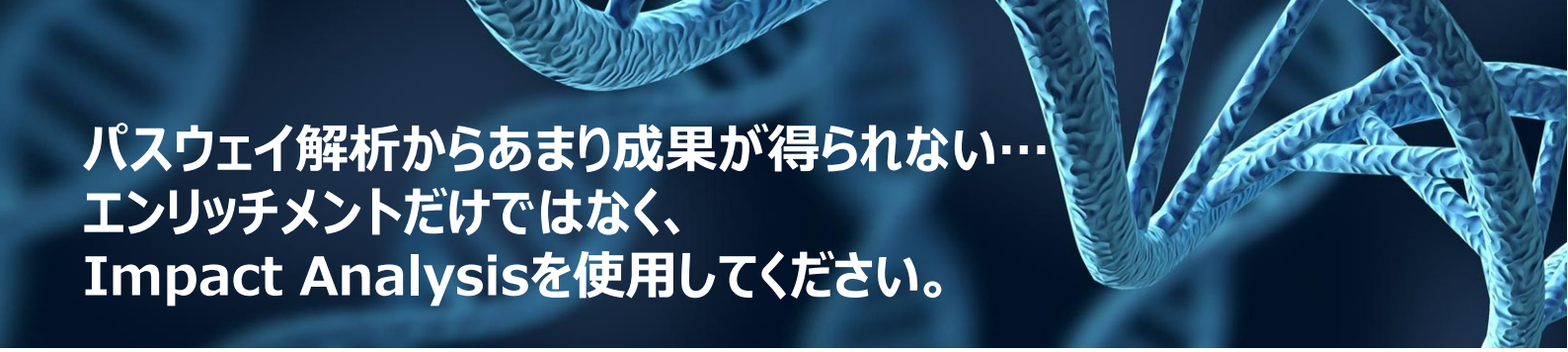


パスウェイ解析からあまり成果が得られない…
エンリッチメントだけではなく、
Impact Analysisを使用してください。

ADVAITA
Bioinformatics for the Extraordinary

パスウェイ解析や、遺伝子変異解析に関するソリューションをラインナップしています。

現在、RNASeq やプロテオミクスデータを どのように解析していますか?

現在、DAVID など、パスウェイ解析に利用できる多くの無料ツールのいずれかを使用していますか? それとも、有償のパスウェイ解析ソフトウェアを使用していますか?もしそうなら、パスウェイエンリッチメント解析のアプローチを使用していることになります。最も古いアプローチの1つであるにもかかわらず（あるいはそれゆえに）、パスウェイエンリッチメント解析は、研究対象の疾患において真に影響を受けているパスウェイを特定する能力を著しく制限することになるのです。また、パスウェイエンリッチメント解析では、表現型に本当に起こっている現象とは関係のない偽陽性を多く出してしまい、多くの行き詰まりを招く可能性があります。この投稿では、実際のデータセットに対するパスウェイエンリッチメント解析とimpact analysisを並べて比較する様子をご紹介します。これは、一般的に用いられるエンリッチメント解析では、表現型の原因に本当に関連するパスウェイを特定できないばかりか、偽陽性をもたらすという明らかな例です。一方、 impact analysis解析では、より多くの真陽性と、より少ない偽陽性が報告されます。

Defining our expectations

優れたパスウェイ解析から期待されることを定義することから始めましょう。パスウェイ解析へのインプットは、2つ以上の表現型 (条件) の複数のサンプルで測定された遺伝子セットで、それらの表現型の違いを理解することを目的としています。これらの2つの条件は、病気 vs. 健康、治療 vs. コントロール、薬剤A vs. 薬剤B などです。単純にするために古典的な 治療 vs. コントロール の比較に焦点を当てましょう。このような研究では、与えられた治療によって大きな影響を受けるパスウェイを特定し、そのパスウェイが特定の表現型効果をどのように生み出すかを明らかにすることを目的としています。例えば、ある治療法が特定の遺伝子を不活化する場合、パスウェイ解析はその特定の遺伝子を含むパスウェイを特定することに役立ちます。

biosupport@filgen.jp https://filgen.jp/ 052-624-4388 052-624-4389 フィルジェン株式会社

実験概要

ほとんどの場合、ある症状に本当に関与しているパスウェイを知ることは困難であることから、様々なパスウェイ解析方法を評価することは困難です。ですが、それはほとんどの場合です。では、実際に表現型の原因が正確に分かっているケースを見てみましょう。GEOのGSE 70302データセットには、特定の遺伝子IL1aがマウスでノックアウト(KO-ed)した実験結果が含まれています。この実験では、4匹のIL1a-KOマウスと4匹のC57BL/6コントロールマウスが比較されました。つまり、このケースでは、IL1a遺伝子の不活性化によって生じた表現型を見ています。このことを知らないふりをして、ゲノム全体で測定された発現変化を解析し、変化を正しく特定できるかどうかを確認します。上記で説明したように、**表現型に関連するパスウェイ**を特定するために、優れたパスウェイ解析が必要です。この場合、ノックアウトされたIL1a遺伝子を含むパスウェイを特定することを意味します。これらのパスウェイは、実際に既定の表現型を引き起こしたパスウェイです。パスウェイのリストは次のとおりです。

Pathway Name
Osteoclast differentiation
Cytokine-cytokine receptor interaction
Fluid shear stress and atherosclerosis
Graft-versus-host disease
Type I diabetes mellitus
MAPK signaling pathway
Pertussis
Tuberculosis
Influenza A
Rheumatoid arthritis
Salmonella infection
Prion diseases
AGE-RAGE signaling pathway in diabetic complications
Leishmaniasis
Inflammatory bowel disease (IBD)
Necroptosis
Cellular senescence
Non-alcoholic fatty liver disease (NAFLD)
Measles

IL1a KO実験における真の陽性パスウェイ

つまり、上記のパスウェイは、KOされた遺伝子を含むパスウェイであるため、解析で特定し報告してもらいたいパスウェイです。

パスウェイエンリッチメント解析結果

パスウェイエンリッチメント解析では、あるパスウェイに該当する差次的発現遺伝子（DEG）の数のみが考慮されます。考え方は非常にシンプルで、遺伝子総数に対するDEGの数の割合と、あるパスウェイに該当するDEGの割合を比較する方法です。あるパスウェイに該当するDEGの割合が、遺伝子セット全体のDEGの割合よりも有意に高い場合、そのパスウェイはDEGが「エンリッチ」されていると見なされ、表現型を引き起こした現象に関連していると考えられます。例えば、測定した1万個の遺伝子のうち、1,000個のDEGがあったとすると、全遺伝子の10％がDEGであることを意味します。このセットからランダムに遺伝子を選ぶとすると、そのうちの約10％がDEGになると予想されます。では、それぞれのパスウェイを一つずつ見ていきます。もし、20個の遺伝子を含むパスウェイに5個のDEGが見つかれば、このパスウェイはエンリッチと見なされます。これは、このパスウェイ上のDEGの割合が25%（20個中5個）と、偶然に予想される10%を大幅に上回っているためです。

このアプローチを実際に見てみましょう。以下は、古典的なパスウェイエンリッチメント解析の結果です。

Pathway Name	#Genes on Pathway	#DE Genes on Pathway	pFisher	pPert
Axon guidance	175	21	0.0067	0.3658
MAPK signaling pathway	293	31	0.0076	0.0160
Synaptic vesicle cycle	62	10	0.0079	Mediat
Leukocyte transendothelial migration	115	15	0.0099	0.2594
Fluid shear stress and atherosclerosis	142	17	0.0143	0.0040
Bacterial invasion of epithelial cells	76	10	0.0304	0.7026
Complement and coagulation cascades	87	11	0.0310	0.0170
Morphine addiction	92	11	0.0442	0.8496
Rap1 signaling pathway	213	21	0.0495	0.8066
Leishmaniasis	67	8	0.0797	0.5287

IL1aノックアウトデータのパスウェイエンリッチメント解析結果を、pFisherでソートした表

注目するべきはpFisherの列です。この列にはパスウェイエンリッチメントのアプローチに基づいて計算されたp値が含まれています。KO-ed遺伝子であるIL1aを含むパスウェイが強調表示されています。まず、このアプローチで同定された最も重要な経路は「Axon guidance（軸索誘導）」で、KO されたIL1a遺伝子とは何の関係もありません。さらに、上位3つのパスウェイのうち2つは**偽陽性**です。さらに下に見ていくと、上位10パスウェイのうち、表現型の原因であるIL1a遺伝子を含むパスウェイは3つしかないことがわかります。言い方を変えると、**トップ3の66%、トップ10の70%のパスウェイが偽陽性**ということです。この分析では、誤検出を報告することで、間違いなく無駄足を踏ませることになります。

Impact analysis解析結果

次に、impact analysisの結果を見てみましょう。簡単に説明すると、 impact analysisでは、パスウェイのトポロジー全体をとらえた数学的モデルを構築し、それを用いて各遺伝子の摂動を算出します。そして、これらの遺伝子の摂動をパスウェイ全体の摂動として合算し、観測値と偶然に予想される値を比較してp値が算出されます。これがpPertの欄に示されるp値です。 impact analysisの結果は次のとおりです。

Pathway Name	#Genes on Pathway	#DE Genes on Pathway	pFisher	pPert
Osteoclast differentiation	128	10	0.3617	0.0025
Cytokine-cytokine receptor interaction	271	23	0.1476	0.0030
Fluid shear stress and atherosclerosis	142	17	0.0143	0.0040
GnRH signaling pathway	89	8	0.2501	0.0095
Graft-versus-host disease	65	2	0.9400	0.0115
Type I diabetes mellitus	70	2	0.9551	0.0150
MAPK signaling pathway	293	31	0.0076	0.0160
Complement and coagulation cascades	87	11	0.0310	0.0170
Olfactory transduction	1122	27	1.0000	0.0185
Taste transduction	88	8	0.2406	0.0210

IL1aノックアウトデータに対するimpact analysis解析の結果を、pPert順に並べた表

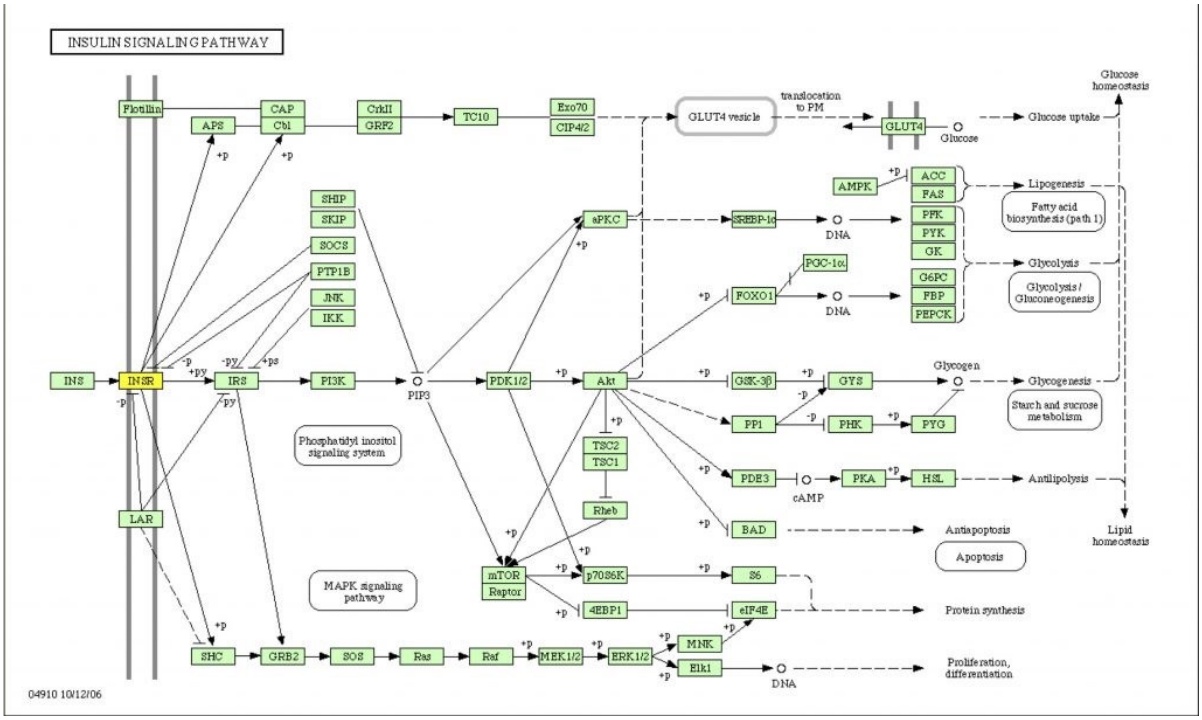
ここで、非常に重要な 2 つのことがわかります。まず、上位3つのパスウェイは、いずれも表現型の真の原因であるIL1a遺伝子が関与する重要なパスウェイです。さらに、pPert列でソートした上位7パスウェイには、6つの真陽性パスウェイと、たった1つの偽陽性が含まれています。

より良いアプローチが必要な場合は、iPathwayGuide の一部としてimpact analysisを利用できます。

なぜこのようなことが起こるのでしょうか？
パスウェイエンリッチアプローチの問題点とは？

シンプルなパスウェイエンリッチメント解析の限界は簡単に説明できます。まず、エンリッチメント解析では、すべての遺伝子が独立であると仮定した統計モデルを用いてp値を算出しています。パスウェイの存在意義は、そのパスウェイに関与する遺伝子間の複雑な依存関係を記述することにあるのです。このように、よく知られた依存関係をすべて無視し、パスウェイ上のすべての遺伝子が独立しているという仮定に基づいて、パスウェイが有意であるかどうかを決定するのは間違いが生じます。

第二に、パスウェイエンリッチメント解析では、パスウェイ上の遺伝子の位置や役割、遺伝子間のシグナルの方向や種類などをすべて完全に無視してしまうことです。各遺伝子がその下流の遺伝子に与える具体的な影響力、すなわち活性化、抑制などを捉えたシグナルがあります。パスウェイで記述された各シグナルと依存関係を証明するためには、おそらく多くの実験と何ヶ月、何年もかかったことでしょう。多くの科学者が何十年もかけて研究してきたこの豊かな知識は、エンリッチメントアプローチでは完全に無視されてしまうのです。パスウェイエンリッチメント解析が捉えるエビデンスは、DEGにおけるパスウェイの相対的なエンリッチメントのみです。たとえば、インスリン経路を見てみましょう。



インスリン受容体INSRは、この経路の重要な入口です。この遺伝子が正しく機能しないと、パスウェイ全体が途切れてしまいます。それにもかかわらず、もしINSRのみがこのパスウェイ上で唯一不活化する遺伝子である場合、パスウェイエンリッチメント解析では有意なp値が得られません。これはパスウェイエンリッチメントアプローチに限り、このパスウェイには、多くの遺伝子のうち、DE遺伝子が1つしかないためです。

一方、impact analysisでは、パスウェイ上の一つ一つの相互作用を捉え、記述する方程式系で構成される数学的モデルを構築します。したがって、このモデルでは、パスウェイ上のすべての遺伝子の位置と役割、ある遺伝子から別の遺伝子へのすべてのシグナルの方向と種類、さらに存在する可能性のあるフィードバックループを考慮します。この方程式系に基づき、impact analysisではすべての遺伝子レベルでの摂動を計算することができます。これらの摂動は、パスウェイ全体のレベルでの摂動を正確に計算し、p値に変換することができるため、本当に影響を受けたパスウェイを特定し、偽陽性を回避するのに役立ちます。

impact analysisを紹介したテクニカルペーパーについては以下を参照してください。

A systems biology approach for pathway level analysis, Genome Research (2007)

A novel signaling pathway impact analysis (SPIA) , Bioinformatics (2009)

Advaita社製品 パスウェイ解析ソフトウェアiPathwayGuideについて(日本語紹介ページ)： https://filgen.jp/Product/BioScience21-software/Advaita/index2.html

【お問い合わせ先】
フィルジェン株式会社 バイオインフォマティクス部
TEL： 052-624-4388 (9:00～17：00)
FAX： 052-624-4389
E-mail： biosupport@filgen.jp