

## Dose0:使用プローブセットとツリー構造

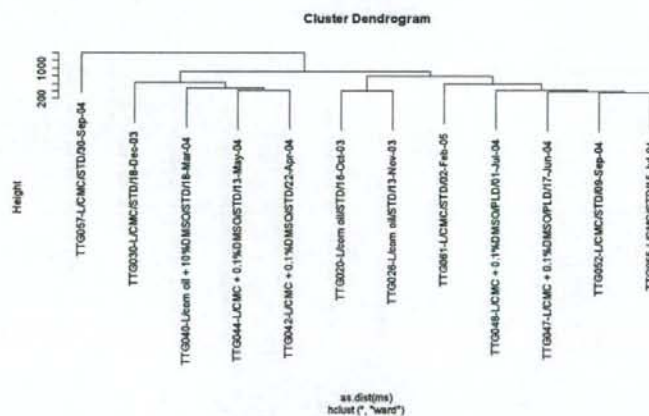
Dose0:距離マトリクス(発現量が何れかの時刻で細胞当たり3コピー以上の遺伝子)

対象プローブ数: 22,122

	ord	pl	nd	1	2	3	4	5	6	7	8	9	10	11	12
				12	17	33	50	42	67	1	2	22	38	40	54
TTG030-L	1	12	0.0	555.7	618.4	516.1	636.6	740.7	498.7	651.5	509.8	663.0	661.2	825.3	
TTG040-L	2	17	555.7	0.0	473.8	424.9	596.8	654.7	533.8	662.7	449.5	587.9	597.6	782.5	
TTG044-L	3	33	618.4	473.8	0.0	471.8	595.6	713.7	637.1	742.6	421.6	580.2	627.9	809.4	
TTG052-L	4	50	516.1	424.9	471.8	0.0	347.5	480.4	380.9	462.6	376.1	370.1	365.5	897.3	
TTG055-L	5	42	636.6	596.8	595.6	347.5	0.0	553.6	498.0	622.2	429.9	387.9	422.3	1150.2	
TTG051-L	6	67	740.7	654.7	713.7	480.4	553.6	0.0	580.0	687.4	524.7	560.6	478.4	1159.2	
TTG020-L	7	1	498.7	533.8	637.1	380.9	498.0	580.0	0.0	403.6	497.1	496.4	535.2	1048.1	
TTG026-L	8	2	651.5	662.7	742.6	462.6	622.2	687.4	403.6	0.0	586.6	610.9	662.9	1202.8	
TTG042-L	9	22	509.8	449.5	421.6	376.1	429.9	524.7	497.1	586.6	0.0	477.1	461.5	896.4	
TTG047-L	10	38	663.0	587.9	580.2	370.1	387.9	560.6	496.4	610.9	477.1	0.0	426.4	1130.9	
TTG048-L	11	40	661.2	597.6	627.9	366.5	422.3	478.4	535.2	662.9	461.5	426.4	0.0	1132.4	
TTG057-L	12	54	825.3	782.5	809.4	897.3	1150.2	1159.2	1048.1	1202.8	896.4	1130.9	1132.4	0.0	

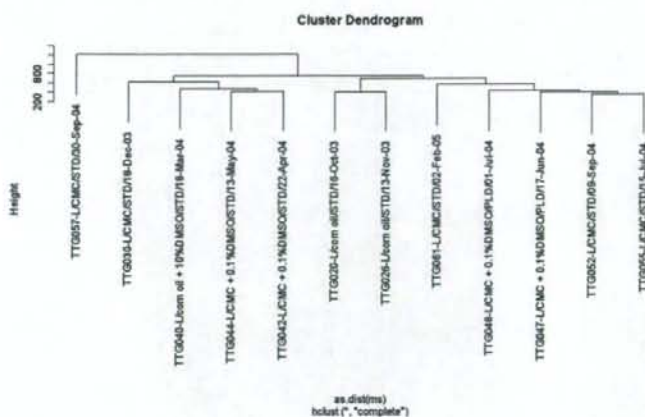
## Dose0:使用プローブセットとツリー構造

Dose0:結果(発現量が何れかの時刻で細胞当たり3コピー以上の遺伝子、Ward法)



## Dose0:使用プローブセットとツリー構造

Dose0:結果(発現量が何れかの時刻で細胞当たり3コピー以上の遺伝子、Complete法)



## Dose0:使用プローブセットとツリー構造

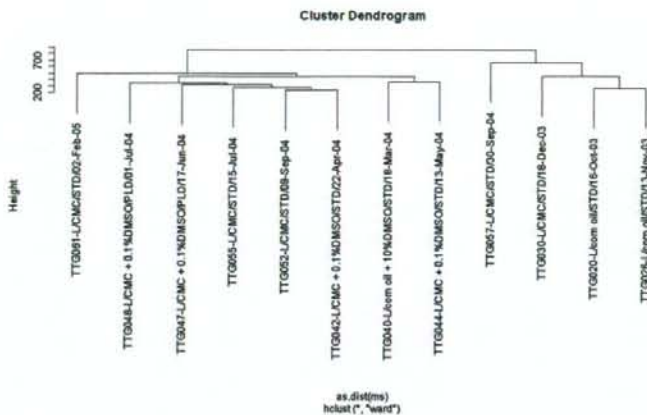
Dose0:距離マトリクス(発現量が何れかの時刻で細胞当たり10コピー以上の遺伝子)

対象プローブ数: 9,659

			1	2	3	4	5	6	7	8	9	10	11	12
	ord	pl no	12	17	33	50	42	67	1	2	22	38	40	54
TTQ030-L	1	12	0.0	411.7	434.7	374.3	508.0	612.6	354.5	454.7	400.4	532.3	545.3	452.1
TTQ040-L	2	17	411.7	0.0	353.8	271.9	418.0	501.4	320.5	412.6	314.6	431.2	437.8	522.2
TTQ044-L	3	33	434.7	353.8	0.0	279.3	371.9	519.8	358.2	437.7	257.7	387.7	430.7	563.7
TTQ052-L	4	50	374.3	271.9	279.3	0.0	244.9	357.6	266.0	326.3	227.2	283.9	276.0	529.3
TTQ055-L	5	42	508.0	418.0	371.9	244.9	0.0	419.6	373.3	453.3	287.7	283.5	329.9	734.0
TTQ061-L	6	67	612.6	501.4	519.8	357.6	419.6	0.0	472.2	559.6	397.1	427.5	342.5	819.0
TTQ020-L	7	1	354.5	320.5	358.2	266.0	373.3	472.2	0.0	266.5	323.8	379.4	420.6	567.6
TTQ026-L	8	2	454.7	412.6	437.7	326.3	453.3	559.6	266.5	0.0	394.7	449.1	503.8	860.5
TTQ042-L	9	22	400.4	314.6	257.7	227.2	287.7	397.1	323.8	394.7	0.0	324.9	330.0	572.6
TTQ047-L	10	38	532.3	431.2	387.7	283.9	283.5	427.5	379.4	449.1	324.9	0.0	326.4	744.4
TTQ048-L	11	40	545.3	437.8	430.7	276.0	329.9	342.5	420.6	503.8	330.0	326.4	0.0	766.6
TTQ057-L	12	54	452.1	522.2	563.7	529.3	734.0	819.0	567.6	860.5	572.6	744.4	766.6	0.0

## Dose0:使用プローブセットとツリー構造

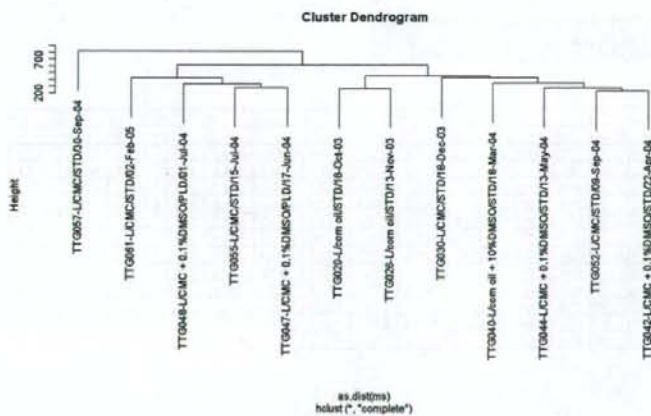
Dose0:結果(発現量が何れかの時刻で細胞当たり10コピー以上の遺伝子、Ward法)



溶媒によりツリー分岐がされているとは言えないような構造が現れた

## Dose0:使用プローブセットとツリー構造

Dose0:結果(発現量が何れかの時刻で細胞当たり10コピー以上の遺伝子、Complete法)



溶媒によりツリー分岐がされているとは言えないような構造が現れた

## Dose0:使用プローブセットとツリー構造

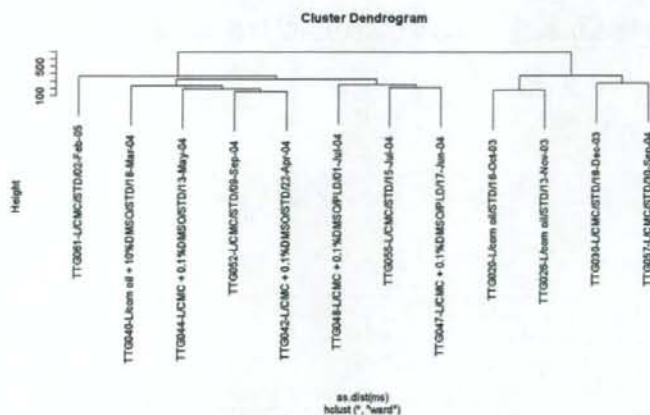
Dose0:距離マトリクス(発現量が何れかの時刻で細胞当たり30コピー以上の遺伝子)

対象プローブ数 4,082

	ord	col no	1	2	3	4	5	6	7	8	9	10	11	12
			12	17	33	50	42	67	1	2	22	38	40	54
TTG030-l	1	12	0.0	298.9	323.5	263.6	373.8	470.6	234.1	287.7	305.6	392.4	405.6	269.8
TTG040-l	2	17	298.9	0.0	247.4	174.1	274.9	352.4	214.0	257.2	214.3	283.9	286.3	323.9
TTG044-l	3	33	323.5	247.4	0.0	180.4	240.2	358.0	243.3	286.0	171.0	263.2	280.3	360.1
TTG052-l	4	50	263.6	174.1	180.4	0.0	171.4	263.7	191.3	223.6	148.2	210.1	195.1	301.1
TTG055-l	5	42	373.8	274.9	240.2	171.4	0.0	309.6	278.8	319.8	187.2	204.2	236.6	452.2
TTG061-l	6	67	470.6	352.4	358.0	263.7	309.6	0.0	372.4	424.9	277.5	311.8	244.5	541.9
TTG020-l	7	1	234.1	214.0	243.3	191.3	278.8	372.4	0.0	169.4	234.7	290.4	315.4	300.0
TTG028-l	8	2	297.7	257.2	286.0	223.6	319.8	424.9	169.4	0.0	273.8	321.9	354.7	350.8
TTG042-l	9	22	305.6	214.3	171.0	148.2	187.2	277.5	234.7	273.8	0.0	215.5	217.9	355.6
TTG047-l	10	38	392.4	283.9	263.2	210.1	204.2	311.8	290.4	321.9	215.5	0.0	226.7	468.1
TTG048-l	11	40	405.6	286.3	280.3	195.1	236.6	244.5	315.4	354.7	217.9	226.7	0.0	482.4
TTG057-l	12	54	269.8	323.9	360.1	301.1	452.2	541.9	300.0	350.8	355.6	468.1	482.4	0.0

## Dose0:使用プローブセットとツリー構造

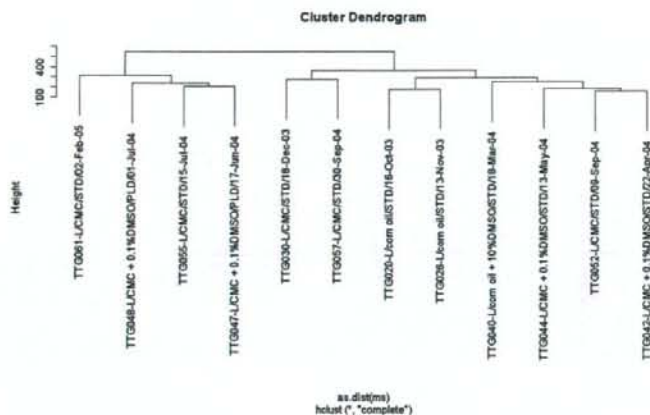
Dose0:結果(発現量が何れかの時刻で細胞当たり30コピー以上の遺伝子、Ward法)



溶媒によりツリー分岐がされているとは言えないような構造が現れた

## Dose0:使用プローブセットとツリー構造

結果(発現量が何れかの時刻で細胞当たり30コピー以上の遺伝子、Complete法)



溶媒によりツリー分岐がされているとは言えないような構造が現れた

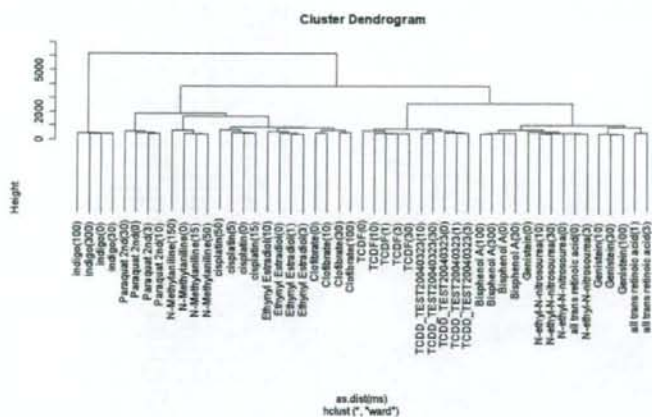
## Doseを含めたクラスタリング

- 化合物&Doseを一つのクラスタリング対象にする



## Doseを含めたクラスタリング

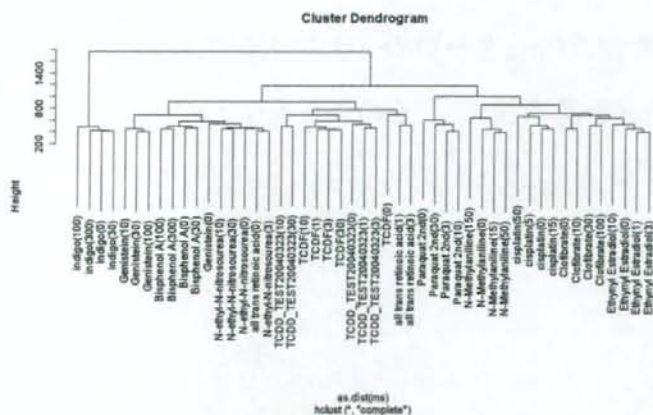
クラスタリング実行結果 Ward法



化合物(Project)ごとに固まり、溶媒群が一つに集まることはない

## Doseを含めたクラスタリング

クラスタリング実行結果 Complete法



化合物(Project)ごとに固まり、溶媒群が一つに集まることはない

## Doseを含めたクラスタリング

化合物間距離の遺伝子ごとの寄与例

- TCDF(1) vs. Bisphenol A(30)

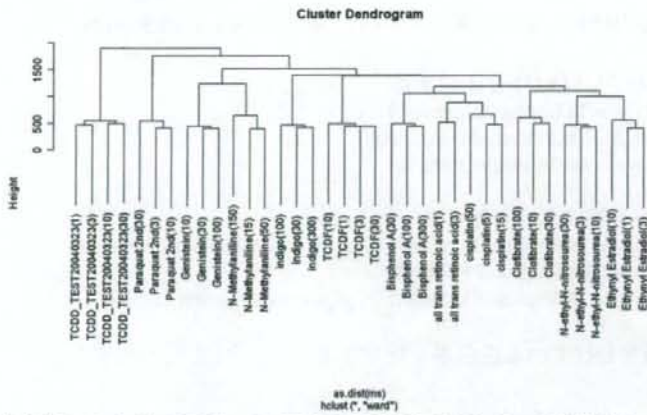
No	orb_no	dist	Gene Symbol
1	8523	95.811	Cyc1a1
2	27453	24.892	?
3	28123	19.995	Hlx2-6
4	7155	18.140	Pabpn1
5	30088	16.246	Rbbp4
6	24472	16.197	Htr1d
7	21190	15.877	Ewer1
8	7231	15.380	M01:2159619
9	8579	15.245	Cys2c70
10	13145	15.173	Wdr53
11	23717	15.143	Xpo7
12	7446	14.851	Lip1
13	19108	14.839	Rbbp4
14	12266	14.599	Ugt2b34
15	34702	14.430	Ars32a
16	45094	14.424	?
17	23240	14.374	Ang1
18	4035	14.271	Cyb3a41
19	35108	14.257	Rad23b
20	42956	14.201	?

## Doseを含めたクラスタリング(Dose0基準)

- 化合物&Doseを一つのクラスタリング対象にする
- Dose0を基準値にし、その差分をクラスタリング対象とする

## Doseを含めたクラスタリング(Dose0基準)

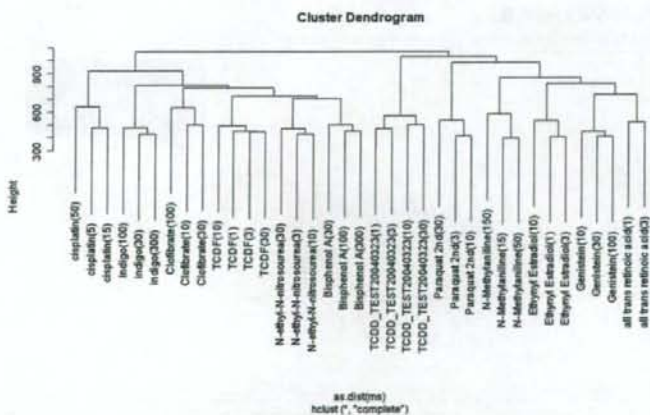
クラスタリング実行結果 Ward法



化合物(Project)ごとに固まり、TCDDとTCDFが離れるなど、Dose0を基準にし、既知の情報と離れる結果となっている

## Doseを含めたクラスタリング(Dose0基準)

クラスタリング実行結果 Complete法



化合物(Project)ごとに固まり、TCDDとTCDFが離れるなど、Dose0を基準にし、既知の情報と離れる結果となっている

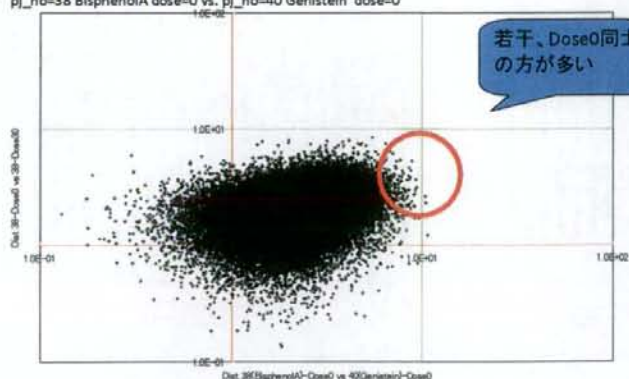


## 詳細調査

- 既知の情報と違う結果となったので個々のチェックを試みた
  - 次の3プロジェクトを比較する
    - pj\_no=38 BisphenolA dose=0
    - pj\_no=38 BisphenolA dose=30
    - pj\_no=40 Genistein dose=0
  - 化合物同士(化合物影響)
  - pj\_no=38 BisphenolA dose=0 vs. pj\_no=38 BisphenolA dose=30
  - Dose0間(プロジェクト間影響)
  - pj\_no=38 BisphenolA dose=0 vs. pj\_no=40 Genistein dose=0
- 実験を分けている遺伝子を特定する

## 詳細調査

- 化合物同士(化合物影響)
  - pj\_no=38 BisphenolA dose=0 vs. pj\_no=38 BisphenolA dose=30
- Dose0間(プロジェクト間影響)
  - pj\_no=38 BisphenolA dose=0 vs. pj\_no=40 Genistein dose=0

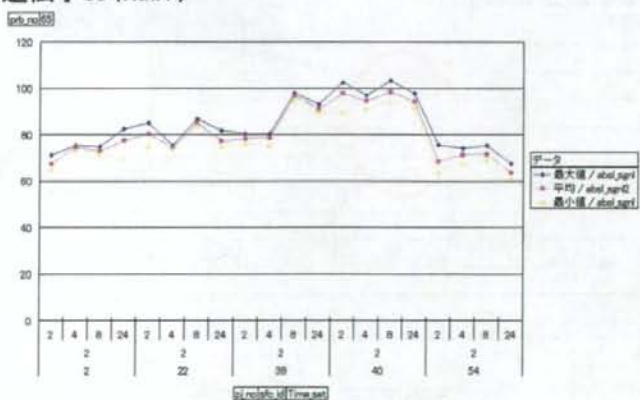


## 詳細調査

- 各遺伝子を5実験の情報を詳細にチェック
  - Pj\_no=2
  - Pj\_no=22
  - Pj\_no=38 BisphenolA dose=0
  - Pj\_no=40 Genistein dose=0
  - Pj\_no=54

## 詳細調査

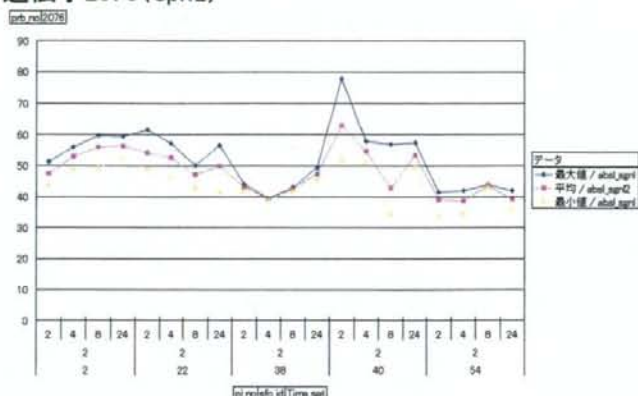
- Dose0間で大きな差異を持つ遺伝子の他の化合物での状況  
遺伝子65 (Rab7)



Pj\_no=38と40はPLD、38の2hrs、4hrsが40と違う理由が不明

## 詳細調査

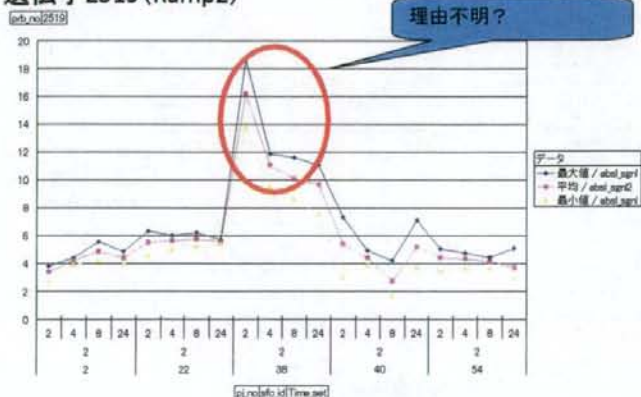
- Dose0間で大きな差異を持つ遺伝子の他の化合物での状況  
遺伝子2076 (Cpn1)



Pj\_no=38で、極端に、偏差が小さい

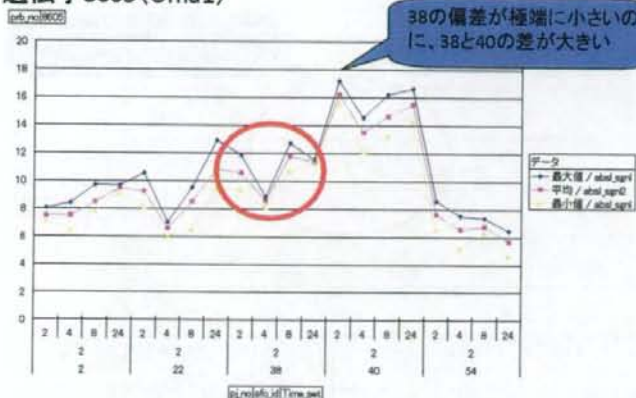
## 詳細調査

- Dose0間で大きな差異を持つ遺伝子の他の化合物での状況  
遺伝子2519 (Ramp2)



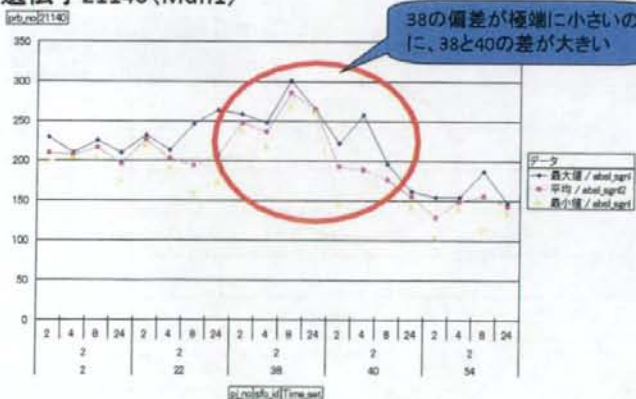
## 詳細調査

- Dose0間で大きな差異を持つ遺伝子の他の化合物での状況  
遺伝子8605(Oma1)



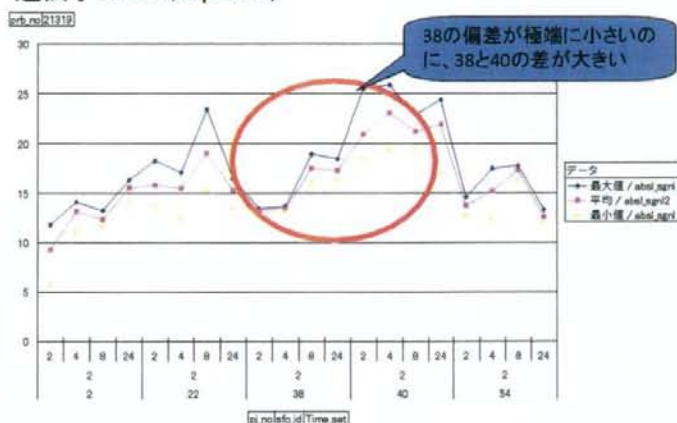
## 詳細調査

- Dose0間で大きな差異を持つ遺伝子の他の化合物での状況  
遺伝子21140(Mdh1)



## 詳細調査

- Dose0間で大きな差異を持つ遺伝子の他の化合物での状況  
遺伝子21319 (Atp6v0b)



## 詳細調査

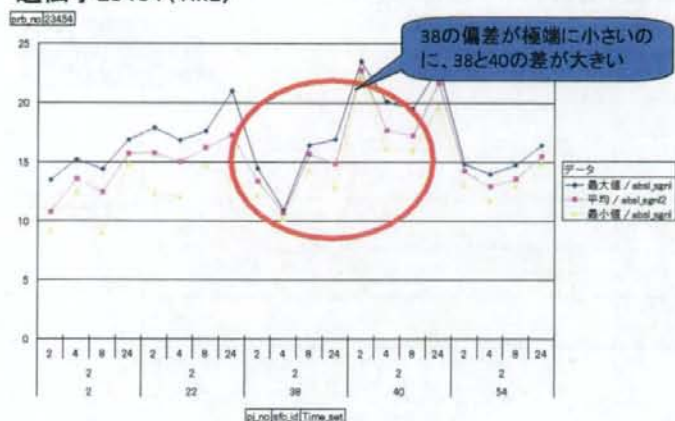
- Dose0間で大きな差異を持つ遺伝子の他の化合物での状況  
遺伝子21376 (Cdc14b)





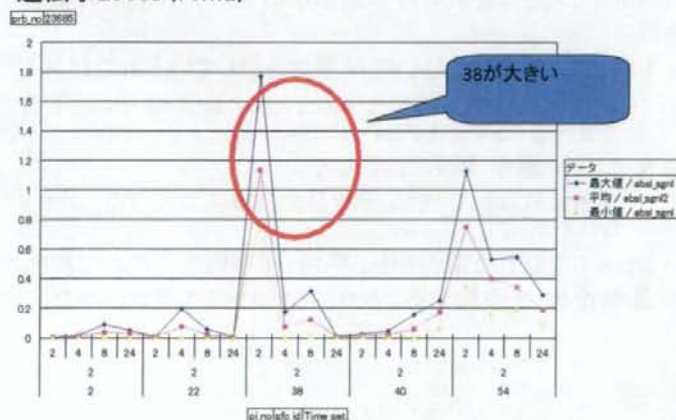
## 詳細調査

- Dose0間で大きな差異を持つ遺伝子の他の化合物での状況  
遺伝子23464 (Tlk1)



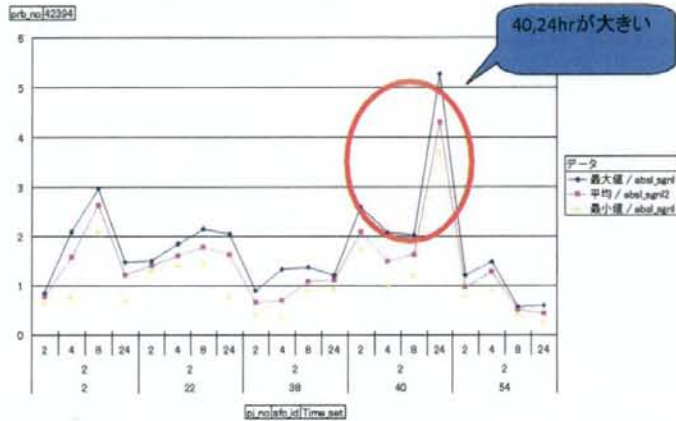
## 詳細調査

- Dose0間で大きな差異を持つ遺伝子の他の化合物での状況  
遺伝子23685 (Prm1)



## 詳細調査

- Dose0間で大きな差異を持つ遺伝子の他の化合物での状況  
遺伝子42394(Mm.214692.1)



## 詳細調査

- Dose0間で大きな差異を持つ遺伝子の他の化合物での状況  
考察
  - N=3でも、非常に小さい標準偏差を示しているものがある
    - 実験の練度・マウスの世代・餌など制御の難しい項目により影響を受けている可能性が高いのではないか
  - N=3では、遺伝子数に比較して少ない
    - 偶然性を排除して、平均・標準偏差を求めるには、少ないのではないか?
  - Dose=0で取れる情報から、平均・標準偏差を求めておき、基準値として使用することで、キャンセルできないか?

## 距離計算の式を再構成

### ● 距離計算の式を再構成した

#### — 偏差により絶対値の差を除する。

- 絶対値と偏差の比を基準にする。ただし、小さな絶対値に対して観測誤差が大きくなるので、補正項を与える
- 同一遺伝子、同一化合物容量内、同一時刻内で標準偏差を計算し、時刻に対して合算
- Dose0内でも偏差を計算し用いる

$$\text{dist}(\text{ChemicalA}, \text{ChemicalB}) = \sqrt{\sum_{\text{全ての遺伝子}} \text{dist}_{\text{遺伝子}}(\text{ChemicalA}, \text{ChemicalB})^2}$$

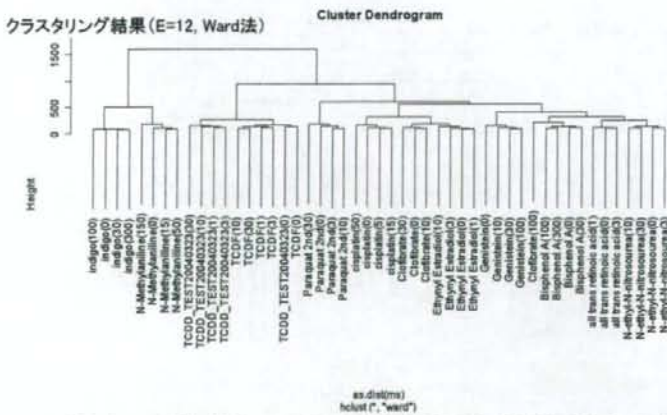
$$\text{dist}_{\text{遺伝子}}(\text{ChemicalA}, \text{ChemicalB}) = \sqrt{\frac{\sum_{\text{re}(2,4,8,24)} (\text{Avg}_{g,A,i} - \text{Avg}_{g,B,i})^2}{\sigma_{g,A}^2 + \sigma_{g,B}^2 + \sigma_{g,dose0}^2 + E}}$$

$$\sigma_{g,A} = \sqrt{\frac{1}{n \cdot 4} \sum_{\substack{\text{re}(2,4,8,24) \\ \text{ce}(1,2,\dots)}} (\text{Absl}_{g,A,i,c} - \text{Avg}_{g,A,i})^2} \quad : n \text{ は実験チップ数 (殆ど } n=3)$$

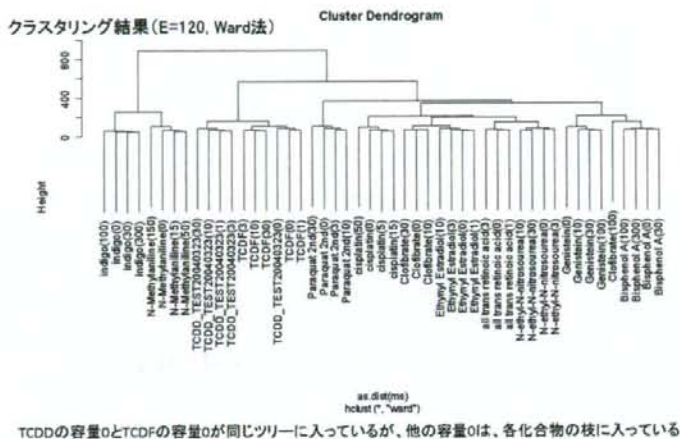
$$\sigma_{g,dose0} = \sqrt{\frac{\sum_{\text{re}(2,4,8,24)} (\text{Absl}_{g,A,i,c} - \text{Avg}_{g,A,i})^2}{\text{ce}(1,2,3)}} \quad : \text{全てのDose0に対して実施}$$

*E*: 観測誤差消去項

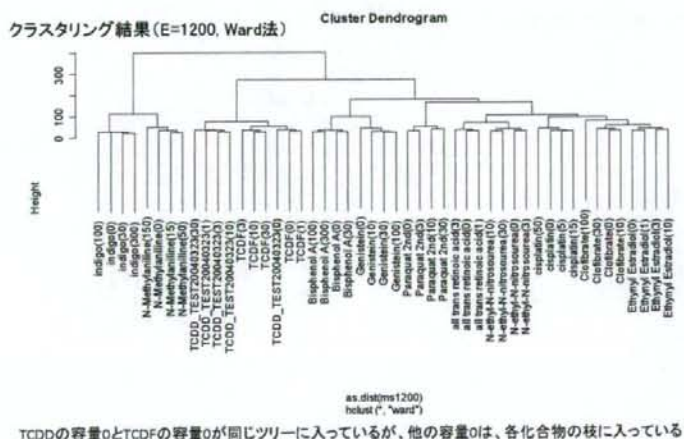
## 距離計算の式を再構成



## 距離計算の式を再構成

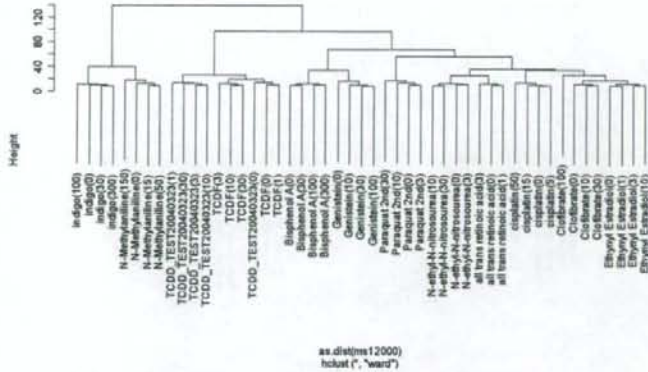


## 距離計算の式を再構成



## 距離計算の式を再構成

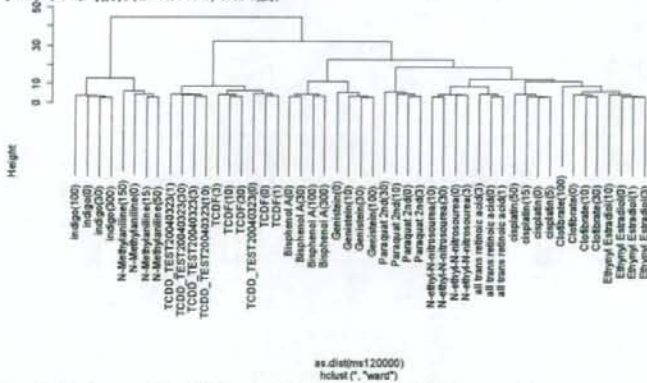
クラスタリング結果 (E=12000, Ward法) Cluster Dendrogram



TCDDの容量0とTCDFの容量0が同じツリーに入っているが、他の容量0は、各化合物の枝に入っている

## 距離計算の式を再構成

クラスタリング結果 (E=120000, Ward法) Cluster Dendrogram

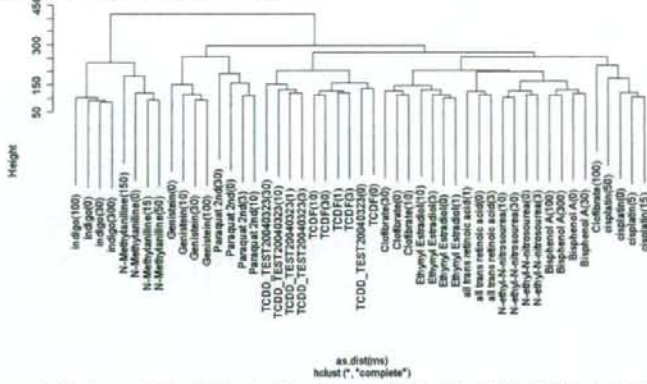


TCDDの容量0とTCDFの容量0が同じツリーに入っているが、他の容量0は、各化合物の枝に入っている



## 距離計算の式を再構成

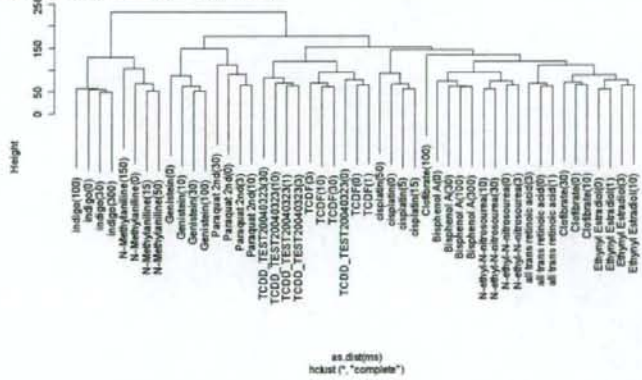
クラスタリング結果 (E=12, Complete法) Cluster Dendrogram



TCDDの容量0とTCDFの容量0が同じツリーに入っているが、他の容量0は、各化合物の枝に入っている

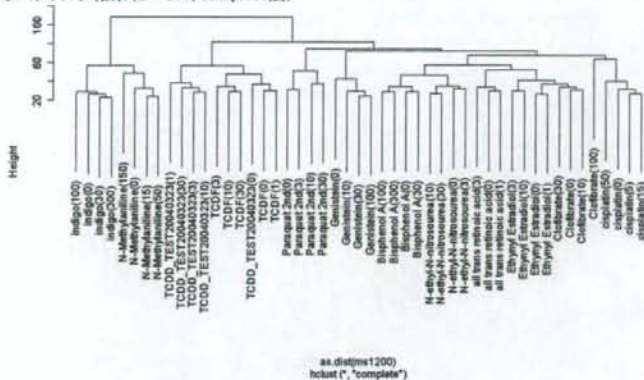
## 距離計算の式を再構成

クラスタリング結果 (E=120, complete法) Cluster Dendrogram



## 距離計算の式を再構成

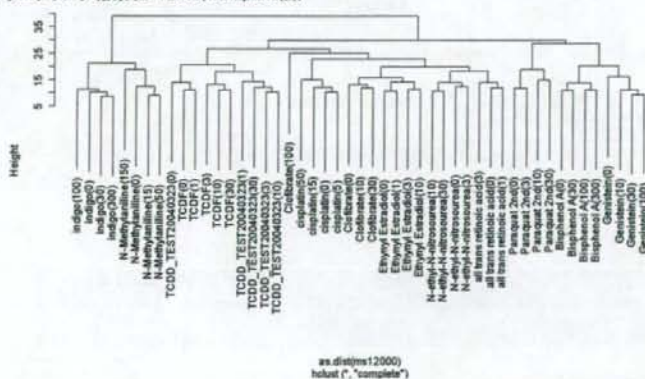
クラスタリング結果 (E=1200, complete法) Cluster Dendrogram



TCDDの容量0とTCDFの容量0が同じツリーに入っているが、他の容量0は、各化合物の枝に入っている

## 距離計算の式を再構成

クラスタリング結果 (E=1200, complete法) Cluster Dendrogram



TCDDの容量0とTCDFの容量0が同じツリーに入っているが、他の容量0は、各化合物の枝に入っている