

# 弊社のタンパク質マイクロアレイについて

---

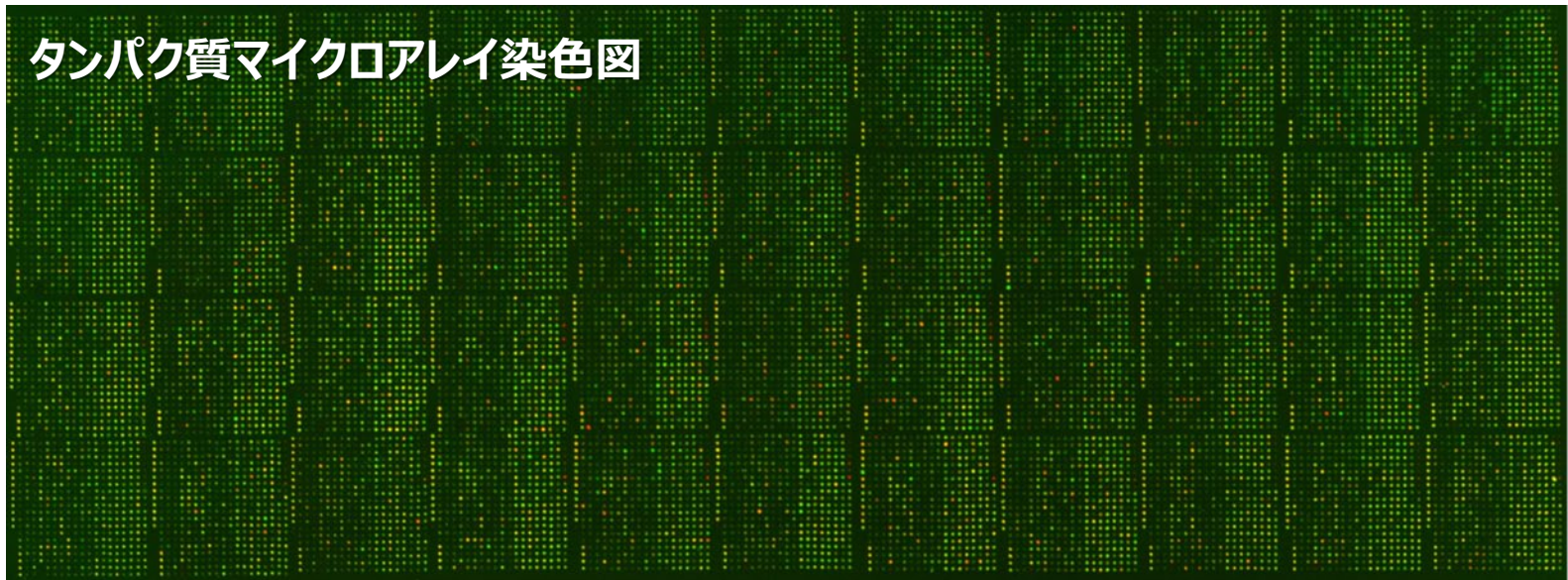
✓ ガラス基板を使用

✓ 2色法を採用

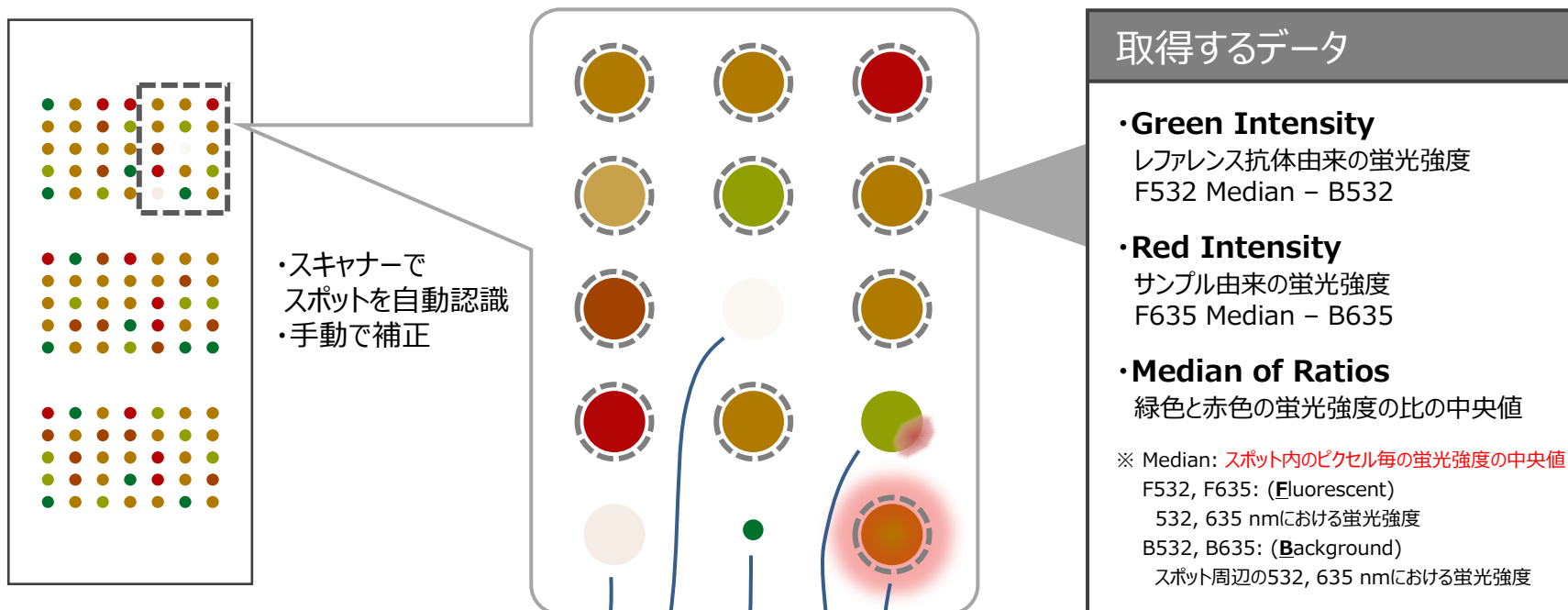
同一アレイ上のサンプル間比較はもとより、異なるアレイ間の比較解析も可能

✓ **Nature Methods (5: 1011-1017, 2008) の方法を基に開発した技術**

タンパク質マイクロアレイ染色図



# スポット検出と蛍光強度について



## データを取得できないスポット

- シグナルが非常に弱く鮮明でないスポット
- スポットング不良
- ゴミが付着したスポット

- バックグラウンドの蛍光が強く、Median of Ratios  
が異常値となったスポット  
(F532 Median - B532, F635 Median - B635の異常値)

データを取得できなかったスポットは欠損値 (0) として処理

# Median of Ratiosについて

## スポット\_1

・大きいスポット

測定データ



ピクセル毎の  
蛍光値を算出



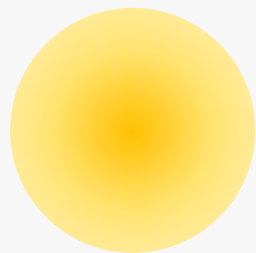
ピクセル毎の  
蛍光値の比率  
を算出



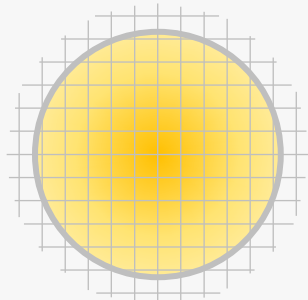
Median of  
Ratiosを用いて  
解析

## スポット\_2

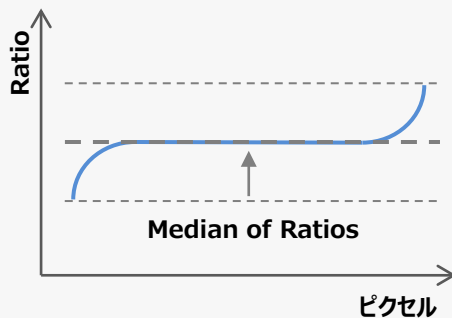
・小さいスポット



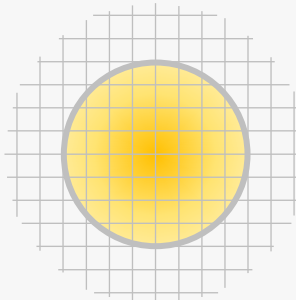
【ピクセル毎に測定】



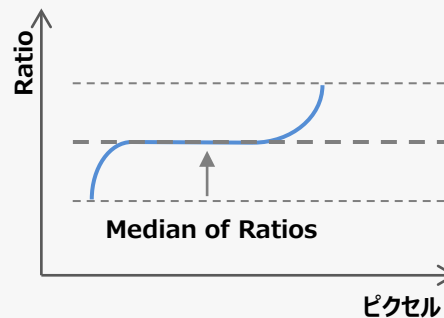
【蛍光値の比率の分布】



【ピクセル毎に測定】

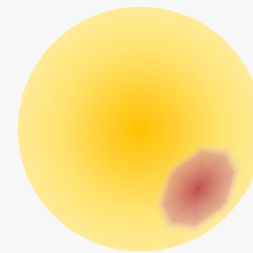


【蛍光値の比率の分布】

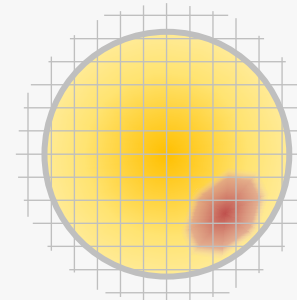


## スポット\_3

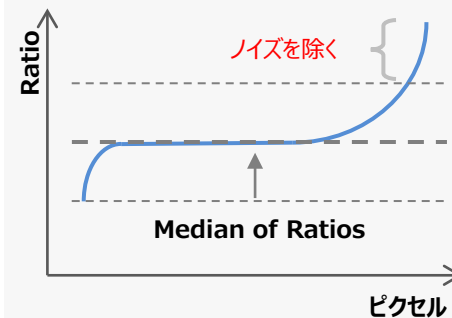
・ノイズが重なっているスポット



【ピクセル毎に測定】



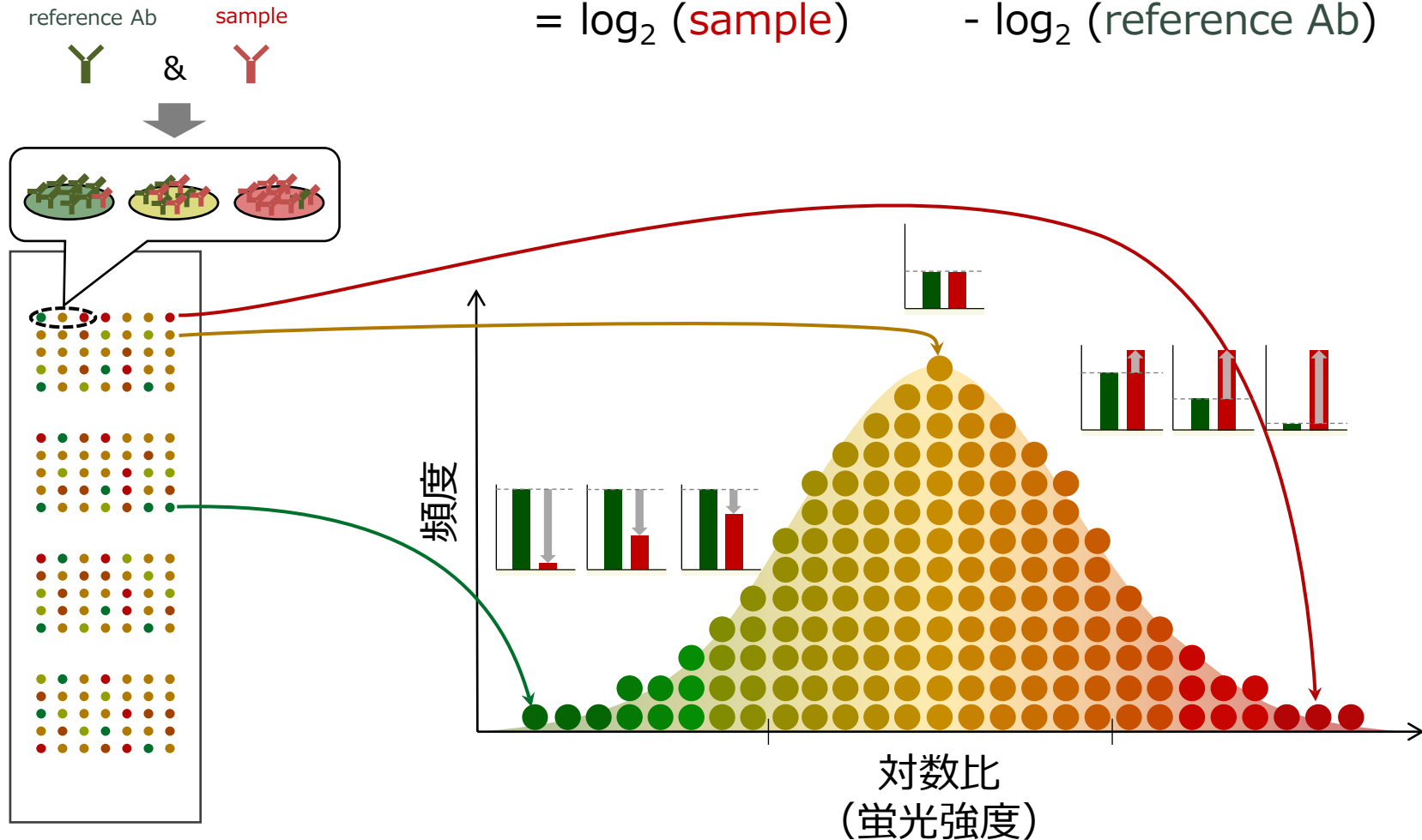
【蛍光値の比率の分布】



# サンプルとレファレンス抗体の比

$$\log \text{ ratio} = \log_2 \left( \frac{\text{sample}}{\text{reference Ab}} \right)$$

$$= \log_2 (\text{sample}) - \log_2 (\text{reference Ab})$$



# 解析データとデータの解説

## 【18K抗原マイクロアレイ】タンパク質マイクロアレイを用いた血中自己抗体プロファイリング

免疫グロブリンのクラス: ヒトIgG

データセット: 18K (18,429スポット) ヒトタンパク質

### 表の見方

- 各行はマイクロアレイ上のスポットをあらわし、1行が1種類のヒト抗原タンパク質（遺伝子）に対応します
- 列「Array 1」は、ネガティブコントロール（NC）としてサンプルを加えず、レファレンス抗体とそれぞれの2次抗体を使って染色して得られたデータです
- 列「Array 2」は、レファレンス抗体、各サンプル（ヒト血清）とそれぞれの2次抗体を使って染色して得られたデータです
- 列「2次比」の数値は、ネガティブコントロール（NC）との相対比で、各サンプル（ヒト血清）由来の反応の強さを示します

### 用語の説明

Green Intensity: レファレンス抗体とその2次抗体由来の緑色の蛍光強度

Red Intensity: サンプルとその2次抗体由来の赤色の蛍光強度

Median of Ratios: 緑色の蛍光強度と赤色の蛍光強度の比の中央値

Raw Data: 演算を加えていない生データ

Normalized Data: アレイ間の比較のために正規化したデータ

2次比: アレイ間の相対比データ ※ネガティブコントロール（NC）との相対比

### 染色条件

アレイフォーマット: 18K抗原マイクロアレイ（ヒトタンパク質）

サンプル: ヒト血清

レファレンス抗体: Goat Reference Antibody Mixture I

2次抗体（Red）: Alexa Fluor 647 anti-Human IgG antibody

2次抗体（Green）: Cy3 anti-Goat IgG antibody

搭載ヒト抗原タンパク質情報				
#	Gene symbol	Gene ID	Gene name	Mutation
1	[A1BG]		1 alpha-1-B glycoprotein	標準型

Green Intensity F532 Median - B532		Red Intensity F635 Median - B635		Median of Ratios (Log2)				
				Raw		Normalized		2次比
Array 1	Array 2	Array 1	Array 2	Array 1	Array 2	Array 1	Array 2	Array 2/1
NC	Sample1	NC	Sample1	NC	Sample1	NC	Sample1	Sample1
237	228	4	381	-5.31	0.77	-0.31	3.92	4.23
 $A_i$		 $B_i$		 $C_i$				

$A_i$ : レファレンス抗体（緑色）のピクセルごとの蛍光強度の中央値

$B_i$ : サンプル（赤色）のピクセルごとの蛍光強度の中央値

$C_i$ : ピクセルごとに緑と赤の比率を算出した後の比率の中央値※

※ Median of Ratios ( $C_i$ )は、 $B_i$ を $A_i$ で割った値ではない

$$C_i \neq \log_2\left(\frac{B_i}{A_i}\right) \quad \text{例) } -5.31 \neq \log_2\left(\frac{4}{237}\right)$$

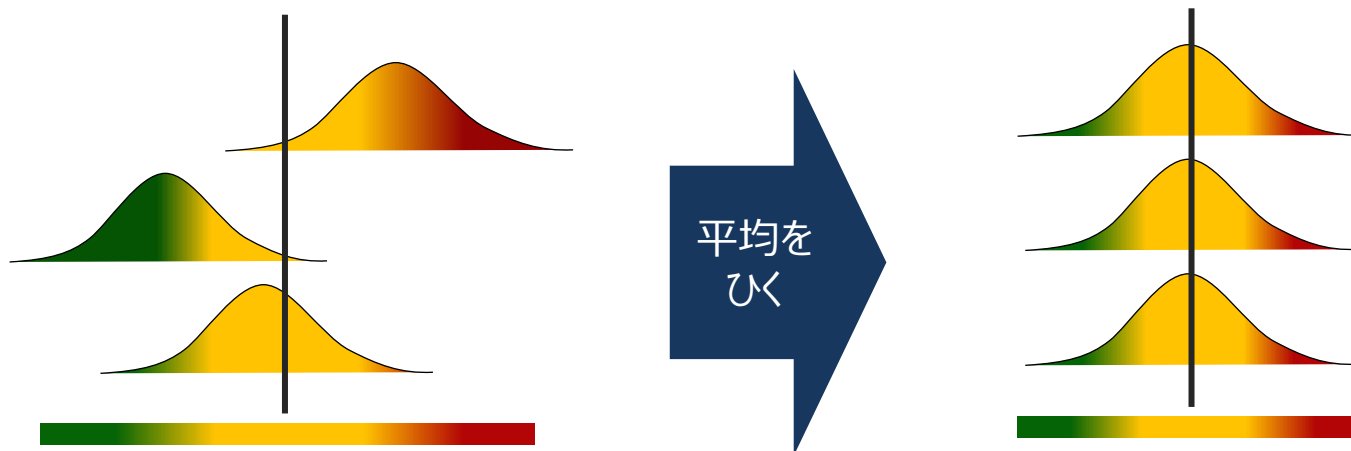
# 正規化について

- ・振れ幅の大きい上下各16%を除いた中央68%の値から平均値を算出
- ・各タンパク質の値と前述の平均値の差を算出

▶ 全体の形を整えた上でサンプル間比較

$$X' = X - \mu$$

$X$ : 変数  
 $\mu$ : 平均値



※ 標準偏差で補正するZ-score ( $Z = \frac{X - \mu}{\sigma}$ )では検体間のばらつきが大きくなる傾向があります。  
このため、弊社では平均値を使用したGlobal normalization ( $X' = X - \mu$ )を採用しています。

# Global normalizationとZ-Score

## 【正規化】

### Global normalization

各スポットの値と平均値の差

$$X' = X - \mu$$

$X$ : 変数  
 $\mu$ : 平均値

### Z-score

各スポットの値と平均値の差を標準偏差で割る

$$Z = \frac{X - \mu}{\sigma}$$

$X$ : 変数  
 $\mu$ : 平均値  
 $\sigma$ : 標準偏差

▶ 正規化には様々な方法があり、目的にあった方法を選択する必要がある。

## 弊社の正規化方法

弊社では下記の理由からGlobal normalizationを採用しています。

- 抗体の反応がなくとも各スポットの値にはマイクロアレイ間でばらつきがある。
- 標準偏差で補正（Z-score）をかけた場合、上記のばらつきが大きくなる傾向にある。  
⇒ 結果的にマイクロアレイ間（検体間）のばらつきも大きくなる。
- Z-scoreを使用することで陽性判定のための閾値が検体によってばらつく。  
⇒ 検体間の比較解析が難しくなる。



# 解析データとデータの解説

## 【18K抗原マイクロアレイ】タンパク質マイクロアレイを用いた血中自己抗体プロファイリング

免疫グロブリンのクラス: ヒトIgG

データセット: 18K (18,429スポット) ヒトタンパク質

### 表の見方

- ・各行はマイクロアレイ上のスポットをあらわし、1行が1種類のヒト抗原タンパク質（遺伝子）に対応します
- ・列「Array 1」は、ネガティブコントロール（NC）としてサンプルを加えず、レファレンス抗体とそれぞれの2次抗体を使って染色して得られたデータです
- ・列「Array 2」は、レファレンス抗体、各サンプル（ヒト血清）とそれぞれの2次抗体を使って染色して得られたデータです
- ・列「2次比」の数値は、ネガティブコントロール（NC）との相対比で、各サンプル（ヒト血清）由来の反応の強さを示します

### 用語の説明

Green Intensity: レファレンス抗体とその2次抗体由来の緑色の蛍光強度

Red Intensity: サンプルとその2次抗体由来の赤色の蛍光強度

Median of Ratios: 緑色の蛍光強度と赤色の蛍光強度の比の中央値

Raw Data: 演算を加えていない生データ

Normalized Data: アレイ間の比較のために正規化したデータ

2次比: アレイ間の相対比データ ※ネガティブコントロール（NC）との相対比

### 染色条件

アレイフォーマット: 18K抗原マイクロアレイ（ヒトタンパク質）

サンプル: ヒト血清

レファレンス抗体: Goat Reference Antibody Mixture I

2次抗体（Red）: Alexa Fluor 647 anti-Human IgG antibody

2次抗体（Green）: Cy3 anti-Goat IgG antibody

搭載ヒト抗原タンパク質情報					Green Intensity F532 Median - B532		Red Intensity F635 Median - B635		Median of Ratios (Log2)				
#	Gene symbol	Gene ID	Gene name	Mutation	Array 1 NC	Array 2 Sample1	Array 1 NC	Array 2 Sample1	Raw		Normalized		2次比
									Array 1 NC	Array 2 Sample1	Array 1 NC	Array 2 Sample1	Array 2/1 Sample1
1	[A1BG]		1 alpha-1-B glycoprotein	標準型	237	228	4	1	-5.31	0.77	-0.31	3.92	4.23

### Normalized data

アレイごとにスポットの値と平均値の偏差を算出

$$X'_i = X_i - \mu \quad \text{例) } -0.31 = -5.31 - (-5.00)$$

平均値  $\mu$

$X_i$

$\vdots$

$X'_i$



# 解析データとデータの解説

## 【18K抗原マイクロアレイ】タンパク質マイクロアレイを用いた血中自己抗体プロファイリング

免疫グロブリンのクラス: ヒトIgG

データセット: 18K (18,429スポット) ヒトタンパク質

### 表の見方

- 各行はマイクロアレイ上のスポットをあらわし、1行が1種類のヒト抗原タンパク質（遺伝子）に対応します
- 列「Array 1」は、ネガティブコントロール（NC）としてサンプルを加えず、レファレンス抗体とそれぞれの2次抗体を使って染色して得られたデータです
- 列「Array 2」は、レファレンス抗体、各サンプル（ヒト血清）とそれぞれの2次抗体を使って染色して得られたデータです
- 列「2次比」の数値は、ネガティブコントロール（NC）との相対比で、各サンプル（ヒト血清）由来の反応の強さを示します

### 用語の説明

- Green Intensity: レファレンス抗体とその2次抗体由来の緑色の蛍光強度
- Red Intensity: サンプルとその2次抗体由来の赤色の蛍光強度
- Median of Ratios: 緑色の蛍光強度と赤色の蛍光強度の比の中央値
- Raw Data: 演算を加えていない生データ
- Normalized Data: アレイ間の比較のために正規化したデータ
- 2次比: アレイ間の相対比データ ※ネガティブコントロール（NC）との相対比

### 染色条件

- アレイフォーマット: 18K抗原マイクロアレイ（ヒトタンパク質）
- サンプル: ヒト血清
- レファレンス抗体: Goat Reference Antibody Mixture I
- 2次抗体（Red）: Alexa Fluor 647 anti-Human IgG antibody
- 2次抗体（Green）: Cy3 anti-Goat IgG antibody

搭載ヒト抗原タンパク質情報					Green Intensity F532 Median - B532		Red Intensity F635 Median - B635		Median of Ratios (Log2)		2次比		
#	Gene symbol	Gene ID	Gene name	Mutation	Raw		Normalized		Array 2/1 Sample1				
					Array 1 NC	Array 2 Sample1	Array 1 NC	Array 2 Sample1					
1	[A1BG]		1 alpha-1-B glycoprotein	標準型	237	228	4	381	-5.31	0.77	-0.31	3.92	4.23

$A_i$        $B_i$        $C_i$

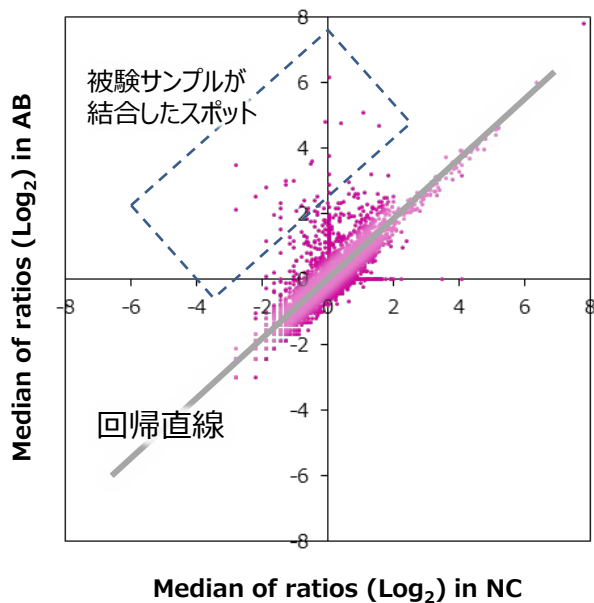
### 二次比

SampleとNCの差を算出

$$C_i = B_i - A_i \quad \text{例) } 4.23 = 3.92 - (-0.31)$$

# 正規化データと二次比データについて

## 正規化データの散布図

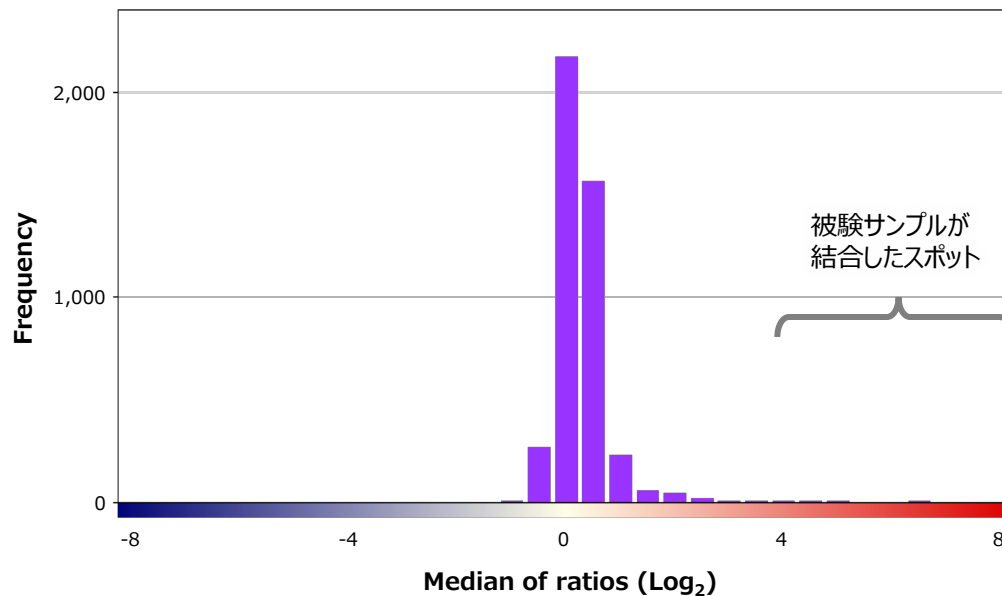


NC: ネガティブコントロール (レファレンス抗体のみで染色)  
AB: 被験サンプル

大部分のスポットが回帰直線近傍に分布

▶ Y軸の+側に移行したスポットが被験サンプルが結合しているスポット

## 二次比データのヒストグラム



二次比データ:  
Median of ratios ( $\text{Log}_2$ ) in AB - Median of ratios ( $\text{Log}_2$ ) in NC

大部分のスポットが0付近に分布

▶ +側に大きく移動したスポットが被験サンプルが結合しているスポット

# データ解析のまとめ

## スポットの検出

## 正規化

## 2次比

### 解析

- ・スキャナーでスポットを自動認識
- ・手動で補正
- ・Median of Ratiosをスポットの値として解析  
(底2の対数に変換し、以降の解析に使用)

### 解析対象

- ・スキャナーで認識されたスポット  
(対数値で0となったスポットは欠損値と分けて解析)

### 欠損値

- ・スキャナーで認識されなかったスポット
  - ・バックグラウンドの蛍光が強く、Median of Ratiosが異常値となったスポット
- ▶ 欠損値 (0) として取扱い

### 解析

- ・振幅の大きい上下各16%を除いた中央68%の値から平均値を算出
- ・各タンパク質の値と前述の平均値の差を算出

$$X' = X - \mu$$

X: 変数  
μ: 平均値

### 解析対象

(スポットの検出と同一)

### 欠損値

(スポットの検出から増減なし)

### 解析

- ・各スポットについてサンプルの測定値(AB)とネガティブコントロール(NC)の測定値の差を算出  
(対数値なので割り算)

$$\text{Median of ratios (Log}_2\text{) in AB} \\ - \text{Median of ratios (Log}_2\text{) in NC}$$

### 解析対象

- ・サンプル、ネガティブコントロールともに検出できたスポット

### 欠損値

- ・サンプル、またはネガティブコントロールのいずれかが検出できなかったスポット

▶ 欠損値 (0) として取扱い

# 解析例

## 抗体の評価

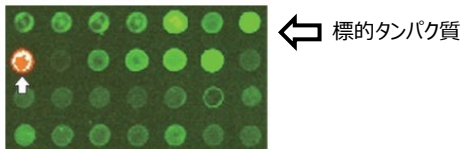
- 抗体の抗原への結合・交差反応を検証
- 抗原未知の抗体の抗原探索

### 交差性評価

- 抗体の交差反応性を評価

#### 【実施例】

市販抗体の結合性、並びに交差反応性評価



Rank	種	分類	Score
<b>On target</b>			
1	West Nile virus	順相	12.29
<b>Off target</b>			
2	Pseudomonas aeruginosa	順相	4.68
3	Candida glabrata	順相	2.93
4	Rhinovirus A	逆相	2.71
5	Human orthopneumovirus	順相	2.59
6	Rift Valley fever phlebovirus	順相	2.55
7	Mycobacterium tuberculosis	順相	2.40
7	Rabies lyssavirus	順相	2.40
9	Mycoplasma pneumoniae	逆相	2.37
10	Lactobacillus collinoides	逆相	2.37
11	Nocardia pseudobrasiensis	逆相	2.32
12	Human mastadenovirus D	逆相	2.31
13	Kluyvera georgiana	逆相	2.30
14	Influenza B virus	順相	2.24
15	Rous sarcoma virus	順相	2.12
16	Avian leukosis virus	逆相	2.09
17	Providencia stuartii	逆相	2.05
18	Salmonella enterica	順相	2.02
19	Rhinovirus A	逆相	1.99
20	Chlamydia pneumoniae	順相	1.99

## 体液中の抗体プロファイリング

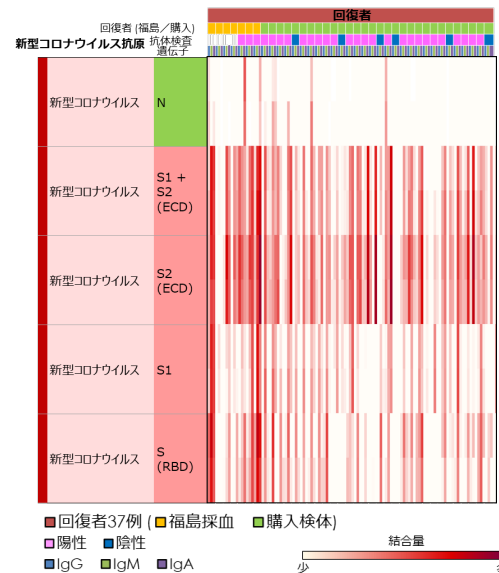
- 全血、血清、血漿、脳脊髄液、尿、唾液、涙液、乳汁、精液などに含まれる抗体のプロファイルを取得
- 疾患に特異的な抗体（疾患マーカー）の探索

### 抗体検査

- 任意の値以上のスコアのタンパク質を抽出
- 標的タンパク質に対する抗体の有無を評価

#### 【実施例】

新型コロナウイルス感染症回復者の血中抗体評価



### 抗体プロファイリング

- サンプル間や群間で差が大きいタンパク質を抽出
- クラスタ分析（階層的クラスタリング）や多群比較など 目的に合わせ様々な解析が可能

#### 【クラスタ分析】

- サンプル間のユークリッド距離を算出

$$d(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$$

- クラスタ間の距離を群平均法で算出

クラスタ (X, Y) の中から、データの一つづつ選択 (X<sub>i</sub>, Y<sub>j</sub>) 距離 {d(X<sub>i</sub>, Y<sub>j</sub>)} の平均値をクラスタ間の距離として採用

$$d(X, Y) = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m d(X_i, Y_j)$$

- ▶ サンプル、クラスタ間の距離により階層化

#### 【多群比較】

- コントロール群で変動が小さく、解析対象群で変動が大きいタンパク質を抽出
- Turkey法により群間の違いを解析