

OpenSWATHによる SWATH-MSデータの解析

- OpenSWATHの使い方
- トランスオミクス解析

東京大学黒田研究室
特任助教

幡野 敦

hatano@bs.s.u-tokyo.ac.jp

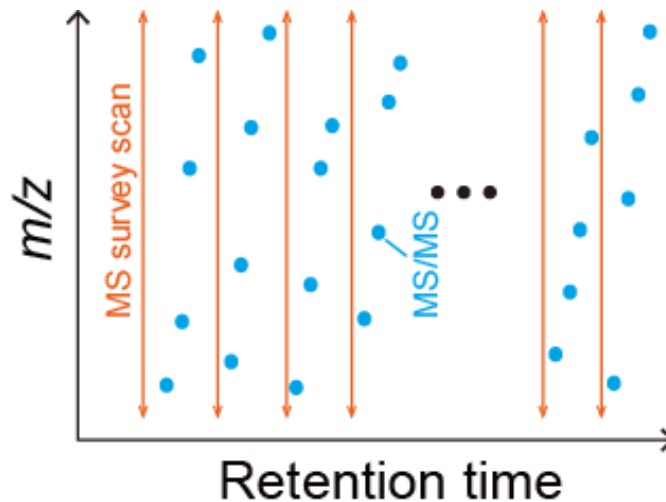
OpenSWATHの使い方

~SWATH-MS~

DDAとDIS (SWATH-MS) の違い

Data-Dependent Acquisition

DDA



データ取得

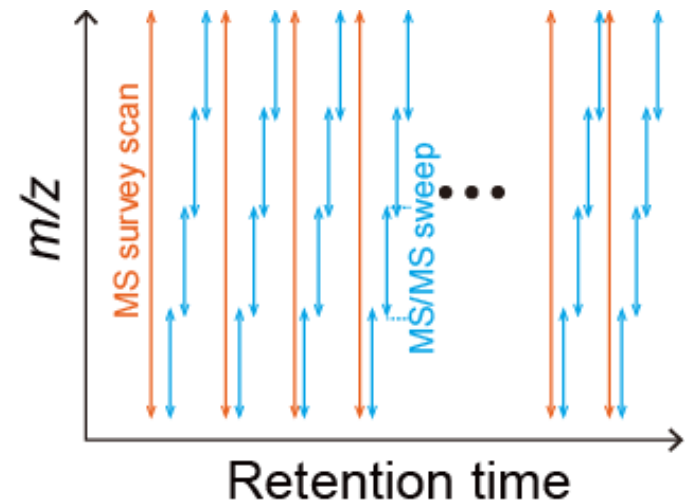
MSおよびMS/MSの
ピークピッキング

解析

▼
サーチエンジンで
リファレンスと照合

Data-Independent Acquisition

DIA (SWATH-MS)



フラグメントイオンの
クロマトグラムの抽出

▼
ライブラリーとの照合

DDAとDIA（SWATH-MS）の違い

Data-Dependent Acquisition DDA

メリット

簡便な解析ツールがある
分画により**Deep proteome**が可能

デメリット

同定の再現性が悪い
プレカーサーイオン定量

適性

少数サンプルで分画してでも
同定数を増やしたい（基質探索など）
さっとデータを取りたい

Data-Independent Acquisition DIA (SWATH-MS)

同定の再現性が高い
複数のフラグメントイオンによる定量

解析パイプラインが煩雑
分画による**Deep proteome**はできない
質の高いライブラリー作りが手間

多サンプルで同じ質のデータを取りたい（システム生物学）
定量性が良いデータを取りたい

特にTTOF5600/6600使用時は
定量的なプロテオーム解析に必要

SWATH-MSの解析の基本的な考え方

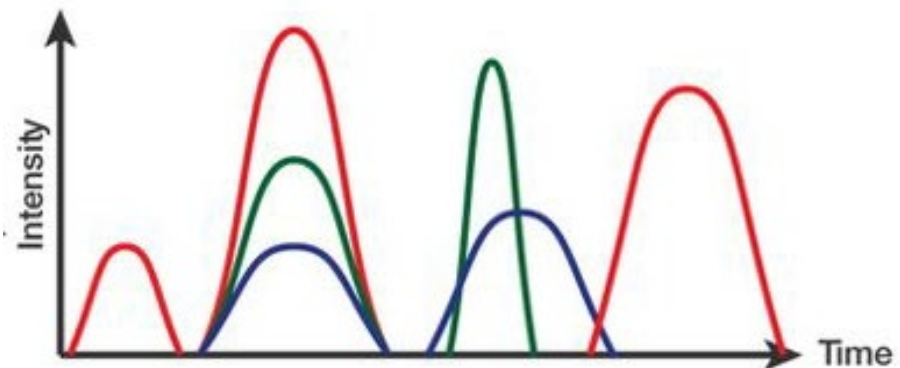
解析

フラグメントイオンの
クロマトグラムの抽出



ライブラリーとの照合

X時間におけるペプチドYに由来する3つの
フラグメントイオンA・B・C（赤・緑・青）



計測条件が一致していればペプチドはライブラリーと同じ溶出時間に溶出する

→“予測された溶出時間のみを検索”

同じペプチド由来のフラグメントイオンは同じ時間に溶出する

→“プレカーサーとフラグメントイオンの共溶出度を定量化”

フラグメントのイオン強度はライブラリーのイオン強度と相関する

→“フラグメントとライブラリーのイオン強度の相関”

この処理を自動で行ってくれるソフトウェアが必要

一般的なDIA解析手法

Software	提供元	Platform	コスト
Peak view	SCIEX社	GUI	有償
Skyline	Washington Univ.	GUI	無償
Spectronaut	Biognosys社	GUI	有償
DIA UNPIRE	Univ. Michigan	Java	無償
OpenSWATH	ETH (Switzerland)	Command prompt	無償

- OpenSWATHは無償で使える解析ソフトウェア
- Trans Proteome Pipelineに組み込まれCommand promptで動く
- 保持時間の補正をマーカーペプチドで行う
- 拡張ツールが存在する
 - TRIC（サンプル間の同定効率の改善）
 - IPF（修飾部位の同定）など

それぞれのソフトの評価はNavarro et al., *Nature biotech*, 2016にて検証されている

OpenSWATHの使い方
~OpenSWATHの使い方~

OpenSWATHの流れと必要な解析ソフト

ライブラリーの構築とDecoyの追加
→ Skylineなど/OpenSWATH

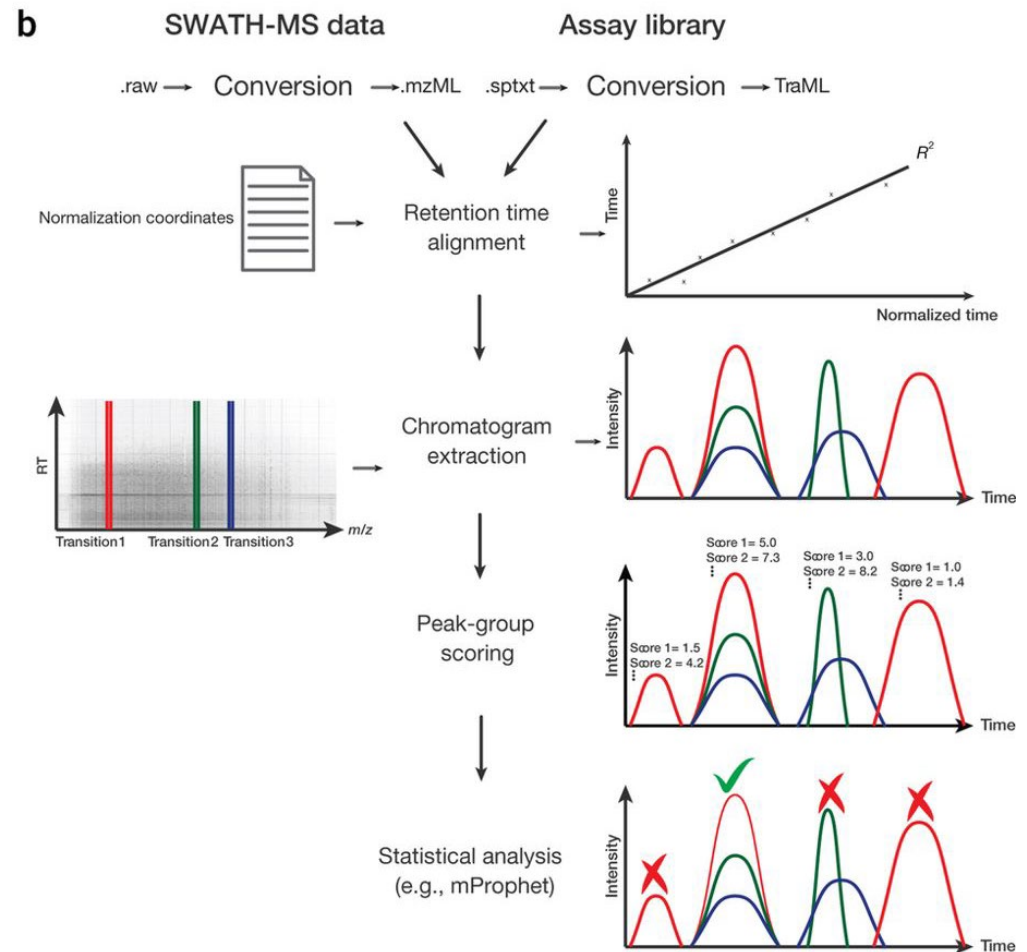
データの取得と変換
→ Analyst/ProteoWizard

保持時間の補正
→ OpenSWATH

クロマトグラムの抽出
→ OpenSWATH

ピークグループのスコア化
→ OpenSWATH

統計的なピーク判定
→ pyProphet



OpenSWATHの流れと必要な解析ソフト

ライブラリーの構築とDecoyの追加
→ Skylineなど/OpenSWATH

必要なソフトウェアの入手先

<https://skyline.ms/project/home/begin.view>

データの取得と変換
→ Analyst/ProteoWizard

<http://proteowizard.sourceforge.net>

保持時間の補正
→ OpenSWATH

クロマトグラムの抽出
→ OpenSWATH

http://ftp.mi.fu-berlin.de/pub/OpenMS/nightly_binaries

※OpenSWATHはOpenMSの中で開発

ピークグループのスコア化
→ OpenSWATH

統計的なピーク判定
→ pyProphet

Python(Anaconda2)を取得

PythonのpipコマンドでpyProphetを取得

プログラミングの基本

今回の話を理解するために最低限必要な基本だけ示します。
あくまで簡易的に示したものです。詳しくはプログラミングの参考書をご参照ください

スクリプトの説明

今回のスクリプトは
『コマンド』 = 作業名
『オプション』 = 条件
からなります。

```
OpenSwathWorkflow.exe コマンド  
-in data.mzML  
-tr library.tsv  
-out_tsv osw_output.tsv } オプション
```

右上のコマンドを読むと以下ようになります

OpenSwathWorkflowを data.mzMLを入力とし、library.tsvに照合し、osw_output.tsvを出力する条件で実行します。

※1 オプションは -in abc.fileや-batchSize 1000のような形のものをあらかじめこちらで打ち込んで指定します。

※2 オプションの説明はそれぞれのコマンド毎に記述があります。OpenSwathWorkflow.exeは-helphelpで表示

※3 全てのファイルからdata.mzMLを検索するのはとても時間がかかります。またもしdata.mzMLがPC内に複数あった場合PCはどのdata.mzMLを参照すればいいかわかりません。そこでどのファイルを明確に指定する必要があります。

そこで-in C:\Users\Atsushi\Desktop\WorkSpace\data.mzMLのように書きます。

ファイルのあるaddress

OpenSWATHの流れと必要な解析ソフト

ライブラリーの構築とDecoyの追加
→ Skylineなど/OpenSWATH

データの取得と変換
→ Analyst/ProteoWizard

保持時間の補正
→ OpenSWATH

クロマトグラムの抽出
→ OpenSWATH

ピークグループのスコア化
→ OpenSWATH

統計的なピーク判定
→ pyProphet

コマンドおよびオプション

説明

OpenSwathWorkflow.exe

実行コマンド

-in data.mzML

ProteoWizard変換後の計測ファイル

-tr library.tsv

transitionのlibrary

-tr_irt iRT_assays.TraML

Retention time補正のペプチドの情報

-swath_windows_file SWATHwindows_analysis.tsv

SWATH windowの情報のファイル

-out_tsv osw_output.tsv

結果を返すファイル名の指定

※オプションは他にもある。特にメモリの負荷を指定するオプションは指定する必要がある。

詳しくはRost et al., *Methods in Molecular Biology*, 2017

ライブラリーの構築とDecoyの追加

OpenSWATHに使用するライブラリーには以下の情報が必要

各ペプチドに対して

- Precursor ionのm/zの値
- Product ionのm/zの値
- 価数
- 保持時間（マーカーで補正）
- 帰属されるタンパク質
- Product ionのintensity（ライブラリー作成のために計測・同定した際の値）

これらの情報は以下から取得可能（作業は非常に煩雑なため以下を参照）

- Trans-Proteomic Pipeline
- Skyline

Trans-Proteomic Pipeline : Schubert et al., *Nature protocol*, 2015

Skyline : https://skyline.ms/wiki/home/software/Skyline/page.view?name=DIA_JA

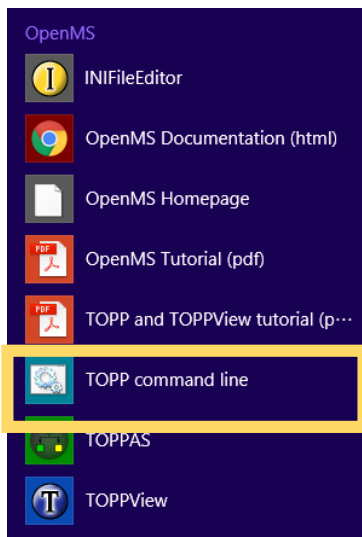
統計的なピーク判定にはライブラリーにDecoyの追加が必要

OpenSWATH内のコマンド"OpenSwathDecoyGenerator"により自動生成

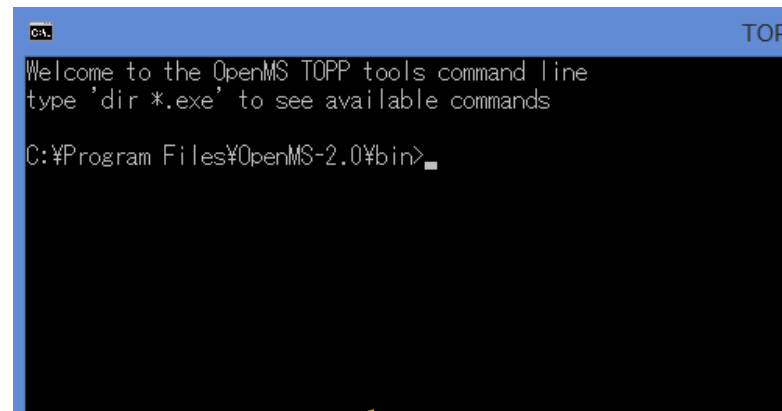
```
OpenSwathDecoyGenerator -in transitionlist_optimized.TraML  
-out transitionlist_optimized_decoys.TraML
```

<http://www.openswath.org/en/latest/docs/ppq.html#decoy-generation>

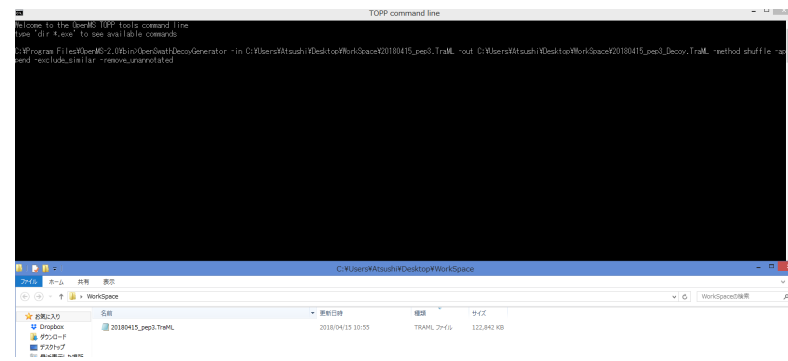
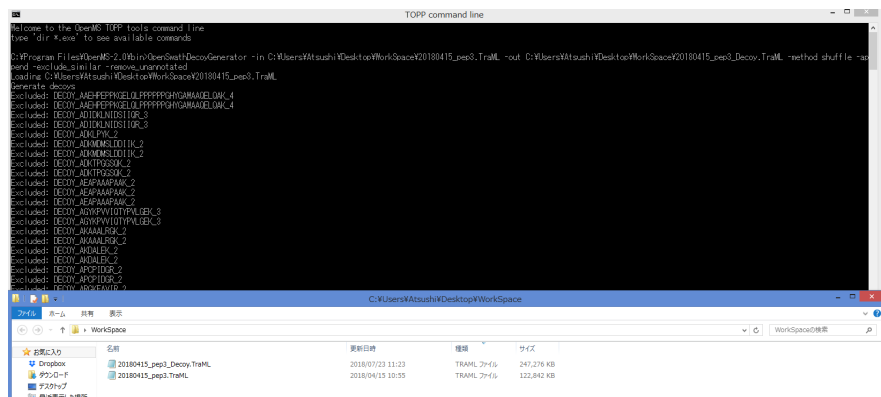
ライブラリーの構築とDecoyの追加（実際）



TOPP command lineを
起動



OpenSwathDecoyGenerator + file name



ライブラリーの構築とDecoyの追加

PrecursorMz	ProductMz	Tr_recalibrated	transition_name	CE	LibraryIntensity	transition_group_id	decoy	PeptideSequence	ProteinName	Annotation
595.3429765	284.1969	47.6929	RS82260.IVALSSSLNAK.2b3_1.light	31.19509097	2491	IVALSSSLNAK_2	0	IVALSSSLNAK	IPI00562876.1	2b3
595.3429765	977.5262	47.6929	RS82260.IVALSSSLNAK.2y10_1.light	31.19509097	9226	IVALSSSLNAK_2	0	IVALSSSLNAK	IPI00562876.1	2y10
595.3429765	1076.5946	47.6929	RS82260.IVALSSSLNAK.2y11_1.light	31.19509097	623	IVALSSSLNAK_2	0	IVALSSSLNAK	IPI00562876.1	2y11
595.3429765	419.2249	47.6929	RS82260.IVALSSSLNAK.2y4_1.light	31.19509097	958.8	IVALSSSLNAK_2	0	IVALSSSLNAK	IPI00562876.1	2y4
595.3429765	532.3089	47.6929	RS82260.IVALSSSLNAK.2y5_1.light	31.19509097	531.7	IVALSSSLNAK_2	0	IVALSSSLNAK	IPI00562876.1	2y5
595.3429765	619.341	47.6929	RS82260.IVALSSSLNAK.2y6_1.light	31.19509097	996	IVALSSSLNAK_2	0	IVALSSSLNAK	IPI00562876.1	2y6
595.3429765	706.373	47.6929	RS82260.IVALSSSLNAK.2y7_1.light	31.19509097	1969	IVALSSSLNAK_2	0	IVALSSSLNAK	IPI00562876.1	2y7
595.3429765	793.405	47.6929	RS82260.IVALSSSLNAK.2y8_1.light	31.19509097	5752	IVALSSSLNAK_2	0	IVALSSSLNAK	IPI00562876.1	2y8
595.3429765	906.4891	47.6929	RS82260.IVALSSSLNAK.2y9_1.light	31.19509097	2428	IVALSSSLNAK_2	0	IVALSSSLNAK	IPI00562876.1	2y9
595.3429765	258.1448341	47.6929	DECOY_RS82260.IVALSSSLNAK.2b3_1.light	31.19509097	2491	DECOY_IVALSSSLNAK_2	1	ASVSISSLALNK	DECOY_IPI00562876.1	2b3
595.3429765	1031.609542	47.6929	DECOY_RS82260.IVALSSSLNAK.2y10_1.light	31.19509097	9226	DECOY_IVALSSSLNAK_2	1	ASVSISSLALNK	DECOY_IPI00562876.1	2y10
595.3429765	1118.641571	47.6929	DECOY_RS82260.IVALSSSLNAK.2y11_1.light	31.19509097	623	DECOY_IVALSSSLNAK_2	1	ASVSISSLALNK	DECOY_IPI00562876.1	2y11
595.3429765	445.2769116	47.6929	DECOY_RS82260.IVALSSSLNAK.2y4_1.light	31.19509097	958.8	DECOY_IVALSSSLNAK_2	1	ASVSISSLALNK	DECOY_IPI00562876.1	2y4
595.3429765	558.360976	47.6929	DECOY_RS82260.IVALSSSLNAK.2y5_1.light	31.19509097	531.7	DECOY_IVALSSSLNAK_2	1	ASVSISSLALNK	DECOY_IPI00562876.1	2y5
595.3429765	645.3930051	47.6929	DECOY_RS82260.IVALSSSLNAK.2y6_1.light	31.19509097	996	DECOY_IVALSSSLNAK_2	1	ASVSISSLALNK	DECOY_IPI00562876.1	2y6
595.3429765	732.4250343	47.6929	DECOY_RS82260.IVALSSSLNAK.2y7_1.light	31.19509097	1969	DECOY_IVALSSSLNAK_2	1	ASVSISSLALNK	DECOY_IPI00562876.1	2y7
595.3429765	845.5090986	47.6929	DECOY_RS82260.IVALSSSLNAK.2y8_1.light	31.19509097	5752	DECOY_IVALSSSLNAK_2	1	ASVSISSLALNK	DECOY_IPI00562876.1	2y8
595.3429765	932.5411278	47.6929	DECOY_RS82260.IVALSSSLNAK.2y9_1.light	31.19509097	2428	DECOY_IVALSSSLNAK_2	1	ASVSISSLALNK	DECOY_IPI00562876.1	2y9

FullUniModPeptideName	MissedCleavages	Replicates	NrModifications	PrecursorCharge	PeptideGroupLabel	LabelType	UniprotID
IVALSSSLNAK	0	0	0	2	NA		IPI00562876.1
IVALSSSLNAK	0	0	0	2	NA		IPI00562876.1
IVALSSSLNAK	0	0	0	2	NA		IPI00562876.1
IVALSSSLNAK	0	0	0	2	NA		IPI00562876.1
IVALSSSLNAK	0	0	0	2	NA		IPI00562876.1
IVALSSSLNAK	0	0	0	2	NA		IPI00562876.1
IVALSSSLNAK	0	0	0	2	NA		IPI00562876.1
IVALSSSLNAK	0	0	0	2	NA		IPI00562876.1
ASVSISSLALNK	0	0	0	2	NA		IPI00562876.1
ASVSISSLALNK	0	0	0	2	NA		IPI00562876.1
ASVSISSLALNK	0	0	0	2	NA		IPI00562876.1
ASVSISSLALNK	0	0	0	2	NA		IPI00562876.1
ASVSISSLALNK	0	0	0	2	NA		IPI00562876.1
ASVSISSLALNK	0	0	0	2	NA		IPI00562876.1
ASVSISSLALNK	0	0	0	2	NA		IPI00562876.1
ASVSISSLALNK	0	0	0	2	NA		IPI00562876.1
ASVSISSLALNK	0	0	0	2	NA		IPI00562876.1
ASVSISSLALNK	0	0	0	2	NA		IPI00562876.1
ASVSISSLALNK	0	0	0	2	NA		IPI00562876.1

- Precursor ionのm/zの値
- Product ionのm/zの値
- 価数
- 保持時間（マーカーで補正）
- 帰属されるタンパク質
- Product ionのintensity

データの取得とデータの変換

データの取得条件

必ず保持時間の補正に使うマーカーペプチドを計測サンプルにも加える

※OpenSWATHではまずこのペプチドを探索することから解析が始まる

SWATH windowは自由に変えて構わない

※ OpenSWATHではSWATH windowの情報を外から与える

計測時間はライブラリー作成時の計測と違って構わない

※保持時間の補正はマーカーペプチドに行う

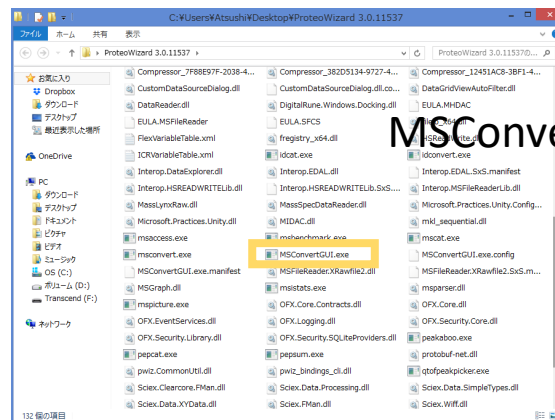
データの変換

ProteoWizard内のMSConvertGUI.exeでwiffファイルをmzMLファイルに変換

※設定はデフォルト。ファイルが大きくなる（20GB）ため注意。

データの取得とデータの変換（実際）

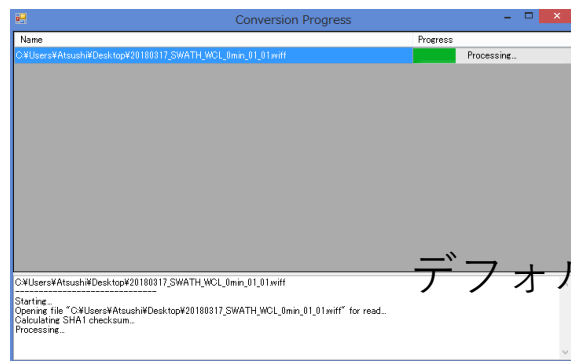
TTOF5600/6600
Analystを用いて計測



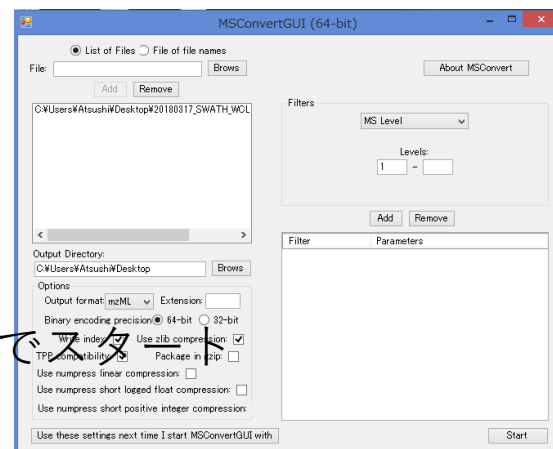
MSConverGUIを起動



wiffファイルを選択



デフォルトでスタート



SWATH windowと保持時間補正のtransitionの指定

SWATH windowsの指定

SWATHwindows_analysis.tsv

SWATH windowの情報をタブ区切りで保存したファイル

※Excelで作りcsvで保存したあと拡張子を.tsvにしても動く

start	end
400	424.5
424.5	449.5
449.5	474.5
474.5	499.5
499.5	524.5
524.5	549.5
549.5	574.5
574.5	599.5

保持時間補正のtransitionの指定

iRT_assays.TraML

Library作成に使用した保持時間補正のペプチドの情報※基本的にlibraryと同じ構造の情報

Precursor	ProductM	Tr_recali	b	transition	CE	LibraryInt	transition	decoy	PeptideSe	ProteinNa	Annotation	FullUniM	MissedCl	Replicates	NrModific	Precursor	GroupLab
432.25	523.29	29.822	25	24.019	517394	9	0	LVQQFTK/AQRT_09	y4	LVQQFTK	0	0	0	0	0	2	light
432.25	651.35	29.822	26	24.019	2809940	9	0	LVQQFTK/AQRT_09	y5	LVQQFTK	0	0	0	0	0	2	light
432.25	750.41	29.822	27	24.019	908259	9	0	LVQQFTK/AQRT_09	y6	LVQQFTK	0	0	0	0	0	2	light
484.26	528.31	18.348	13	26.30744	601587	5	0	AEAPAAA/AQRT_05	y6	AEAPAAA	0	0	0	0	0	2	light
484.26	599.35	18.348	14	26.30744	782514	5	0	AEAPAAA/AQRT_05	y7	AEAPAAA	0	0	0	0	0	2	light
484.26	696.4	18.348	15	26.30744	1512390	5	0	AEAPAAA/AQRT_05	y8	AEAPAAA	0	0	0	0	0	2	light
489.78	531.35	45.932	31	26.55032	354654	11	0	AQTFTLV/AQRT_11	y5	AQTFTLV	0	0	0	0	0	2	light
489.78	678.42	45.932	32	26.55032	389251	11	0	AQTFTLV/AQRT_11	y6	AQTFTLV	0	0	0	0	0	2	light
489.78	770.47	45.932	33	26.55032	1256120	11	0	AQTFTLV/AQRT_11	y7	AQTFTLV	0	0	0	0	0	2	light

TraMLにはOpenMS内のコマンド
ConvertTSVToTraMLが使用可能

ConvertTSVToTraML -in iRT_assays.tsv -out iRT_assays.TraML

保持時間の補正・クロマトグラムの抽出 ・ピークグループのスコア化

OpenSWATHの実行

OpenSWATH内のコマンド”OpenSwathWorkflow.exe”で保持時間の補正・クロマトグラムの抽出・ピークグループのスコア化の3つが同時に行われる。

※コマンドの実行必要なファイル

・mzML化した測定データ・ライブラリー・マーカーペプチドの情報・SWATH windowの情報

※処理は1計測ずつ行うためバッチ処理推奨

```
OpenSwathWorkflow.exe -in data.mzML -tr library.tsv -sort_swath_maps -readOptions cache -tempDirectory  
C:\Temp -batchSize 1000 -tr_irt iRT_assays.TraML -swath_windows_file SWATHwindows_analysis.tsv -out_tsv  
osw_output.tsv
```

Rost et al., *Methods in Molecular Biology*, 2017

保持時間の補正

保持時間のズレの許容範囲はオプションで任意に指定可能

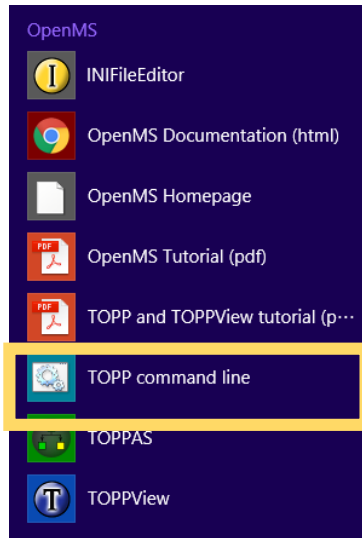
クロマトグラムの抽出

outputのファイルには”potential outlier”と予測されたフラグメントが提示される

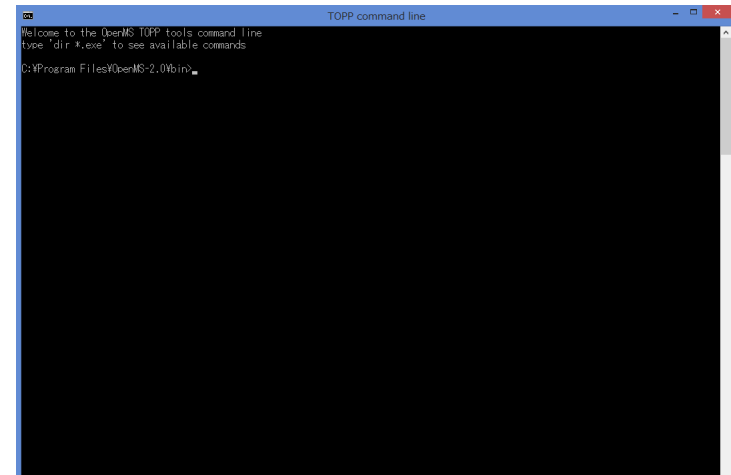
ピークグループのスコア化

dscoreとしてピークの特徴がスコア化され、次の統計的なピーク判定が可能となる

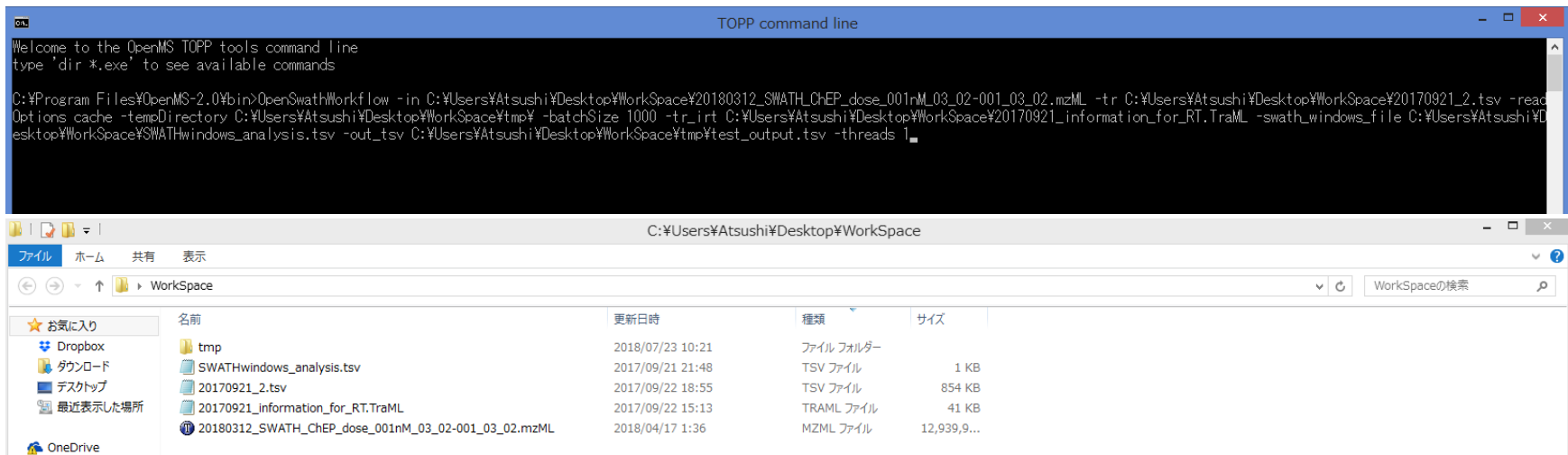
保持時間の補正・クロマトグラムの抽出 ・ピークグループのスコア化（実際）



TOPP command lineを
起動



OpenSwathWorkflow + file name



統計的なピーク判定

Pythonのインストール（提示しているコマンドは**Anaconda2**）が必要

Pythonのパッケージである**pyprophet**のインストールが必要

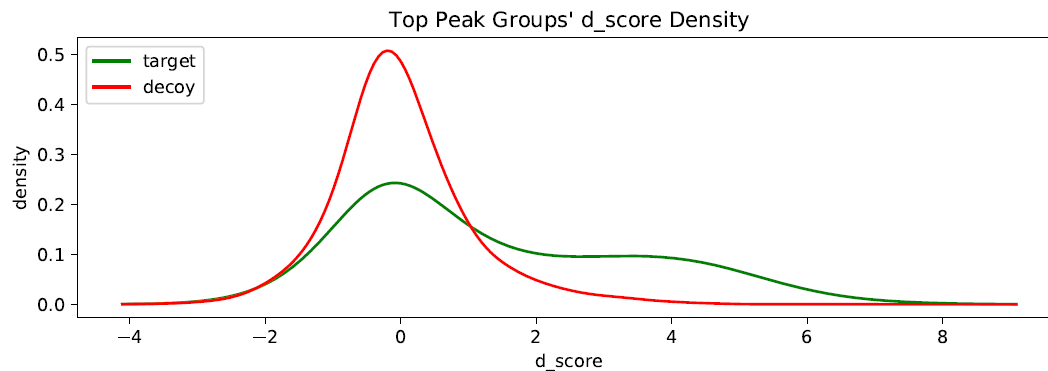
```
C:\Anaconda2\Scripts\pyprophet.exe --ignore.invalid_score_columns --d_score.cutoff=0.5 osw_output.tsv
```

Rost et al., *Methods in Molecular Biology*, 2017

ピークグループのスコア化で算出されたdscoreのcutoffを指定

→結果にdscoreとFDRの関係が出るため、最終的なdscoreの設定は結果を見て行う

解析を行うとpdf, csvファイルが複数出てくる。



qvalue	FDR	cutoff
0	0	8.258003
0.01	0.009995	2.632961
0.02	0.019994	2.361086
0.05	0.05	1.953323
0.1	0.099979	1.581219
0.2	0.20001	1.119368
0.3	0.299957	0.77724
0.4	0.400023	0.435916
0.5	0.499987	0.011703

解析の注意点およびパラメータの設定

ライブラリーの構築とDecoyの追加
→ Skylineなど/OpenSWATH

データの取得と変換
→ Analyst/ProteoWizard

保持時間の補正
→ OpenSWATH

クロマトグラムの抽出
→ OpenSWATH

ピークグループのスコア化
→ OpenSWATH

統計的なピーク判定
→ pyProphet

ライブラリーは大きいほどいいわけではない
質のよいフラグメントイオンのみ選択するとよい
同定数向上を目指すのであれば最も注力すべき

計測時には必ずマーカープепチドを使用する
mzMLファイルが大きいためストレージを圧迫する

versionによっては動かないことがある
保持時間は補正があるため絞ることが可能
Noisyなフラグメント情報の除去の活用

保持時間のずれの許容幅は狭い方が誤同定に効果あり
補正が**library**の質が高ければ狭くする（**150**秒程度）

一般的な**FDR**は**0.01**～**0.1**に設定
※**output**ファイルに**FDR**と**cutoff value**が示される

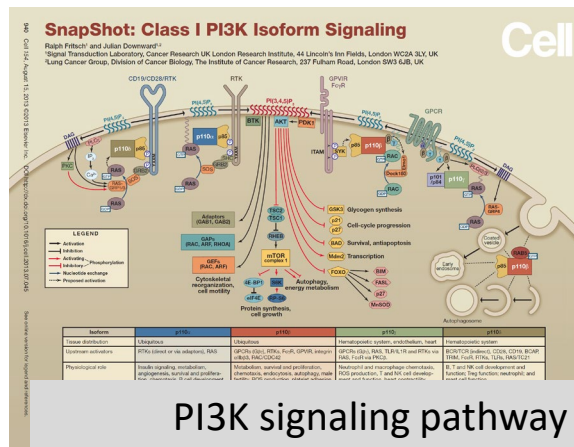
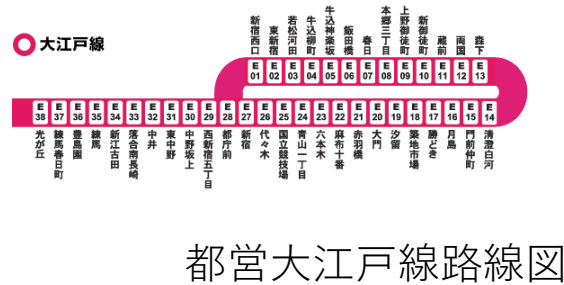
Trans-omics解析

細胞の地図を作る

“個別”

“包括的”

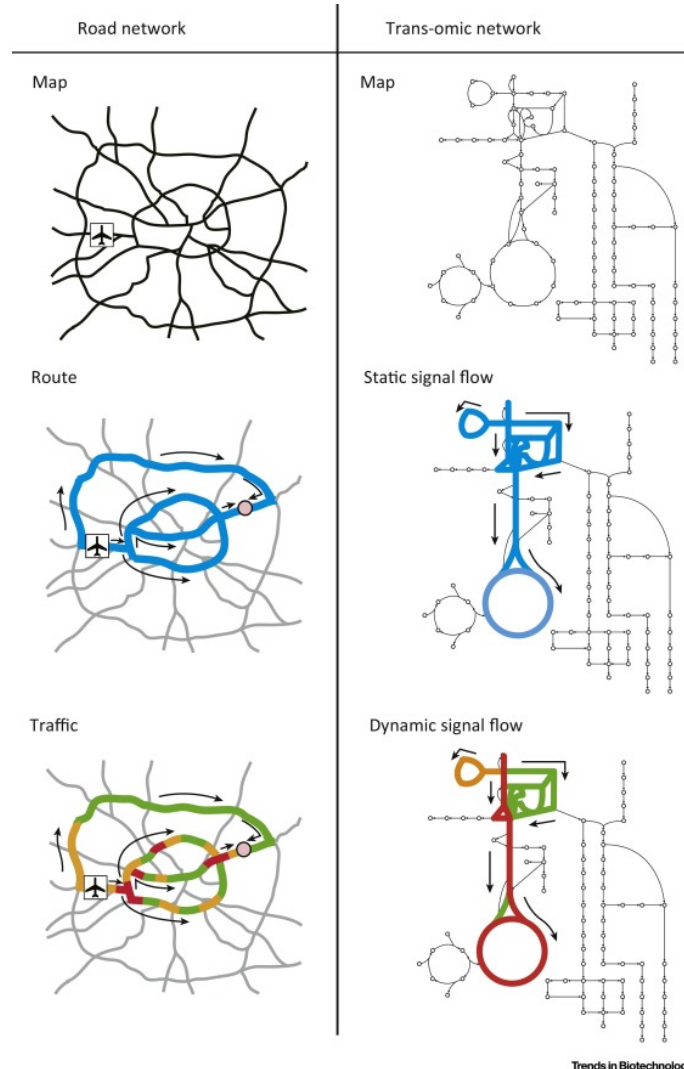
東京の地下鉄



Fritsch R. and Downward J. *Cell* 2013

分子ネットワーク

細胞の地図を作ると“流れ”が分かる

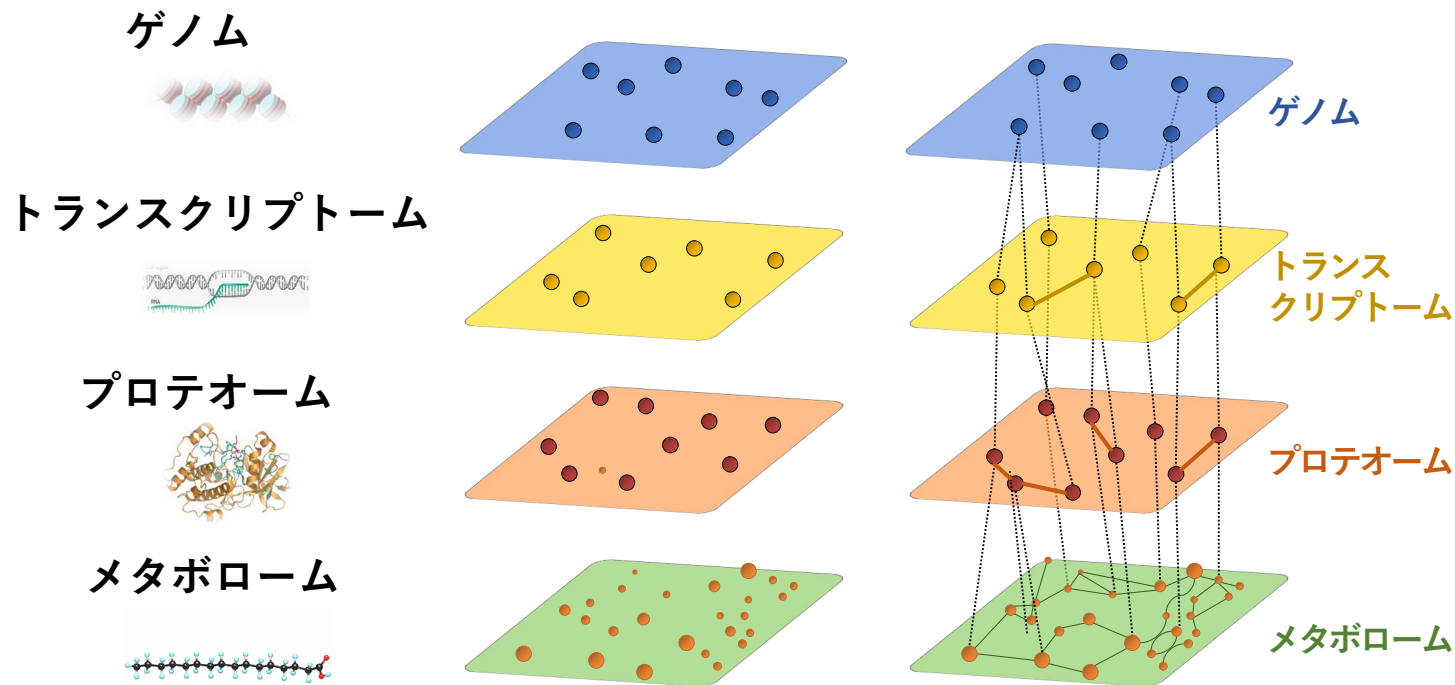


MAP: 道路図

ROUTE: 実際に使っているルート

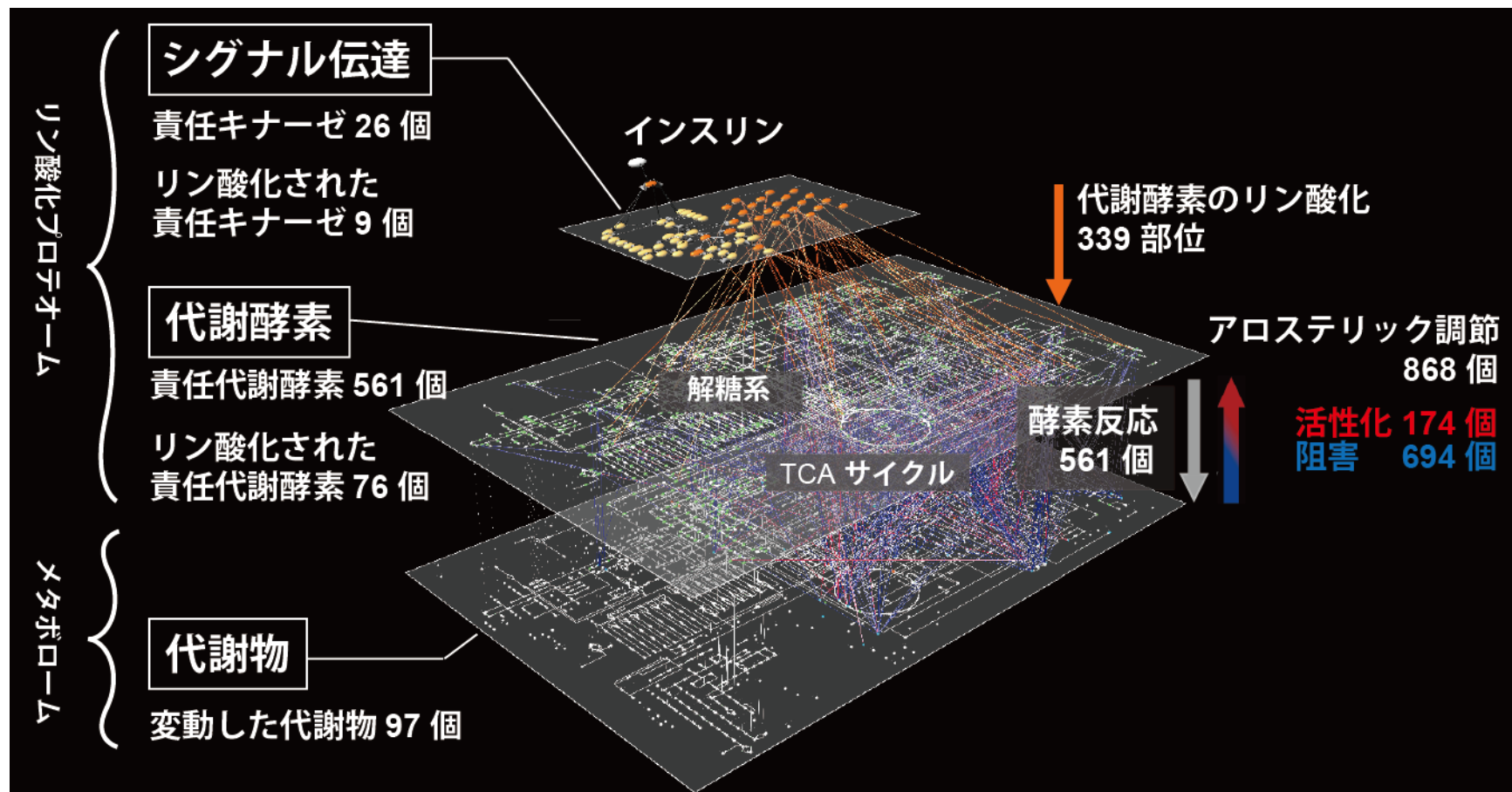
TRAFFIC: 実際の“流れ”

細胞の構成要素は多階層ネットワーク “トランスオミクネットワーク”からなる



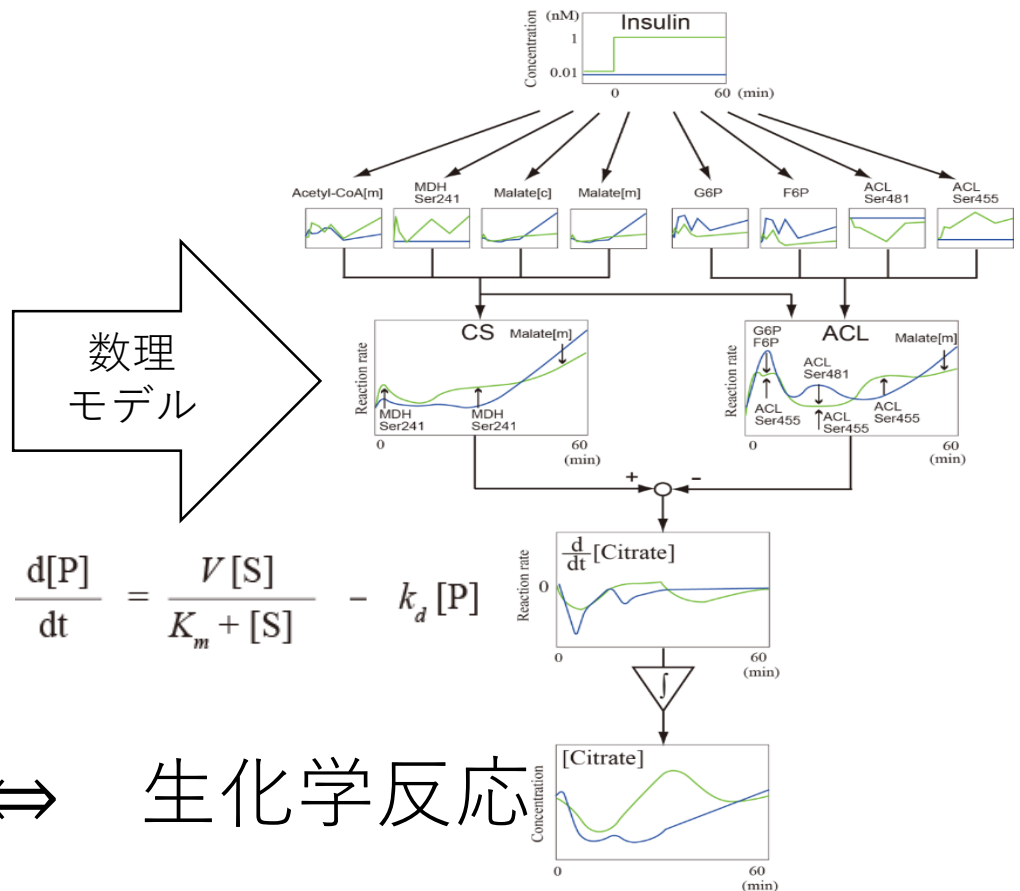
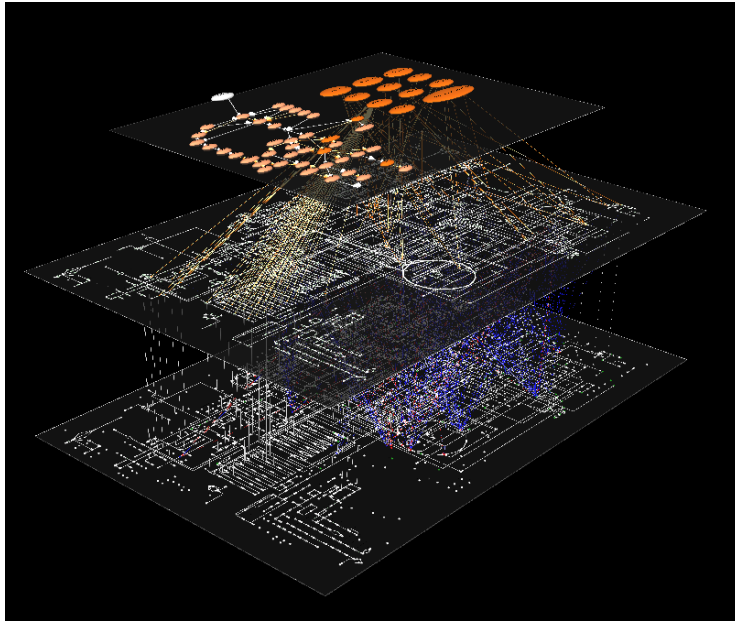
トランスオミクス

リン酸化プロテオームとメタボロームデータを用いた インスリンシグナル伝達のトランスオミクネットワーク



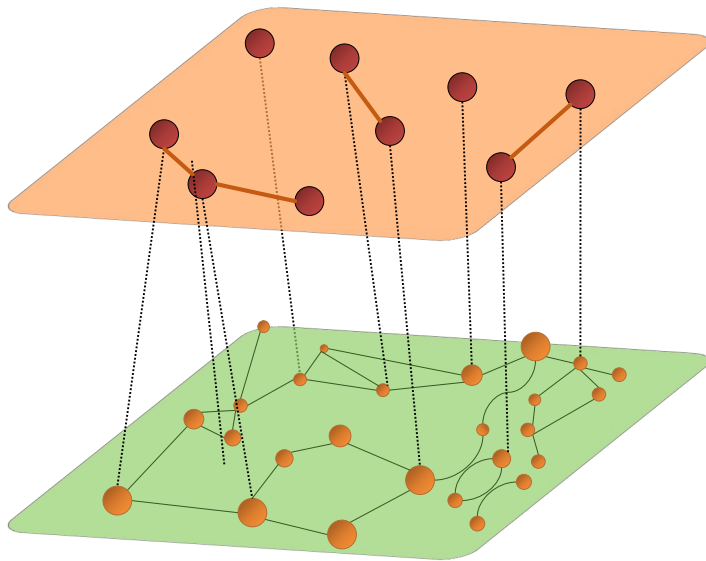
静的トランスオミクス
⇒ 定性的関係

動的トランスオミクス
⇒ 定量的因果関係



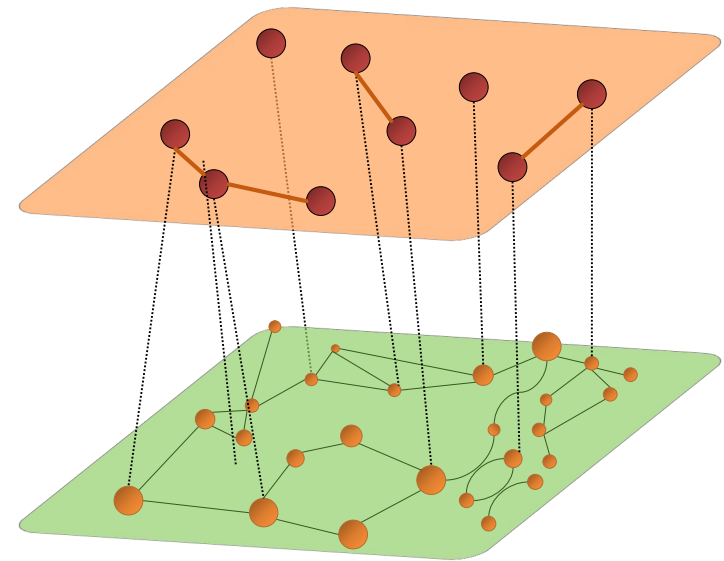
“トランスオミクネットワーク”は 制御・被制御の関係

プロテオーム



メタボローム

制御階層



被制御階層

1. 制御階層・被制御階層・両者の紐づけ
2. 制御階層（・被制御階層）の定量

制御階層＝機能的プロテオミクス Functional proteomics

機能的プロテオミクスとは

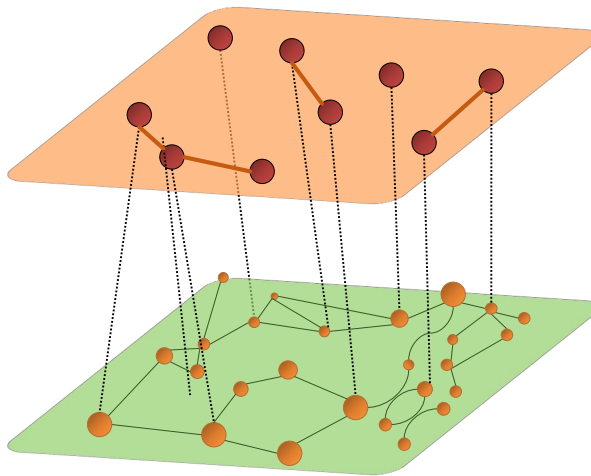
従来の発現量を定量するだけでなく活性あるいは活性に近い情報を網羅的に取得するプロテオミクス

具体的には

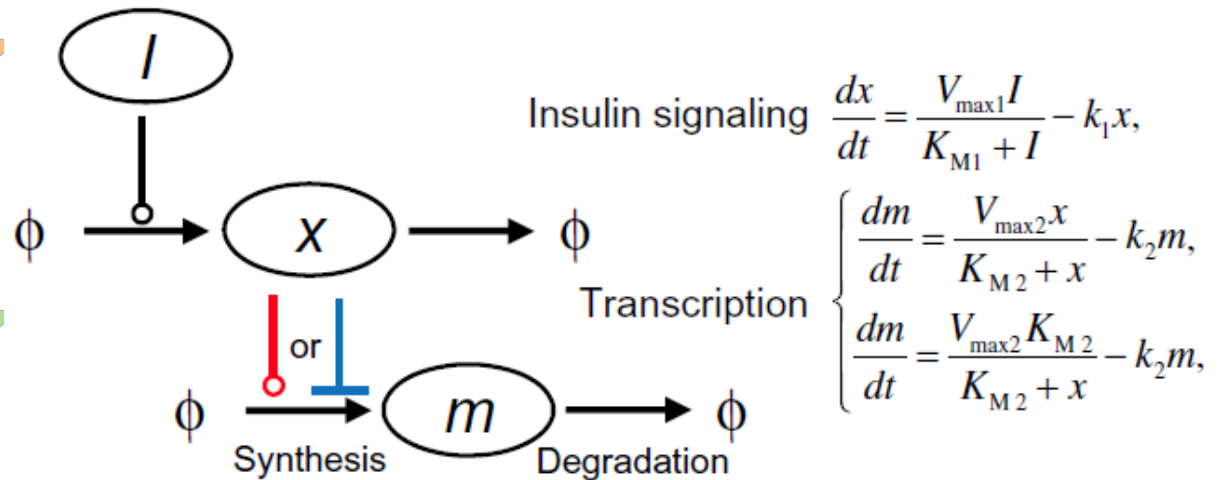
- ・ 翻訳後修飾プロテオミクス (Phospho, Acetylなど)
- ・ 結合プロテオミクス (IP-MS, AP-MS, BioIDなど)
- ・ 構造プロテオミクス (LiP, Thermal proteome profiling)
- ・ 細胞分画プロテオミクス (Organelle, ChEP)

Trans-omics

ChEPで定量



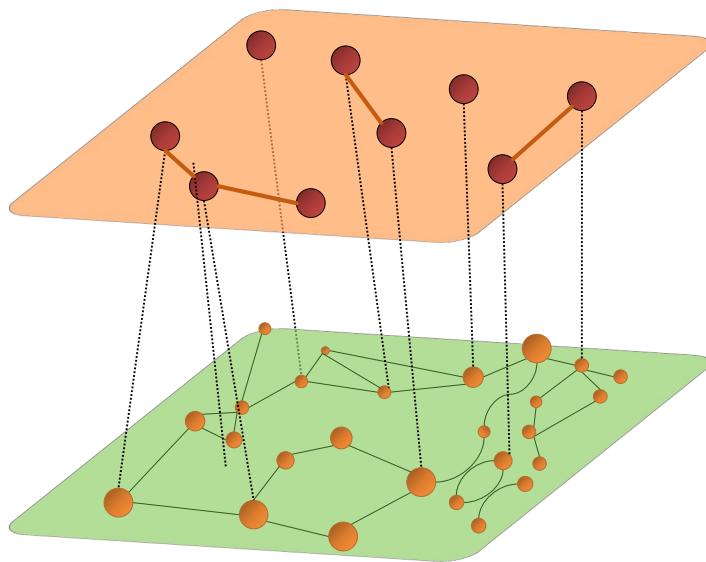
RNA-seq



転写反応をグローバルにモデル化

制御階層の活性を機能的プロテオミクスにより取得する

制御階層



被制御階層

タンパク質は様々な活性を持つ



現在網羅的に計測できる活性は限定的



活性が測れる経路をいかに増やすか

機能的プロテオミクスが今後のトランスオミクス解析の鍵を握る