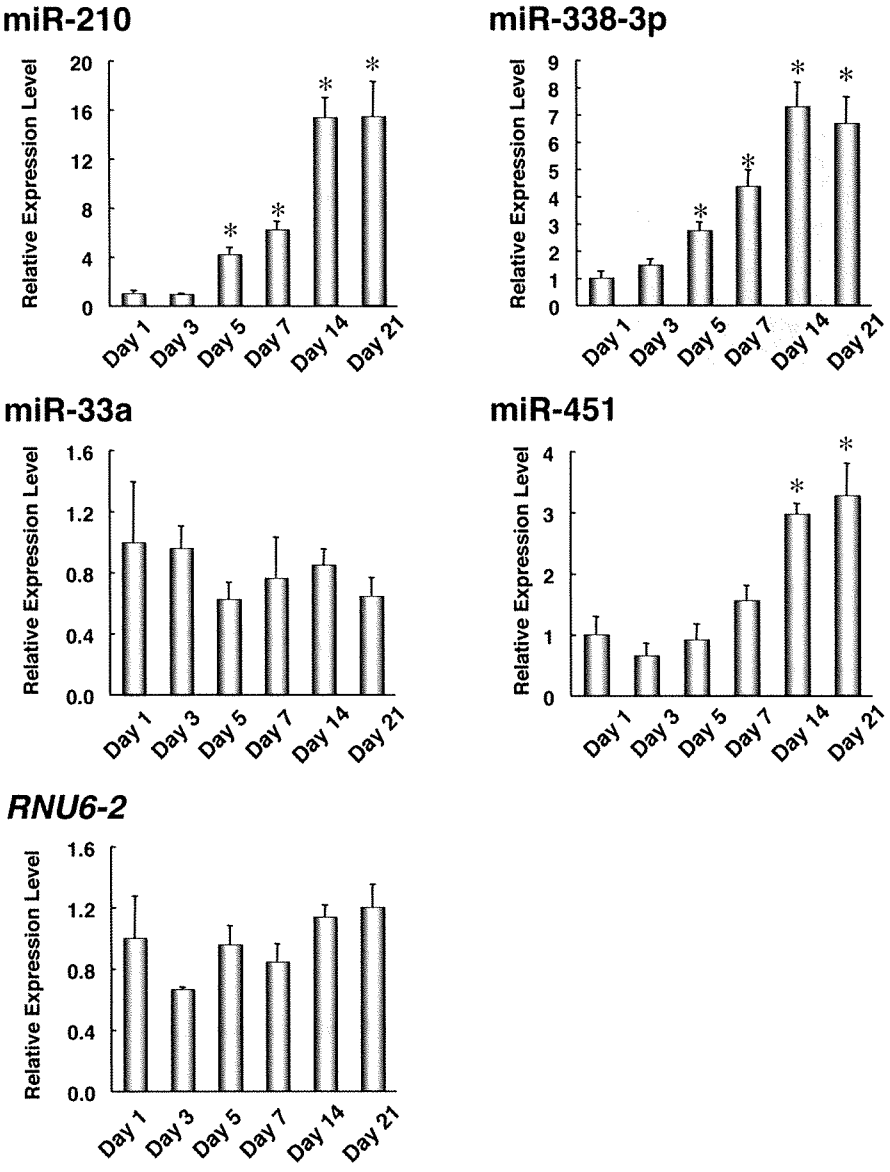
Supplementary Figure 2. Transfection efficiency and stability of synthetic microRNAs in T84 cells. (**A**) Just after plating $5x10^5$ cells on a polycarbonate filter, cells were transfected with cy3-labeled control miRNA (Green; Ambion). After 24 hours, Filters were harvested and fixed in 4% formaldehyde. For nuclear stain, filters

were incubated with bisbenzimide H 33342 trihydrochloride (Blue; Sigma), and were observed by confocal microscopy. *Bars*, 20 μm. (**B-F**) Time course of miR-210 (**B**), miR-338-3p (**C**), miR-33a (**D**), miR-451 (**E**) and *RNU6-2* (**F**) amounts in T84 cells transfected with none, NC or combinations of synthetic miRNAs was analyzed by qRT-PCR. The values are shown as Ct value.

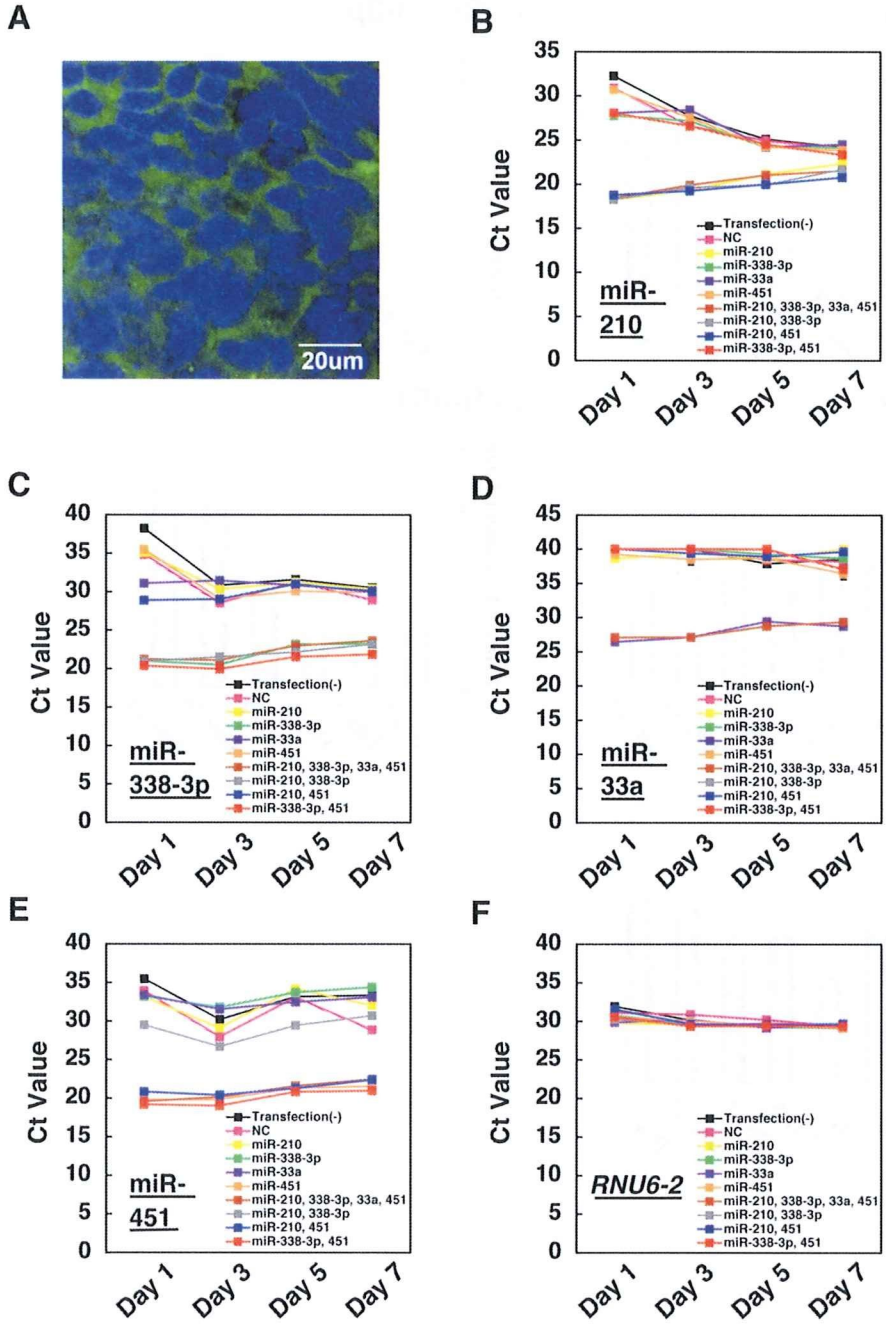
Supplementary Figure 3. Effects of overexpression and functional inhibition of four miRNAs on the formation and function of tight junctions. (**A**) Time course of the development of transepithelial electrical resistance (TER) in Caco-2 cells seeded in the transwell chamber. The formation and function of tight junctions were monitored by measurement of TER, which was measured using Millicell-ERS (Millipore, Bedford, MA). The graph is an average of three experiments; error bars indicate SE (n=3). (**B,C**) The TER at the day 7 in Caco-2 cells transfected with antisense-miRNA olidonucleotides (**B**) or synthetic miRNAs (**C**) are shown. Caco-2 cells were transfected at day 0, when Caco-2 cells were seeded in transwell chambers. Mean ± SE are shown (n=3). Significant difference was not detected (Student *t* test).

Supplementary Figure 4. Effects of overexpression and functional inhibition of four miRNAs on the ALPI mRNA expresion levels. The ALPI mRNA expression levels in T84 cells transfected with synthetic miRNAs (**A**) or antisense-miRNA olidonucleotides (**B**) were investigated at the day 7 by qRT-PCR. Mean ± SE are shown (n=3). Significant difference was not detected (Student *t* test).

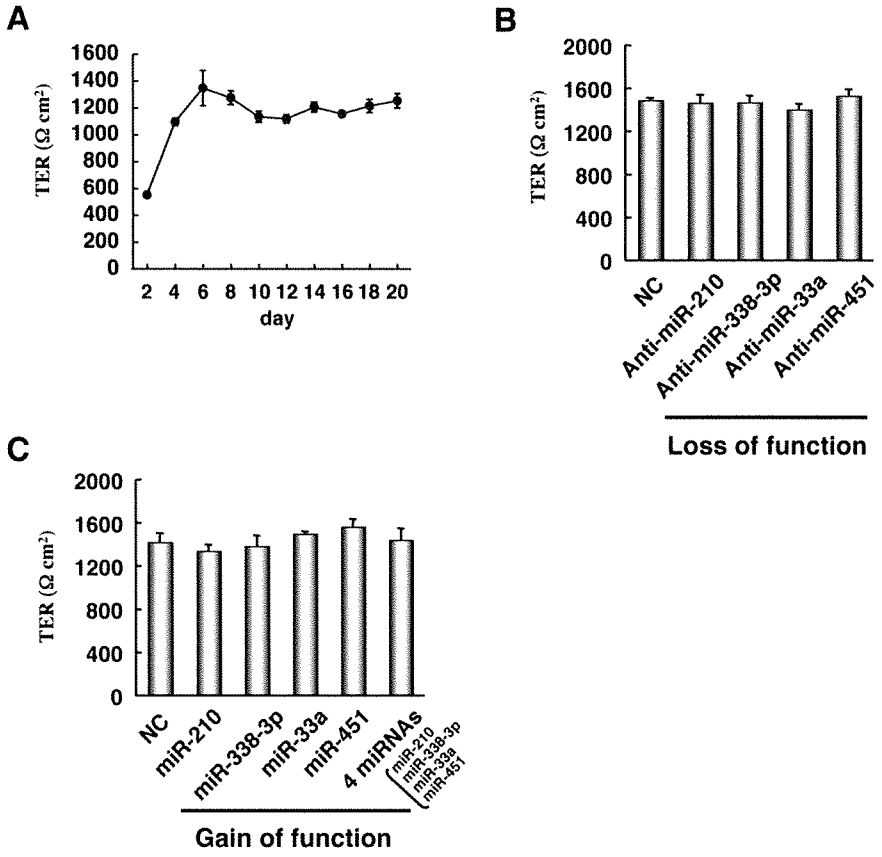
Tsuchiya *et al.* Supplementary Figure 1



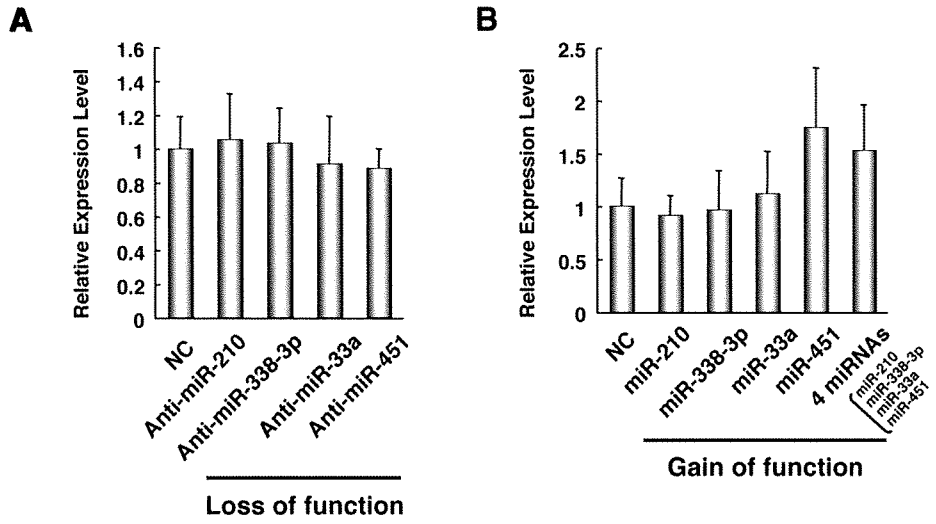
Tsuchiya *et al.* Supplementary Figure 2



Tsuchiya *et al.* Supplementary Figure 3



Tsuchiya *et al.* Supplementary Figure 4



Supplementary Table 1:

miRNA	Day 0 (Ct value)	Day 1 (Ct value)	Day 7 (Ct value)
let-7a	21.26	21.3	20.49
let-7b	21.35	21.5	20.96
let-7c	25.63	25.98	25.5
let-7d	24.66	25.39	24.09
let-7e	28.56	28.02	27.48
let-7f	23.42	23.3	22.73
let-7g	21.32	21.68	21.08
let-7i	24	24.16	24
miR-1	40	40	34.41
miR-100	29.05	29.3	28.55
miR-101	27.24	27.11	26.15
miR-103	22.1	22.4	21.37
miR-106a	22.72	22.35	22.07
miR-106b	21.24	21.21	20.56
miR-107	26.36	26.72	25.86
miR-10a	22.41	22.51	21.48
miR-10b	24.24	24.66	23.56
miR-122a	40	40	40
miR-124a	40	40	40
miR-125a	24.26	24.49	24.22
miR-125b	31.03	32.13	31.47
miR-126	30.14	30.1	28.83
miR-127	35.42	36.99	34.42
miR-128a	33.21	31.7	33.19
miR-128b	40	40	34.27
miR-129	39.93	39.9	39.14
miR-130a	30.72	31.14	30.92
miR-130b	25.3	26.07	25.52
miR-132	27.52	26.8	28.04
miR-133a	37.4	33.25	35.76
miR-133b	32.13	32.04	31.19
miR-134a	36.41	38.35	36.85
miR-135a	28.4	28.73	29.12
miR-135b	25.14	24.89	24.71
miR-137	40	40	40
miR-138	38.44	39.78	36.11
miR-139	40	36.55	36.48
miR-140	27.1	26.7	26.52
miR-141	22.06	22.35	21.29
miR-142-3p	21.18	20.93	20.33
miR-142-5p	24.6	24.89	24.24
miR-143	37.22	40	36.1
miR-145	32.97	33.29	32.83
miR-146a	28.82	28.81	28.17
miR-146b	27.59	27.89	27.01
miR-147	34.47	35.24	35.29
miR-148a	24.12	24.61	23.48
miR-148b	25.19	25.28	25.02
miR-149	29.69	30.25	30.13
miR-150	40	36.36	35.71
miR-151	25.83	26.27	25.48
miR-152	30.53	31.28	30.02
miR-153	40	40	40
miR-154	40	40	40
miR-155	25.25	25.21	24.38
miR-15a	24.28	24.54	23.91
miR-15b	21.33	21.48	20.82
miR-16	19.52	19.05	19.17
miR-17-3p	27.75	27.9	27.21
miR-17-5p	21.61	22.48	21.65
miR-181a	24.66	24.91	24.45
miR-181b	22.69	22.66	22.57
miR-181c	29.31	29.56	28.27
miR-181d	23.02	23.13	23.1
miR-182	24.78	24.59	24.23
miR-183	27.71	27.1	26.05
miR-184	40	40	40

Supplementary Table 1:

miRNA	Day 0 (Ct value)	Day 1 (Ct value)	Day 7 (Ct value)
miR-186	25.28	24.99	24.44
miR-187	35.45	40	36.45
miR-188	29.69	29.59	30.16
miR-18a	25.48	25.02	24.35
miR-190	28.38	28.52	27.7
miR-191	23.74	23.76	23.29
miR-192	19.57	19.69	18.47
miR-193a	27.36	27.84	26.93
miR-193b	28.18	28.67	27.7
miR-194	21	21.28	20.05
miR-195	26	26.85	25.58
miR-196a	24.07	24.9	24.04
miR-196b	21.47	22.47	21.64
miR-197	25.72	27.02	25.83
miR-198	35.63	38.39	35.7
miR-199a	39.71	40	40
miR-199b	40	35.09	40
miR-19a	23.19	22.38	22.58
miR-19b	20.34	20.47	19.42
miR-200a*	27.05	27.34	27.32
miR-200c	19.53	20.29	19.17
miR-202	40	38.23	40
miR-203	22.59	22.99	22.54
miR-204	34.21	34.63	32.61
miR-205	35.22	40	37.04
miR-206	34.52	36	33.97
miR-208	40	39.87	40
miR-20a	19.79	20.33	19.48
miR-20b	27.42	27.61	26.58
miR-21	19.23	18.17	18.49
miR-210	27.25	26.14	22.41
miR-211	37.36	34.66	36.77
miR-212	31.86	31.63	32.34
miR-213	30.08	29.64	30.19
miR-214	35.26	34.05	34.8
miR-215	25.11	24.88	23.65
miR-216	38.86	38.7	36.53
miR-217	40	40	35.79
miR-218	34.73	35.07	40
miR-219	33.55	34.08	33.66
miR-22	27.28	27.05	26.89
miR-220	40	40	40
miR-221	23.03	22.41	22.36
miR-222	21.45	21.23	21.19
miR-223	34.84	33.55	34.64
miR-224	25.54	25.25	24.94
miR-23a	23.59	25.02	23.91
miR-23b	24.58	24.56	23.32
miR-24	21.05	20.62	20.32
miR-25	21.84	22.16	21.3
miR-26a	19.92	20.31	19.02
miR-26b	21.76	21.62	20.68
miR-27a	21.6	21.29	21.24
miR-27b	23.07	23.62	22.06
miR-28	25.45	25.59	24.71
miR-296	31.29	31.51	30.83
miR-299-5p	40	40	40
miR-29a	19.64	19.83	19.37
miR-29b	23.22	23.5	21.75
miR-29c	24.38	24.82	24.14
miR-301	25.7	25.78	25.03
miR-302a*	40	40	40
miR-302b*	40	40	39.61
miR-302c*	40	40	40
miR-302d	40	31.42	36.83
miR-30a-3p	28.72	28.76	28.04
miR-30a-5p	23.37	23.16	22.28

Supplementary Table 1:

miRNA	Day 0 (Ct value)	Day 1 (Ct value)	Day 7 (Ct value)
miR-30b	22.36	22.57	21.15
miR-30c	22.14	21.89	20.71
miR-30d	24.25	24.03	23.38
miR-30e-3p	25.3	25.11	24.67
miR-30e-5p	27.39	27.31	26.9
miR-31	23.27	22.51	22.43
miR-32	27.02	26.41	25.64
miR-320	23.35	24.11	23.56
miR-323	40	39.54	40
miR-324-3p	27.33	27.93	26.89
miR-324-5p	26.51	26.69	26.01
miR-325	40	40	40
miR-325	32.58	33.1	32.47
miR-328	32.07	31.61	31.05
miR-33a	40	35.29	31.09
miR-330	32.86	33.15	33.03
miR-331	26.02	26.45	25.87
miR-335	27.8	28.48	27.78
miR-337	40	40	38.87
miR-338-3p	33.33	32.83	29.07
miR-339	25.71	26.66	25.2
miR-340	32.1	31.94	31.19
miR-342	39.95	40	40
miR-345	27.45	27.35	27.28
miR-346	34.64	33.26	34.25
miR-34a	27.06	27.71	26.06
miR-34b	40	32.49	38.32
miR-34c	35.53	35.49	35.5
miR-361	26.47	26.62	25.95
miR-361	26.02	25.55	25.56
miR-367	40	40	40
miR-368	40	40	40
miR-369-3p	40	37.06	36.51
miR-369-5p	40	40	40
miR-370	34.49	34.09	32.25
miR-371	40	39.98	40
miR-372	40	40	39.87
miR-373*	40	40	39.97
miR-374	24.73	24.55	24.46
miR-375	23.44	23.49	23.09
miR-376a	37.89	37.06	37.16
miR-378	28.28	28.21	28.25
miR-379	36.59	35.23	40
miR-380-3p	40	40	40
miR-381	40	40	40
miR-382	38.96	37.28	40
miR-383	35.74	36.38	36.14
miR-409-5p	40	38.02	40
miR-422a	30.66	30.19	29.26
miR-422b	25.27	25.19	23.99
miR-423	24.74	24.07	23.17
miR-424	33.29	39.19	32.25
miR-425	26.24	26.52	25.41
miR-429	23.86	23.56	23.1
miR-432	33.48	33.46	33.13
miR-433	35.05	34.99	34.14
miR-449	29	29.82	27.81
miR-450	40	40	37.53
miR-451	33.5	30.91	30.56
miR-452	28.71	28.45	28.56
miR-485-5p	34.36	33.24	34.35
miR-489	32.28	34.1	31.91
miR-490	37.54	37.12	36.83
miR-491	31.26	31.4	31.03
miR-494	37.12	38.16	35.26
miR-496	35.26	39.37	36.31
miR-497	31.68	32.1	30.91
miR-500	27.21	27.53	27.53

Supplementary Table 1:

miRNA	Day 0 (Ct value)	Day 1 (Ct value)	Day 7 (Ct value)
mIR-501	27.04	27.56	27.14
miR-502	29.46	29.61	28.87
miR-505	24.47	25.05	24.21
miR-506	37.05	38.02	39.01
miR-508	34.3	36.22	37.66
miR-509	32.35	32.84	31.95
miR-510	36.7	35.48	35.75
miR-511	39.54	37.78	40
miR-512-5p	37.84	38.04	38.23
miR-513	40	37.45	38.26
miR-514	34.26	34.32	37.95
miR-515-3p	33.02	33.52	35.05
miR-515-5p	40	40	40
miR-516-3p	31.31	31.92	30.55
miR-517a	40	40	38.55
miR-517b	40	38.95	39.01
miR-517c	39.98	35.37	33.63
miR-518a	37.07	36.22	35.35
miR-518b	34.29	33.32	32.72
miR-518c	37.28	38.17	39.07
miR-518d	33.35	33.71	33.18
miR-518e	33.61	34.93	33.94
miR-519b	40	40	40
miR-519c	32.48	31.39	31.88
miR-519d	36.4	34.63	35.2
miR-519e	35.49	34.82	34.39
miR-520a	33.12	32.87	32.1
miR-520b	35.35	35.86	35.5
miR-520c	38.32	39.1	36.9
miR-520d	37.86	38.41	35.58
miR-520e	40	39.35	40
miR-520f	36.34	38.33	39.12
miR-520g	35.47	39.27	40
miR-520h	37.71	39.77	35.14
miR-521	40	40	40
miR-522	38.77	40	37.4
miR-523	40	40	40
miR-526a	36.38	40	36.66
miR-526b	40	40	40
miR-7	24.55	24.38	24.35
miR-9	31.92	31.23	30.88
miR-92	19.52	19.43	19.07
miR-93	20.15	20.15	20.09
miR-95	28.37	28.64	27.9
miR-96	27.83	28.72	27.65
miR-98	25.59	24.48	24.08
miR-99a	29.1	29.14	28.21
miR-99b	27.15	26.71	27.27
RNU6-2	27.15	27.57	27.21

Supplementary Table 2:

Affymetrix Number	"NC" at day 5	"miR-338-3p and miR-451" at the day5	"four miRNA" at the day5	"NC" at day 7	Gene Title	Gene Symbol
1554178_a_at	4.11548801	7.130905465	6.385671653	5.144875183	family with sequence similarity 126, member B	FAM126B
1568634_a_at	4.589534786	5.697253778	6.036500852	5.752213536	similar to hypothetical protein MGC38937	LOC339977
1569532_a_at	11.44797662	12.80008443	13.03232262	13.17809588	MSFL2541	UNQ2541
200602_at	8.077331321	10.33837173	10.33981392	9.315642273	amyloid beta (A4) precursor protein (peptidase nexin-II, Alzheimer disease)	APP
200632_s_at	10.9112862	12.30968749	12.70221762	12.68460718	N-myc downstream regulated gene 1	NDRG1
200864_s_at	5.92418717	8.323806144	8.504071752	7.312231797	RAB11A, member RAS oncogene family	RAB11A
200884_at	11.03669433	12.27821633	12.43081856	12.29524792	creatine kinase, brain	CKB
201117_s_at	4.18943128	6.472159876	6.226529492	5.572593444	carboxypeptidase E	CPE
201133_s_at	7.481537796	9.570529403	9.490627636	8.495587467	praja 2, RING-H2 motif containing	PJA2
201427_s_at	4.749835378	6.898059267	6.332444852	5.948607496	selenoprotein P, plasma, 1	SEPP1
201848_s_at	9.690484677	10.89202396	11.04495292	10.7325944	BCL2/adenovirus E1B 19kDa interacting protein 3	BNIP3
201849_at	10.84644956	12.61777068	12.64597681	12.10628219	BCL2/adenovirus E1B 19kDa interacting protein 3	BNIP3
201963_at	7.31066512	9.396758082	9.694253265	8.610155052	acyl-CoA synthetase long-chain family member 1	ACSL1
202022_at	8.638318561	9.951297425	10.15875337	10.11098757	aldolase C, fructose-bisphosphate	ALDOC
202269_x_at	4.888217758	5.909130523	6.254611173	6.073798028	guanylate binding protein 1, interferon-inducible, 67kDa	GBP1
202619_s_at	7.07921316	8.506417987	8.50047286	8.49830617	procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2	PLOD2
202620_s_at	7.854450343	10.30866695	10.27724972	9.412501227	procollagen-lysine, 2-oxoglutarate 5-dioxygenase 2	PLOD2
202688_at	7.604870636	9.361466547	9.527479362	8.957609885	tumor necrosis factor (ligand) superfamily, member 10	TNFSF10
202843_at	5.022547191	6.653230563	6.792741965	6.033649482	DnaJ (Hsp40) homolog, subfamily B, member 9	DNAJB9
202855_s_at	6.554349718	7.679086748	8.501810701	7.917178631	hypoxanthine phosphoribosyltransferase 1 (Lesch-Nyhan syndrome)	HPRT1
202856_s_at	7.247816037	8.726959198	9.395124825	8.856918189	solute carrier family 16, member 3 (monocarboxylic acid transporter 4)	SLC16A3
202887_s_at	10.17958574	11.63983557	11.94647799	11.45657207	protein phosphatase 2 (formerly 2A), regulatory subunit A, beta isoform	PPP2R1B
202901_x_at	4.471674118	7.256964105	7.378820872	5.876186161	cathepsin S	CTSS
202973_x_at	7.443917371	9.633233684	9.881823251	9.625193008	family with sequence similarity 13, member A1	FAM13A1
203123_s_at	9.640517993	11.10661505	11.430023	11.77424571	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2	SLC11A2
203124_s_at	9.287091792	11.344019	11.80452127	11.60299306	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2	SLC11A2
203178_at	5.034403888	7.316677742	6.589087805	6.271798231	glycine amidinotransferase (L-arginine:glycine amidinotransferase)	GATM
203282_at	8.604965572	10.88396517	10.89012837	9.978399076	glucan (1,4-alpha-), branching enzyme 1 (glycogen branching enzyme, Andersen disease, glycogen storage disease type IV)	GBE1
203739_at	7.18404445	9.190316226	8.907544428	8.429271091	zinc finger protein 217	ZNF217
203780_at	7.714926098	9.639256515	9.641474451	9.137281012	myelin protein zero-like 2	MPZL2
203845_at	5.389813317	6.833882835	6.79026835	6.513522165	p300/CBP-associated factor	PCAF
203882_at	6.810305842	8.518939771	8.4619363	7.933472457	interferon regulatory factor 9	IRF9
204032_at	5.80172848	7.11828647	7.21826657	6.979980838	breast cancer anti-estrogen resistance 3	BCAR3
204614_at	4.465537616	5.76368556	5.506134207	5.479825247	serpin peptidase inhibitor, clade B (ovalbumin),	SERPINB2
205097_at	10.10340784	11.83014815	11.61053669	12.05313041	solute carrier family 26 (sulfate transporter), member 2	SLC26A2
205112_at	6.419153754	7.509577012	7.55933798	7.467844969	phospholipase C, epsilon 1	PLCE1
205198_s_at	5.586016982	6.734307764	6.934148049	6.614372118	ATPase, Cu++ transporting, alpha polypeptide (Menkes syndrome) /// similar to ATPase, Cu++ transporting, alpha polypeptide	ATP7A /// LOC644732
205199_at	7.856874495	8.954634644	9.651890483	9.408790283	carbonic anhydrase IX	CA9
205289_at	7.706460421	9.077458049	9.299993038	8.940160693	bone morphogenetic protein 2	BMP2
205290_s_at	6.949333002	8.227092187	8.582592556	8.35505017	bone morphogenetic protein 2	BMP2
205597_at	6.010001253	7.172551162	8.068396101	7.699018389	solute carrier family 44, member 4	SLC44A4
205625_s_at	5.952760972	8.684153965	8.835426218	7.098979374	calbindin 1, 28kDa	CALB1
205626_s_at	5.047012777	8.312614213	8.676700083	6.293466762	calbindin 1, 28kDa	CALB1
205771_s_at	8.389825017	10.33453272	10.19008661	10.13586234	A kinase (PRKA) anchor protein 7	AKAP7
206084_at	5.976669815	8.357528818	8.390029566	7.766570038	protein tyrosine phosphatase, receptor type, R	PTPRR
206098_at	4.041760191	5.450828629	5.362373989	5.106651977	zinc finger and BTB domain containing 6	ZBTB6
206143_at	6.014979455	7.225241187	8.233819778	9.683436866	solute carrier family 26, member 3	SLC26A3
206667_s_at	4.726513364	5.809279531	5.892387609	5.761768702	secretory carrier membrane protein 1	SCAMP1
206784_at	7.046211082	8.33135891	8.91975344	10.38158479	aquaporin 8	AQP8
207014_at	3.347204163	5.350927212	6.019901367	5.469339938	gamma-aminobutyric acid (GABA) A receptor, alpha 2	GABRA2
207052_at	7.972751661	9.479389266	9.518003157	9.323768474	hepatitis A virus cellular receptor 1	HAVCR1
207405_s_at	5.939486238	8.131607956	8.361158284	6.947483743	RAD17 homolog (S. pombe)	RAD17
207543_s_at	9.439890547	10.77896041	10.79615014	10.68582492	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), alpha polypeptide I	P4HA1

Affymetrix Number	"NC" at day 5	"miR-338-3p and miR-451" at the day5	"four miRNA" at the day5	"NC" at day 7	Gene Title	Gene Symbol
208383_s_at	5.631920662	6.815265124	7.264951016	6.897658951	phosphoenolpyruvate carboxykinase 1 (soluble)	PCK1
208550_x_at	3.627886931	4.980551699	4.791039928	4.908893059	potassium voltage-gated channel, subfamily G,	KCNG2
208892_s_at	7.076469371	9.826496985	10.20910318	8.50747382	dual specificity phosphatase 6	DUSP6
208933_s_at	5.976669815	7.374757698	7.527994554	7.059160441	lectin, galactoside-binding, soluble, 8 (galectin 8)	LGALS8
209211_at	6.263508389	9.717519383	10.07604776	7.633860305	Kruppel-like factor 5 (intestinal)	KLF5
209357_at	6.539103384	8.998276422	8.907168448	7.582838871	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2	CITED2
209546_s_at	6.564123423	7.567643151	7.587690732	7.60742213	apolipoprotein L, 1	APOL1
209598_at	6.566521251	7.790380316	8.00054991	7.89815097	paraneoplastic antigen MA2	PNMA2
210174_at	5.30359129	6.622955139	6.420356404	6.309010135	nuclear receptor subfamily 5, group A, member 2	NR5A2
210479_s_at	4.009012944	5.114179943	5.346837128	5.195431966	RAR-related orphan receptor A	RORA
210675_s_at	6.654537906	8.589991191	8.969781633	8.25308997	protein tyrosine phosphatase, receptor type, R	PTPRR
212195_at	4.320760089	7.135357859	7.260364461	5.975897909	Interleukin 6 signal transducer (gp130, oncostatin M	IL6ST
212294_at	8.324560521	9.426791874	9.553548038	9.360227275	guanine nucleotide binding protein (G protein), gamma	GNG12
212476_at	6.254624742	7.936225812	8.055417194	7.324476249	centaurin, beta 2	CENTB2
212499_s_at	6.197704866	9.028207265	8.802599377	7.474530483	homolog (S. cerevisiae) /// chromosome 14 open reading frame 32	C14orf32 /// FCF1
212560_at	10.24360825	11.60834563	11.74373344	11.65801622	chromosome 11 open reading frame 32	C11orf32
212593_s_at	7.285096802	9.787666854	9.768565506	8.383756824	programmed cell death 4 (neoplastic transformation	PDCD4
212689_s_at	9.468873016	10.53073165	10.6523603	10.62974761	jumonji domain containing 1A	JMJD1A
212870_at	7.561809986	8.961417087	9.06070942	8.659360076	son of sevenless homolog 2 (Drosophila)	SOS2
212900_at	7.225346225	8.921041078	8.672273705	8.302896752	SEC24 related gene family, member A (S. cerevisiae)	SEC24A
213317_at	5.735199655	7.029147063	7.496140801	6.780779614	chloride intracellular channel 5	CLIC5
213349_at	8.977484033	10.4769866	10.72533836	10.3470371	transmembrane and coiled-coil domain family 1	TMCC1
213351_s_at	6.850862634	8.232238388	8.543702682	8.215514894	transmembrane and coiled-coil domain family 1	TMCC1
213352_at	6.168437402	7.762776625	7.979066948	7.346120262	transmembrane and coiled-coil domain family 1	TMCC1
213397_x_at	7.703179243	9.59180828	9.539803195	8.842370008	ribonuclease, RNase A family, 4	RNASE4
213510_x_at	5.640196541	7.368729909	7.365185014	6.754484046	TL132 protein	LOC220594
213552_at	6.34828151	8.761950219	8.775948058	7.471411736	glucuronic acid epimerase	GLCE
213929_at	5.976711236	8.130184094	8.355439777	7.149790477	CDNA clone IMAGE:4733238	---
214835_s_at	8.410629048	10.67504632	10.54420869	9.529855269	succinate-CoA ligase, GDP-forming, beta subunit	SUCLG2
214855_s_at	4.733113599	6.942246172	6.578552704	6.180773377	GTPase activating Rap/RanGAP domain-like 1	GARNL1
216733_s_at	6.59747408	8.05181529	7.979989584	7.646883162	glycine amidinotransferase (L-arginine:glycine amidinotransferase)	GATM
217047_s_at	7.939972582	9.141281837	9.514216867	9.484409309	family with sequence similarity 13, member A1	FAM13A1
217954_s_at	7.871363356	9.834032309	9.710572314	8.960673984	PHD finger protein 3	PHF3
218158_s_at	5.143658866	6.877173655	7.104461602	6.30532525	adaptor protein, phosphotyrosine interaction, PH domain and leucine zipper containing 1	APPL1
218326_s_at	7.954597579	9.876587751	9.878552701	9.028149777	leucine-rich repeat-containing G protein-coupled	LGR4
218490_s_at	6.221317658	8.12712908	7.651300166	7.53532776	zinc finger protein 302	ZNF302
218498_s_at	8.014493129	9.290754207	9.49387991	9.167731131	ERO1-like (S. cerevisiae)	ERO1L
218521_s_at	3.774561094	5.365392995	5.245911861	5.039634888	ubiquitin-conjugating enzyme E2W (putative)	UBE2W
218983_at	5.751310988	7.371845206	7.675080297	6.885080783	complement component 1, r subcomponent-like	C1RL
219190_s_at	4.372865508	5.575611937	5.687396967	5.854943421	eukaryotic translation initiation factor 2C, 4	EIF2C4
219232_s_at	5.887250375	7.714395485	7.745223433	7.482719944	egl nine homolog 3 (C. elegans)	EGLN3
219739_at	6.931919587	8.211645335	8.775781894	9.429879085	ring finger protein 186	RNF186
220724_at	3.406558651	4.550020012	4.714369431	4.709323808	hypothetical protein FLJ21511	FLJ21511
221478_at	8.65216326	10.65611691	10.56796973	9.945376378	BCL2/adenovirus E1B 19kDa interacting protein 3-like	BNIP3L
221530_s_at	6.868824404	8.321501498	8.746123858	8.322663149	basic helix-loop-helix domain containing, class B, 3	BHLHB3
221589_s_at	5.013690065	8.635717805	8.498469898	6.041640471	aldehyde dehydrogenase 6 family, member A1	ALDH6A1
222408_s_at	6.683182543	8.059326589	8.453892143	8.094005407	yippee-like 5 (Drosophila)	YPEL5
222646_s_at	9.034021943	10.90836829	11.0662356	10.2446226	ERO1-like (S. cerevisiae)	ERO1L
222847_s_at	7.403190628	8.502827798	8.826787216	8.688979017	egl nine homolog 3 (C. elegans)	EGLN3
223044_at	8.209295373	10.04351695	10.34910455	9.854872652	solute carrier family 40 (iron-regulated transporter),	SLC40A1
224604_at	7.911197735	9.925135421	9.936590221	8.950461375	HCV F-transactivated protein 1	LOC401152
224797_at	6.67095383	9.115016611	9.012833396	8.066793497	arrestin domain containing 3	ARRDC3
224953_at	6.241490017	7.600790805	7.720243284	7.34845668	Yip1 domain family, member 5	YIPF5
224959_at	10.3221387	11.80776225	11.60690807	12.06818646	solute carrier family 26 (sulfate transporter), member 2	SLC26A2
225537_at	5.390904461	7.633451809	7.229951063	7.055080453	trafficking protein particle complex 6B	TRAPPC6B
225622_at	4.880651916	6.635469075	6.685568328	6.465939561	phosphoprotein associated with glycosphingolipid microdomains 1	PAG1
225626_at	6.675458589	7.740474417	7.949961061	7.708897189	phosphoprotein associated with glycosphingolipid microdomains 1	PAG1
225707_at	8.362855553	10.86585669	10.59347236	9.931852221	ADP-ribosylation-like factor 6 interacting protein 6	ARL6IP6
225892_at	5.653254033	9.053837895	8.726959198	6.790759102	iron-responsive element binding protein 2	IREB2

Supplementary Table 2:

Affymetrix Number	"NC" at day 5	"miR-338-3p and miR-451" at the day5	"four miRNA" at the day5	"NC" at day 7	Gene Title	Gene Symbol
226275_at	5.922976694	7.613730767	7.867903295	7.694194749	MAX dimerization protein 1	MXD1
226333_at	5.675801987	7.137833873	7.328416421	6.871515198	interleukin 6 receptor	IL6R
226347_at	6.235473412	7.681914049	7.855614513	7.391617347	---	---
226381_at	5.676130891	7.947293301	8.233438459	7.483757926	HBV preS1-transactivated protein 4	PS1TP4
226520_at	7.868346203	9.324983194	9.175980212	8.969290149	Primary neuroblastoma cDNA, clone:Nbla11485	---
226576_at	6.364669013	7.910684628	7.985669214	7.681535549	Rho GTPase activating protein 26	ARHGAP26
226622_at	8.490704248	9.57914271	10.17689812	10.46259815	mucin 20, cell surface associated	MUC20
226675_s_at	9.378616475	11.97824232	11.50777574	10.43651804	metastasis associated lung adenocarcinoma transcript 1 (non-protein coding)	MALAT1
226752_at	6.894339681	8.927196776	8.87832124	8.248769179	transmembrane protein 157	TMEM157
227226_at	5.901081794	7.758213143	8.181061069	8.21035859	chromosome 6 open reading frame 117	C6orf117
227337_at	7.858845414	9.615691149	9.656813795	9.554526665	ankyrin repeat domain 37	ANKRD37
227354_at	4.463063351	5.659657075	6.010730807	5.62702037	phosphoprotein associated with glycosphingolipid microdomains 1	PAG1
227777_at	5.49319106	6.90962684	6.992002014	6.52623483	chromosome 10 open reading frame 18	C10orf18
228094_at	4.854988897	6.399741139	6.451505643	6.274853488	adhesion molecule, interacts with CXADR antigen 1	AMICA1
228483_s_at	7.212676328	8.69855694	8.394244236	8.217031465	TAF9B RNA polymerase II, TATA box binding protein (TBP)-associated factor, 31kDa	TAF9B
228937_at	4.82543288	6.679053992	7.256859273	6.485574096	chromosome 13 open reading frame 31	C13orf31
229546_at	4.344203799	8.31952575	8.24182868	5.880836215	hypothetical LOC653602	LOC653602
229810_at	5.754795088	7.109311472	7.341948729	6.94852399	Transcribed locus	---
230043_at	5.757699112	7.069906111	7.729340612	7.394798481	mucin 20, cell surface associated	MUC20
230083_at	6.570428963	8.37693326	8.207295919	7.672297139	ubiquitin specific peptidase 53	USP53
230492_s_at	6.633125258	8.119973297	8.118342256	7.679886156	hypothetical protein KIAA1434	RP5-1022P6.2
230710_at	6.614372118	7.72068119	7.96559722	8.092585837	CDNA FLJ41489 fis, clone BRTHA2004582	---
230746_s_at	3.893856334	5.985461645	5.860107445	5.074459235	Transcribed locus	---
231033_at	4.935669216	6.159648181	6.909103313	6.319716024	Full length insert cDNA clone YI40A07	---
231941_s_at	7.684624801	8.93082416	9.562565544	9.777599423	mucin 20, cell surface associated	MUC20
231982_at	7.255643425	8.653518111	8.758115826	8.705637167	similar to HSPC323	LOC284422
232628_at	5.921435738	8.459866179	7.48750351	7.490595182	CDNA FLJ13464 fis, clone PLACE1003478	---
233329_s_at	9.18503364	10.48530119	10.37213879	10.24691658	lysine-rich coiled-coil 1	KRCC1
234989_at	7.704633832	9.192728669	8.790663213	8.901870892	trophoblast-derived noncoding RNA	TncRNA
236224_at	6.831204353	8.449875855	8.372723988	7.927170095	Ras-like without CAAX 1	RIT1
237521_x_at	5.856859707	6.976247301	6.931810353	6.94018831	Transcribed locus	---
238103_at	6.612148119	8.157183132	8.439884722	8.038375578	CDNA FLJ37936 fis, clone CTONG2005468	---
238215_at	4.252021071	5.510220861	6.359643711	5.308148646	solute carrier family 6, member 18	SLC6A18
238476_at	5.434237641	6.668320818	6.68865952	6.56497712	chromosome 5 open reading frame 41	C5orf41
238756_at	8.641677267	10.77165504	10.68732062	9.671488485	Growth arrest-specific 2 like 3	GAS2L3
239843_at	4.927784396	7.332063988	7.362341334	6.507542826	Ras-like without CAAX 1	RIT1
240991_at	5.546596914	6.991391263	6.861463527	7.305491841	Transcribed locus	---
242727_at	5.038757429	7.92462005	8.070967914	6.636267302	ADP-ribosylation factor-like 5B	ARL5B
243702_at	5.026079173	6.113012851	6.340472483	6.139552257	---	---
243774_at	4.451986626	5.61127716	5.834331116	5.517815263	Programmed cell death 2	PDCD2
244567_at	7.389740474	10.05973282	9.945547726	8.949541276	Transcribed locus	---
244811_at	5.137220094	6.882207185	6.591245474	6.294519393	Full-length cDNA clone CS0DF025YM09 of Fetal brain of Homo sapiens (human)	---

OPEN ACCESS Freely available online

PLoS one

Intra-Platform Repeatability and Inter-Platform Comparability of MicroRNA Microarray Technology

Fumiaki Sato[1]*, Soken Tsuchiya[1], Kazuya Terasawa[2], Gozoh Tsujimoto[2]

1 Department of Nanobio Drug Discovery, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto, Kyoto, Japan, **2** Department of Pharcogenomics, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto, Kyoto, Japan

Abstract

Over the last decade, DNA microarray technology has provided a great contribution to the life sciences. The MicroArray Quality Control (MAQC) project demonstrated the way to analyze the expression microarray. Recently, microarray technology has been utilized to analyze a comprehensive microRNA expression profiling. Currently, several platforms of microRNA microarray chips are commercially available. Thus, we compared repeatability and comparability of five different microRNA microarray platforms (Agilent, Ambion, Exiqon, Invitrogen and Toray) using 309 microRNAs probes, and the Taqman microRNA system using 142 microRNA probes. This study demonstrated that microRNA microarray has high intra-platform repeatability and comparability to quantitative RT-PCR of microRNA. Among the five platforms, Agilent and Toray array showed relatively better performances than the others. However, the current lineup of commercially available microRNA microarray systems fails to show good inter-platform concordance, probably because of lack of an adequate normalization method and severe divergence in stringency of detection call criteria between different platforms. This study provided the basic information about the performance and the problems specific to the current microRNA microarray systems.

Citation: Sato F, Tsuchiya S, Terasawa K, Tsujimoto G (2009) Intra-Platform Repeatability and Inter-Platform Comparability of MicroRNA Microarray Technology. PLoS ONE 4(5): e5540. doi:10.1371/journal.pone.0005540

Editor: Kumar Selvarajoo, Keio University, Japan

Received February 10, 2009; **Accepted** April 8, 2009; **Published** May 14, 2009

Funding: Grants-in-aid from Japanese Ministry of Education, Culture, Sports, Science, and Technology, Japan (#20591569 for F. Sato, #19109001 for G. Tsujimoto, # 19770171 for K. Terasawa), Uehara Memorial Foundation, New Energy and Industrial Technology Development Organization (NEDO) in Japan, and endowed fund from TORAY Industries Inc. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: fsato@pharm.kyoto-u.ac.jp

Introduction

Since the first DNA microarray paper demonstrated that microarray technology can monitor multiple gene expression profile in 1995 [1], DNA microarray technology has been developed steadily. After the Human Genome Project was finished, the ability of DNA microarray expanded to genome-wide analysis of not only gene expression profiling, but also, genome variation, epigenetics, DNA-protein interaction, and so on. In the research field, these genome-wide analyses using microarray technology have been providing deeper biological insights for a decade. In the clinical field, the US Food and Drug Administration (FDA) approved MammaPrint® as the first *in vitro* diagnostic multivariate index assay (IVDMIA) in February, 2007. Thus, microarray-based transcriptome devices started to be utilized to stratify patients for personalized medicine. For the quality control and standardization of microarray chips, the US FDA initiated the MicroArray Quality Control project (MAQC) in 2005. A series of reports regarding the first phase of the MAQC project was published in 2006 [2–7]. The MAQC report showed intra platform consistency across test sites as well as a high level of inter-platform concordance in terms of genes identified as differentially expressed.

MicroRNAs are a class of small non-coding RNAs [19–23 nucleotides (nt)] that have been found in animal and plant cells. As of today, 718 human microRNAs are registered in the miRBase database (Release 13, March, 2009) [8–11]. MicroRNA genes are transcribed as non-coding transcripts, and processed through a series of sequential steps involving the RNase III enzymes, Drosha and Dicer. The processed microRNAs are finally incorporated into the RNA-induced silencing complex (RISC) to mediate target mRNA repression of translation and/or degradation. It is reported that microRNAs are involved in physiological and pathological functions, such as the regulation of developmental timing and pattern formation [12], restriction of differentiation potential [13], chromatin rearrangements [14], and carcinogenesis [15]. Many of the mechanistic details still remain unknown.

Recently, microarray technology has been utilized to analyze a comprehensive microRNA expression profiling. Currently, several platforms of microRNA microarray chips are commercially available. As mentioned above, the MAQC Project is currently underway for quality control and standardization of mRNA expression microarray. However, no comparative and quality control study of microRNA microarray platforms has been reported yet. Therefore, we compared repeatability and comparability of microRNA microarray using five different platforms (Agilent, Ambion, Exiqon, Invitrogen and Toray). In addition, we compared quantitivity of microarray data generated from five different platforms with that of quantitative RT-PCR (Taqman) method, which is the golden standard method of microRNA measurement.

Results

Experimental design

This project repeatedly assayed two RNA sample types on a variety of microRNA expression platforms at one laboratory. Our preliminary experiments showed that the amount of microRNA obtained from the same amount of total RNA depends on the tissue types of the samples (data not shown). This finding suggested that repeatability or comparability of microRNA microarray analysis might depend on the amount of microRNA contained in total RNA. To assess the reproducibility of microRNA microarray data using the different tissue types, we chose both tissue samples, which contain relatively small and large amounts of microRNA. Our preliminary data shows that mouse liver tissue contains relatively small amounts of microRNAs. Therefore, we used two types of total RNA, FirstChoice® Human Liver Total RNA (Ambion, lot no. 040000129) and FirstChoice® Human Prostate Total RNA (Ambion, lot no. 050500710), in this study. In fact, the amount of microRNAs in Human Liver Total RNA was smaller than that of Human Prostate Total RNA (Figure 1).

Five commercially available microRNA microarray platforms were tested: Agilent Technologies (AGL); Ambion Inc. (AMB); Exiqon (EXQ); Invitrogen (IVG) and Toray Industries Inc. (TRY) (Table 1). Four of the microarray providers used one-color protocols where one labeled RNA sample was hybridized to each microarray. The Invitrogen array was tested using a two-color and dye-swapping protocol so that, at first, two RNA samples were divided and differently labeled in red-green and green-red combinations, and each combination of the RNA sample set was simultaneously hybridized to a microarray.

Agilent and Toray used its own method or software to generate a quantitative signal value and a qualitative detection call for each probe on the microarray, whereas Ambion, Exiqon, and Invitrogen did not specify the scanner or software to quantify the signals of probes in the manufacturer's protocol booklet. To generate a qualitative call for probes, we asked the technical support centers of Ambion, Exiqon, and Invitrogen about the method of detection call. We followed the methods recommended by their technical support center.

Probe mapping

The MAQC project had a probe mapping problem in that each gene was detected by a differently designed probe between the different microarray platforms [6]. In contrast to the MAQC project, this cross-platform study of microRNA microarray has much less variability of probe mapping, because of the short length (18–23 nucleotides) of microRNAs. Instead of this probe mapping problem, we faced a different kind of annotation problem, due to the database version. The frequent update of the miRBase microRNA database [16] causes the situation that different microRNA platforms were designed based on A different version of miRBase database. Between the versions, names of some microRNAs were changed, and the sequence of some microRNAs bearing the same names were slightly changed in length. Therefore, we compared the sequences in the annotation list provided by the manufacturers. The 309 microRNAs which had the complete identical sequences probed in all different platforms were included in this study to simplify the inter-platform comparison and to avoid a bias based on miRBase version.

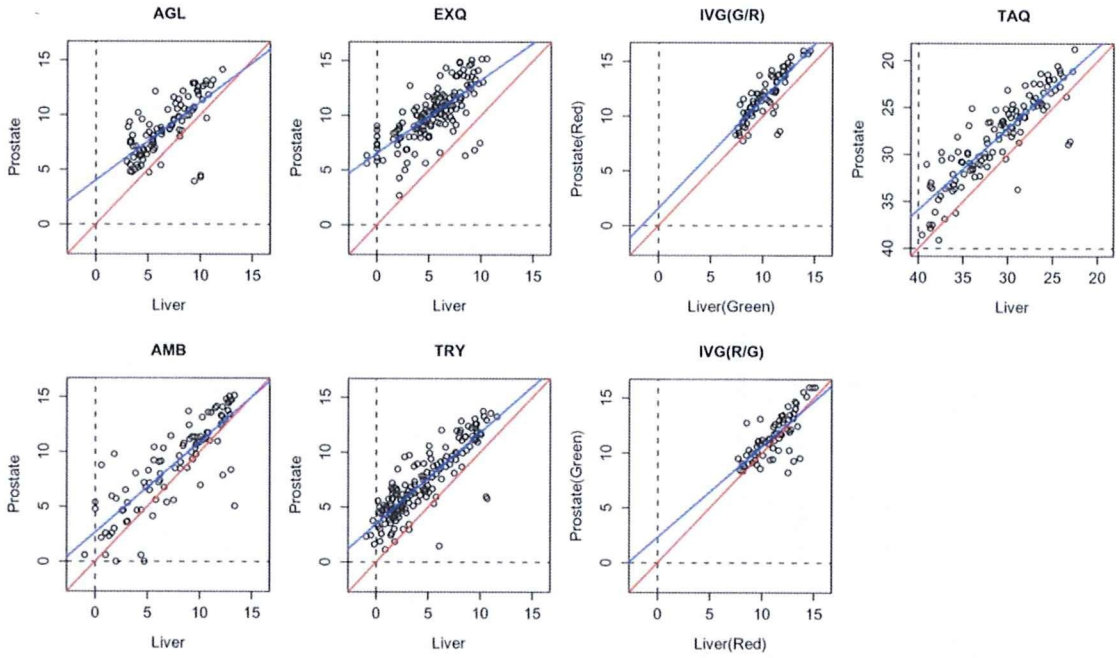


Figure 1. MicroRNA expression level in human liver and prostate tissues. For the microarray platforms, log2 transformed values of representative signal intensity for detection call-positive microRNAs were plotted. For the Taqman analysis, Ct values of microRNAs were plotted. Red and blue lines indicate Y = X line and regressed linear line, respectively. In all scatter plots. blue lines were shifted upward, which indicated that the general microRNA expression level in human prostate was higher than in human liver.
doi:10.1371/journal.pone.0005540.g001

Table 1. microRNA expression platforms and experimental procedures.

Code	Protocol	Platform	miRBase release	# of miRNAs	RNA (ug)	small RNA enrichment	Labeling Kit	Dye	Agitation	# of replicates
AGL	one color	Human miRNA V2 Oligo Microarray	10.1	723	0.1	no	Agilent	Cy3	yes	3
AMB	one color	mirVana miRNA Bioarray V9.2	9.2	312	20	yes	Ambion	Cy5	no	3
EXQ	one color	miRCURY LNA microRNA Array v.10.0	10.0	704	1	no	Exiqon	Hy5	no	3
IVG	two color	Ncode Human miRNA Microarray V3	10.0	699	5	no	Invitrogen	Alexa5 Alexa3	no	3×2
TRY	one color	3D-Gene Human miRNA oligo chip	10.1	723	0.5	no	Exiqon	Hy5	yes	3

doi:10.1371/journal.pone.0005540.t001

Distribution profile of microRNA microarray data

It will be important to know whether all data follows a specific distribution, e.g. Gaussian or not. Thus, we checked the distribution profile of data used in this study (Figure S1 and Table S1). MicroRNA microarray data have various distribution profiles between different platforms, although microarray data tend to have positive skewness (a right-side longer tail). It has been reported that the number of genes that are expressed at a similar level is approximately exponentially distributed in typical biological samples [17]. However, the skewness and kurtosis of microRNA microarray data were far smaller than those of the exponential distribution (skewness = 4, and kurtosis = 9) (Table S1). We also checked whether non-zero log2 data were normally distributed, or not. However, non-zero log2 data did not fit to normal distribution (Figure S1B). On the other hand, the log-ratio data between two samples were approximately normally distributed (Figure S1C).

Intra-platform data repeatability

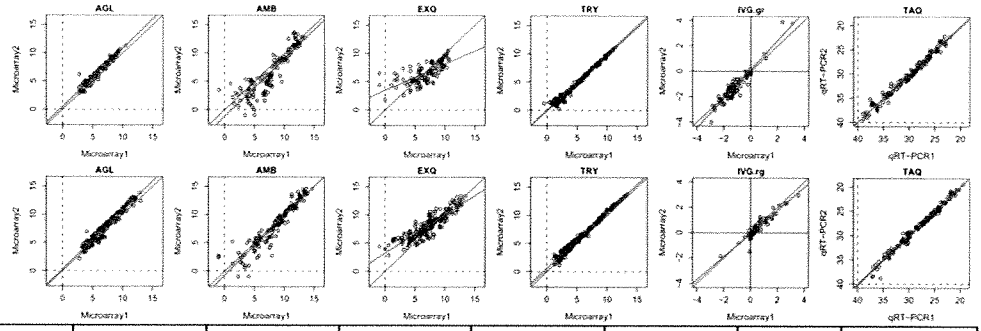
We examined microarray data for consistency within each platform by reviewing the repeatability at two levels: the quantitative signal values and the qualitative microRNA list agreement. To assess the data consistency of quantitative signal values, rank-correlation analysis and coefficient of variation (CV) analysis were performed. In this analysis, only data of microRNAs with positive detection call were used. Representative scatter plots of microarray platforms and the Taqman system are displayed in Figure 2A (scatter plots for all possible combinations between three replicates were shown in Figure S4). The Spearman's correlation coefficients (Rs), and the coefficient of variation (CV) between the three replicates was calculated using the 309 common microRNAs. Different platforms had various ranges of Rs values (liver: 0.82-0.96, prostate: 0.89-0.99, respectively). Thus, the 2-sample t-test and Mann-Whitney did not detect any significant difference between liver and prostate using whole data sets. However, the Rs values for prostate samples were constantly better than those for liver samples (Paired t-test: $p = 0.0013$, and Wilcoxon's signed-rank test: $p = 0.0005$). It is reasonable that Rs values of liver were lower than those of prostate, because higher signals in microarray data tend to have smaller data variability in general.

The distribution of CV for each platform was displayed in Figure 2B. Two platforms (AMB and EXQ) have low stringent criteria for detection call, in that all microRNAs with positive signal values after subtraction of background are considered as detected. It is also reasonable that these two platforms have higher CV values (both t-test and Mann-Whitney test: $p<0.0001$), because these platforms include microRNAs with near-zero values. In addition, the CV values of microRNA microarray platforms ranged in equivalent level to those of the Taqman assay.

Next, we assessed the variation in log-ratio measurement. For each platform, we performed triplicate experiments using human liver and prostate samples. Thus, we can generate 9 (= 3×3) log-ratios (prostate/liver) for each microRNA. Then, we calculated the Spearman's correlation coefficients (Rs) between 9 sets of log-ratios for the detected microRNAs, and visualized these Rs values inside of green squares in a blue-white heat map (Figure 3). The means and 95% confidence intervals (95%CI) of Rs values were listed in Table S2. The Rs values were high and consistent in two platforms (AGL, and TRY), in which protocol hybridization were performed with agitation. Another reason for the inconsistency of log-ratio values in AMB and EXQ might be the low stringent criteria of detection call, which included microRNAs with near-zero values.

To assess variation in the qualitative measures, the percentage of 309 microRNAs with concordant detection calls between

A



	Sample Type	AGL	AMB	EXQ	TRY	TAQ	Sample Type	IVG
# of probes	L	98+-3.6	142+-13.9	160+-48.3	173+-4.6	118+-1.2	L/P (G/R)	75.3+-1.1
Rs		0.95+-0.01	0.93+-0.02	0.83+-0.08	0.97+-0.003	0.99+-0.002		0.88+-0.01
Slope		1.06+-0.02	1.03+-0.02	0.68+-0.24	0.98+-0.002	0.96+-0.01		1.00+-0.11
Int		0.34+-0.24	-0.77+-0.53	2.23+-1.56	-0.06+-0.10	1.5+-0.42		-0.32+-0.27
# of probes	P	155+-8.7	146+-19.4	242+-43.7	217+-2.1	131+-2.5	L/P (R/G)	62+-5.6
Rs		0.98+-0.01	0.97+-0.01	0.93+-0.03	0.99+-0.003	0.99+-0.0004		0.91+-0.02
Slope		1.03+-0.02	1.11+-0.04	0.81+-0.14	0.99+-0.003	0.99+-0.02		0.97+-0.04
Int		0.21+-0.10	-1.5+-0.49	3.17+-1.16	0.53+-0.18	0.48+-0.62		-0.72+-0.40

MEAN +- SE

B

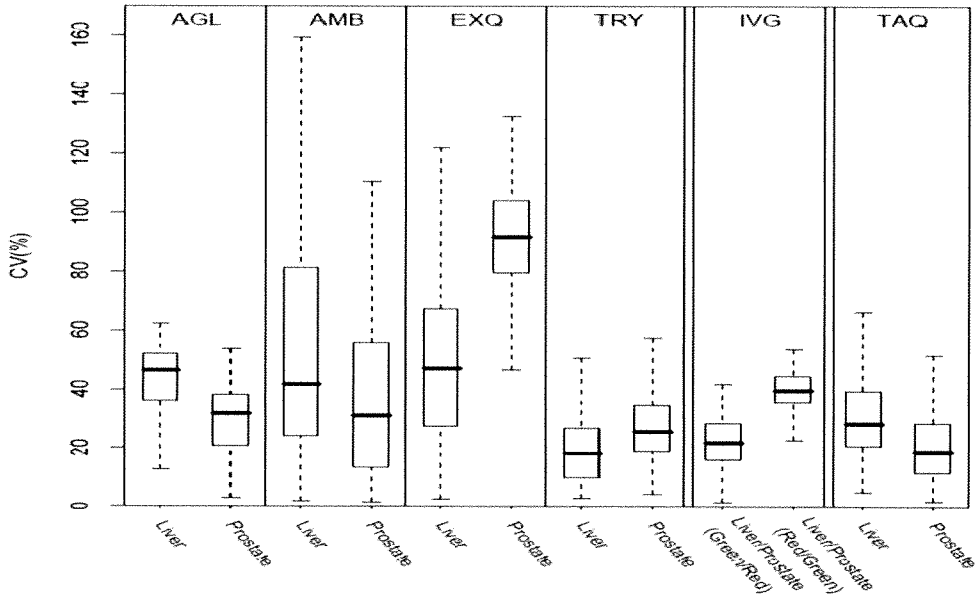


Figure 2. Intra-platform repeatability of quantitative assessment of microRNA expression. MicroRNA measurement of the same sample (L: human liver, P: human prostate) was replicated three times. Only data of microRNAs with positive detection call were used for analysis. **2A**: Scatter plots show the correlation between replicate 1 and 2 (scatter plots for all possible combinations between three replicates were shown in Figure S4). Spearman's correlation coefficients (Rs) for replicates 1 vs. 2, 1 vs. 3, and 2 vs. 3 were calculated and summarized in lower table. Rs for the prostate sample were generally better than those for the liver sample ($p = 0.0005$, paired T-test). This finding suggests that repeatability of microRNA would depend on the sample cell type, and that repeatability in the case of samples expressing A higher amount of microRNAs would be better. TAQ (Taqman analysis) obtained the best Rs values despite a slightly wider spread of data. It might be a result from wider range of microRNA detection (microarray: 2^{16}, Taqman: about 2^{20}). **2B**: Box plot of coefficient of variation (CV) for microRNA detection platforms. The coefficient of variation for each microRNA assessment was calculated by a formula, CV = (standard deviation/mean) ×100, and the distribution of CV was plotted in the box plot diagram. Bold line: median, bottom and top line of the box: first and third quantile, respectively.
doi:10.1371/journal.pone.0005540.g002

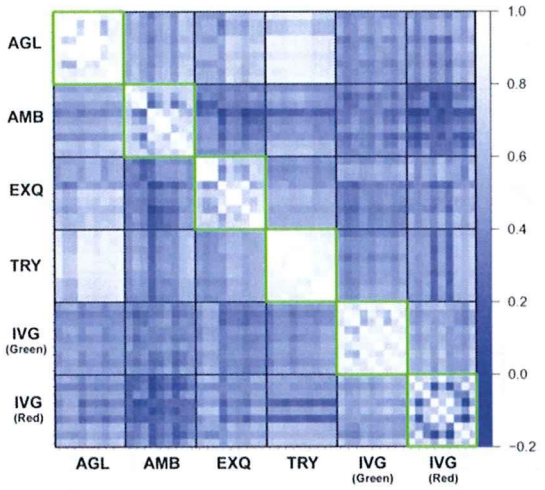


Figure 3. Rank correlation of log-ratios between intra- and inter-platform replications. For each platform, microRNA expression profiles in the liver and the prostate were measured three times by independent microarray chips. Therefore, 9 (=3×3) combinations of log-ratios (liver/prostate) for each microRNA was calculated. Then, 81 (=9×9) Spearman's correlation coefficients (Rs) values were calculated, and visualized in blue-white heat map. White indicates high correlation, whereas blue means low correlation. Heatmaps by Pearson's and Kendall's correlation coefficients were available in Figure S4.
doi:10.1371/journal.pone.0005540.g003

replicates of the same sample type was calculated on each platform (line graphs in Figure 4A). As expected, microarray signals from liver samples were generally weaker than those of prostate samples (Figure 1). Thus, the percent of detected microRNA subset in liver samples was significantly smaller than that in prostate samples (Figure 4A, paired T-test: $p = 0.0003$, and Wilcoxon's signed rank test: $p = 0.0005$). In the current study, we used criteria of detection call of microRNAs that the manufacturers recommended. However, the stringency of these detection call criteria was very different. For AMB and EXQ array, all microRNAs with positive signal were handled as detected microRNAs, whereas other manufacturers provided their own formula as detection call criteria. This difference in the detection call stringency may result in the divergence of detected microRNA percentage. Thus, detected microRNA percentage of AMB and EXQ array were less stable in three replicates (t-test and Mann-Whitney test of standard deviation, $p = 0.0011$ and 0.004, respectively) than the others.

Intra-platform concordance in detected microRNA list was shown inside of green squares in Figure 4B and 4C. It is reasonable that AMB and EXQ with instable percentage of detected microRNAs also had higher inconsistency in the detected microRNA list than the others. Intra-platform concordance in a list of differentially expressed microRNAs was illustrated inside of green squares in Figure 5. The means and 95%CIs of agreement percentages were listed in Table S3. AGL and TRY had more than 90% concordance of differentially expressed microRNAs list within intra-platform replicates.

Inter-platform data comparability

MicroRNA expression values generated on different platforms cannot be directly compared because unique labeling methods and probe sequences will result in variable signal distributions for probes that hybridize to the same target microRNAs. (Figure S1) Alternatively, the relative expression between a pair of sample types should be maintained across platforms. For this reason, we examined the microarray data for comparability between platforms by reviewing liver sample to prostate sample expression values with two different levels: rank correlation of the log-ratio as qualitative assessment, and the microRNA list agreement (detection call and identification of differentially expressed microRNAs) as qualitative assessment.

To show the inter-platform concordance in the detected microRNA list, the percentage of 309 microRNAs with concordant detection calls between replicates on different platforms was calculated and visualized outside of green squares in Figure 4B and 4C. The median percentages of inter-platform detection concordance were 74.0% and 72.1% for liver and prostate sample, respectively. There was no statistical difference in detection call concordance between liver and prostate samples. For both samples, these percentages were widely distributed, ranging 56.3–97.9% and 58.2–95.9%, respectively, because the difference in detection call stringency lead to a divergence in detection call rate across the platforms.

The comparability of results across the platforms was also examined using a rank correlation metric. For rank correlation, only detected microRNAs from the common 309 gene list were included in the analysis. Log-ratios for the differential expression observed between liver sample replicates and prostate sample replicates were calculated for the generally detected common microRNAs and then compared across the platforms. The rank correlations of the log-ratios are displayed visually in Figure 5A. Good agreement was not observed between the platforms, compared to the original MAQC report. In fact, the best correlation was obtained between AGL and TRY ($Rs = 0.8717$), and the median rank correlation was 0.55 between the microarray platforms.

For the list overlap of differentially expressed microRNAs, all 309 common genes were considered. A list of differentially expressed microRNAs was generated for each platform and compared to lists from the other platform. A percent score was calculated to indicate the number of microRNAs in common between each pair of platforms. The percentage of overlap for each comparison is displayed in Figure 5. Note the graphic comparisons are asymmetrical indicating the analysis is performed in two directions. That is, the percentage of platform Y microRNAs on the list from platform X can be different from the percentage of platform X microRNAs on the platform Y list. In contrast with one color platforms, IVG (two-color method) identified A much lower number of differentially expressed microRNAs, probably due to log-ratio compression (Figure 6). Therefore, percentages of list overlap between IVG and one-color platforms were generally low. AGL, EXQ and TRY had a good concordance in terms of identifying differentially expressed microRNAs.

Correlation to Taqman assay

In the MAQC project, the quantitative accuracy of several non-microarray devices was checked, then quantitative RT-PCR (Taqman system) was selected as a validation method of microarray data. In the microRNA research field, several different types of quantitative RT-PCR (qRT-PCR) methods are in use, such as qRT-PCR using stem-loop shaped RT-primer, Taqman system, Applied Biosystems) [18], qRT-PCR using locked nucleic acid primers (Exiqon) [19], and qRT-PCR with poly-A tailing (QIAGEN, Stratagene). In this study, we also used the Taqman

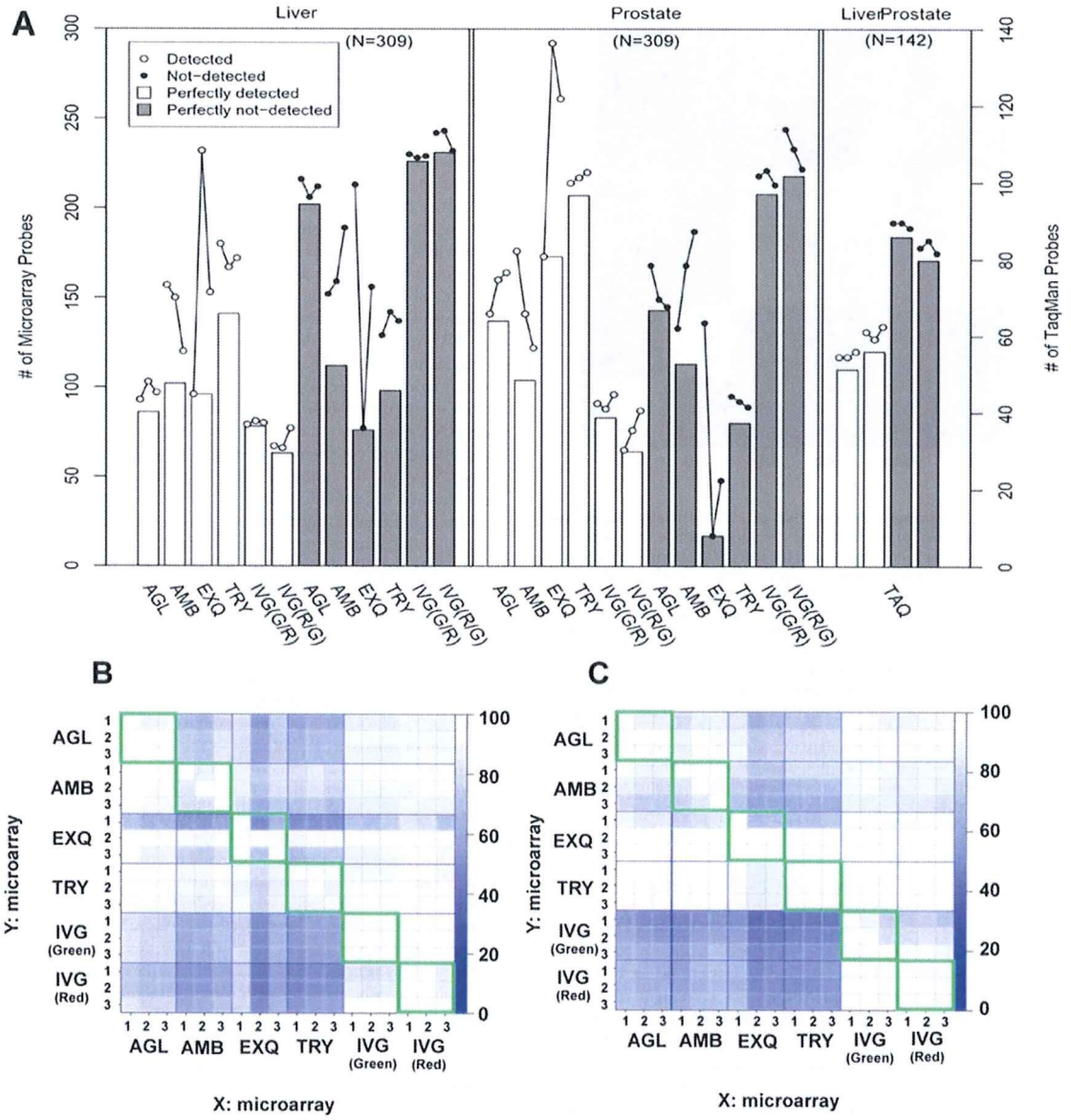


Figure 4. Repeatability and agreement of detection call. As a qualitative assessment of microRNA, A list of detected microRNAs should agree between different platforms. Detection call of microRNA for each platform was performed according to different criteria recommended by the manufacturer. **4A**: The number of detected microRNAs. Closed circles: detected microRNAs, open circles: not detected microRNAs, white bar: perfectly detected microRNAs, which were detected in all three replications, gray bar: perfectly not-detected microRNAs, which were not detected in all three replications. For the Taqman analysis, amplified microRNAs within 40 cycles were considered as detected. **4B** & **C**: Agreement rate of detection call list between intra- (inside of green squares) and inter-platform (outside of green squares) replications using liver (4B) and prostate (4C) samples. The percent agreement of detected microRNAs was calculated as the number of microRNAs detected by platform Y relative to the number of microRNAs detected by platform X. Therefore, two blocks in A diagonally symmetric position are not always the same color, because the denominators are different.
doi:10.1371/journal.pone.0005540.g004

microRNA assay system as a validation method, which is a method most widely used. Further comparisons between each microarray platform relative to the TaqMan assays are presented as scatter plots in Figure 6. One hundred forty two microRNAs were randomly selected from 309 common microRNAs to the microRNA platforms, then the expression levels of these 142 microRNA in the human liver and prostate were measured by Taqman system. Good correlation coefficients (Rs = 0.85,0.86)