



Agilent GeneSpring および Mass Profiler Professional の相関分析

「変数 X と変数 Y の間に関係があるか否かは、データ分析者にとって重要な問題です。変数 X と変数 Y の 2 つの相関を検証することで答えは得られます」

Chen, P; Popovich, P. 「Correlation: Parametric and Nonparametric Measures」 Sage Publications, 2002

技術概要

著者

Pritha Aggarwal, Durairaj Renu, and
Pramila Tata
Strand Life Sciences
Bangalore, India

Michael Rosenberg
Agilent Technologies, Inc.
Santa Clara, California, USA

はじめに

相関分析を使用すると、研究対象のサンプル間の関係を求めるように、遺伝子や代謝物などの関連し合う分子を推定することができます。相関分析の機能は、マイクロアレイや質量分析計、次世代シーケンサー (NGS) などのハイスループットオミクスプラットフォームの装置のデータをサポートしている GeneSpring/Mass Profiler Professional (MPP) 13.0 ソフトウェアで導入されました。相関分析の機能ではペアワイズ相関をサポートしています。ペアワイズ相関は、同じテクノロジーで測定した結果で行うこともできますし、2 つの異なるテクノロジーで測定した結果で行うこともできます。

この技術概要では、GeneSpring/MPP 13.0 でサポートする相関分析の詳細について説明します。



Agilent Technologies

相関についての重要な概念

相関分析は、最もよく使用されている統計手法の1つです。相関では、2つの定量的変数間の線形関係の強度と方向性を測定します (<http://ja.wikipedia.org/wiki/相関係数>)。

- 変数 X が大きくなるのに伴って変数 Y が大きくなることを正の相関があるといいます。変数 X が小さくなるのに伴って変数 Y が大きくなることを負の相関あるいは反相関があるといいます。
- 相関値は -1~+1 です。2つの変数が正の相関にあるとき、相関係数の値は +1 に近づきます。+1 は完全な相関を示します。同様に、2つの変数が負の相関にあるとき、相関係数は -1 に近づきます。相関係数がゼロに近い場合、2つの変数の間に大きな依存性はありません。
- 相関は2つの変数間の線形関係の度合いのみを表します。変数間の原因と結果の関係を意味するものではありません。
- 直接的な直線回帰の場合、適合度は二乗ピアソン相関係数 (図1の R^2 と R) と等しくなります。このため、相関は、回帰直線からの Experiment データの偏差より求めた値として考えられます。

- 相関分析するには、最低3つのデータポイントが必要です。定義より、任意の2つのデータポイントの相関係数は+1、-1、または未定義です。
- 相関の計算には多くの方法がありますが、最もよく使用されるのは、ピアソンとスピアマンの相関係数です。ピアソンの相関係数 (http://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient) は、正規分布するランダムな変数における直線関係の強さを見るために広く使用されています。スピアマンの相関係数 (<http://ja.wikipedia.org/wiki/スピアマンの順位相関係数>) は、順位相関のアルゴリズムで、データセットの外れ値に対して許容度が高くなります。ピア

ソン指標とスピアマン指標の両方を、GeneSpring/MPP 相関分析で使用することができます。

- 生物系では、正の相関が転写活性化因子とターゲット遺伝子との間に観察されます。一方、miRNA のような阻害剤とその mRNA ターゲットは負の相関を示します。

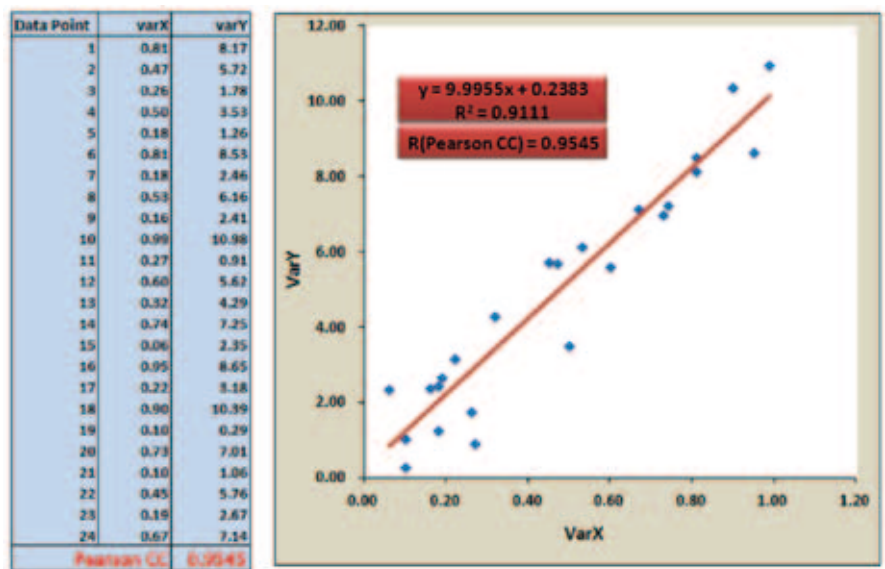


図1. 変数 X と変数 Y の間のピアソン相関係数 (CC)。Experiment によるデータの散布図と回帰直線が X と Y の間の正の相関を表しています。

エンティティ間の相関

相関分析はペアワイズ法で実行され、アバンドンスレベルで対の生物学的エンティティ間の依存性を特定および観察します。GeneSpring/MPP でのエンティティは、遺伝子、代謝物、タンパク質、発現アレイ内のプローブです。

相関分析を実行するオプションは、ソフトウェアの実験 (Experiment) の Workflow Browser に含まれています。相関分析は、単一の Experiment 内のエンティティまたは 2 つの異なる Experiment

にわたるエンティティに対して実行できません。クロス Experiment 相関分析の場合、相関のためにエンティティが選択された 2 つの Experiment を使用して、マルチオミックス分析 (MOA) Experiment が GeneSpring に作成されます。表 1 は、相関分析をサポートする GeneSpring の Experiment のタイプをまとめています。

単一の Experiment からのエンティティの相関分析と 2 つの Experiment からのエンティティの相関分析については、それぞれ図 2 と図 3 で説明しています。

表 1. 相関分析に使用可能な GeneSpring での分析タイプ。

分析タイプ	単一の Experiment	2 つの Experiment
mRNA 発現	○	○
エクソン発現	○	○
miRNA	○	○
RTPCR	○	○
DNA-Seq	×	×
RNA-Seq	×	○*
smallRNA-Seq	×	○*
メタボロミクス	○	○
プロテオミクス	○	○

* RNA-Seq および smallRNA-Seq Experiment で相関を実行するには、Strand NGS v2.1ソフトウェア (<http://www.strand-ngs.com/>) で作成したデータを GeneSpring にインポートしてください。

相関分析の入力

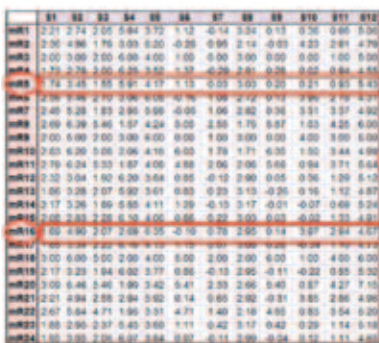
単一 Experiment 内または MOA Experiment 内で実行される相関分析の結果は、分析のための入力として選択するエンティティリスト、解析、相関係数のタイプによって決定されます。

エンティティリスト: 選択したエンティティリスト内のエンティティ間のペアワイズ相関が計算されます。GeneSpring/MPP では、アクティブな Experiment から 1 つのエンティティリストを、または MOA Experiment 内の 2 つの Experiment から異なる 2 つのエンティティリストを選択できます。

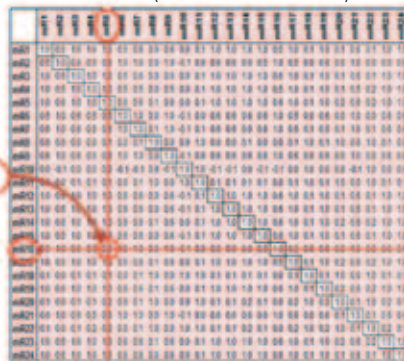
解析: GeneSpring で任意の他の分析として、平均化された解析が選択された場合、複数の Experiment にわたる各エンティティの平均強度値が分析に使用されます。平均化されない解析では、各サンプル内のエンティティの強度値が解析に使用されます。

サポートされる相関のタイプ: GeneSpring 13.0 の相関分析フレームワークでは、ピアソンとスピアマンの相関係数をサポートしています。他のタイプの相関係数は今後のリリースで追加される予定です。GeneSpring では、エンティティの組み合わせ間の相関を計算するために、最低 3 つの有効なデータポイントが必要です。平均化されない解析では、最低 3 つのサンプルが解析の一部として含まれる必要があります。平均化された解析では、最低 3 つの条件が必要です。

エンティティの発現量 (アバンドンス)



相関係数 (Correlation coefficient)



相関ヒートマップ (Correlation heatmap)

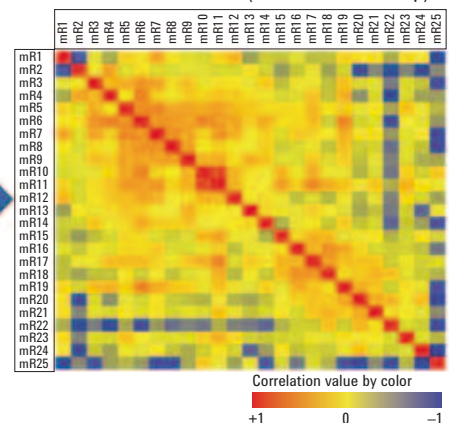


図 2. 単一の Experiment のエンティティ間の相関係数の計算。エンティティの発現量の表は、サンプルのセット内で選択されたエンティティの正規化後の強度 (アバンドンス) を基に作成されています。ペアワイズの相関係数 (ピアソンまたはスピアマン) は、エンティティの発現量の行の各組み合わせの間で計算され、ヒートマップで表示されます。

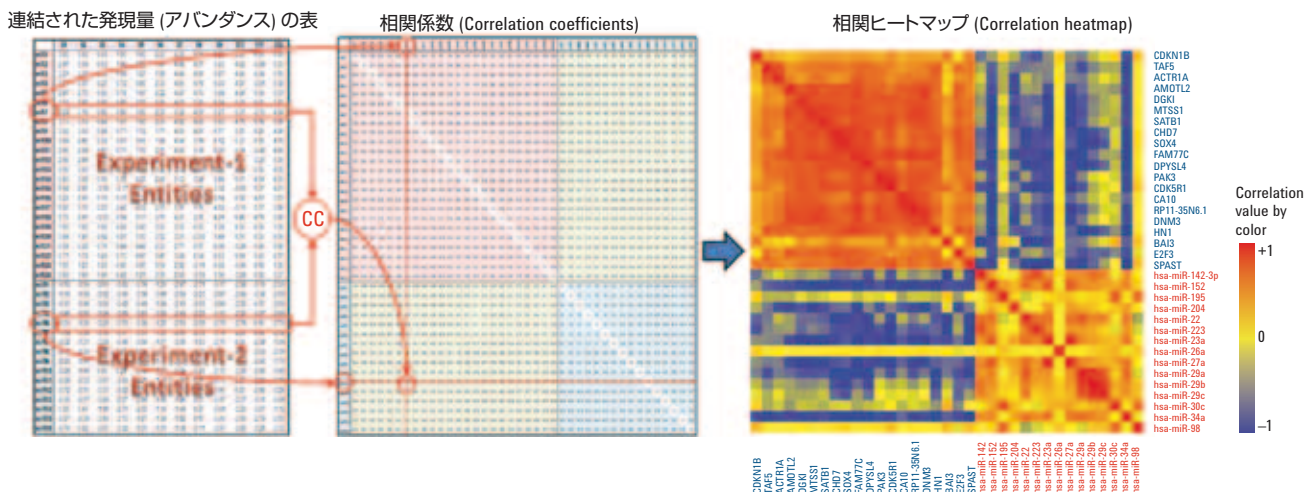


図 3. 2つの異なる Experiment のエンティティ間でのペアワイズの相関の計算。単一の Experiment の相関分析と同様に、エンティティの発現量の各表は Experiment 1 と Experiment 2 から作成されています。2つのエンティティの発現量の表は、ユーザが提供するサンプルのペアリング情報を基に作成されます。クロス Experiment での相関分析では、2つの異なるプラットフォームによって同一または類似の生体サンプルが測定されることとなります。m エンティティが Experiment 1 から、n エンティティが Experiment 2 から選択される場合、連結されたヒートマップには m+n のエンティティがそれぞれの軸にあります。計算された相関係数の値は、ヒートマップとして表されます。

相関ヒートマップ

選択したデータベースの相関係数は、ヒートマップとして表示されます。相関ヒートマップは複数のエンティティ間の関係を理解するために便利なツールです。ヒートマップ内の各セルの色は、ヒートマップの X 軸と Y 軸でのエンティティ間のペアワイズ相関係数の値によって定義されます。

ヒートマップ内のエンティティの順序は、X と Y の次元で同じです。この関係は、単一の Experiment のエンティティまたは MOA

Experiment での 2つの異なる Experiment のエンティティ間での相関にも適用されます。結果のヒートマップは正方形で対角線は各エンティティ自身 (すなわち、対角線上のすべての値は +1) の相関を表しています。単一の Experiment と MOA Experiment でのエンティティの相関のヒートマップ表示を、それぞれ図 2 と図 3 に示します (右端のパネル)。MOA での表示は図 4 に示すように、すべてのエンティティのヒートマップとクロス Experiment エンティティのヒートマップとを切り替えることができます。

相関ヒートマップ内の各セルについてのデータは、散布図として確認できます。散布図は、そのエンティティの組み合わせ間での相関係数の計算に使用する、Experiment からのアブダンスの値をグラフ化したものです。散布図内の回帰線と回帰式は、エンティティの組み合わせ間での依存関係の方向と強さを表示します (図 5)。

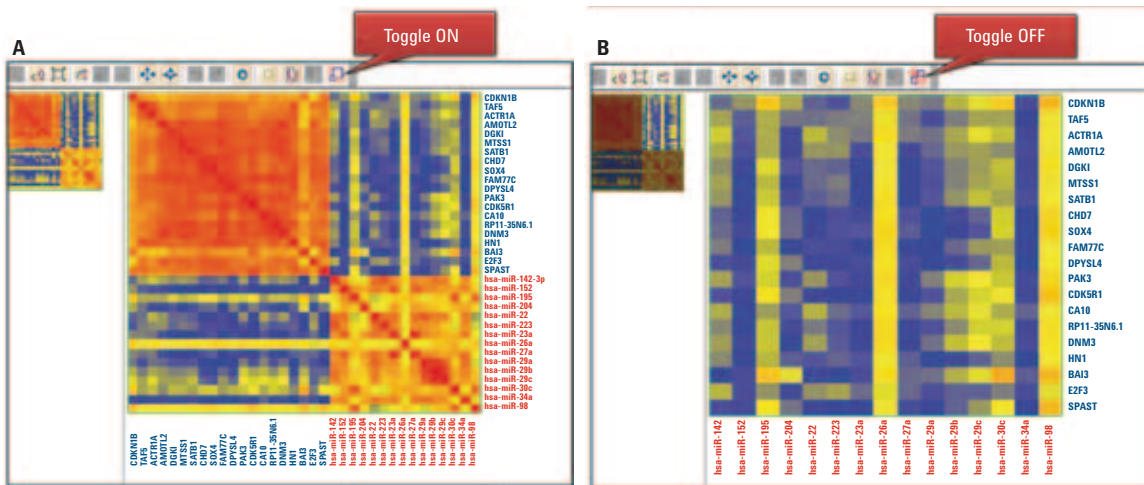


図 4. MOA 相関ヒートマップ。メニューバーの ON/OFF 切り替えを使用して、表示を (A) すべての入力エンティティまたは (B) クロス Experiment のみに切り替えることができます。

相関ヒートマップ内のエンティティのフィルタリング

GeneSpring/MPP で実行した膨大な統計解析の結果は、エンティティリストと関係するデータ、例えば、倍率変化 (Fold change)、 p 値、レギュレーションなどとして保存されます。相関分析では、研究者は任意の関係データを使用して相関ヒートマップをフィルタリングできます。フィルタリングは、Fold change または p 値のような数値データとアップ/ダウンレギュレーションなどの分類上のデータの両方がサポートされています (図 6A)。複数のフィルタが適用されると、適用されるフィルタのすべてを通るエンティティが表示に保持されます (ブール演算 AND)。

相関分析は、外部属性を含むエンティティと関係する任意の属性を基にしたヒートマップのフィルタリングをサポートします。外部エンティティの属性は、すべての関係する値を含むエンティティリストをタブ区切りテキストや Excel ファイル形式でアップロードすることによって GeneSpring に読み込まれます。図 6A に示した例で読み込まれた各遺伝子 (図 6B) には、関連した値として属するパスウェイがあります。パスウェイ名 (例えば、図 6 の受容体型チロシンキナーゼ (RTK)) によってヒートマップをフィルタリングすることに

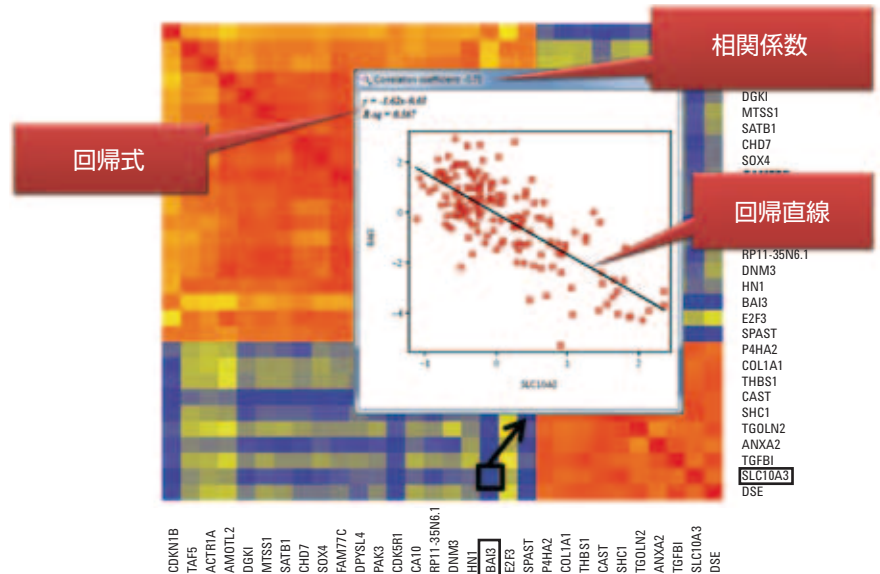


図 5. Experiment によるデータ、回帰直線、相関係数の表示を含む、エンティティの強調表示された組み合わせ間で負の相関を示す散布図。

よって、選択したパスウェイのメンバーの発現量間の関係を調べることができます。

相関分析では、1 つ以上のフィルタをパスしたエンティティを保存することができます。Experiment ナビゲータで、新しいエンティティリストは、作成された相関分析のノードの下に表示されます。例えば、図 6 で、ヒートマッ

プはフィルタリングされて RTK パスウェイのメンバーのみを表示しています。RTK パスウェイのメンバー間の相関は保存され、以降の分析やリファレンスに使用できます。

MOA 相関では、Experiment ごとに 1 つ、つまり 2 つのフィルタリングタブが提供されます。

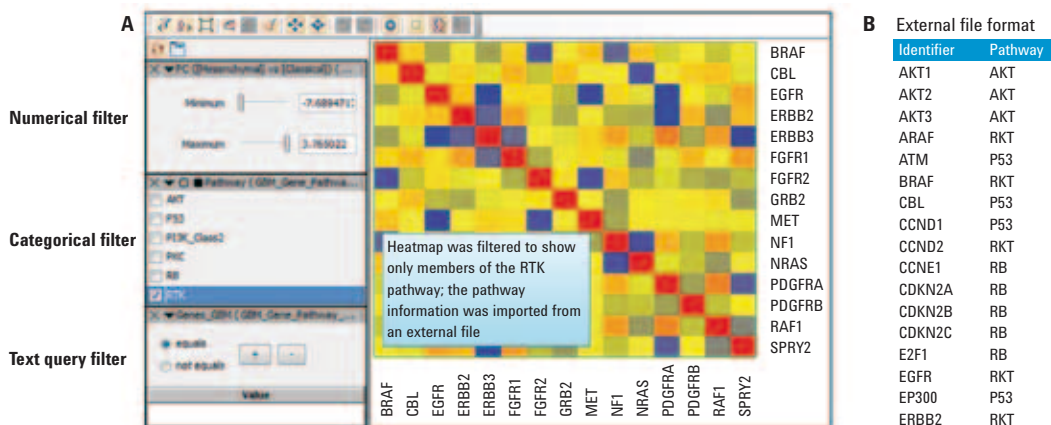


図 6. 数値および分類でのフィルタリング。A) の例で、Fold Change (FC) は数値フィルタ、Pathway および Genes_GBM は分類フィルタです。分類数が 30 未満の場合、フィルタパネルにチェックボックスとして表示されます (図の Categorical filter を参照)。30 以上になるとテキストボックスが表示されます (図の Text Query filter を参照)。B) は、外部エンティティリストとフィルタリングに使用する値の例です。

相関ヒートマップでのエンティティのクラスタリング

GeneSpring/MPP 相関分析を使用すると、発現量の値ではなく相関係数を基に、階層的にエンティティをクラスタリングすることができます。同時発現するエンティティは共通の生物学的機能を共有する可能性が高いことから、クラスタリングの結果は新しい仮説を示唆するか既存の仮定を裏付けることになるため、階層クラスタリングは重要な機能です。相関分析でサポートされるクラスタリングパラメータを、表 2 にまとめています。

単一の Experiment では、クラスタリングは、所定のエンティティリスト内のすべてのエンティティ間のペアワイス相関係数を基に

実行されます。MOA Experiment では、クラスタリングは、クロス Experiment 相関係数のみを基に実行されます (図 7)。クラスタリングより前にフィルタリングオプションが適用されていると、単一のテクノロジーまたは MOA Experiment のいずれの場合も、クラスタリングはフィルタをパスしたエンティティのみで実行されます。クラスタリングの後でフィルタリングが適用されると、クラスタリングデンドログラムは削除され、ヒートマップでのエンティティの順序はデフォルトにリセットされます。

表 2. パラメータのサマリは、GeneSpring 内での階層型クラスタリングのためにサポートされています。

パラメータ	値
距離メトリック	Euclidean, Squared Euclidean, Manhattan (Cityblock), Maximum (Chebychev), Minimum, Differential, Canberra, Harmonic, Pearson's Centered, Pearson's Uncentered (Cosine), Pearson's Centered - Absolute, Pearson's Uncentered - Absolute
リンケージルール	Average, Centroid, Ward's, Median, Single, Complete

このデンドログラムは行間の距離(類似度)に基づいて作成されています。

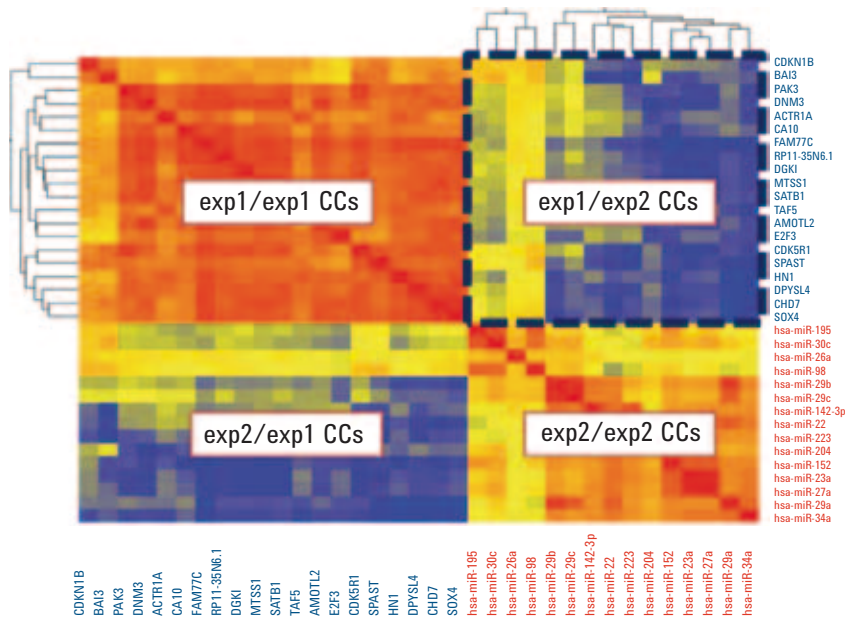


図 7. MOA 内のクラスタリングは、クロス Experiment の相関値で定義されたプロフィールを使用しています。

相関係数のエクスポート

相関分析は、すべてのエンティティ、フィルタリング済みのエンティティ、選択済みのエンティティの相関係数のエクスポートをサポートします。関係するデータとアノテーションをエクスポートすることもできます (図 8)。

エンティティの強調表示と選択

GeneSpring/MPP 内の相関分析では、選択されたエンティティリストと整合性のある相関ヒートマップ内のエンティティのサブセットを強調表示できます。選択したリストと整合性のあるエンティティは、対応する行と列の隣にある赤線によって強調表示されます。デフォルトでは、整合性のないエンティティは透明度レベルが 0 に設定され、見えなくなります (図 9)。透明度はカスタマイズ可能で、ヒートマップのプロパティダイアログで変更できます。

相関ヒートマップ内のエンティティを選択して、エンティティリストの選択肢として保存することができます。保存したエンティティリストは、他のエンティティリストと同様に GeneSpring でさらなる分析に使用できます。例えば、あるクラスタを構成するエンティティを選択し、エンティティリストとして保存します。保存したエンティティリストに対して遺伝子オントロジー解析またはパスウェイ解析を実行すると、対象のクラスタで改善される生物学的機能またはパスウェイを特定することができます。

Identifier	CC's										COLUMN ANNOTATIONS									
	E1_2882	E1_3882-2	E1_MVT1	E1_CRMP1	E1_MVT1	E1_SOX11	E1_2882	E2_hsa-miR-10a	E2_hsa-miR-143	E2_hsa-miR-155	E2_p[Down]	E2_Regulat	E2_FC [abs]	E1_chr	E1_EntrezGene	E1_map_location	E1_Symbol	E1_Ortholog	chr2882	
E1_2882	1.000	0.583	0.707	0.524	0.603	0.526	0.507	-0.218	-0.232	-0.100	-0.282			2	3825	2q27	2882		Hs.34871	
E1_3882-2	0.583	1.000	0.670	0.681	0.729	0.706	0.614	-0.100	-0.207	-0.163	-0.309			20	4821	11p15.5	3882-2		None	
E1_MVT1	0.707	0.670	1.000	0.595	0.764	0.633	0.578	-0.195	-0.165	-0.127	-0.378			8	9795	9p22	MVT1		Hs.34694	
E1_CRMP1	0.524	0.681	0.595	1.000	0.735	0.757	0.701	-0.262	-0.223	-0.238	-0.402			4	1800	4p16.3	CRMP1		Hs.135270	
E1_MVT1	0.603	0.729	0.764	0.735	1.000	0.729	0.727	-0.183	-0.119	-0.148	-0.418			20	4861	11p15.5	MVT1		Hs.279162	
E1_SOX11	0.526	0.706	0.633	0.757	0.729	1.000	0.783	-0.100	-0.233	-0.211	-0.252			2	4664	2p25	SOX11		Hs.433488	
E1_2882	0.507	0.614	0.578	0.701	0.727	0.783	1.000	-0.132	-0.065	-0.150	-0.228			5	3627	5q35.3	2882		Hs.138816	
E2_hsa-miR-10a	-0.218	-0.200	-0.195	-0.262	-0.283	-0.300	-0.132	1.000	0.065	0.015	0.214	1.73E-04	down		3,433209					
E2_hsa-miR-143	-0.100	-0.207	-0.165	-0.223	-0.119	-0.218	-0.065	0.065	1.000	0.758	0.085	2.11E-02	down		3,12976					
E2_hsa-miR-155	-0.163	-0.163	-0.127	-0.238	-0.188	-0.211	-0.150	0.015	0.758	1.000	0.091	4.00E-03	down		1,844095					
E2_p[Down]								1.71E-04	2.11E-02	4.00E-03	5.94E-10									
E2_Regulation								down	down	down	down									
E2_FC [abs]								1.451098417	1.329760313	1.446395139	2.045527779									
E1_chromosome	2	20	8	4	20	2	5													
E1_EntrezGene	3825	4821	9795	1800	4861	4664	3627													
E1_map_location	2q27	11p15.5	9p22	4p16.3	11p15.5	2p25	5q35.3													
E1_Symbol	2882	NXK2-2	MVT1	CRMP1	MVT1	SOX11	2882													
E1_Ortholog	Hs.34871	None	Hs.34694	Hs.135270	Hs.279162	Hs.433488	Hs.138816													

図 8. 相関係数をエクスポートした後の出力ファイルの表示。ペアワイス相関係数の値とともに、エンティティと関係するデータおよびアノテーションが行と列にエクスポートされます。相関係数は橙色の部分に、列アノテーションは青色の部分に、行アノテーションは緑色の部分に示されています。

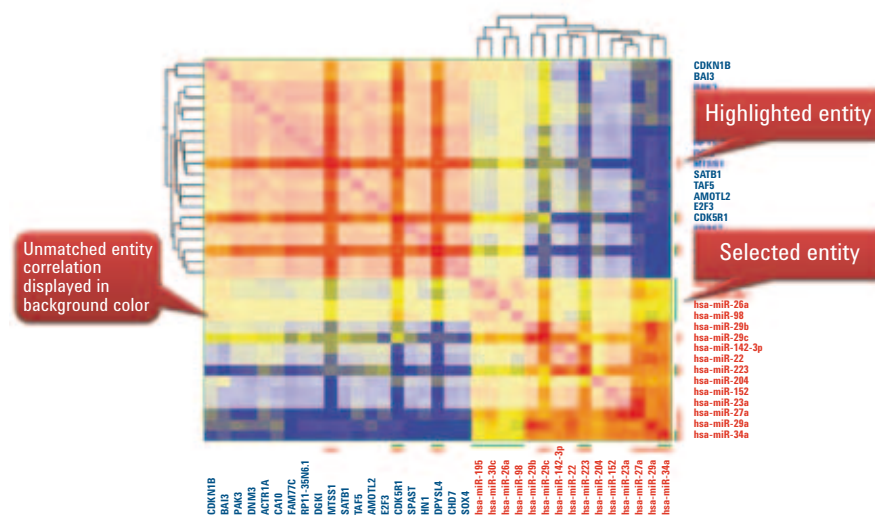


図 9. 強調表示されたエンティティは赤線で、選択されたエンティティは緑色の線で示されています。強調表示や選択を示すために使用する色は、ヒートマップダイアログで変更できます。

サンプル間の相関

エンティティ間の相関に加え、GeneSpring/MPP 相関分析は、その Experiment での複数の生体サンプル間のペアワイズ相関分析をサポートしています。サンプル相関により、研究内のサンプル間に存在するコンディションワイズの関係を特定できます。

サンプル相関は、エンティティ相関でサポートされているものと同じ Experiment タイプで実行されます (表 1)。サンプル相関は、単一 Experiment 内のサンプル間でのみサポートされ、MOA Experiment ではサポートされません。

サンプル相関分析は、分析のために入力、選択するエンティティリスト、解析、相関係数のタイプによって定義されます。表 3 は、Gene 1~12 の信号強度値を基に、サンプル 1 とサンプル 2 の間で計算された相関の例を示し

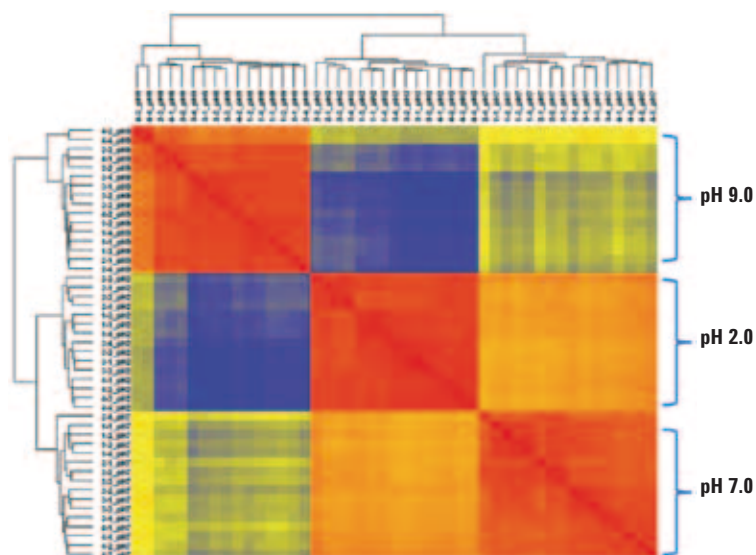
ています。選択されたエンティティリストと解析は、相関分析用のサンプルとエンティティを定義するために使用されます。

サンプル間の相関分析は、Experiment 作成中に実行されるベースラインの変化に影響されます。分析から導き出される生物学的推定は、ベースラインが変更されたデータを使用して変更することができます。ベースラインが変更されていないデータでのサンプル間相関の実行を推奨します。

サンプル間の相関の結果はヒートマップとして表示されます。ヒートマップ内のサンプルは、サンプルの相関プロファイルを基に階層型クラスタリングを使用して計算されず (図 10)。

表 3. 2つの指定サンプル間の相関の計算例。

	サンプル 1	サンプル 2
Gene 1	0.651	1.372
Gene 2	0.818	1.590
Gene 3	0.945	1.716
Gene 4	0.578	0.643
Gene 5	0.464	1.186
Gene 6	0.675	0.947
Gene 7	0.323	0.642
Gene 8	0.304	0.774
Gene 9	0.043	0.783
Gene 10	0.943	1.452
Gene 11	0.908	1.686
Gene 12	0.415	0.808
ピアソン相関係数		0.822



サンプル-サンプルの相関ヒートマップでは、熱帯熱マラリア原虫 (*Plasmodium falciparum*) の感染で pH 間での強い相関があることを示しています。

Tube no.	NRBC	IRBC at 10 %	SLO 250 stock units	Set A	Set B	Set C
1-1	500 µL			pH 2	pH 7	pH 9
1-2	500 µL			pH 2	pH 7	pH 9
1-3	500 µL			pH 2	pH 7	pH 9
1-4	500 µL			pH 2	pH 7	pH 9
2-1	500 µL		10 µL	pH 2	pH 7	pH 9
2-2	500 µL		10 µL	pH 2	pH 7	pH 9
2-3	500 µL		10 µL	pH 2	pH 7	pH 9
2-4	500 µL		10 µL	pH 2	pH 7	pH 9
3-1		500 µL		pH 2	pH 7	pH 9
3-2		500 µL		pH 2	pH 7	pH 9
3-3		500 µL		pH 2	pH 7	pH 9
3-4		500 µL		pH 2	pH 7	pH 9
4-1		500 µL	10 µL	pH 2	pH 7	pH 9
4-2		500 µL	10 µL	pH 2	pH 7	pH 9
4-3		500 µL	10 µL	pH 2	pH 7	pH 9
4-4		500 µL	10 µL	pH 2	pH 7	pH 9

図 10. メタボロミクス研究でのサンプル-サンプル相関ヒートマップ²。相関係数でのクラスタリングにより、感染状況 (NRBC = 非感染の RBC、IRBC = 感染した RBC) ではなく pH 値を基にサンプルがグループ化していることが明らかに示されています。

参考文献

1. 本ドキュメントで示した結果は、TCGA Research Network (<http://cancergenome.nih.gov/>) が作成したデータの全体または一部に記載されています。
2. Sana, T. R 他 Global Mass Spectrometry Based Metabolomics Profiling of Erythrocytes Infected with Plasmodium falciparum. PLoS ONE **2013**, 8(4): e60840. doi:10.1371/journal.pone.0060840.

ノート

ノート

製品情報 (2014年9月現在)

製品番号	製品概要
Mass Profiler Professional	
G3835AA	Mass Profiler Professional (MPP) Perpetual
G9274AA	Mass Profiler Professional (MPP) Perpetual のアップグレード
G3836AA	MPP 用パスウェイ機能 Perpetual
G9275AA	MPP 用パスウェイ機能 Perpetual のアップグレード
G9277AA	サンプルクラス予測 (Perpetual)。Agilent MSD ChemStation または MassHunter を使用し MPP で生成したクラス予測モデルを使用できます。
G9281AA	Mass Profiler Pro (MPP) Concurrent ライセンス。制限なしでインストールできますが、プログラムにアクセスできるのは1度に1ユーザのみです。
G9282AA	Mass Profiler Pro (MPP) Concurrent ライセンスのアップグレード。G9281AA の事前購入が必要です。
GeneSpring	
G5886AA	GeneSpring GX Standard Perpetual Academic + 1 年 SMA
G5887AA	GeneSpring GX Standard Perpetual Commercial + 1 年 SMA
G5888AA	GeneSpring GX Standard アップグレード - Academic
G5889AA	GeneSpring GX Standard アップグレード - Commercial
G5890AA	GeneSpring GX Concurrent Perpetual Academic + 1 年 SMA
G5891AA	GeneSpring GX Concurrent Perpetual Commercial + 1 年 SMA
G5892AA	GeneSpring GX Concurrent Perpetual アップグレード - Academic
G5893AA	GeneSpring GX Concurrent Perpetual アップグレード - Commercial
G3784AA	GeneSpring GX スタンドアロン 1 年 - Academic
G3782AA	GeneSpring GX スタンドアロン 2 年 - Academic
G3780AA	GeneSpring GX スタンドアロン 3 年 - Academic
G3783AA	GeneSpring GX Concurrent 1 年 - Academic
G3781AA	GeneSpring GX Concurrent 2 年 - Academic
G3779AA	GeneSpring GX Concurrent 3 年 - Academic
G3778AA	GeneSpring GX スタンドアロン 1 年 - Commercial
G3776AA	GeneSpring GX スタンドアロン 2 年 - Commercial
G3774AA	GeneSpring GX スタンドアロン 3 年 - Commercial
G3777AA	GeneSpring GX Concurrent 1 年 - Commercial
G3775AA	GeneSpring GX Concurrent 2 年 - Commercial
G3773AA	GeneSpring GX Concurrent 3 年 - Commercial

www.agilent.com/chem/jp

本文書に記載の情報、説明、製品仕様等は
予告なしに変更されることがあります。

アジレント・テクノロジー株式会社
© Agilent Technologies, Inc., 2014
Published in Japan, September 29, 2014
5991-5165JAJP



Agilent Technologies