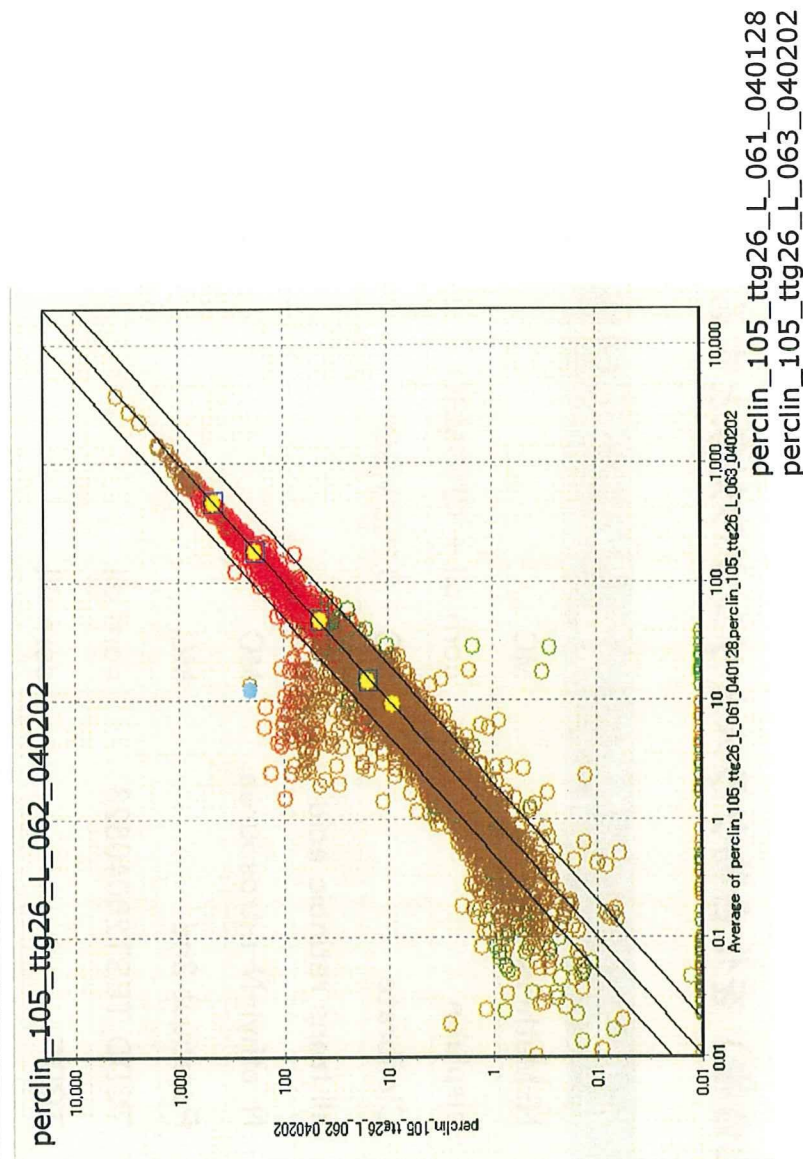


3. 実験結果の精度

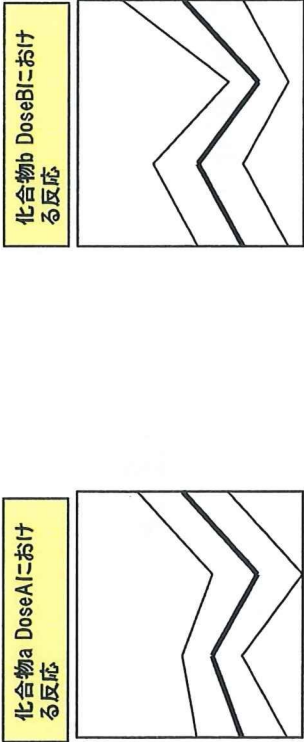
他の器官の試料が混入にした可能性が高い



算術平均、標準偏差は、混入による影響を受けやすい。
中央値やMAD（中央絶対偏差：Median Absolute Deviation）を用いる。

4. 1. 化合物間距離の定義

Mlang補正により、定量的な取り扱いが可能となった。差の量が重要になるような式を構成する。



各遺伝子
 E_{ab} : 化合物 ab 間の 4時刻を用いた距離
 σ_a : 化合物 a の 4時刻全体の中央絶対偏差
 σ_b : 化合物 b の 4時刻全体の中央絶対偏差
 σ_v : 全 Vehicle の中央絶対偏差
 c : 定数

中央絶対偏差MAD

$$MAD = median_{i=1}^n (|x_i - \bar{x}|) \quad SD = \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}$$

データのちらばり具合を表す指標で、他の指標と比較して異常値の影響を受けにくい。

遺伝子間の非類似度(距離)

$$d_{abi} = \sqrt{\frac{E_{abi}^2}{\sigma_a + \sigma_b + \sigma_v + const.}}$$

$const. = 10$

偶然近いものを無視

$$\hat{d}_{abi} = \begin{cases} d_{abi} - const. & \text{for } d_{abi} \geq const. \\ 0 & \text{for } d_{abi} < const. \end{cases}$$

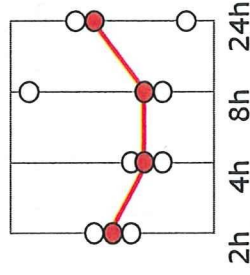
$const. = 2$

化合物間の距離=全遺伝子の合計

$$d_{ab} = \left(\sum_{i=1}^n \hat{d}_{abi}^2 \right)^{\frac{1}{2}}$$

4. 1. 化合物間距離の定義

次のように化合物間距離を定義する。



化合物a、時刻iにおける代表値

$$x_{ai} = \text{median}_{j=1}^3(x_{aij})$$

化合物aにおける中央絶対偏差

$$\sigma_a = \text{median}_{i=1, j=1}^{4,3}(|x_{aij} - x_{ai}|)$$

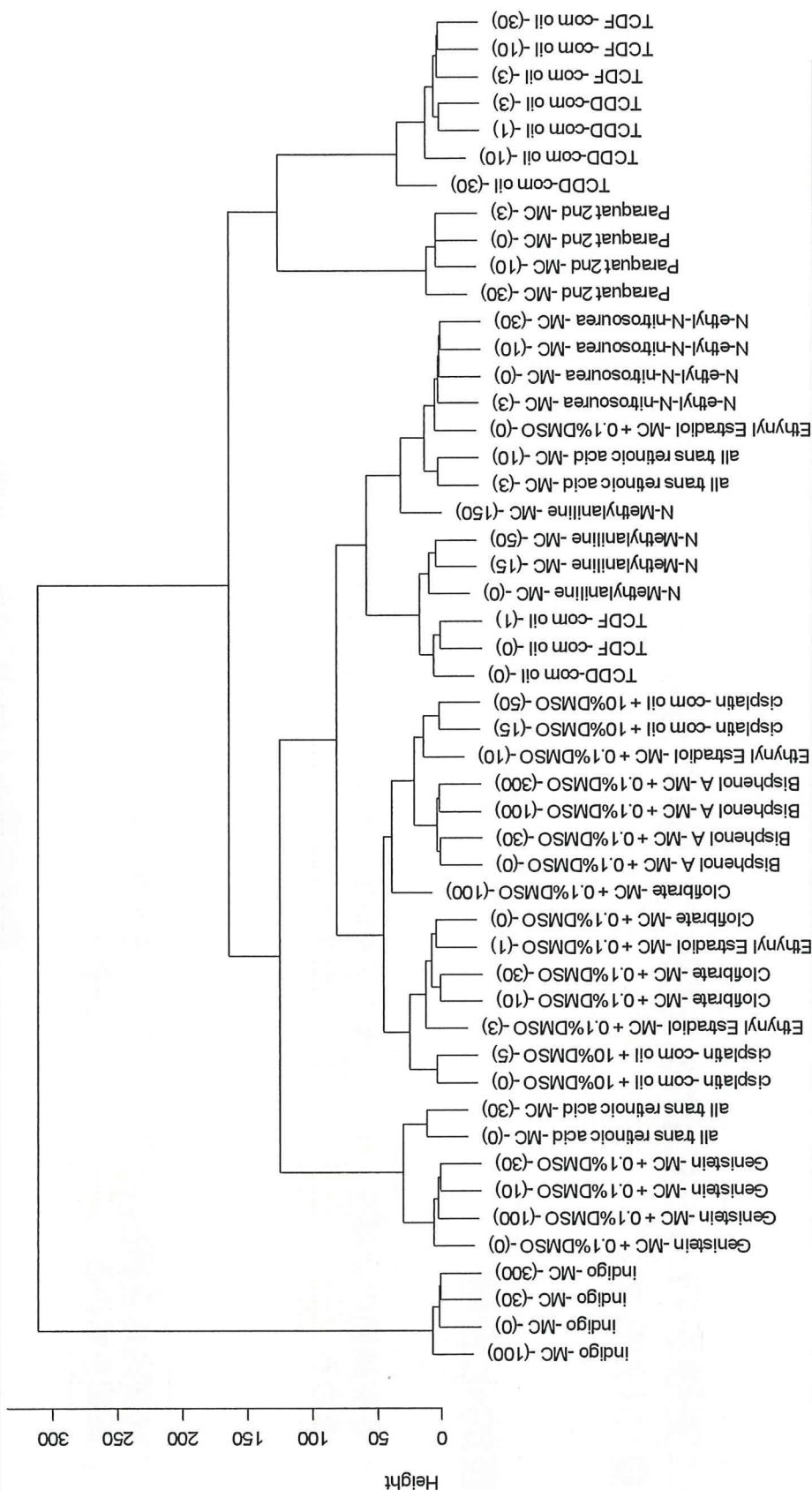
全Vehicleの標準偏差

$$\sigma_0 = \sqrt{\frac{1}{n \cdot 4 \cdot 3} \sum_{a \in V} \sum_{i=1}^4 \sum_{j=1}^3 (x_{aij} - \bar{x})^2}$$

日内変動が大きな遺伝子やプロジェクト間の差の大きな遺伝子を見捨てる。

4. 2. 化合物間距離による階層クラスタリング

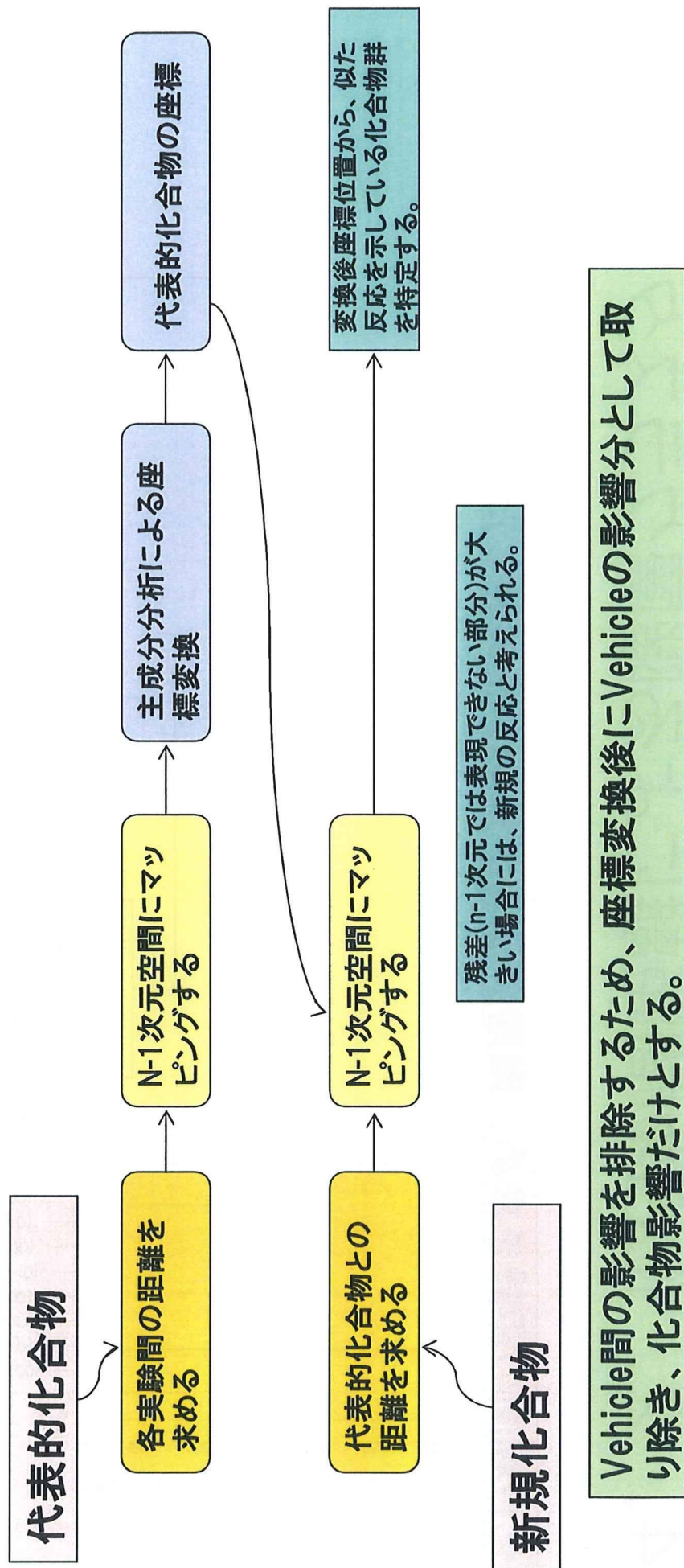
化合物間の距離を求め、階層クラスタリングを実施した。



Vehicleは、プロジェクト内で固まってしまう、溶媒により枝分かれしたように思われる。

4. 3. 化合物遺伝子応答評価方法

すでに実験を行った化合物を用いて、似た遺伝子が発現する化合物群を見つけ出す。



4. 3. 新規化合物の計算方法

新規化合物に対する座標位置の求め方を構成する。

一般形

$$r_i = \text{dist}(\mathbf{p}_i, \mathbf{x}) = \sqrt{\sum_{j=1}^n (p_{ij} - x_j)^2}$$

$$r_i^2 = \sum_{j=1}^n (p_{ij} - x_j)^2$$

3点（2次元空間）に対する解

$$\begin{cases} r_1^2 = (p_{11} - x_1)^2 + (p_{12} - x_2)^2 \\ r_2^2 = (p_{21} - x_1)^2 + (p_{22} - x_2)^2 \\ r_3^2 = (p_{31} - x_1)^2 + (p_{32} - x_2)^2 \end{cases}$$

$$\begin{cases} r_1^2 = p_{11}^2 - 2p_{11}x_1 + x_1^2 + p_{12}^2 - 2p_{12}x_2 + x_2^2 \\ r_2^2 = p_{21}^2 - 2p_{21}x_1 + x_1^2 + p_{22}^2 - 2p_{22}x_2 + x_2^2 \\ r_3^2 = p_{31}^2 - 2p_{31}x_1 + x_1^2 + p_{32}^2 - 2p_{32}x_2 + x_2^2 \end{cases}$$

$$\begin{cases} r_1^2 - r_2^2 = (p_{11}^2 - p_{21}^2) - 2(p_{11} - p_{21})x_1 + (p_{12}^2 - p_{22}^2) - 2(p_{12} - p_{22})x_2 \\ r_2^2 - r_3^2 = (p_{21}^2 - p_{31}^2) - 2(p_{21} - p_{31})x_1 + (p_{22}^2 - p_{32}^2) - 2(p_{22} - p_{32})x_2 \\ r_3^2 - r_1^2 = (p_{31}^2 - p_{11}^2) - 2(p_{31} - p_{11})x_1 + (p_{32}^2 - p_{12}^2) - 2(p_{32} - p_{12})x_2 \end{cases}$$

$$\begin{cases} 2(p_{11} - p_{21})x_1 + 2(p_{12} - p_{22})x_2 = (p_{11}^2 - p_{21}^2) + (p_{12}^2 - p_{22}^2) - (r_1^2 - r_2^2) \\ 2(p_{21} - p_{31})x_1 + 2(p_{22} - p_{32})x_2 = (p_{21}^2 - p_{31}^2) + (p_{22}^2 - p_{32}^2) - (r_2^2 - r_3^2) \\ 2(p_{31} - p_{11})x_1 + 2(p_{32} - p_{12})x_2 = (p_{31}^2 - p_{11}^2) + (p_{32}^2 - p_{12}^2) - (r_3^2 - r_1^2) \end{cases}$$

$$\begin{cases} (p_{11} - p_{21}) & (p_{12} - p_{22}) & (x_1) \\ (p_{21} - p_{31}) & (p_{22} - p_{32}) & (x_2) \end{cases} = \frac{1}{2} \begin{cases} (p_{11}^2 - p_{21}^2) - (r_1^2 - r_2^2) \\ (p_{21}^2 - p_{31}^2) - (r_2^2 - r_3^2) \\ (p_{31}^2 - p_{11}^2) - (r_3^2 - r_1^2) \end{cases}$$

$$\begin{pmatrix} p_1 - p_2 \\ p_2 - p_3 \\ p_3 - p_1 \end{pmatrix} \mathbf{x} = \frac{1}{2} \begin{pmatrix} p_1^2 - p_2^2 - (r_1^2 - r_2^2) \\ p_2^2 - p_3^2 - (r_2^2 - r_3^2) \\ p_3^2 - p_1^2 - (r_3^2 - r_1^2) \end{pmatrix}$$

過剰な条件を有する連立1次方程式に変形できる。

連立方程式を解くことにより、既存化合物で作成された空間での座標と、残差を求めることができる。

4. 3. 新規化合物の計算方法

一般形：N点(N-1次元空間)に対する解

$$\begin{cases} r_1^2 = \sum_{j=1}^{n-1} (p_{1j} - x_j)^2 \\ r_2^2 = \sum_{j=1}^{n-1} (p_{2j} - x_j)^2 \\ \vdots \\ r_n^2 = \sum_{j=1}^{n-1} (p_{nj} - x_j)^2 \end{cases} \quad \begin{cases} r_1^2 - r_2^2 = \sum_{j=1}^{n-1} ((p_{1j}^2 - p_{2j}^2) - (2p_{1j}x_j - 2p_{2j}x_j)) \\ r_2^2 - r_3^2 = \sum_{j=1}^{n-1} ((p_{2j}^2 - p_{3j}^2) - (2p_{2j}x_j - 2p_{3j}x_j)) \\ \vdots \\ r_n^2 - r_1^2 = \sum_{j=1}^{n-1} ((p_{nj}^2 - p_{1j}^2) - (2p_{nj}x_j - 2p_{1j}x_j)) \end{cases}$$

$$\begin{cases} r_1^2 = \sum_{j=1}^{n-1} (p_{1j}^2 - 2p_{1j}x_j + x_j^2) \\ r_2^2 = \sum_{j=1}^{n-1} (p_{2j}^2 - 2p_{2j}x_j + x_j^2) \\ \vdots \\ r_n^2 = \sum_{j=1}^{n-1} (p_{nj}^2 - 2p_{nj}x_j + x_j^2) \end{cases} \quad \begin{cases} \sum_{j=1}^{n-1} (2p_{1j}x_j - 2p_{2j}x_j) = \sum_{j=1}^{n-1} (p_{1j}^2 - p_{2j}^2) - (r_1^2 - r_2^2) \\ \sum_{j=1}^{n-1} (2p_{2j}x_j - 2p_{3j}x_j) = \sum_{j=1}^{n-1} (p_{2j}^2 - p_{3j}^2) - (r_2^2 - r_3^2) \\ \vdots \\ \sum_{j=1}^{n-1} (2p_{nj}x_j - 2p_{1j}x_j) = \sum_{j=1}^{n-1} (p_{nj}^2 - p_{1j}^2) - (r_n^2 - r_1^2) \end{cases}$$

$$\begin{cases} p_1 - p_2 \\ p_2 - p_3 \\ \vdots \\ p_n - p_{n-1} \end{cases} \cdot \mathbf{x} = \frac{1}{2} \begin{pmatrix} (p_1^2 - p_2^2) - (r_1^2 - r_2^2) \\ (p_2^2 - p_3^2) - (r_2^2 - r_3^2) \\ \vdots \\ (p_n^2 - p_{n-1}^2) - (r_n^2 - r_1^2) \end{pmatrix}$$

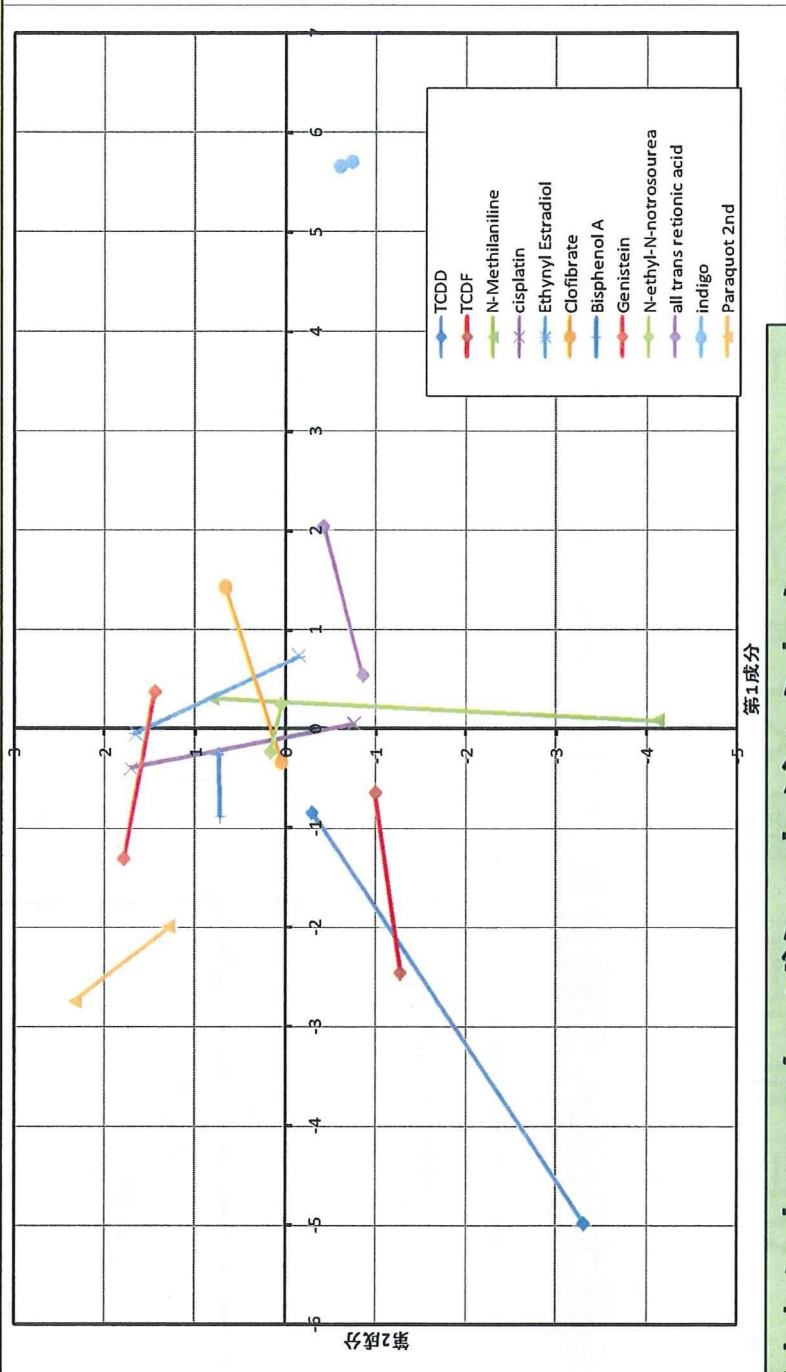
過剰な条件を有する連立1次方程式に変形できる。

連立方程式を解くことにより、既存化合物で作成された空間での座標と、残差を求めることができる。

4. 4. Vehicleと最大Doseによる代表化①

12化合物のVehicleと最大Doseを用いて、14次元空間に割り当てた。変動が少ない次元が与えられた実験数より若干減少した。

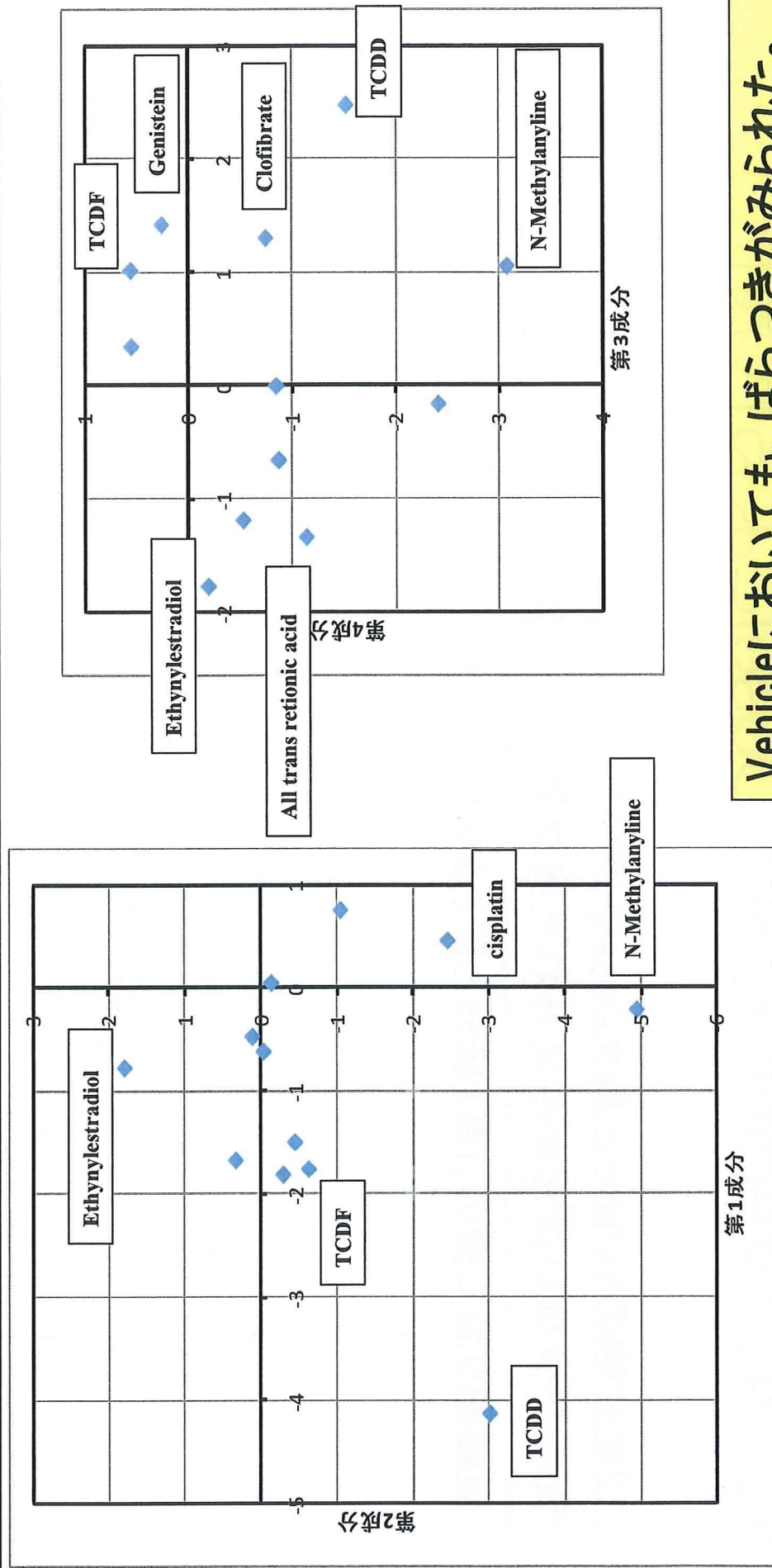
座標割り当て後の、第1成分と第2成分による各化合物の分布。



Vehicleにおいても、ばらつきがみられた。

4. 3. Vehicleと最大Doseによる代表化②

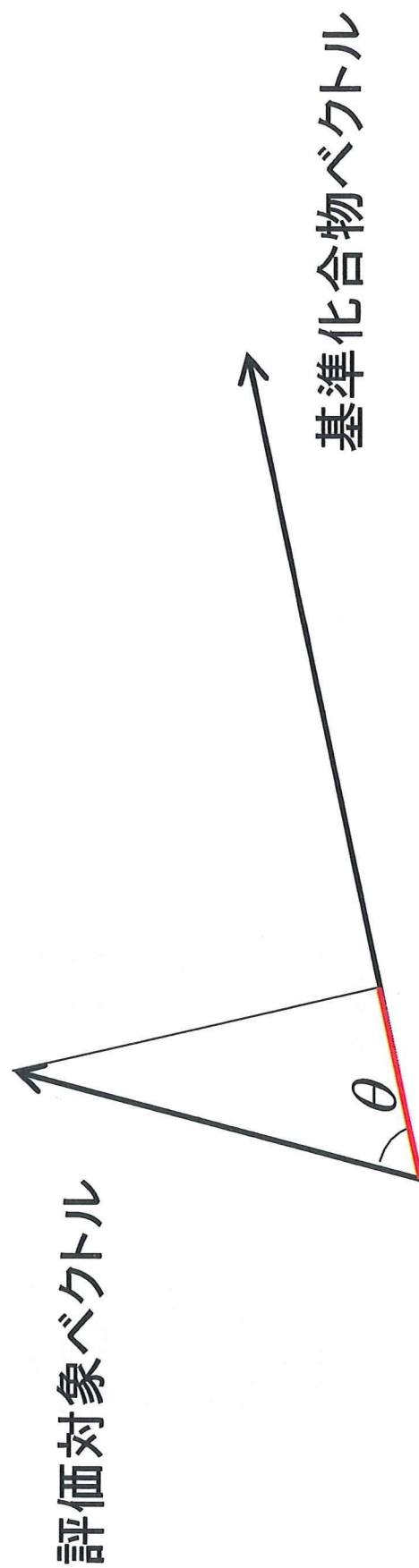
Vehicleから最大Doseへの差分により、化合物特有の情報を抽出した。



Vehicleにおいても、ばらつきがみられた。

4. 3. 新規化合物の評価

ベクトル空間における内積を用いて、新規化合物に対する評価指標を作成する。



$\cos \theta$ 基準化合物と方向の差異

$\frac{b \cdot \cos \theta}{a}$ 基準化合物に落とした時の比

4. 3. 新規化合物の評価

TCDD(MidHigh, MidLow, Low)の割り当て結果

	方向の差異			基準化合物に対する比		
	MidHigh	MidLow	Low	MidHigh	MidLow	Low
TCDD	0.7632	0.5931	0.8823	0.4163	0.3440	0.4813
TCDF	0.7326	0.6919	0.7684	0.7915	0.7948	0.8302
N-Methylaniline	0.2945	0.2337	0.4144	0.1627	0.1373	0.2290
cisplatin	-0.1180	-0.0643	-0.1297	-0.0836	-0.0484	-0.0919
Ethinyl Estradiol	-0.1871	0.0123	0.1226	-0.1540	0.0108	0.1009
Clofibrate	0.1253	0.0936	0.1838	0.0843	0.0670	0.1237
Bisphenol A	-0.0911	-0.1772	-0.1976	-0.1981	-0.4096	-0.4295
Genistein	0.5799	0.1927	0.3561	0.7987	0.2823	0.4905
N-ethyl-N-nitrosourea	0.1501	0.0444	0.0330	0.3573	0.1123	0.0784
all trans retionic acid	0.2476	0.0409	0.1018	0.2575	0.0452	0.1058
indigo	0.3788	0.2859	0.3527	1.0380	0.8329	0.9666
Paraquat 2nd	-0.2849	-0.0016	0.0095	-0.2640	-0.0016	0.0088

方向の差異から、TCDDは、TCDDとTCDFと似た反応を示している
 基準化合物に対する比で、indigoは小さな反応なので、indigoの反応を含
 んでいるのではないだろうか？

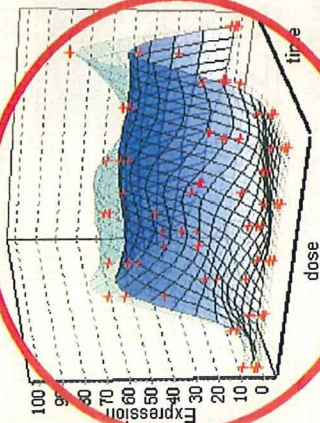
4. 3. 新規化合物の評価（まとめ）

- ・ 本評価方法は、基準とする化合物の選定が難しい
 - 化合物間の差異を表す「距離」が数学上の距離の定義を満たしていない。（三角不等式）
 - 少数の実験だけでは、遺伝子の複雑な動きを網羅できない。

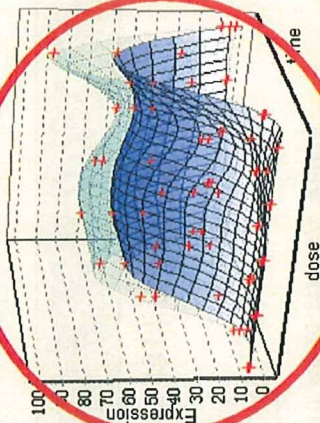
4. 4. 距離定義と異常遺伝子

- ・ Indigoの実験における異常
 - Indigo実験において、他の実験より発現のベースレベルが低い。
 - Indigo実験のVehicleの動きをみると、corn oilに近い遺伝子が存在する。

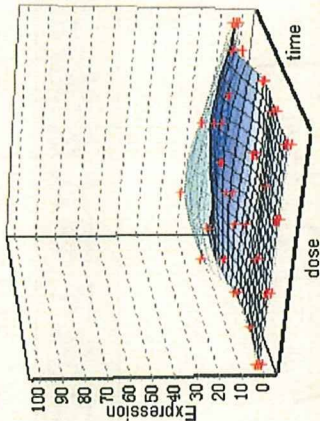
TTG020 - TCDD
corn oil - 2003/10/16



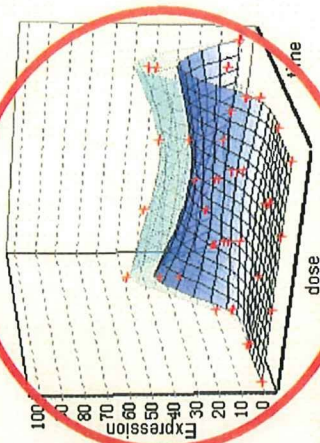
TTG026 - TCDF
corn oil - 2003/11/13



TTG030 - N-Mthylaniline
MC - 2003/12/18

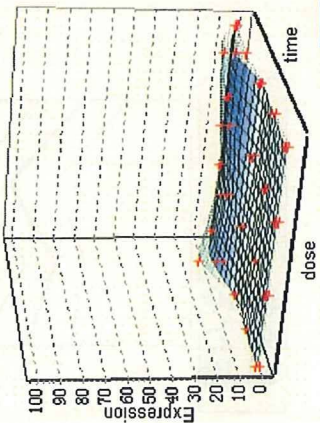


TTG040 - cisplatin
corn oil + 10% DMSO - 2004/3/18

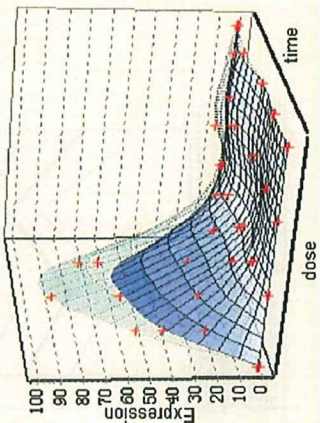


TTG057がなければ、corn-oilによるものと推測可能な遺伝子

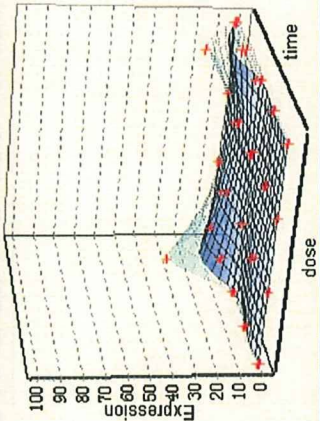
TTG042 - Ethynyl Estradiol
MC+0.1%DMSO - 2004/4/22



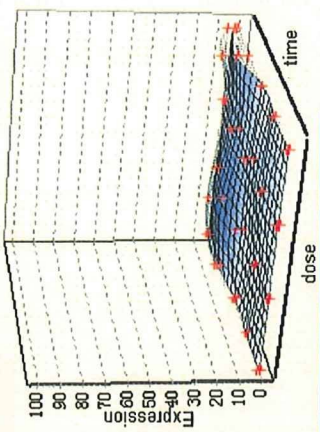
TTG044 - Clofibrate
MC+0.1%DMSO - 2004/5/13



TTG047 - Bisphenol A
MC+0.1%DMSO - 2004/6/17

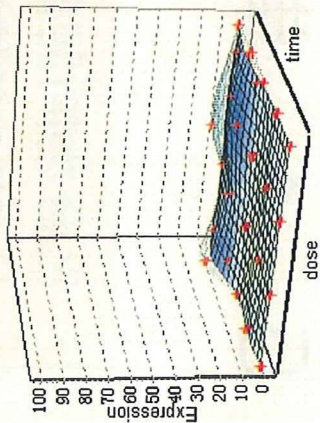


TTG048 - Genistein
MC+0.1%DMSO - 2004/7/1

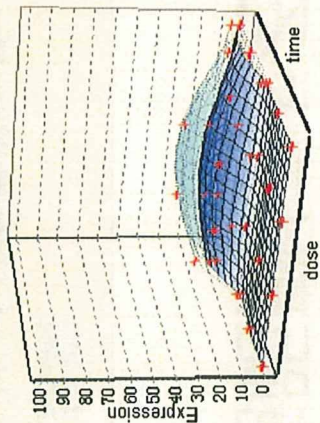


この遺伝子を活性化させている原因は？

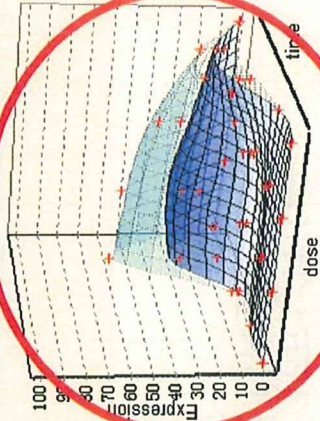
TTG055 - N-ethyl-N-nitrosourea MC
MC - 2004/7/15



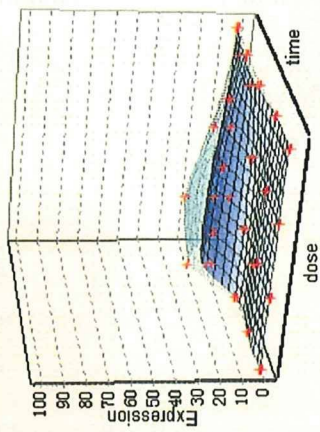
TTG052 - all trans retinoic acid
MC - 2004/9/9



TTG057 - indigo
MC - 2004/9/30



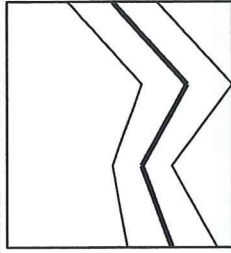
TTG061 - Paraquat 2nd
MC - 2005/2/2



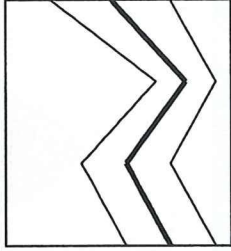
5. 1. 単純距離による座標変換

- ・ 化合物間距離の定義
 - MLANG補正により、より定量的な取り扱いができるようになったので、発現量の差のみで距離を定義する。

化合物a DoseAにおける反応



化合物b DoseBにおける反応



遺伝子 i

E_{abi} : 化合物 ab 間の 4時刻を用いた距離

$$E_{abi} = \sqrt{\sum_{t \in (2,4,8,24)} (x_{ait} - x_{bit})^2}$$

x_{ait} : 化合物 a の時刻 t における発現量

遺伝子間の非類似度(距離)

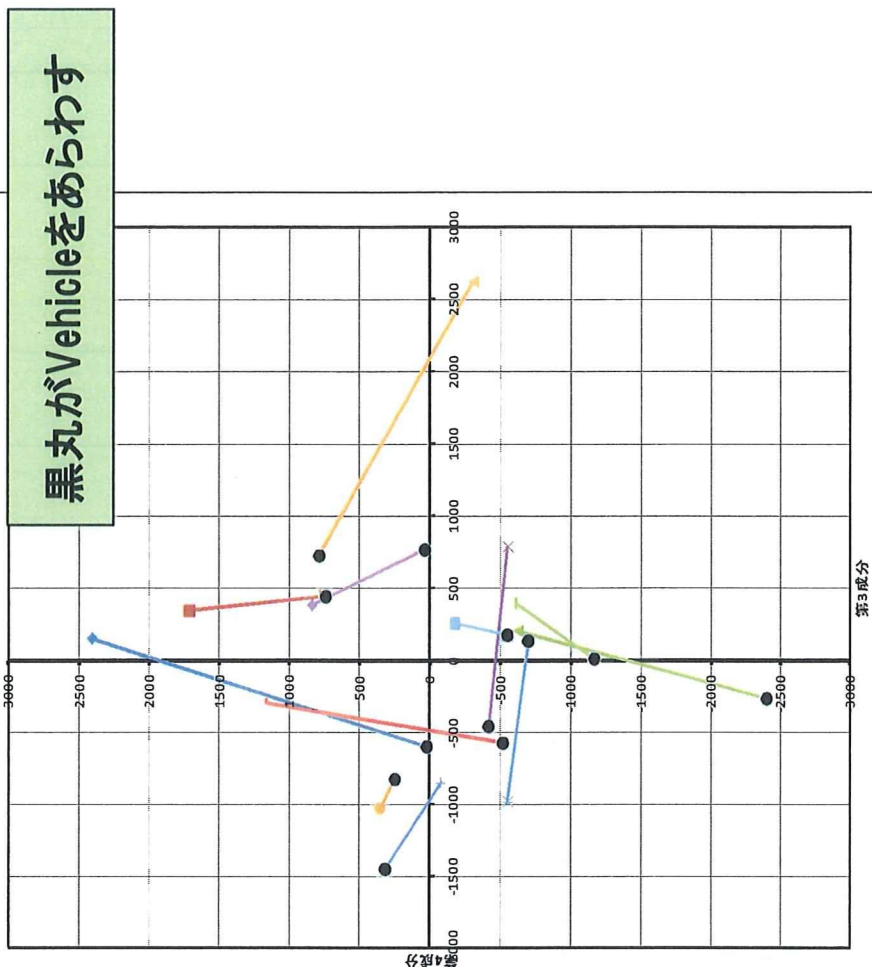
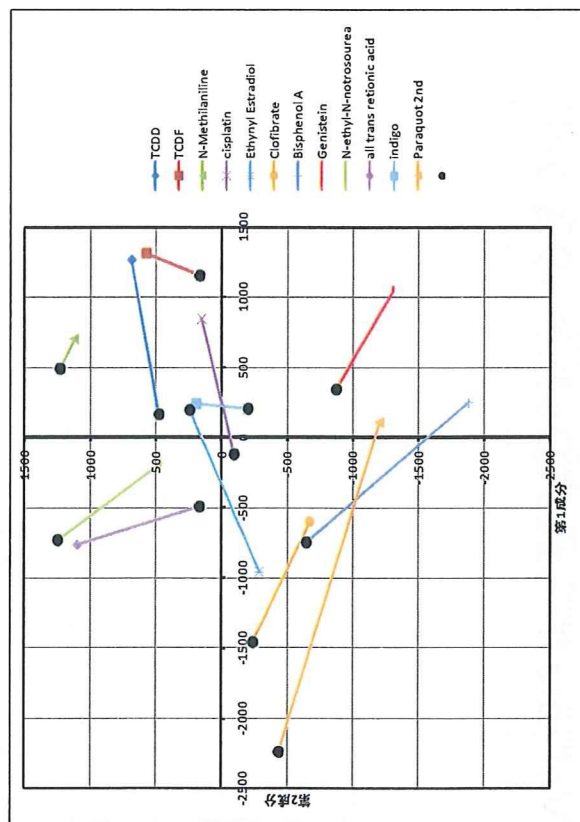
$$E_{ab} = \left(\sum_{i=1}^n E_{abi}^2 \right)^{\frac{1}{2}}$$

距離表示の式変形がないため、三角不等式を満たす。

5. 2. 単純距離 MLANG Vehicleと最大Doseによる代表化①

12化合物のVehicleと最大Doseを用いて、23次元空間に割り当てた。

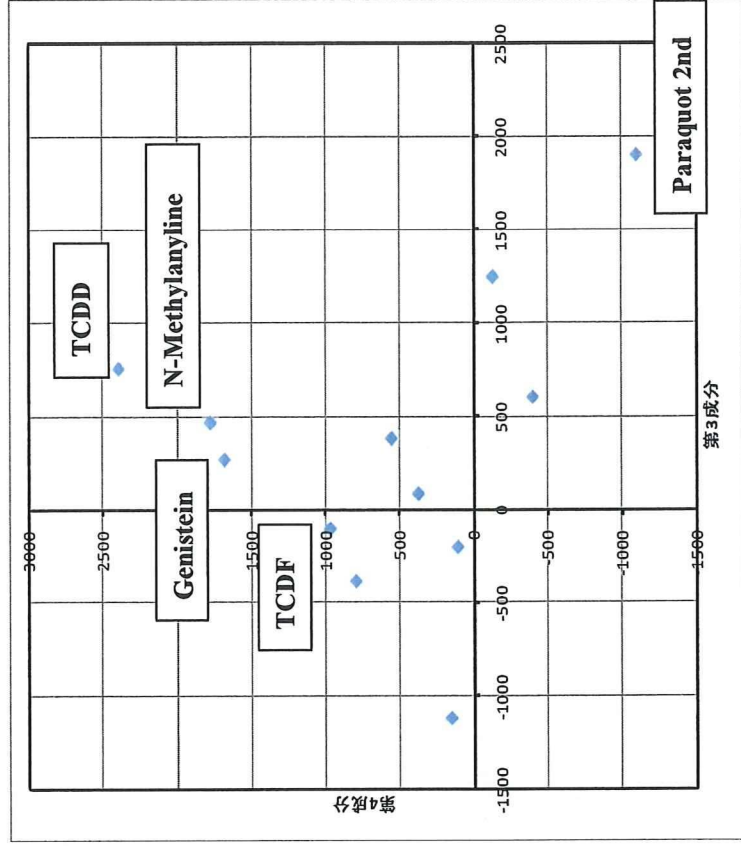
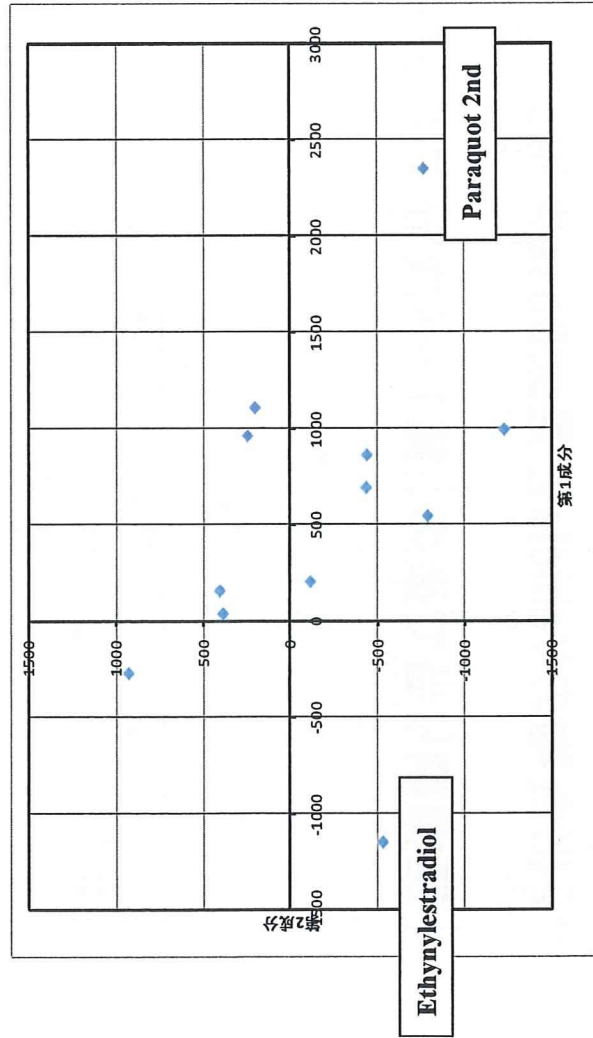
座標割り当て後の、第1～4成分による各化合物の分布



第4軸にTCDD, TCDFが強く現れている

5. 2. 単純距離 MLANG Vehicleと最大Doseによる代表化②

Vehicleから最大Doseへの差分により、化合物特有の情報を抽出した。



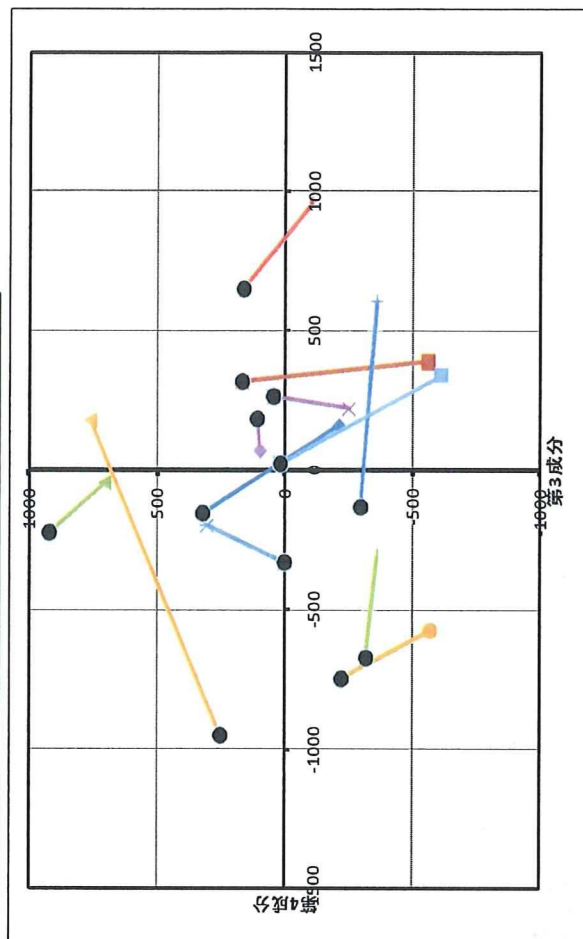
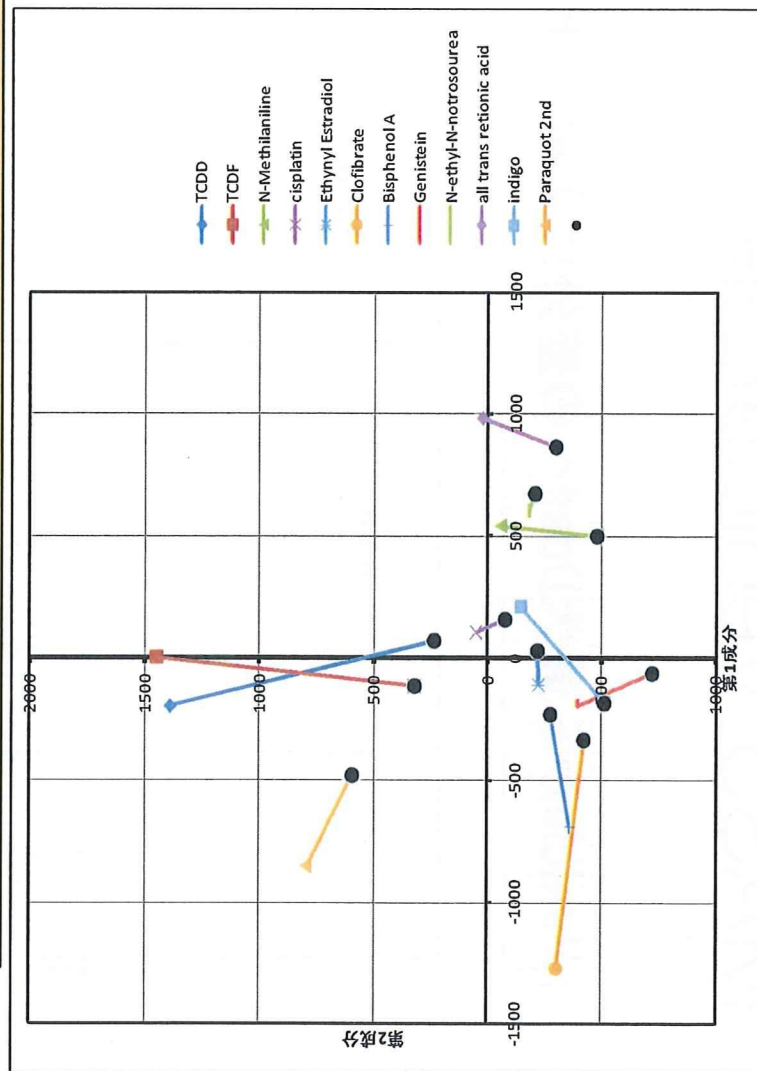
Vehicleにおいても、ばらつきがみられた。

5. 2. 単純距離

MAS5 Vehicleと最大Doseによる代表化①

12化合物のVehicleと最大Doseを用いて、21次元空間に割り当てた。変動が少ない次元が与えられた実験数より若干減少した。

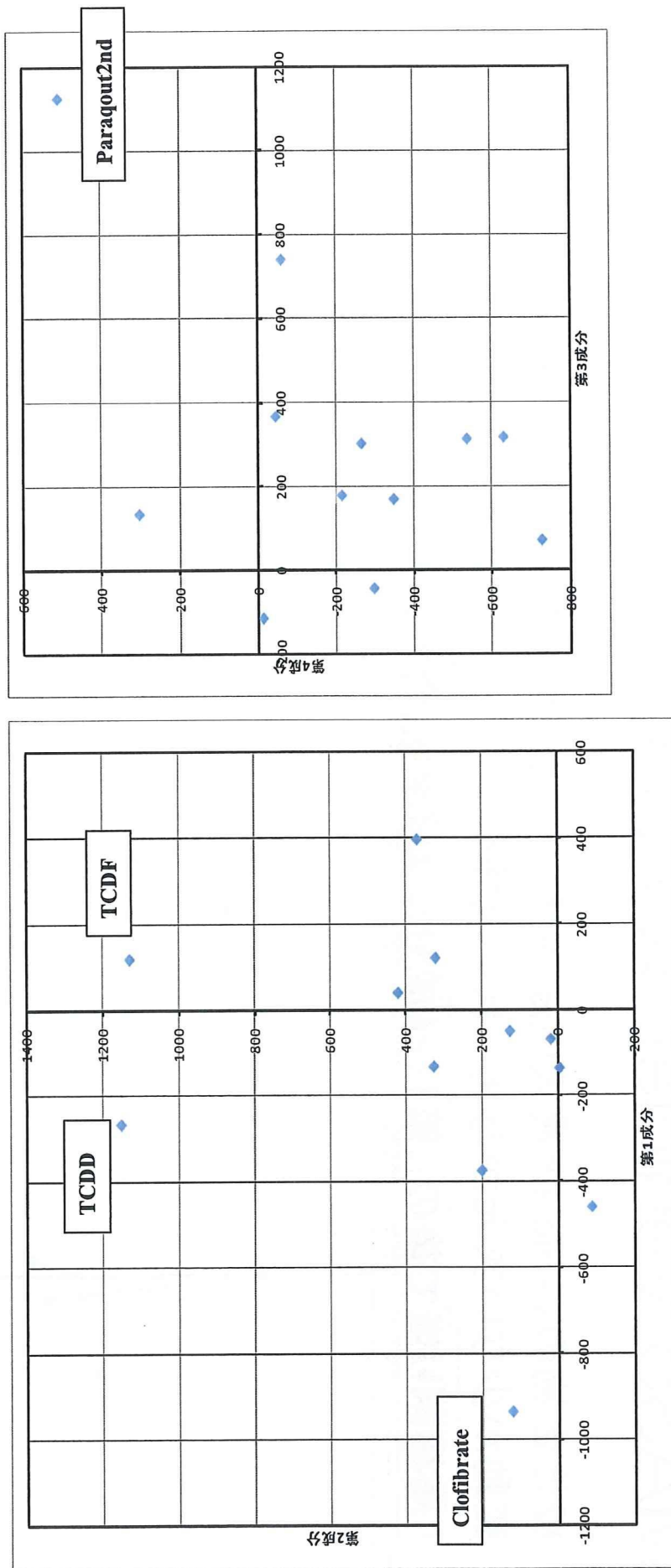
座標割り当て後の、第1～4成分による各化合物の分布



Clofibrateのみが変動し、他の変動は小さくなっている。

5. 2. 単純距離 MAS5 Vehicleと最大Doseによる代表化②

Vehicleから最大Doseへの差分により、化合物特有の情報を抽出した。



Vehicleにおいても、ばらつきがみられた。

6. 各プロジェクトのVehicle間で最も距離を生み出す要素となっている遺伝子

各プロジェクトのVehicleグループ同士の距離を大きくしている遺伝子を特定し、その遺伝子同士の距離順に、その形状を確認した。