

TLDAによる発現解析結果-2 (48時間-2)

48 hr

Non-genotoxic
Genotoxic

Gene	4h	48h	Total	Gene	4h	48h	Total
Phf23a3	30	46	76	Hamp2	-1	9	8
Ddit4l	32	22	54	Lpp	9	2	11
Col1a1a	14	42	56	Sdc	7	0	7
Oyp4a14	31	20	51	Sulf1d1	2	2	4
Cong1	20	30	50	Fabp5	3	1	4
Rtg2	24	19	43	Gstm3	-4	7	3
Ddit4	31	11	42	Ccrf	-1	2	1
Dscr1	26	13	39	Dscr1	1	0	1
Trp53inp1	16	18	34	Nono	1	0	1
Bag2	16	16	32	Cyp2b10	0	0	0
Acox	22	5	27	Evi1	0	0	0
Tubb2c	16	8	24	Stxip5	0	0	0
Jun	13	9	22	Hr23b	0	0	0
Mib2	2	18	20	Gsta2	-5	4	-1
Slx	6	13	19	Hhex	-3	0	-3
Mib1	19	0	19	Car3	-6	-2	-8
Cerm1d	12	4	16	Miox1	-14	0	-14
Lrp1	12	2	15	Cad44g	-9	0	-9
Lyla	0	14	14	Irfp1	-17	-3	-20
Phf4d1	14	0	14	Irfp1	-21	-6	-27
Mdm2	-1	12	11				
Cyp9l	4	6	10				
Egr	10	0	10				
Rhhb2	3	6	9				
Ephr1	0	9	9				

Genes in black; control, red; up-regulated blue; down-regulated in GeneChip

Distinguishable genes
No data
Increased
Decreased

方向性、変化率を考慮した各遺伝子の採点

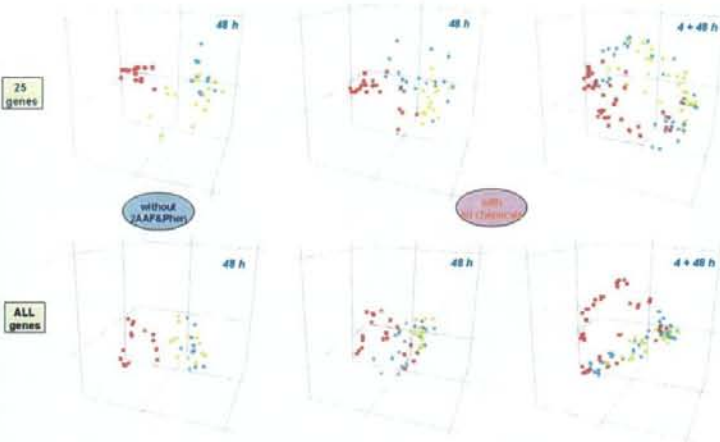
Gene	SCORE			Gene	SCORE		
	4h	48h	Total		4h	48h	Total
Phf23a3	30	46	76	Hamp2	-1	9	8
Ddit4l	32	22	54	Lpp	9	2	11
Col1a1a	14	42	56	Sdc	7	0	7
Oyp4a14	31	20	51	Sulf1d1	2	2	4
Cong1	20	30	50	Fabp5	3	1	4
Rtg2	24	19	43	Gstm3	-4	7	3
Ddit4	31	11	42	Ccrf	-1	2	1
Dscr1	26	13	39	Dscr1	1	0	1
Trp53inp1	16	18	34	Nono	1	0	1
Bag2	16	16	32	Cyp2b10	0	0	0
Acox	22	5	27	Evi1	0	0	0
Tubb2c	16	8	24	Stxip5	0	0	0
Jun	13	9	22	Hr23b	0	0	0
Mib2	2	18	20	Gsta2	-5	4	-1
Slx	6	13	19	Hhex	-3	0	-3
Mib1	19	0	19	Car3	-6	-2	-8
Cerm1d	12	4	16	Miox1	-14	0	-14
Lrp1	12	2	15	Cad44g	-9	0	-9
Lyla	0	14	14	Irfp1	-17	-3	-20
Phf4d1	14	0	14	Irfp1	-21	-6	-27
Mdm2	-1	12	11				
Cyp9l	4	6	10				
Egr	10	0	10				
Rhhb2	3	6	9				
Ephr1	0	9	9				

選択した遺伝子

Genes in red; up-regulated in GeneChip
Genes in blue; down-regulated in GeneChip

遺伝子発現データ(2回目)を用いた主成分分析

Genotoxic
Non genotoxic
Control



全遺伝子を用いた統計解析(t-testにて有意となる遺伝子数およびスコアリング値による判定)

	4h				48h				4h + 48h				
	non genotoxic	genotoxic	p		non genotoxic	genotoxic	p		non genotoxic	genotoxic	p		
	DEFP	DEFP	The	Fam	DEFP	DEFP	The	Fam	DEFP	DEFP	The	Fam	
遺伝子発現変動 LAD1発現	p0001	9	12	9	11	20	22	14	20	47	20	22	21
	p0002	7	10	1	18	18	19	13	13	23	18	13	23
	p0003	4	9	3	12	17	18	7	12	23	20	17	18
	p0004	3	8	2	9	17	16	6	21	12	20	17	18
	p0005	1	9	2	12	17	4	9	14	23	18	13	23
遺伝子発現変動 LAD2発現	p0006	8	9	12	13	18	9	11	20	47	20	22	21
	p0007	8	1	11	10	17	6	10	20	46	24	18	23
	p0008	9	8	9	10	18	8	14	20	45	18	17	20
	p0009	4	8	9	10	18	7	14	20	45	18	17	20
	p0010	1	9	2	10	18	7	14	20	45	18	17	20
遺伝子発現変動 LAD3発現	p0011	9	12	9	11	20	22	14	20	47	20	22	21
	p0012	8	11	9	10	17	15	13	18	46	18	17	20
	p0013	9	1	11	10	17	6	10	20	46	24	18	23
	p0014	9	8	9	10	18	8	14	20	45	18	17	20
	p0015	4	8	9	10	18	7	14	20	45	18	17	20
	p0016	1	9	2	10	18	7	14	20	45	18	17	20
遺伝子発現変動 LAD4発現	p0017	9	12	9	11	20	22	14	20	47	20	22	21
	p0018	8	11	9	10	17	15	13	18	46	18	17	20
	p0019	9	1	11	10	17	6	10	20	46	24	18	23
	p0020	9	8	9	10	18	8	14	20	45	18	17	20
	p0021	4	8	9	10	18	7	14	20	45	18	17	20
	p0022	1	9	2	10	18	7	14	20	45	18	17	20
遺伝子発現変動 LAD5発現	p0023	9	12	9	11	20	22	14	20	47	20	22	21
	p0024	8	11	9	10	17	15	13	18	46	18	17	20
	p0025	9	1	11	10	17	6	10	20	46	24	18	23
	p0026	9	8	9	10	18	8	14	20	45	18	17	20
	p0027	4	8	9	10	18	7	14	20	45	18	17	20
	p0028	1	9	2	10	18	7	14	20	45	18	17	20
遺伝子発現変動 LAD6発現	p0029	9	12	9	11	20	22	14	20	47	20	22	21
	p0030	8	11	9	10	17	15	13	18	46	18	17	20
	p0031	9	1	11	10	17	6	10	20	46	24	18	23
	p0032	9	8	9	10	18	8	14	20	45	18	17	20
	p0033	4	8	9	10	18	7	14	20	45	18	17	20
	p0034	1	9	2	10	18	7	14	20	45	18	17	20
遺伝子発現変動 LAD7発現	p0035	9	12	9	11	20	22	14	20	47	20	22	21
	p0036	8	11	9	10	17	15	13	18	46	18	17	20
	p0037	9	1	11	10	17	6	10	20	46	24	18	23
	p0038	9	8	9	10	18	8	14	20	45	18	17	20
	p0039	4	8	9	10	18	7	14	20	45	18	17	20
	p0040	1	9	2	10	18	7	14	20	45	18	17	20
遺伝子発現変動 LAD8発現	p0041	9	12	9	11	20	22	14	20	47	20	22	21
	p0042	8	11	9	10	17	15	13	18	46	18	17	20
	p0043	9	1	11	10	17	6	10	20	46	24	18	23
	p0044	9	8	9	10	18	8	14	20	45	18	17	20
	p0045	4	8	9	10	18	7	14	20	45	18	17	20
	p0046	1	9	2	10	18	7	14	20	45	18	17	20
遺伝子発現変動 LAD9発現	p0047	9	12	9	11	20	22	14	20	47	20	22	21
	p0048	8	11	9	10	17	15	13	18	46	18	17	20
	p0049	9	1	11	10	17	6	10	20	46	24	18	23
	p0050	9	8	9	10	18	8	14	20	45	18	17	20
	p0051	4	8	9	10	18	7	14	20	45	18	17	20
	p0052	1	9	2	10	18	7	14	20	45	18	17	20
遺伝子発現変動 LAD10発現	p0053	9	12	9	11	20	22	14	20	47	20	22	21
	p0054	8	11	9	10	17	15	13	18	46	18	17	20
	p0055	9	1	11	10	17	6	10	20	46	24	18	23
	p0056	9	8	9	10	18	8	14	20	45	18	17	20
	p0057	4	8	9	10	18	7	14	20	45	18	17	20
	p0058	1	9	2	10	18	7	14	20	45	18	17	20
遺伝子発現変動 LAD11発現	p0059	9	12	9	11	20	22	14	20	47	20	22	21
	p0060	8	11	9	10	17	15	13	18	46	18	17	20
	p0061	9	1	11	10	17	6	10	20	46	24	18	23
	p0062	9	8	9	10	18	8	14	20	45	18	17	20
	p0063	4	8	9	10	18	7	14	20	45	18	17	20
	p0064	1	9	2	10	18	7	14	20	45	18	17	20
遺伝子発現変動 LAD12発現	p0065	9	12	9	11	20	22	14	20	47	20	22	21
	p0066	8	11	9	10	17	15	13	18	46	18	17	20
	p0067	9	1	11	10	17	6	10	20	46	24	18	23
	p0068	9	8	9	10	18	8	14	20	45	18	17	20
	p0069	4	8	9	10	18	7	14	20	45	18	17	20
	p0070	1	9	2	10	18	7	14	20	45	18	17	20
遺伝子発現変動 LAD13発現	p0071	9	12	9	11	20	22	14	20	47	20	22	21
	p0072	8	11	9	10	17	15	13	18	46	18	17	20
	p0073	9	1	11	10	17	6	10	20	46	24	18	23
	p0074	9	8	9	10	18	8	14	20	45	18	17	20
	p0075	4	8	9	10	18	7	14	20	45	18	17	20
	p0076	1	9	2	10	18	7	14	20	45	18	17	20
遺伝子発現変動 LAD14発現	p0077	9	12	9	11	20	22	14	20	47	20	22	21
	p0078	8	11	9	10	17	15	13	18	46	18	17	20
	p0079	9	1	11	10	17	6	10	20	46	24	18	23
	p0080	9	8	9	10	18	8	14	20	45	18	17	20
	p0081	4	8	9	10	18	7	14	20	45	18	17	20
	p0082	1	9	2	10	18	7	14	20	45	18	17	20
遺伝子発現変動 LAD15発現	p0083	9	12	9	11	20	22	14	20	47	20	22	21
	p0084	8	11	9	10	17	15	13	18	46	18	17	20
	p0085	9	1	11	10	17	6	10	20	46	24	18	23
	p0086	9	8	9	10	18	8	14	20	45	18	17	20
	p0087	4	8	9	10	18	7	14	20	45	18	17	20
	p0088	1	9	2	10	18	7	14	20	45	18	17	20
遺伝子発現変動 LAD16発現	p0089	9	12	9	11	20	22	14	20	47	20	22	21
	p0090	8	11	9	10	17	15	13	18	46	18	17	20
	p0091	9	1	11	10	17	6	10	20	46	24	18	23
	p0092	9	8	9	10	18	8	14	20	45	18	17	20
	p0093	4	8	9	10	18	7	14	20	45	18	17	20
	p0094	1	9	2	10	18	7	14	20	45	18	17	20
遺伝子発現変動 LAD17発現	p0095	9	12	9	11	20	22	14	20	47	20	22	21
	p0096	8	11	9	10	17	15	13	18	46	18	17	20
	p0097	9	1	11	10	17	6	10	20	46	24	18	23
	p0098	9	8	9	10	18	8	14	20	45	18	17	20
	p0099	4	8	9	10	18	7	14	20	45	18	17	20
	p0100	1	9	2	10	18	7	14	20	45	18	17	20
遺伝子発現変動 LAD18発現	p0101	9	12	9	11								

選択した25遺伝子による統計解析結果

種	遺伝子				p	種	遺伝子				p	種	遺伝子				p	
	Gen1	Gen2	Gen3	Gen4			Gen1	Gen2	Gen3	Gen4			Gen1	Gen2	Gen3	Gen4		
新設セット L25遺伝子 Reference	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000
	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000
	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000
	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000
新設セット L25遺伝子 Reference	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000
	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000
	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000
	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000
新設セット L25遺伝子 Reference	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000
	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000
	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000
	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000
新設セット L25遺伝子 Reference	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000
	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000
	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000
	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000
新設セット L25遺伝子 Reference	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000
	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000
	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000
	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000
新設セット L25遺伝子 Reference	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000
	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000
	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000
	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000
新設セット L25遺伝子 Reference	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000
	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000
	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000
	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000
新設セット L25遺伝子 Reference	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000
	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000
	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000
	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000
新設セット L25遺伝子 Reference	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000
	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000
	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000
	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000
新設セット L25遺伝子 Reference	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000
	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000
	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000
	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000
新設セット L25遺伝子 Reference	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000
	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000
	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000
	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000
新設セット L25遺伝子 Reference	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000
	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000
	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000
	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000
新設セット L25遺伝子 Reference	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000
	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000
	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000
	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000
新設セット L25遺伝子 Reference	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000	Gen1	2	1	1	1	0.000
	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000	Gen2	1	1	1	1	0.000
	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000	Gen3	1	1	1	1	0.000
	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000	Gen4	1	1	1	1	0.000

結果-2のまとめ

- 新規セットによるバリデーションの結果、前回に比べ良いパフォーマンスが得られた。
- 前回同様、遺伝子の選択により判別性は向上した。
- 発現変化の方向性、変化率を考慮した統計処理により、遺伝子傷害性の有無を判別するためのアルゴリズムを確立できた。
- TLDAと平行して、SYBR Greenによる通常のRT-PCR法により、同様の結果が得られた。

考察

- GeneChipによる遺伝子の選択には、偶然のばらつきによる擬陽性遺伝子が含まれてていた可能性がある。
- 化合物が変われば、最適な遺伝子セットは変化する。
- 変化する遺伝子のみを選択しており、データの標準化の問題等で本来変化しないが逆の動きとして現れた遺伝子があると考えられる(セット1)。
- 確立した判別アルゴリズムに関しては、新規の化合物セットを用いて検証を行う必要がある。有効性が確認できれば、新たな遺伝毒性試験法として活用できる。

結論

- 標的遺伝子が絞りこまれた場合には、簡便、迅速かつ信頼性の高い手法としてTLDAを用いたRT-PCR解析法が有効であることが確認できた。
- バリデーションのため化合物セットを解析した結果、遺伝子傷害性肝発癌物質の判別を行うためのアルゴリズムを設定できた。
- 最終的には、反応性の良い25遺伝子を用いることで、パフォーマンスを向上できることがわかった。

厚生労働科学研究費補助金(化学物質リスク研究事業)
分担研究報告書

【トキシコゲノミクスのためのインフォマティクス開発】
非形質依存的インフォマティクス開発研究

研究分担者 相崎 健一

国立医薬品食品衛生研究所・安全性生物試験研究センター・毒性部・主任研究官

研究要旨

非形質依存的に Percellome データベースを有効活用するための基本技術開発を行っている。今年度は引き続きマイクロアレイの飽和現象およびクロスハイブリダイゼーション現象に対するデータ補正アルゴリズムの開発や、遺伝子の発現パターンによる並べ替え・抽出アルゴリズムの開発・改良を進めると共に、網羅的解析の実用化に即した Percellome データベースの最適化とナレッジデータ強化を実施した。

A. 研究目的

全遺伝子情報からなる大量データから生物学的に有意な情報を効率的に抽出する技術を生成し、安全性評価におけるトキシコゲノミクスの有機的活用を促進させる。特に、症状として現れる毒性反応自体の分子機序だけではなく、その毒性反応を引き起こす発端となった分子標的から最終的な毒性反応を誘導するまでの過程の解明を最終目標とした非形質依存的解析手法の開発を目指す。

B. 研究方法

我々は遺伝子発現データの絶対量化技術である Percellome 手法を考案し、これに最適化した”Millefeuille”(ミルフィーユ)データ処理システムを開発している。

開発・改良したアルゴリズムは、研究者本人がソフトウェアを作成し、実際に本研究計画

で得られた生データをサンプルに用いて、ワークステーション上で性能検証を行う。これは本研究におけるプログラム開発はアルゴリズムの改良・微調整を兼ねた try and error 方式で進めなければならない、職業プログラマに開発委託する際に必要となる仕様の事前決定が出来ないためである。開発言語には、研究的なプログラミングに適した、つまり生成したプログラムの実行速度が早いだけでなく、実行プログラムの生成(コンパイル速度)も速い Delphi 言語(=object pascal 言語, 開発環境としては Borland Delphi7 もしくは Delphi2006, Delphi2007)を使用した。また一般的なグラフ描画には TeeChart コンポーネント(Steema 社)を利用し、100 万件以下の小規模なデータベース操作には組み込み型リレーショナルデータベースコンポーネントである DBISAM (ElevateSoftware 社)を利用して、プログラム開

発の効率化を図った。

大規模計算が必要な場合は、専用計算サーバーシステム(MFサーバーシステム)にアプリケーションソフトウェアを移植して、大規模計算用グリッド PC や高速データベースエンジンである Teradata データベース(日本 Teradata 社)を活用して計算処理を行った。

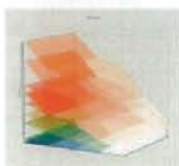
C. 研究結果

1. マイクロアレイのデータ補正

マイクロアレイデータの誤差の主要な原因の1つとして、以前からプローブの飽和による測定値の頭打ち現象を確認しているが、本年度の研究により、染色・洗浄前のハイブリダイゼーション過程におけるクロスハイブリダイゼーション影響が、多数のプローブの測定値に誤差(飽和現象も含む)を発生させている可能性が示唆された。現在、全プローブ間のクロスハイブリダイゼーション係数を求め、補正計算モデルを構築して、試験計算を行っている。

2. エクソンアレイの性能評価

Affymetrix エクソンアレイについては、H18年度に性能評価用の基礎データを採取し、エクソンレベルの定量性能に問題があることを見いだしている(下図)が、本年度はメーカーの協力の下、ジャンクションアレイ(2つのエクソンにまたがるプローブを搭載した技術開発用アレイ。非売品)なども用いて定量的数値化アルゴリズムの開発を検討した。



同じ遺伝子に所属するエクソンであるにも関わらず発現レベルが一致しない。

現在はまだエクソンアレイおよびジャンクショ

ンアレイを利用して基礎データを再採取した段階だが、1のExpressionアレイのデータ補正技術をエクソンアレイに拡張することで、大幅な性能改善が見込まれる。

3. 発現パターンの特徴抽出

遺伝子発現を3次元平面波で表現する手法(Surface グラフ)を元に、その凹凸の多寡を評価して単純なものから複雑なものへ並び替える RSort アルゴリズムを改良し、凹凸の少ない単純なパターンを示す遺伝子、つまり、より強力もしくは単純な発現制御を受ける遺伝子を網羅的に自動抽出できるように改良を加えた。

また独自に開発した変動レベル別の有意差指標 Fx Curve や、偶然、特徴的な発現パターンを呈するケースを除外するために変動 CV 値やt検定などの統計手法と組み合わせた結果、絞り込み効率が飛躍的に向上した。

またこれを応用して遺伝子発現波の出現時間を特定し、自動的に候補遺伝子を時系列順に並び替えるアルゴリズムも生成した。

4. 遺伝子クラスタの意味抽出

GO キーワード等の意味を持つノード(意味ノード)が配置された疑似空間に、複数の候補遺伝子群(遺伝子クラスタ)を意味ノードとの相関関係に応じた位置へプロットする技術、GoPlot の開発・改良を進めているが、今年度はさらに、定型的な意味スクリーニング処理を行うための可視化ツール、MFGoMap や関連性の高い意味キーワードを直接総覧するための可視化ツール、MFGoCloudを作成した。



MF GoMap

MF GoCloud

5. Percellome データベースの最適化とナレッジデータ強化

Percellome データベースに登録する基本解析データとして、倒立 Surface グラフを元にした tmf Clustering (独自に開発した発現変動パターンによる教師無しクラスタリングアルゴリズム。以下、倒立 tmf Clustering) データを追加採用した。原理的に遺伝子発現誘導波の検出に強い正立 Surface グラフを元にした tmf Clustering (以下、正立 tmf Clustering) と遺伝子発現抑制波の検出に強い倒立 tmf Clustering を併用することで、有意なパターンの見落とし回避が徹底された。

既に活用されている RSort と正立・倒立 tmf Clustering データの3つを併せて解析することで、化合物投与実験で生じた遺伝子発現パターンを効率よく、確実に網羅できるようになった。

また pukiwiki を利用して構築されている PercellomeWiki ナレッジデータベースも順調に稼働中で、登録データ数は Wiki ページ400、データファイル2000を超え、解析作業において日常的に活用されている。

D. 考察

1. マイクロアレイのデータ補正

今のところ飽和補正やクロスハイブリダイゼーション補正は完全ではなく、直接、数値比較を行えるレベルに達していないが、現状においても、化合物クラスタリングなど、飽和現象等による誤差の影響を受けやすい解析精度の向上には有効であると考えている。

飽和補正式の定数項を求めるにあたっては、基礎データが不足しているため、補正精度をさらに向上させるために LBM サンプルと同様の試料を用意し、定数項の取得数拡大を目指す。

2. エクソンアレイの性能評価

エクソンアレイの数値化アルゴリズムについては、基本的に 1 で Expression アレイ用に開発したデータ補正技術が応用できる見込みであるが、スプライシングバリエーションを考慮したプローブ選択が必須であり、これを実現する追加アルゴリズムの開発に着手した。

3. 発現パターンの特徴抽出

RSort の開発・改良は極めて順調に進み、既に日常的な解析作業に活用され、網羅的な遺伝子発現解析の効率向上に役立っている。

今後は RSort 計算結果のデータベース化を推し進め、串刺し検索機能や遺伝子の発現順序関係抽出機能の実現など、さらなる解析効率の向上を目指す。

4. 遺伝子クラスタの意味抽出

現状の MFGoMap や MFGoCloud は、直感的な情報把握にはかなりの慣れが必要で、実用的とはいえない状態にある。今後、表示形態やユーザーインターフェースの改良を進め、遺伝子クラスタの意味把握の作業効率向上を目指す。

また一連の研究・開発では、意味データのモデルとして GO を用いたが、遺伝子クラスタの意味抽出目的には情報量が不足していた。より効率的な意味抽出を行うためには、Percellome データベースにおける化合物別の発現情報など独自に遺伝子アノテーション情報を生成すると共に、既存の様々な遺伝子アノテーション情報を収集・集積し、複合的に利用する必要があるだろう。

5. Percellome データベースの最適化とナレッジデータ強化

取得済みの Percellome データは膨大で、データベース化はまだ完了していないが、作業は順調に進んでいる。

現在ナレッジデータベース機能は pukiwiki エンジンを利用しているが、既存の全データが登録されただけでも wiki ページ数が倍増すると予測されていることから、より大きなデータを取り扱うことの出来る wiki エンジン(例えば MediaWiki)などへの移行も検討する。

E. 結論

RSort の性能向上や、正立/倒立 tnf Clustering データの整備を進めた結果、世界有数の規模を誇る Percellome データベースから毒性に関わる遺伝子群を効率良く、しかも網羅的に抽出できる様になった。

抽出した候補遺伝子群の意味抽出については、基本アルゴリズム構築は進んだものの、活用の際に必須のアノテーションデータベースが不足しており、その構築や収集が急務である。

今後は遺伝子クラスタの意味抽出に重点を移し、毒性反応シグナルネットワーク研究を進める。

F. 研究発表

1. 論文発表

雑誌

Sanosaka T, Namihira M, Asano H, Kohyama J, Aisaki K, Igarashi K, Kanno J, Nakashima K. Identification of genes that restrict astrocyte differentiation of midgestational neural precursor cells. *Neuroscience*. 155:780-8, 2008.

相崎健一, 菅野 純 Percellome システム

細胞 40: 176-180 2008

2. 学会発表

種村健太郎, 五十嵐勝秀, 相崎健一, 北嶋 聡, 菅野 純 発生期ドーモイ酸暴露によるマウス脳微細構造異常と情動-認知行動障害 第 35 回日本トキシコロジー学会学術年会、2008 年 6 月 26-28 日、東京

種村健太郎, 五十嵐勝秀, 相崎健一, 北嶋 聡, 菅野 純 エストロゲン受容体 (α 型) ノックダウンマウスの神経行動解析 第 35 回日本トキシコロジー学会学術年会、2008 年 6 月 26-28 日、東京

北嶋 聡, 相崎健一, 五十嵐勝秀, 児玉幸夫, 高木篤也, 関田清司, 今井 清, 菅野 純 Percellome 手法を用いたフタル酸エステル DEHP とその活性代謝産物 MEHP の腎に及ぼす遺伝子発現変動の比較 第 35 回日本トキシコロジー学会学術年会、2008 年 6 月 26-28 日、東京

五十嵐勝秀, 小川幸男, 笠井辰也, 長野嘉介, 北嶋 聡, 相崎健一, 菅野 純 シックハウス指針値レベルの経気道暴露による遺伝子発現変化の Percellome 解析 第 35 回日本トキシコロジー学会学術年会、2008 年 6 月 26-28 日、東京

関根司, 吉田功, 相崎健一, 桐生悠史, 大草敏史, 菅野純, 岡安 勲 潰瘍性大腸炎 in vitro モデルにおける酪酸による p53 活性化機序の解明: CITED2 の関与 第 97 回日本病理学会総会、2008 年 5 月 15-17 日、金沢

Polouliakh N, Matsuoka Y, Ghosh S, Nock R,
Nielsen F, Kitajima S, Takagi A, Aisaki K.I,
Kanno J, Kitano H, Signaling network in
Mouse Embryonic Stem Cells The 9th
International Conference on Systems Biology,
Aug. 22-28, 2008, Sweden

G. 知的財産所有権の出願・登録情報

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

分担研究項目 非形質依存的インフォマティクス 開発研究

相 崎 健 一
国立医薬品食品衛生研究所
安全性生物試験研究センター
毒性部

2009.2.23.



目的

全遺伝子の発現変動情報からなる大量データから生物学的に有意な情報を効率的に抽出する技術を生成し、安全性評価におけるトキシコゲノミクスの有機的活用を促進させる。特に、症状として現れる毒性反応自体の分子機序だけではなく、その毒性反応を引き起こす発端となった分子標的から最終的な毒性反応を誘導するまでの過程の解明を最終目標とした非形質依存的な解析手法の開発を目指す。



H20年度の進捗

- マイクロアレイのデータ補正
- エクソンアレイの性能評価
- 発現パターンの特徴抽出
- 遺伝子クラスタの意味抽出
- Percellomeデータベースの最適化とナレッジデータ強化

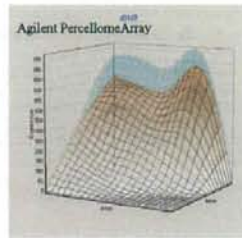
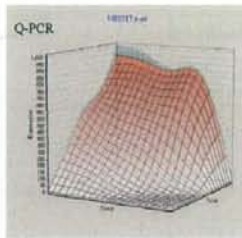
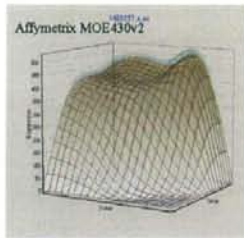
マイクロアレイデータの補正

Probe飽和の例

Affymetrix

Q-PCR

Agilent



AZA_ByArrayed
 * MOE430v2
 * 1422217_a_at
 * AgilentID
 * A_01_P279693
 * AgilentID44K
 * 42187

Cytlai

PercelemArray Project

5

Langmuirの方程式

- Irving Langmuirによって1918年に導出された理論的な吸着等温式である。以下のよう
な仮定を持っている。
 - 吸着媒には有限な数Nの吸着サイトがあり、そこだけで吸着質分子と結合する。
 - すべての吸着サイトは等価である。
 - 1つの吸着サイトは1つの吸着質分子としか結合しない。
 - 空の吸着サイトM、気相中の吸着質S、吸着サイトに結合した吸着質M-Sの間に
 $M + S \rightleftharpoons M-S$ の化学平衡が成立する。
- 各プローブが飽和している場合には、Langmuirの方程式に従うと仮定する。

$$\log(PM_{pjl}) = \log\left(I_p \frac{k_p c_j}{1 + k_p c_j} + bg_p\right) + \varepsilon_{pjl}$$

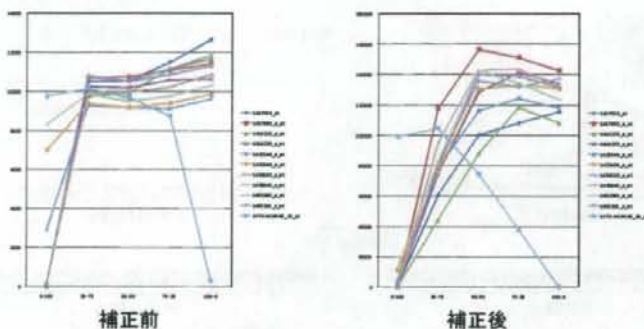
p = 1, ..., *P*: probe_index
j: concentration_index
c: concentration
l = 1, ..., *L*: replicate_index
I: Saturation_Intensity
k: equilibrium_constant
bg: background_component

PercelemArray Project

6

亜飽和データの補正

- 飽和補正結果(LBM: 単Langmuir(H19度実施))
- 高発現域(50%:50%で、10000以上)



>>>> 用量反応相関が改善している。

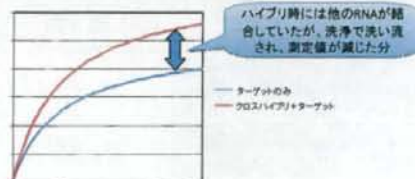


クロスハイブリダイゼーション影響

最大飽和値のバラツキが予想以上に大きい

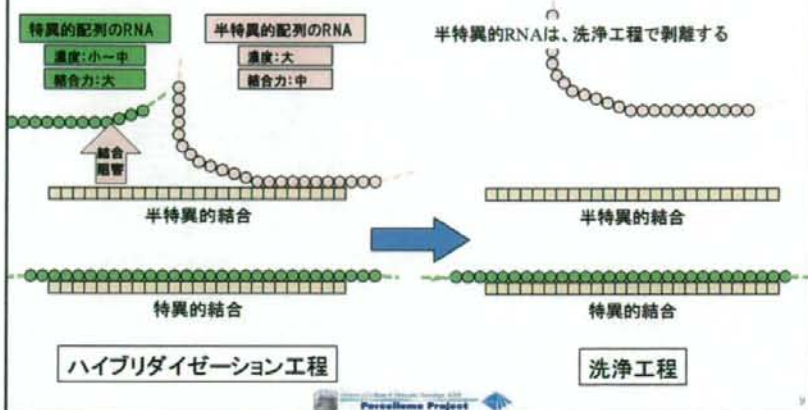
仮説

- ハイブリダイゼーション工程では、半特異的なRNAもプローブに結合しているが、洗浄工程で洗い流され、測定時には本来のターゲットRNAの結合量のみが測定される。
- プローブに、本来のターゲット以外のRNAが結合してしまうと、本来のターゲットRNAの結合量が減り、最大飽和値が見掛け上、小さく推定される。



クロスハイブリダイゼーションと洗浄除去による 見かけ上の最大飽和値の変化

半特異的結合 (Semi-specific Hybridization) という概念を導入し、モデルを構築する
半特異的な配列を持つRNAは、プローブと比較的強い力で結合するため、ハイブリダイゼーション工程では外れにくく、特異的な配列を持つRNA(ターゲットRNA)に置き換わらない。



Langmuir吸着等温式の多溶質への拡張

多溶質による吸着のモデル化
Langmuir吸着等温式と同じ仮定を設定

吸着速度は、吸着体の残り個数と溶質の濃度に比例すると仮定する

$$V' = K \cdot \left(N - \sum_{i=1}^n N_i \right) \left(\sum_{i=1}^n C_i \right)$$

個々の溶質の吸着速度は、溶質の濃度の比に比例する

$$V'_i = \frac{C_i}{\sum_{j=1}^n C_j} V'$$

個々の溶質の離脱速度は、吸着した個数に比例する

$$V_i = \alpha_i \cdot N_i$$

式変形

プローブに吸着したターゲットの個数

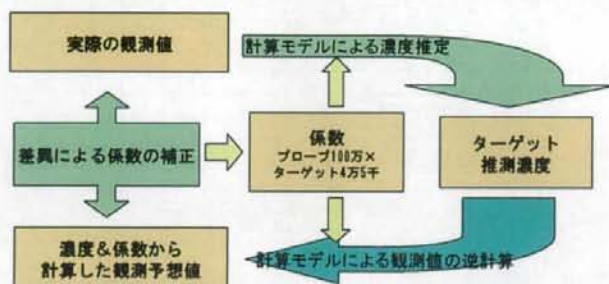
$$N_y = \frac{C_y}{\alpha_y} \cdot K \cdot \left(N - \sum_{i=1}^n N_i \right) + \alpha_y$$

プローブに吸着したターゲットの離脱速度係数
吸着速度係数は1とし、 α で調整する

$$\alpha_{yy} \geq 0$$

塩基配列に基づいたプローブ・ターゲット結合力係数の算出

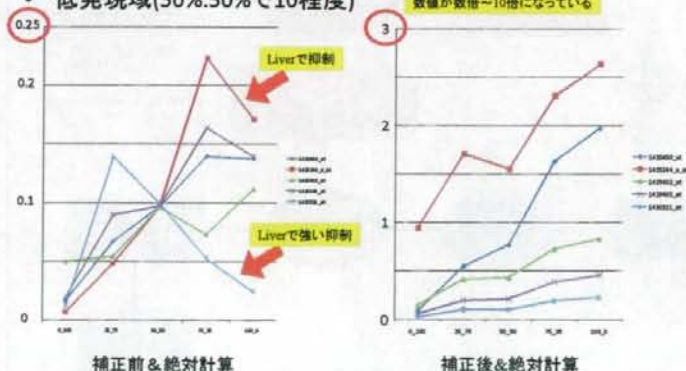
プローブとターゲットのペアの塩基配列に基づいた結合力の算出は、未知の要素(ターゲット設計配列範囲外での結合など)が多いため困難である。そこで観測値から仮に結合力係数の推測値を求め、この推定値を使用した場合の観測予想値を計算し、実際の観測値との誤差で係数を補正する方法を試みた。



・様々な実験条件のデータを用いて、本過程を繰り返し実施することで本来あるべき係数に近づいていく。
・ただし計算に時間がかかるうえ、計算が収束しないケースもありうる。

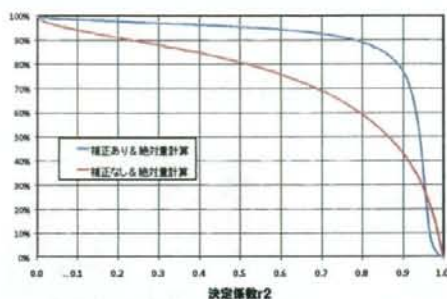
クロスハイブリダイゼーションの影響を受けたデータの補正

- ・ 飽和補正結果(LBM: 多Langmuir&絶対量化)
- ・ 低発現域(50%:50%で10程度)



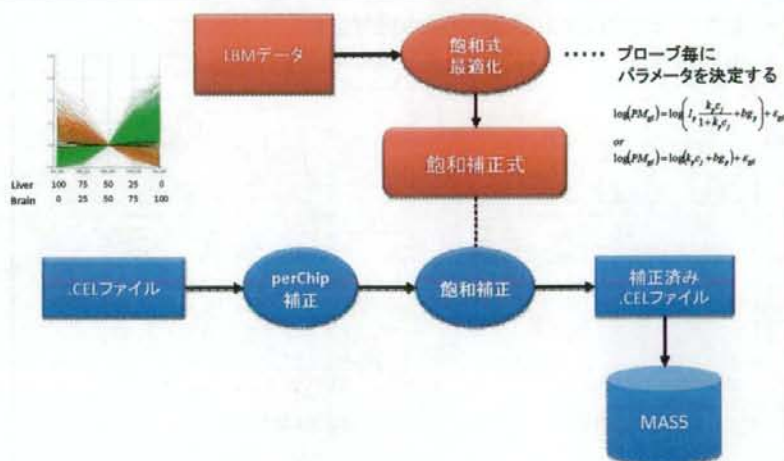
クロスハイブリダイゼーション補正による効果

- 補正によるLBMデータの直線性向上

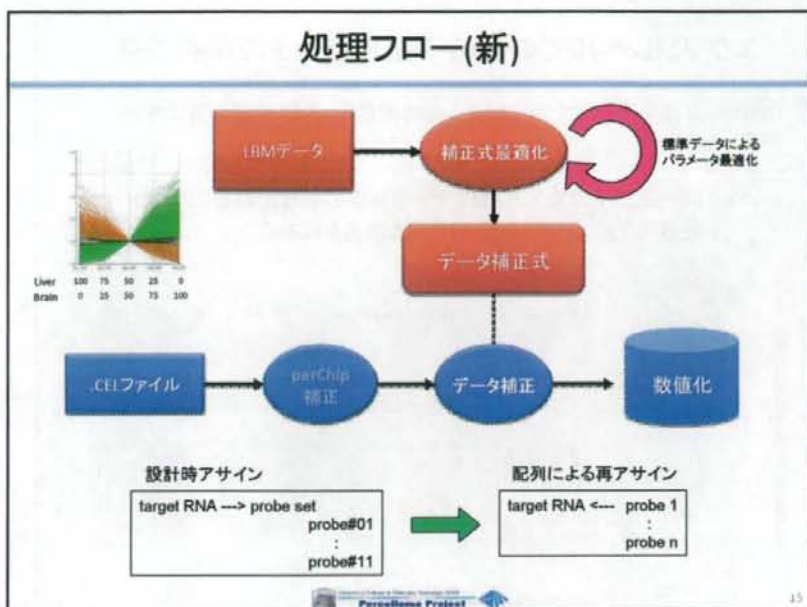


- 補正により直線性が格段に改善している
- 決定係数1付近が少ないのは、control RNA値を用いた前処理の影響が考えられる

処理フロー(旧)



処理フロー(新)

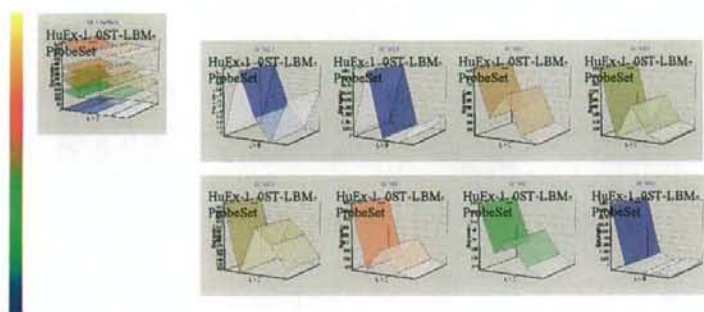


エクソンアレイの性能評価

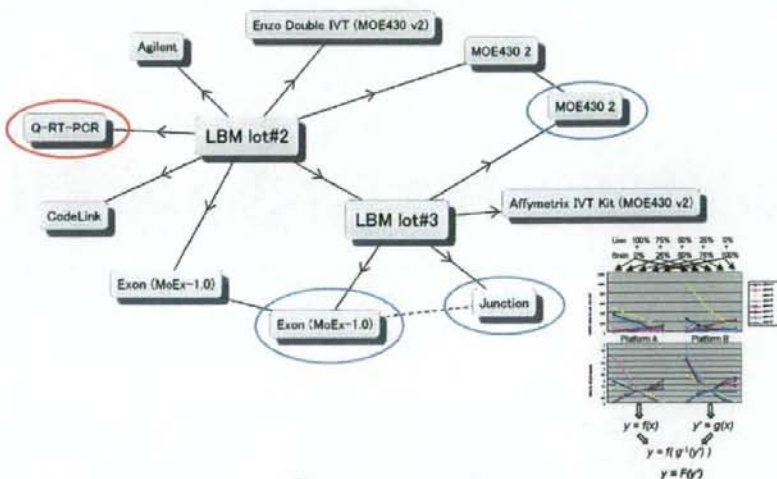
エクソンレベルでの遺伝子発現データの解析技術

LBMサンプルを用いてプローブセット毎の定量性・再現性等の基本性能を評価した。

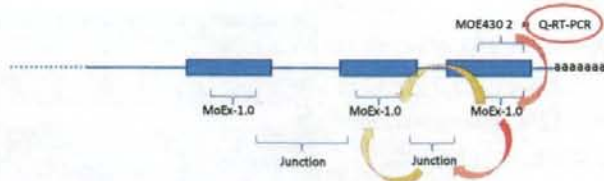
>>>> 同一遺伝子のエクソン間でデータ数値の乖離が顕著であった。
数値化アルゴリズムの改良による対処を試みる。



エクソンアレイと3'アレイの間でデータ相互変換を実現するための基本データの取得



エクソンプローブセット毎のデータの評価方針



- >>> (1) エクソンアレイにおいて、3'末端エクソンの理想値として、MOE430 2の測定値を用い、補正モデルを作成する。
(2) 隣接するJunctionプローブやエクソンプローブセットの理想値を(1)から求め、補正式を最適化する。
(3) 塩基配列の類似性でプローブをグループ分け(クラスタリング)し、バックグラウンド値やクロスハイブリダイゼーション影響、飽和影響を推定し、補正を試みる。
(4) 3'アレイ用のデータ補正アルゴリズムをスプライシングバリエーションの存在等を加味して拡張し、エクソンアレイデータの補正アルゴリズムを作成する。

発現パターンの特徴抽出