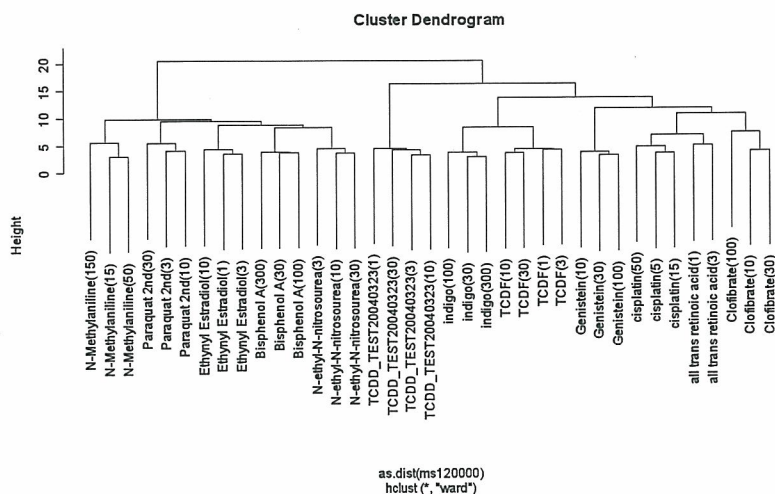


2. 8. 距離計算の式を再構成 (Dose=0をキャンセル)



クラスタリング結果 (E=120000, ward法)



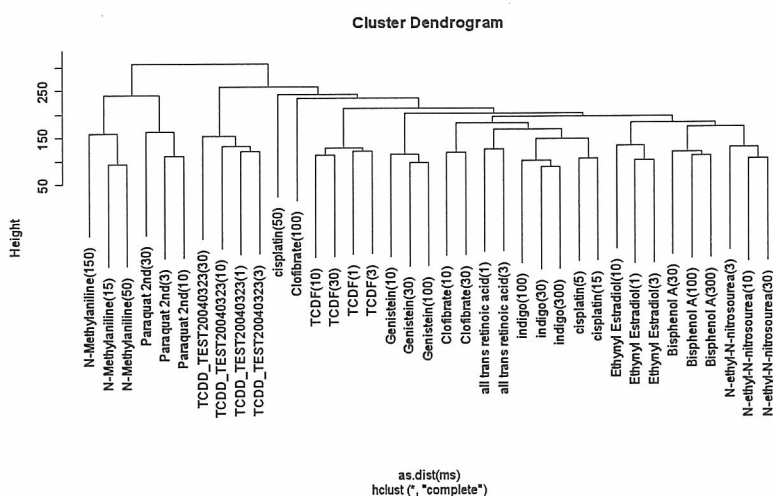
TCDDとTCDFが離れるなど、より既知の知見と異なる結果となった

Copyright © NTT COMWARE 2007

2. 8. 距離計算の式を再構成 (Dose=0をキャンセル)



クラスタリング結果 (E=12, Complete法)



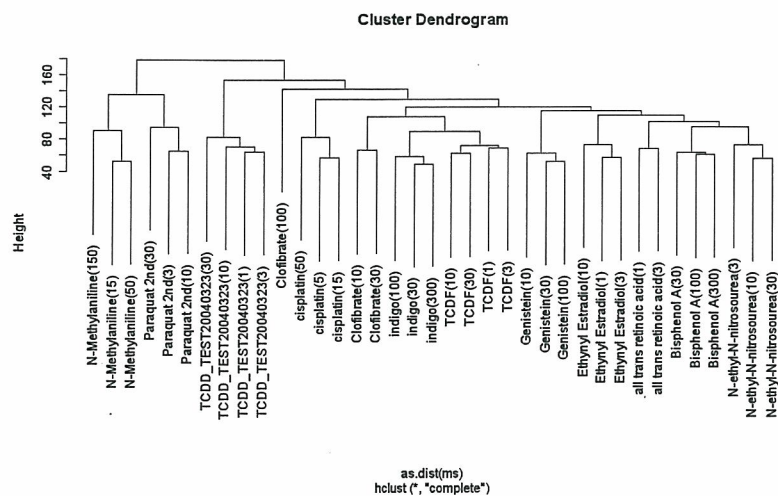
TCDDとTCDFが離れるなど、より既知の知見と異なる結果となった

Copyright © NTT COMWARE 2007

2. 8. 距離計算の式を再構成 (Dose=0をキャンセル)



クラスタリング結果 (E=120, Complete法)



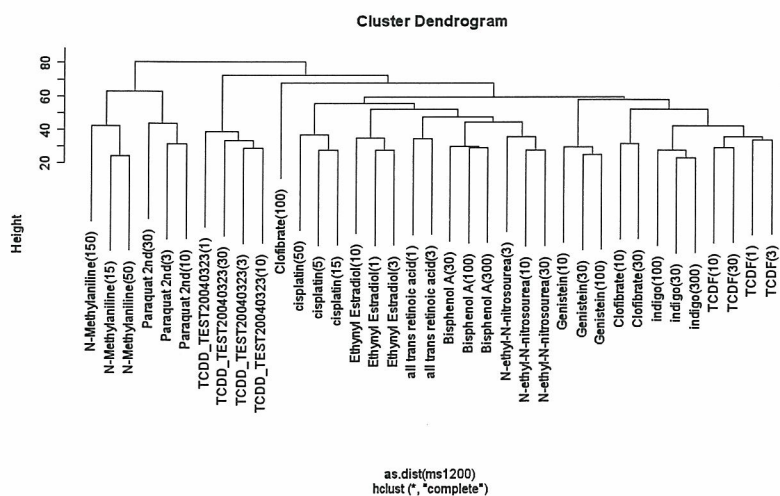
TCDDとTCDFが離れるなど、より既知の知見と異なる結果となった

Copyright © NTT COMWARE 2007

2. 8. 距離計算の式を再構成 (Dose=0をキャンセル)



クラスタリング結果 (E=1200, Complete法)



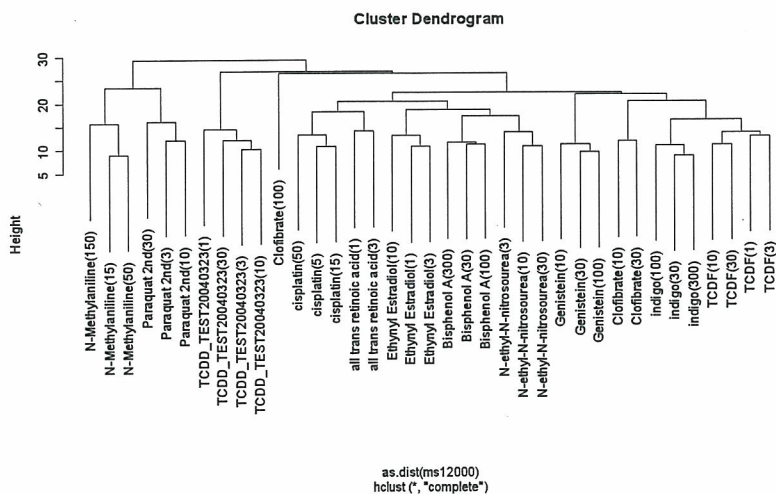
TCDDとTCDFが離れるなど、より既知の知見と異なる結果となった

Copyright © NTT COMWARE 2007

2. 8. 距離計算の式を再構成 (Dose=0をキャンセル)



クラスタリング結果 (E=12000, Complete法)



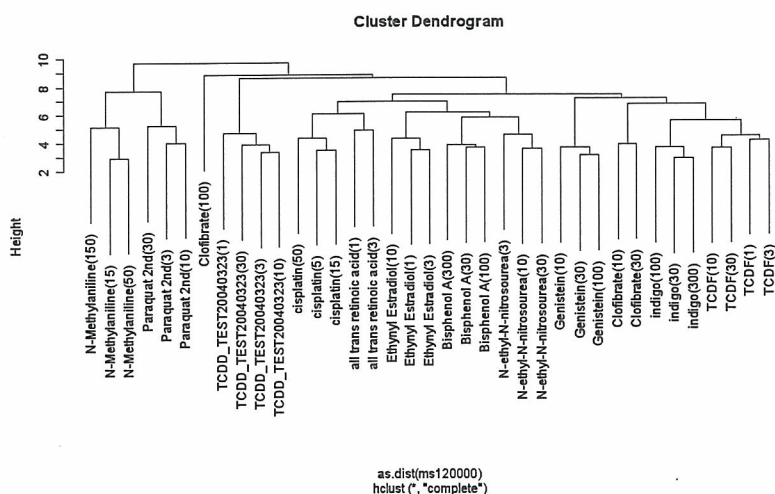
TCDDとTCDFが離れるなど、より既知の知見と異なる結果となった

Copyright © NTT COMWARE 2007

2. 8. 距離計算の式を再構成 (Dose=0をキャンセル)



クラスタリング結果 (E=120000, Complete法)



TCDDとTCDFが離れるなど、より既知の知見と異なる結果となった

Copyright © NTT COMWARE 2007

2. 8. 距離計算の式を再構成 (Dose=0をキャンセル)



- まとめ
 - 特性の似ていると考えられる化合物が離れてしまった
 - TCDD vs. TCDF
- 調査方法
 - 原因を作り出していると考えられる遺伝子を特定
 - Dose0間の距離とProject内の距離の平均値を算出して、Dose0間の距離が大きな遺伝子を特定する

Copyright © NTT COMWARE 2007

2. 9. Dose0内距離と化合物内距離の詳細



- Dose0内距離と化合物内距離の差の上位30プローブ

No	prb no	Prb ID	Gene Symbol	化合物内距離の平均	Dose0内距離の平均	差	比
1	45092	AFFX-r2-P1-cre-5 at	?	0.7783	3.8072	3.0289	4.8917
2	45091	AFFX-r2-P1-cre-3 at	?	0.7890	3.4568	2.6677	4.3810
3	19908	1435602 at	Sepha2	0.6564	2.6911	2.0347	4.0998
4	45051	AFFX-CreX-3 at	?	0.7354	2.7305	1.9950	3.7127
5	45052	AFFX-CreX-5 at	?	0.7567	2.5536	1.7969	3.3747
6	12266	1427960 at	Ust2b34	0.6065	2.3571	1.7506	3.8865
7	7563	1423257 at	Cyp4a14	1.1469	2.7055	1.5585	2.3589
8	751	1416420 a at	Rol9	0.5690	2.0797	1.5108	3.6551
9	40164	1455869 at	Gamk2b	0.8039	2.3060	1.5021	2.8685
10	9159	1424853 s at	Cyp4a10 /// BC013476	1.0650	2.5424	1.4773	2.3971
11	8579	1424273 at	Cyp2c70	0.8237	2.2872	1.4634	2.7523
12	8391	1424085 at	Ndufa4	0.4928	1.8545	1.3617	3.7634
13	45041	AFFX-b-ActinMur/M12	Actb	0.6499	1.9799	1.3300	3.0484
14	7155	1422849 a at	Fabp1	0.3355	1.6484	1.3129	4.9135
15	45082	AFFX-r2-Bs-thr-5 s at	?	0.5136	1.6260	1.1124	3.5551
16	4771	1420455 s at	Mup1 /// Mup2	0.8958	2.1875	1.2917	2.4420
17	33064	1448764 a at	Fabp1	0.8701	2.1538	1.2837	2.4753
18	45094	AFFX-ThrX-5 at	?	0.4738	1.7407	1.2668	3.6735
19	38903	1454608 x at	Ttr	0.6910	1.9215	1.2305	2.7809
20	35808	1451513 x at	Serpina1b	0.6411	1.8577	1.2166	2.8978
21	24633	1440327 at	Al195470	0.6659	1.8705	1.2046	2.8091
22	1748	1417417 a at	Cox6a1	0.5118	1.7141	1.2023	3.3489
23	32980	1448680 at	Serpina1a	0.7456	1.9438	1.1982	2.6070
24	20997	1436691 x at	Ptdx1	0.5552	1.7627	1.1975	3.1189
25	10853	1426547 at	Gc	0.6523	1.8379	1.1856	2.8175
26	8172	1423866 at	Serpina3k	0.7069	1.8797	1.1728	2.6592
27	3563	1419232 a at	Apoa1	0.7270	1.8872	1.1602	2.5958
28	228	1415897 a at	Mest1	0.6517	1.7955	1.1438	2.7551
29	1892	1417561 at	Apoa1	0.6855	1.8285	1.1430	2.6675
30	8250	1423944 at	Hoxn	0.6583	1.7975	1.1392	2.7305

AFFXで始まるものが多い。

Copyright © NTT COMWARE 2007

2.9. Dose0内距離と化合物内距離の詳細



- Dose0内距離と化合物内距離の差の下位30プローブ

No	prb no	Prb ID	Gene Symbol	化合物内距離の平均	Dose0内距離の平均	差	比
1	6523	1422217 a at	Cyp1a1	2.1254	0.0531	-2.0723	0.0250
2	35010	1450715 at	Cyp1a2	1.8326	1.4844	-0.3482	0.8100
3	35781	1451486 at	1200006F02Rik	0.5021	0.2493	-0.2528	0.4965
4	874	1416545 at	Nfe2l2	0.5765	0.3272	-0.2494	0.5675
5	11027	1426721 s at	Tiparp	0.3231	0.1251	-0.1981	0.3871
6	36109	1451814 a at	Httat2	0.4954	0.3035	-0.1919	0.6127
7	36455	1452160 at	Tiparp	0.2281	0.0601	-0.1681	0.2634
8	1604	1417273 at	Pdk4	0.1943	0.0630	-0.1313	0.3242
9	3480	1419149 at	Serpine1	0.1339	0.0130	-0.1209	0.0970
10	36456	1452161 at	Tiparp	0.1504	0.0559	-0.0946	0.3714
11	639	1416308 at	Ugdh	0.7161	0.6332	-0.0830	0.8942
12	11432	1427126 at	Hspa1b	0.3644	0.3024	-0.0620	0.8299
13	943	1416612 at	Cyp1b1	0.0890	0.0288	-0.0599	0.3268
14	11433	1427127 x at	Hspa1b	0.4175	0.3618	-0.0556	0.8567
15	37208	1452913 at	Pcp4l1	0.0751	0.0362	-0.0389	0.5089
16	33825	1449525 at	Fmo3	0.0778	0.0438	-0.0341	0.5621
17	19183	1434877 at	Nptx1	0.0420	0.0163	-0.0258	0.3869
18	19765	1435459 at	Fmo2	0.0823	0.0577	-0.0245	0.7018
19	36613	1452318 a at	Hspa1b	0.5120	0.4892	-0.0228	0.9555
20	12612	1428306 at	Ddit4	0.2030	0.1812	-0.0217	0.8929
21	3849	1419518 at	Tuba8	0.0621	0.0445	-0.0176	0.7161
22	10143	1425837 a at	Corn4l	0.2819	0.2651	-0.0168	0.9404
23	10254	1425948 a at	Slc25a30	0.1089	0.0936	-0.0153	0.8595
24	944	1416613 at	Cyp1b1	0.0386	0.0265	-0.0121	0.6854
25	15086	1430780 a at	Pmm1	0.0205	0.0085	-0.0120	0.4156
26	3267	1418936 at	Maff	0.0476	0.0357	-0.0120	0.7491
27	12790	1428484 at	Osbpl3	0.0341	0.0222	-0.0119	0.6500
28	42641	1458346 x at	2810442122Rik	0.0626	0.0512	-0.0114	0.8177
29	8473	1424167 a at	Pmm1	0.0336	0.0223	-0.0113	0.6626
30	24516	1440210 at	Gacn2	0.0299	0.0209	-0.0090	0.7002

既知の遺伝子Cyp1a1,Cyp1a2が上位に挙がっている。しかし、その後ろに続く遺伝子がない

Copyright © NTT COMWARE 2007

2.9. Dose0内距離と化合物内距離の詳細



- Dose0内距離と化合物内距離の比の上位30プローブ(参考)

No	prb no	Prb ID	Gene Symbol	化合物内距離の平均	Dose0内距離の平均	差	比
1	39096	1454791 a at	Rbbp4	0.1096	0.5423	0.4326	4.9455
2	7155	1422849 a at	Pabpn1	0.3355	1.6484	1.3129	4.9135
3	45092	AFFX-r2-P1-cre-5 at	?	0.7783	3.8072	3.0289	4.8917
4	23872	1439566 at	Q030038J10Rik	0.0419	0.2027	0.1608	4.8428
5	34702	1450407 a at	Anp32a	0.1888	0.9142	0.7254	4.8415
6	2519	1418188 a at	Ramp2	0.0565	0.2679	0.2114	4.7392
7	13183	1428877 at	Srp72	0.0943	0.4301	0.3358	4.5617
8	45091	AFFX-r2-P1-cre-3 at	?	0.7890	3.4568	2.6677	4.3810
9	23717	1439411 a at	Xpo7	0.1071	0.4525	0.3454	4.2249
10	21052	1436746 at	Prkwnk1	0.0384	0.1596	0.1213	4.1617
11	19908	1435602 at	Seph2	0.6564	2.6911	2.0347	4.0998
12	3587	1419256 at	Spnb2	0.0354	0.1419	0.1065	4.0133
13	22513	1438207 at	Gbf1	0.0332	0.1327	0.0996	4.0033
14	22709	1438403 s at	Ramp2	0.0919	0.3666	0.2747	3.9895
15	37109	1452814 at	Cpna3	0.0362	0.1442	0.1080	3.9794
16	2520	1418189 s at	Ramp2	0.0941	0.3731	0.2790	3.9653
17	31635	1447329 at	?	0.1591	0.6309	0.4717	3.9640
18	23711	1438405 x at	1700051C09Rik	0.2120	0.8330	0.6210	3.9299
19	13145	1428839 at	Wdr53	0.0368	0.1439	0.1070	3.9054
20	12266	1427960 at	Ugt2b34	0.6065	2.3571	1.7506	3.8865
21	18584	1434278 at	Mtrn1	0.1326	0.5143	0.3817	3.8795
22	21258	1436952 at	Kir9	0.0537	0.2081	0.1544	3.8732
23	45089	AFFX-PvruCarbMur/L0	Pcv	0.0713	0.2755	0.2042	3.8648
24	832	1416501 at	Fdrk1	0.0309	0.1186	0.0878	3.8440
25	5520	1421214 at	Cmah	0.1068	0.4052	0.2985	3.7957
26	30014	1445708 x at	3110021A11Rik	0.1304	0.4943	0.3638	3.7892
27	7446	1423140 at	Lip1	0.3092	1.1673	0.8582	3.7755
28	44929	1460637 s at	Pfdn5	0.1128	0.4245	0.3117	3.7638
29	8391	1424085 at	Ndufa4	0.4928	1.8545	1.3617	3.7634
30	21089	1436783 x at	Ywhab	0.0966	0.3615	0.2649	3.7407

Copyright © NTT COMWARE 2007

NTT COMWARE

- Dose0内で差が大きい



Copyright © NTT COMWARE 2007

NTT COMWARE

Cluster Dendrogram



Copyright © NTT COMWARE 2007

NTT COMWARE

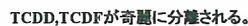
E=1200 上位1000プローブセット Ward法



Copyright © NTT COMWARE 2007

MIT COMWARE

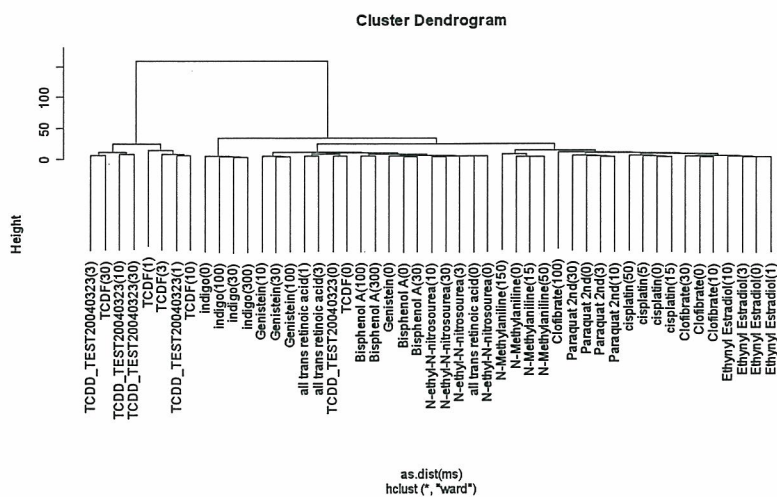
E=1200 上位10000プローブセット Ward法



Copyright © NTT COMWARE 2007

MIT COMWARE

E=1200 上位30000プローブセット Ward法

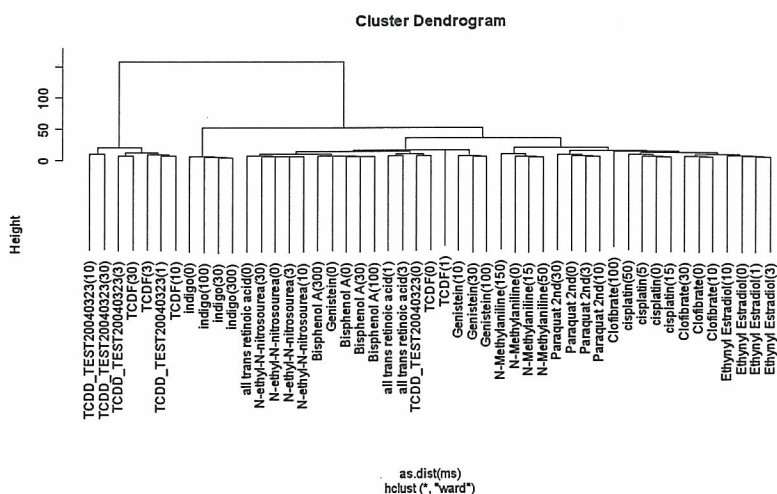


TCDD, TCDFが奇数に分離される。

Copyright © NTT COMWARE 2007

NII COMWARE

E=1200 上位35000プローブセット Ward法



TCDD,TCDFが奇麗に分離される。

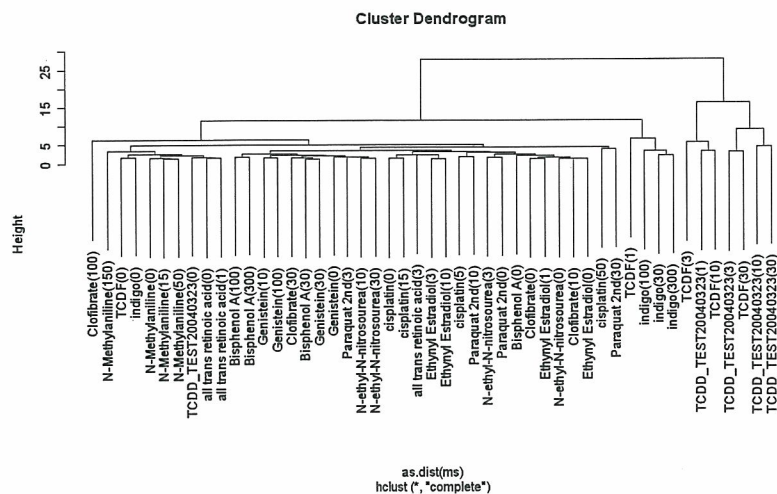
Copyright © NTT COMWARE 2007

2. 9. Dose0内距離と化合物内距離の詳細



Dose内距離の大きな遺伝子を用いたクラスタリング

E=1200 上位100プローブセット Complete法



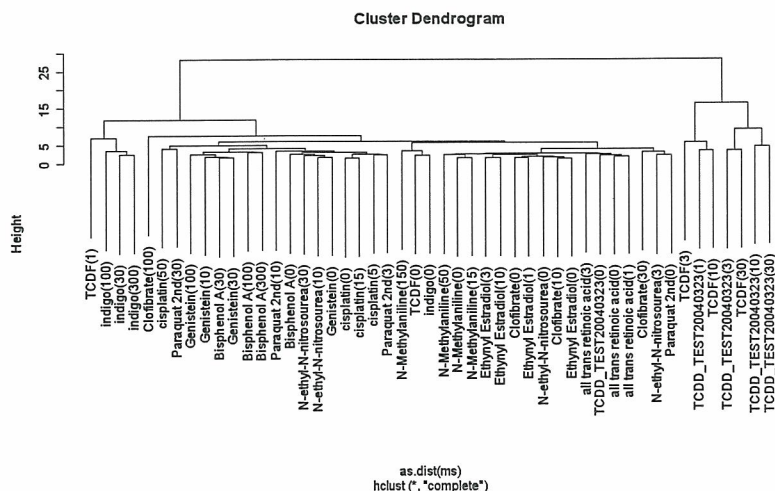
Copyright © NTT COMWARE 2007

2. 9. Dose0内距離と化合物内距離の詳細



Dose内距離の大きな遺伝子を用いたクラスタリング

E=1200 上位1000プローブセット Complete法



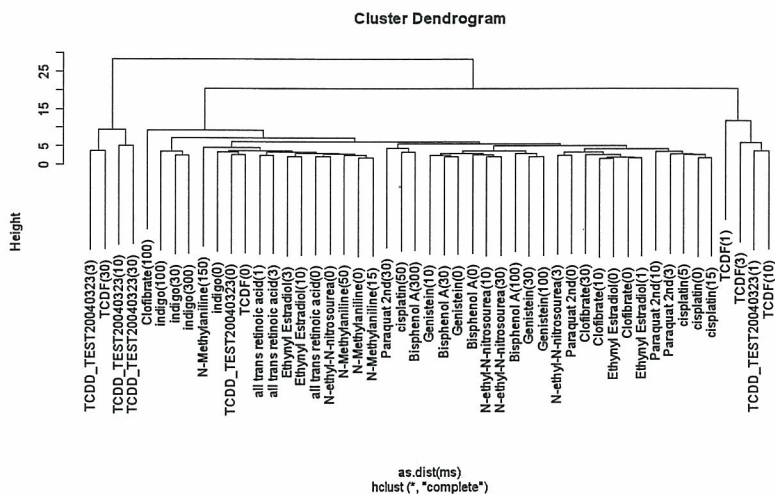
Copyright © NTT COMWARE 2007

2. 9. Dose0内距離と化合物内距離の詳細



Dose内距離の大きな遺伝子を用いたクラスタリング

E=1200 上位10000プローブセット Complete法



TCDD,TCDFが奇麗に分離される。ただし、Complete法の欠点からか、薄れている。

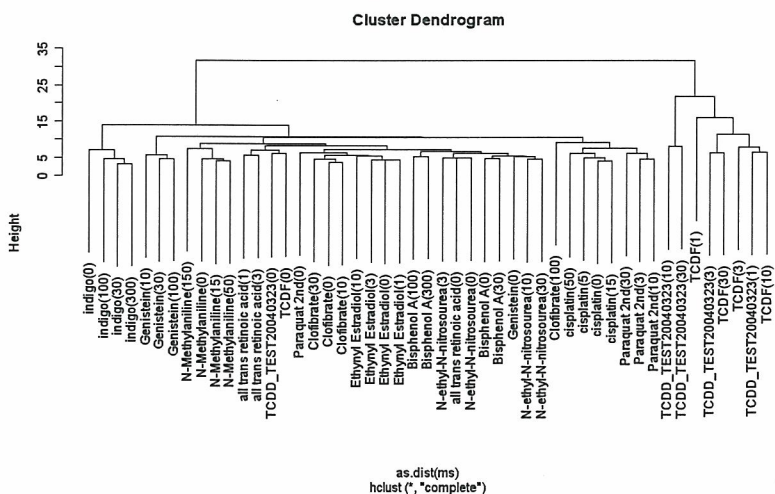
Copyright © NTT COMWARE 2007

2. 9. Dose0内距離と化合物内距離の詳細



Dose内距離の大きな遺伝子を用いたクラスタリング

E=1200 上位30000プローブセット Complete法



TCDD,TCDFが奇麗に分離される。ただし、Complete法の欠点からか、薄れている。

Copyright © NTT COMWARE 2007

MIT COMWARE

E=1200 上位35000プローブセット Complete法



Copyright © NTT COMWARE 2007

MIT COMWARE

Dose0キャンセル E=1200 上位100プローブセット Ward法

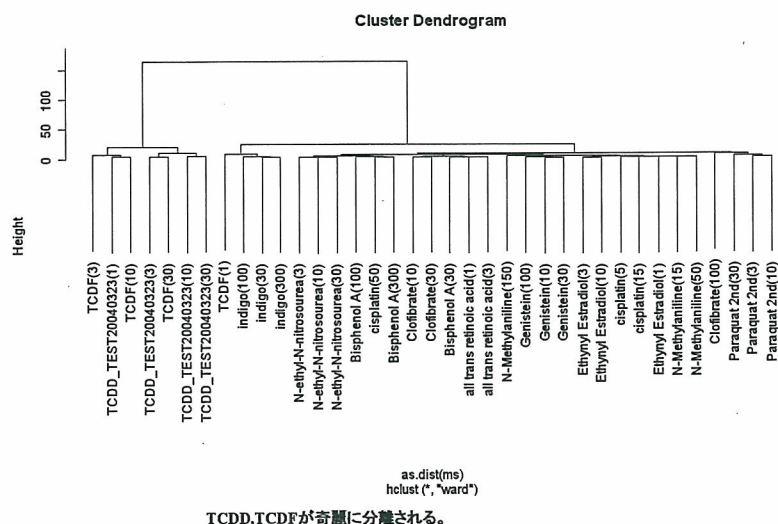


Copyright © NTT COMWARE 2007

2. 9. Dose0内距離と化合物内距離の詳細



Dose内距離の大きな遺伝子を用いたクラスタリング
Dose0キャンセル E=1200 上位1000プローブセット Ward法

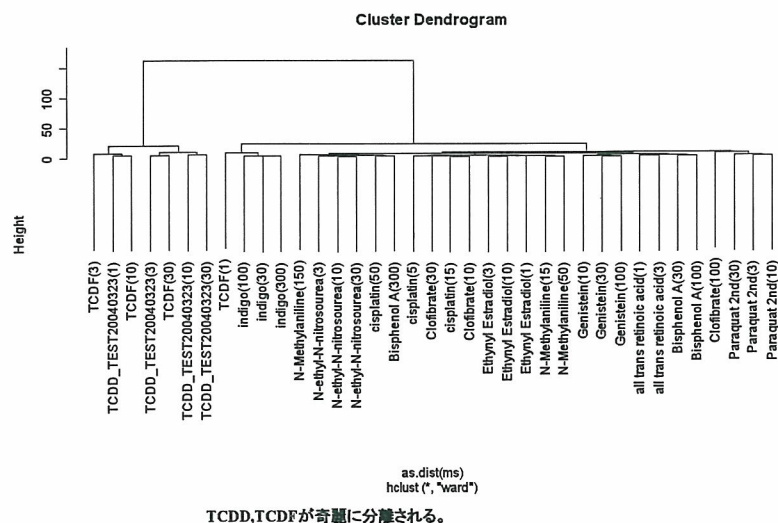


Copyright © NTT COMWARE 2007

2. 9. Dose0内距離と化合物内距離の詳細



Dose内距離の大きな遺伝子を用いたクラスタリング
Dose0キャンセル E=1200 上位10000プローブセット Ward法

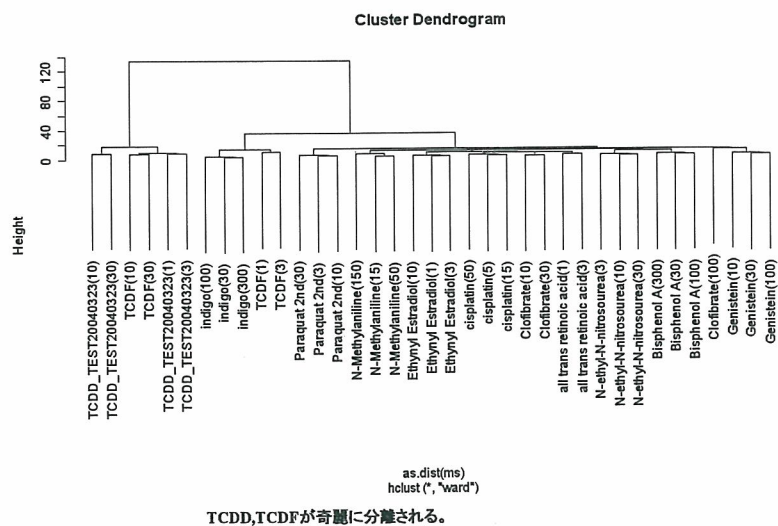


Copyright © NTT COMWARE 2007

2. 9. Dose0内距離と化合物内距離の詳細



Dose内距離の大きな遺伝子を用いたクラスタリング
Dose0キャンセル E=1200 上位30000プローブセット Ward法

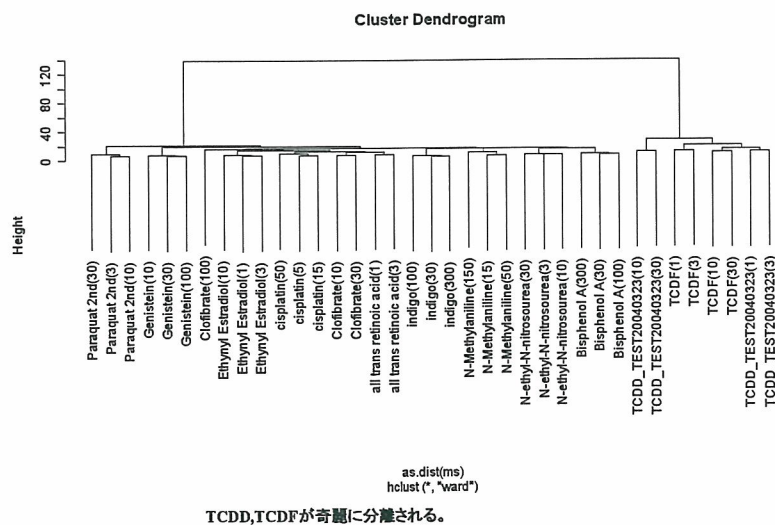


Copyright © NTT COMWARE 2007

2. 9. Dose0内距離と化合物内距離の詳細



Dose内距離の大きな遺伝子を用いたクラスタリング
Dose0キャンセル E=1200 上位35000プローブセット Ward法

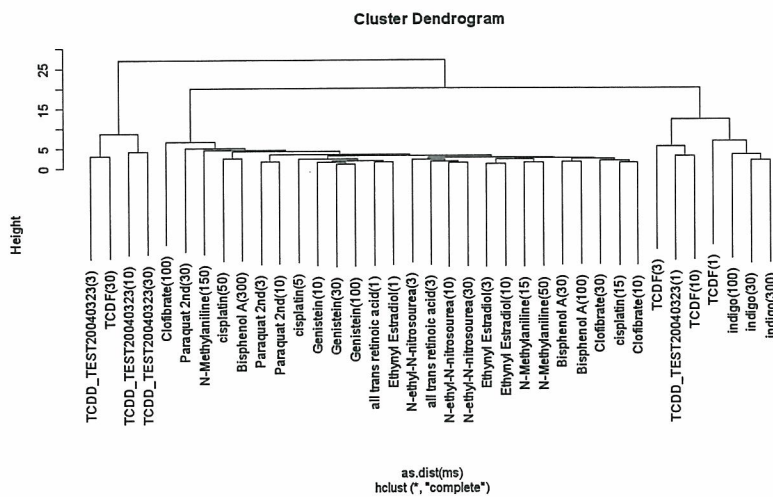


Copyright © NTT COMWARE 2007

2. 9. Dose0内距離と化合物内距離の詳細



Dose内距離の大きな遺伝子を用いたクラスタリング
Dose0キャンセル E=1200 上位100プローブセット Complete法



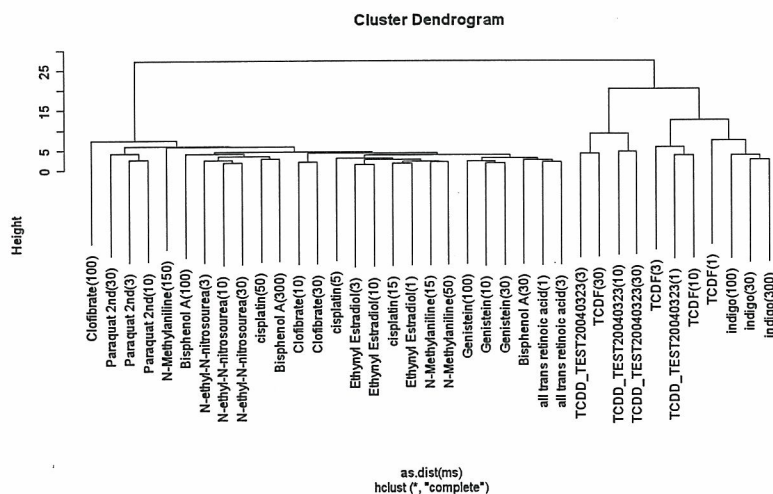
TCDD,TCDFが他と分離される。しかし、Complete法の特徴か、indigoも入ってくる。

Copyright © NTT COMWARE 2007

2. 9. Dose0内距離と化合物内距離の詳細



Dose内距離の大きな遺伝子を用いたクラスタリング
Dose0キャンセル E=1200 上位1000プローブセット Complete法



TCDD,TCDFが他と分離される。しかし、Complete法の特徴か、indigoも入ってくる。

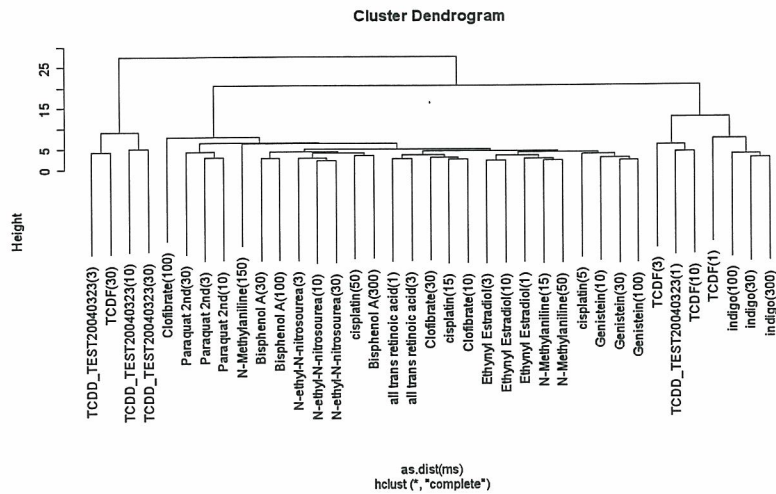
Copyright © NTT COMWARE 2007

2.9. Dose0内距離と化合物内距離の詳細



Dose内距離の大きな遺伝子を用いたクラスタリング

Dose0キャンセル E=1200 上位10000プローブセット Complete法



TCDD,TCDFが他と分離される。しかし、Complete法の特徴が、indigoも入ってくる。

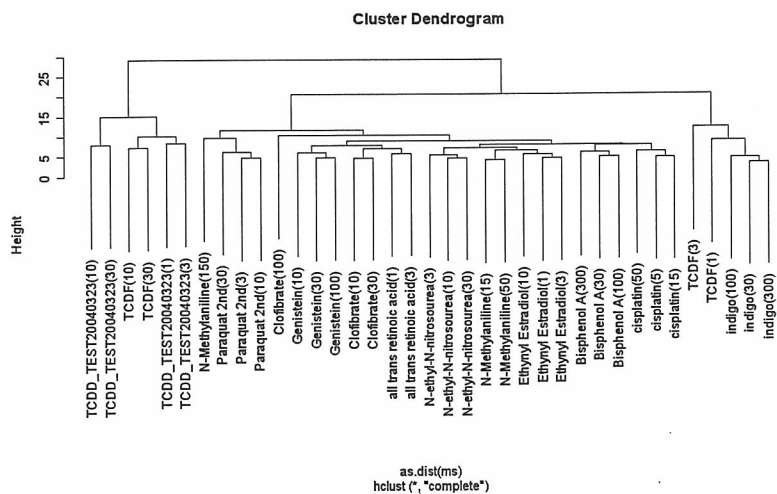
Copyright © NTT COMWARE 2007

2.9. Dose0内距離と化合物内距離の詳細



Dose内距離の大きな遺伝子を用いたクラスタリング

Dose0キャンセル E=1200 上位30000プローブセット Complete法



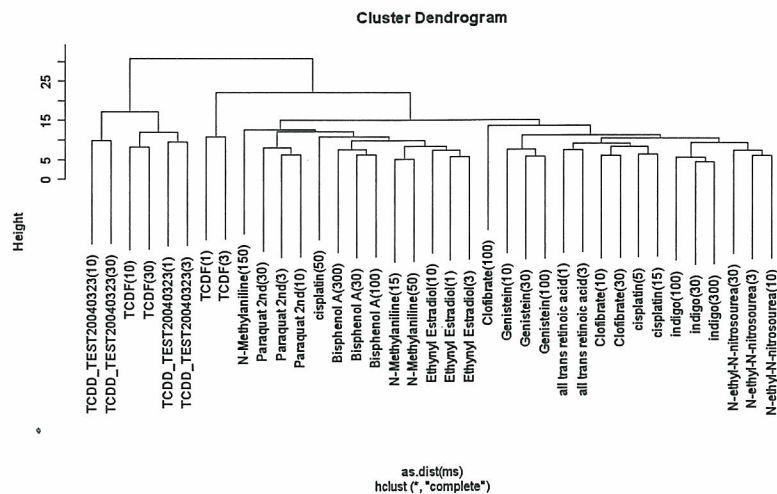
TCDD,TCDFが他と分離される。しかし、Complete法の特徴が、indigoも入ってくる。

Copyright © NTT COMWARE 2007

2. 9. Dose0内距離と化合物内距離の詳細



Dose内距離の大きな遺伝子を用いたクラスタリング
Dose0キャンセル E=1200 上位35000プローブセット Complete法



Complete法の特徴か、あまり既知の情報と一致しない。

Copyright © NTT COMWARE 2007

2. 9. Dose0内距離と化合物内距離の詳細



- 化合物の影響(プロジェクト内の差)よりも、プロジェクト間の影響の方が強い遺伝子が多い
 - 発現値の小さな遺伝子は、観測誤差消去項で影響を打ち消されている。発現値の大きな遺伝子でプロジェクト間の影響がある。
 - 遺伝子チップのプローブが飽和に近い状況になっていることが考えられる。

Copyright © NTT COMWARE 2007

2.10. クラスタリング対象実験（対象追加）



・ クラスタリング対象実験(対象追加)

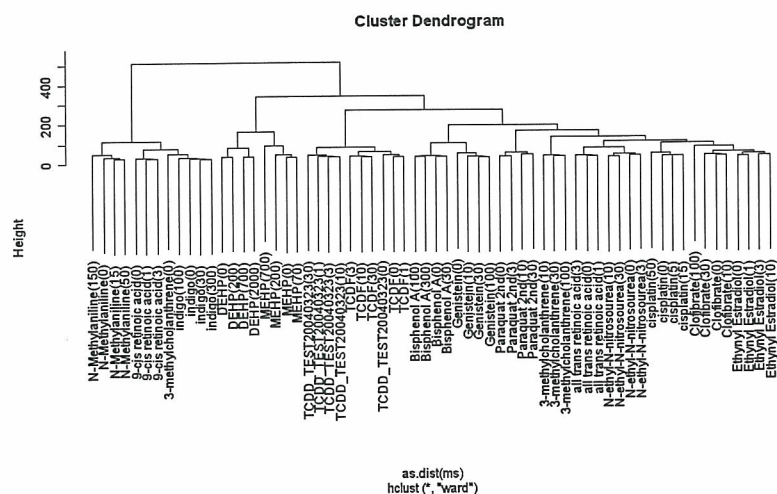
	化合物	溶媒	餌種類	実験日	Ord No	Prj No
TTG030-L	N-Methylaniline	MC	STD	2003/12/18	1	12
TTG040-L	cisplatin	corn oil + 10%DMSO	STD	2004/3/18	2	17
TTG044-L	Clofibrate	MC + 0.1%DMSO	STD	2004/5/13	3	33
TTG052-L	all trans retinoic acid	MC	STD	2004/9/9	4	50
TTG055-L	N-ethyl-N-nitrosourea	MC	STD	2004/7/15	5	42
TTG061-L	Paraquat 2nd	MC	STD	2005/2/2	6	67
TTG020-L	TCDD_TEST20040323	corn oil	STD	2003/10/16	7	1
TTG026-L	TCDF	corn oil	STD	2003/11/13	8	2
TTG042-L	Ethynyl Estradiol	MC + 0.1%DMSO	STD	2004/4/22	9	22
TTG047-L	Bisphenol A	MC + 0.1%DMSO	PLD	2004/6/17	10	38
TTG048-L	Genistein	MC + 0.1%DMSO	PLD	2004/7/1	11	40
TTG057-L	indigo	MC	STD	2004/9/30	12	54
TTG051-L	9-cis retinoic acid	MC + 0.1% DMSO		2004/9/2	13	48
TTG056-L	3-methylcholanthrene	Corn oil		2004/9/16	14	52
TTG098-L	DEHP	MC + 0.1% DMSO		2006/5/9	15	107
TTG104-L	MEHP	MC + 0.1% DMSO		2006/5/16	16	108

Copyright © NTT COMWARE 2007

2.10. クラスタリング対象実験（対象追加）



・ クラスタリング結果



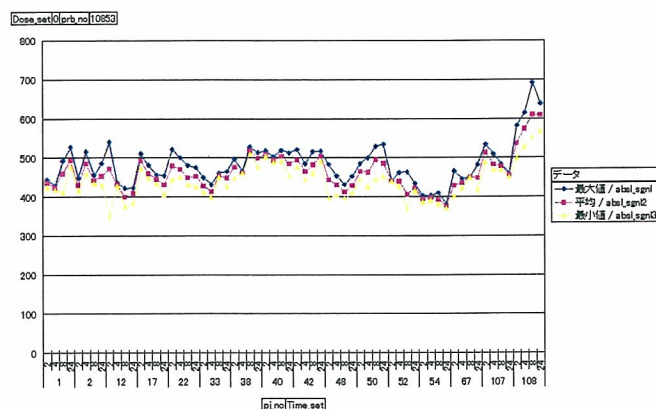
追加した実験は、階層構造の上位で分解されている。これらの実験もVehicleがその集団に入っている

Copyright © NTT COMWARE 2007

2.11. 化合物間クラスタリングの課題



- 化合物間 (Dose0同士) の距離を決定付けている遺伝子群の中からGc (Prb_No=10853) をグラフ化してみた。発現量が大きく、飽和していると考えられる



このようなプローブセットは排除する必要がある

Copyright © NTT COMWARE 2007

2. まとめ



- 課題
 - 遺伝子チップのプローブが飽和に近い状況になっていることが考えられる。飽和に近い状況に関して、絶対値変換の際の対策が必要である。
- 本クラスタリングの可能性
 - 今回は、階層型クラスタリングを行った。前段階として、全化合物同士の距離を計算している。これを用いることにより、MADICなどの距離を使用したクラスタリング手法を用いることも可能である。
 - また、距離計算の前段階として各遺伝子間の距離を計算している。これを用いることにより、化合物間の距離に対する各遺伝子の寄与を計算することができ、遺伝子に順位をつけることができる。逆に遺伝子リストを与えることで、飽和等の影響を排除した距離を計算することもできる。

Copyright © NTT COMWARE 2007

3. 化合物間クラスタリング（クラスタリング結果の山脈構造を用いた方法）

Copyright © NTT COMWARE 2007

127

3. 化合物間クラスタリング（山脈構造）

◆同期率計算の山脈構造を用いる。ただし、計算量が多い

各遺伝子で

化合物aのクラスタリング結果を山脈化

化合物bのクラスタリング結果を山脈化



課題

1. クラスタリング対象外の取扱
2. ユニークとなった遺伝子の影響度
3. 計算量

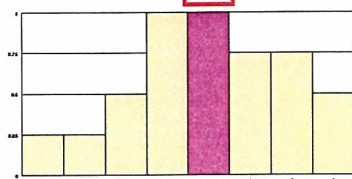
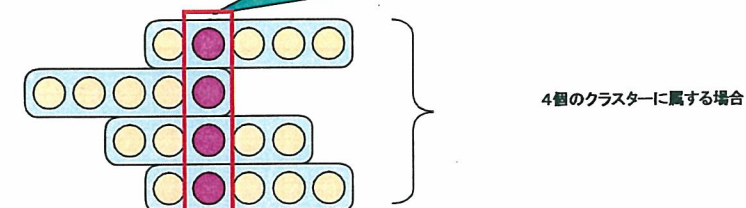
Copyright © NTT COMWARE 2007

3.1. 同期率計算の山脈作成方法



着目する遺伝子に関わるクラスターを全部

着目遺伝子



属する個数の割合から、
山脈が描ける

この山脈の形状が似ていれば、近くの遺伝子の存在状況が似ているといえる

Copyright © NTT COMWARE 2007

3.2. クラスタリング対象実験



pj_no	prj_name	chemical	sfc_no	sim_no	clst_no	山脈定義用データ数
12	TTG030-L	N-Methylaniline	2	1	1	99,957,189
17	TTG040-L	cisplatin	1	2	1	49,920,803
33	TTG044-L	Clofibrate	1	1	1	6,386,957
42	TTG055-L	N-ethyl-N-nitrosourea	1	1	1	9,802,013
50	TTG052-L	all trans retinoic acid	1	1	1	6,418,489
67	TTG061-L	Paraquat 2nd	1	1	1	54,690,275

Copyright © NTT COMWARE 2007