

# Nextflowによるオミクス研究の仕組化 ～空間的遺伝子発現解析への対応事例～

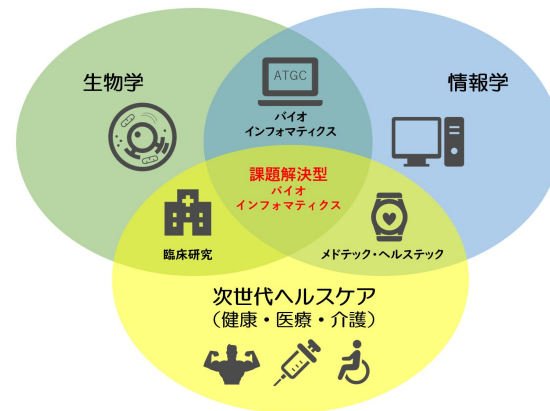
2024年11月13日  
アメリエフ株式会社  
代表取締役社長 山口 昌雄  
<https://amelieff.jp/>

# 会社紹介

社名	アメリエフ株式会社
本社	東京都港区西新橋3丁目7番1号 ランディック第2新橋ビル2階
従業員数	32名(アルバイト含む)
株主	PHCホールディングス株式会社(東京証券取引所 プライム市場上場)
沿革	<p>2009年～ 個人事業時代に SNPアレイ解析受託</p> <p>2010年03月 <b>次世代シーケンス(NGS)</b>データ解析開始</p> <p>2010年03月 トレーニングサービス・システム開発サービス開始</p> <p>2013年01月 京都大学 iPS細胞研究所と共同研究契約締結</p> <p>2015年12月 <b>クリニカルシーケンスシステム提供開始</b></p> <p>2016年01月 <b>第三者割当増資</b></p> <p>2017年03月 熊本大学との遺伝子検査の拠点形成プロジェクトを開始</p> <p>2017年08月 <b>網羅的遺伝子検査支援システム AmeliCure</b>が 三重大学病院に採用</p> <p>2018年04月 東京大学との共同研究を開始</p> <p>2018年09月 日本マイクロソフトとヘルスケアクラウド賛同パートナーとして連携</p> <p>2018年12月 <b>ISO27001/ISMS認証(情報セキュリティー)</b>取得</p> <p>2020年09月 PHCホールディングス株式会社の連結子会社 となる</p>



## 課題解決型 バイオインフォマティクス



生命現象を解明するための情報科学であるバイオインフォマティクスをベースに、健康に関わる研究開発で利用できるソリューションを提供しています。

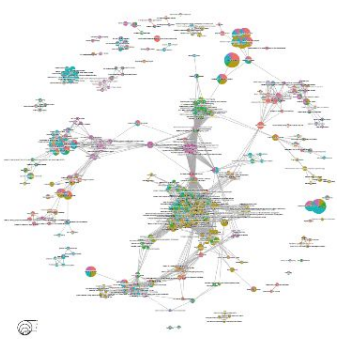
# アメリエフのバイオインフォマティクスサービス

## 受託解析

【累計実績：200 件以上】  
研究テーマやご要望に応じた  
データ解析を実施します。

### 【PRODUCT】

#### 次世代シーケンサデータ解析

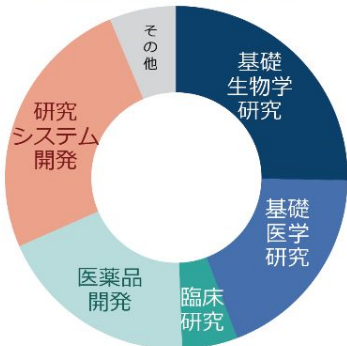


## コンサルティング

【累計実績：240 件以上】  
効果的な手法選択と  
柔軟な対応により、  
データ解析を推進します。

### 【SERVICE】

#### 技術調査・解析検証

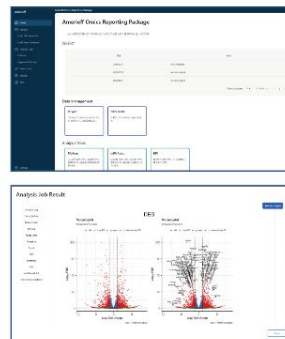


## 解析システム ソフトウェア

【累計実績：250 件以上】  
バイオデータを研究や  
製品開発に活用するための  
解析環境を整備します。

### 【PRODUCT】

#### データ解析基盤システム



## 人材育成

【累計実績：1,100 名以上】  
データ解析しながら学ぶ  
実践型のトレーニングを通して、  
初学者や人材育成を支援します。

### 【PROJECT】

#### 厚生労働省委託事業「がんの全ゲノム 解析に関する人材育成推進事業」



# アジェンダ

- Nextflowとは
  - 内製解析フローの整備 クラウドプラットフォームの利用
  - Nextflowを利用することによるメリット
- Nextflowの活用例
  - amelieffでの活用事例
  - 空間的発現解析での対応

# Nextflowとは

- バイオインフォマティクスなどの科学計算分野で利用されるワークフロー環境  
複雑なステップの解析、多サンプルでの解析処理の**並列実行**や**分散処理**フレームワークへの適用をシンプルに実装可能
- Basepairなど、多くのクラウドプラットフォームにおいてもカスタムパイプラインの実行環境として利用可能

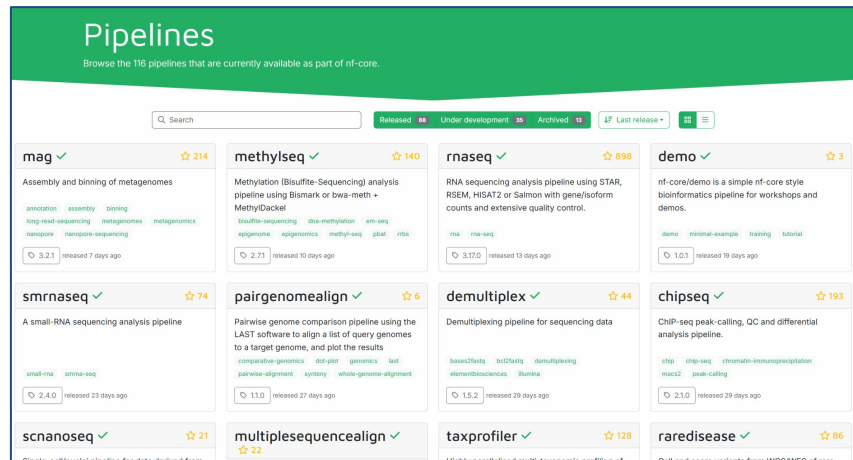


<https://www.nextflow.io/index.html>

# Nextflowを利用することによるメリット

- 複数環境への適用と再現性
  - ・ジョブスケジューラを使用した分散処理など  
シンプルに記述可能
  - ・Docker等の仮想環境を利用し、再現性確保

- nf-core(<https://nf-co.re/>)という  
コミュニティから多くの解析パイプラインを  
利用可能



<https://nf-co.re/pipelines/>

- パフォーマンス計測やエラー対応 などの共通処理が対応済み
  - 実行した際のCPU実行時間、出力ディレクトエラーメッセージなどが表示されたレポートが出力される
  - エラーで中断した場合、修正して中断箇所から再実行可能

# 内製解析フローの整備 クラウドプラットフォームの利用

## 内製解析ワークフローを保有しているラボでのニーズ

- ラボで保有しているパイプラインをクラウドプラットフォームで動かしたい
- 短期間で大量の解析処理を実行したい
- GUI実行、マネージド環境でのメンテナンス、ランニングの負担軽減など、クラウドプラットフォームのメリットをなるべく早く享受したい



## Nextflow ワークフロー環境のご紹介

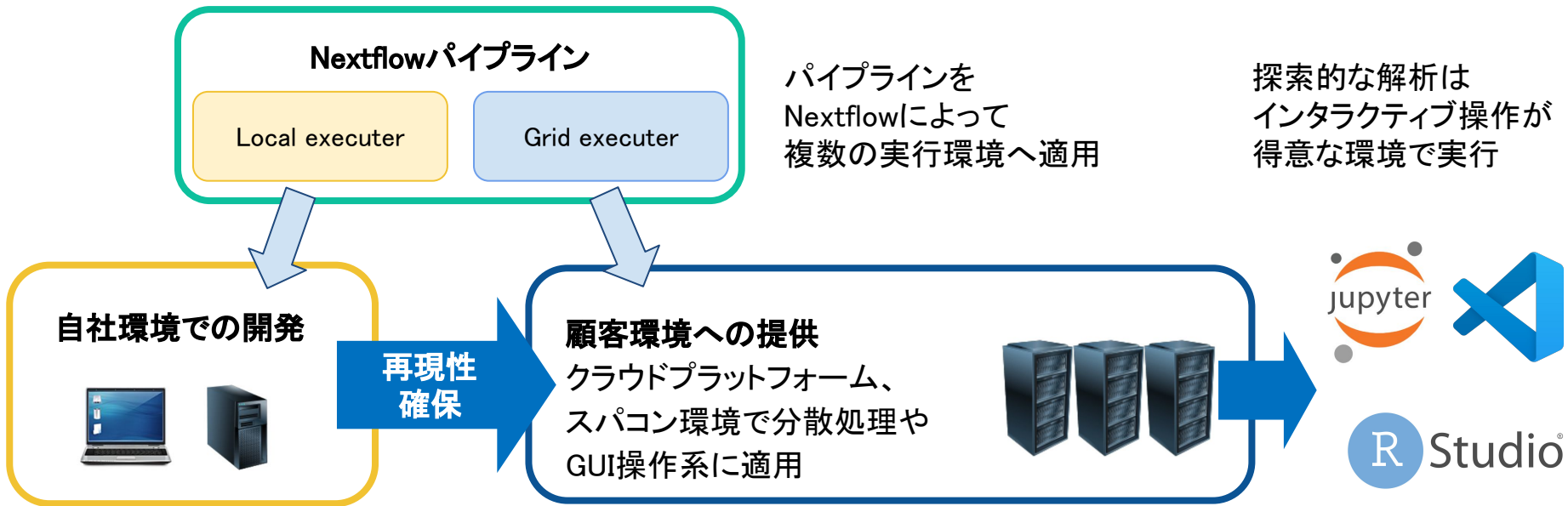
- 自社のカスタムパイプラインの再現性の確保  
様々な環境で、同じ処理を実現できる
- 並列実行、分散処理環境へ即座に適用  
大規模処理やマネージド環境が効果的に、すばやく利用可能になる

## Nextflowの活用例 空間的遺伝子発現解析への対応



# amelieffでの活用事例

弊社ではNextflowを用いてパイプラインを開発  
環境の再現性を利用して、パイプラインの**自社受託サービス活用** や**顧客提供**

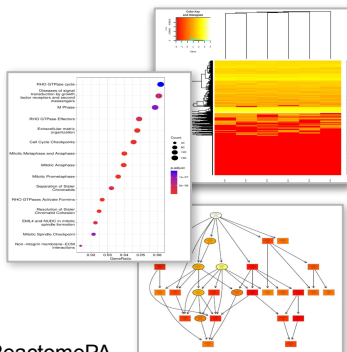


# パイプラインの開発例、公開パイプラインの利用例

## パイプラインの開発例

### RNA-seq解析

- ・発現定量
- ・主成分分析、階層的クラスタリング
- ・発現変動遺伝子の検出
- ・二群間比較解析
- ・Gene Ontology解析
- ・パスウェイエンリッチメント解析

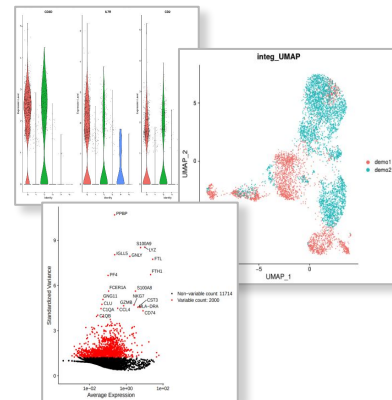


利用ソフト

FastQC, PRINSEQ, Trimmomatic, STAR, featureCounts, edgeR, EnhancedVolcano, ReactomePA

### シングルセル RNA-seq解析

- ・次元削減とクラスタリング
- ・高発現遺伝子の同定
- ・Feature plot 作図
- ・Gene Ontology解析、パスウェイ解析
- ・全クラスターにおける二群間発現比較
- ・擬似系譜解析



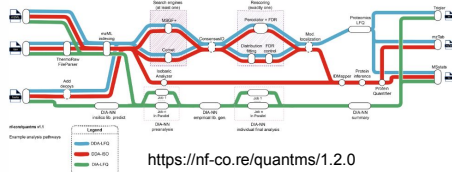
利用ソフト

Seurat, EnhancedVolcano, ReactomePA, monocle3

## 公開パイプラインの利用例

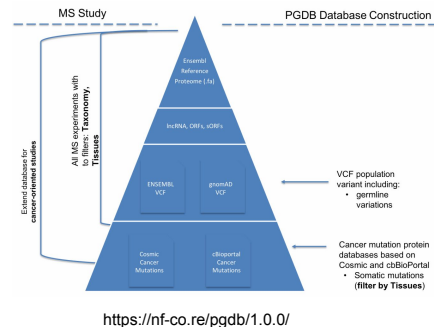
### プロテオミクス定量 (quantms)

- ・たんぱく質同定
- ・発現定量
- ・DIA、DDAの統合フロー

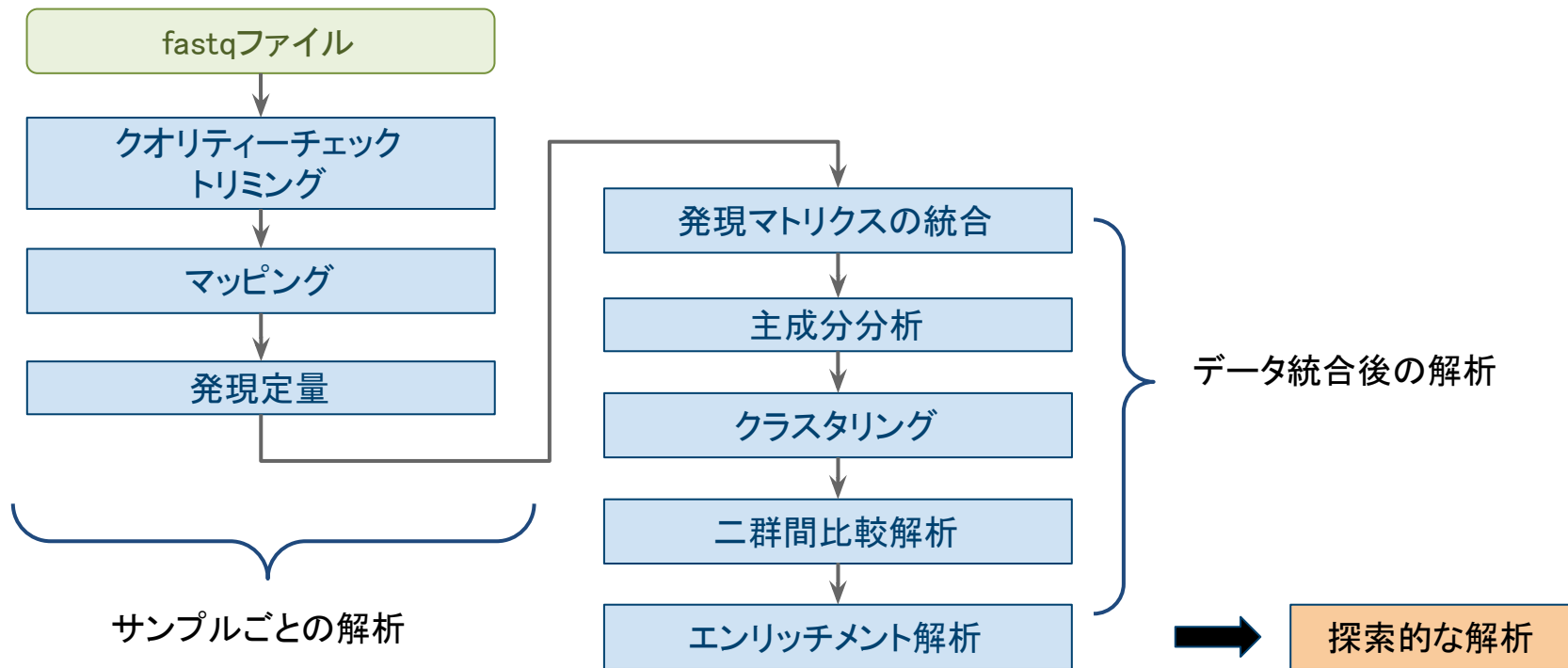


### プロテオミクス DB構築 (pgdb)

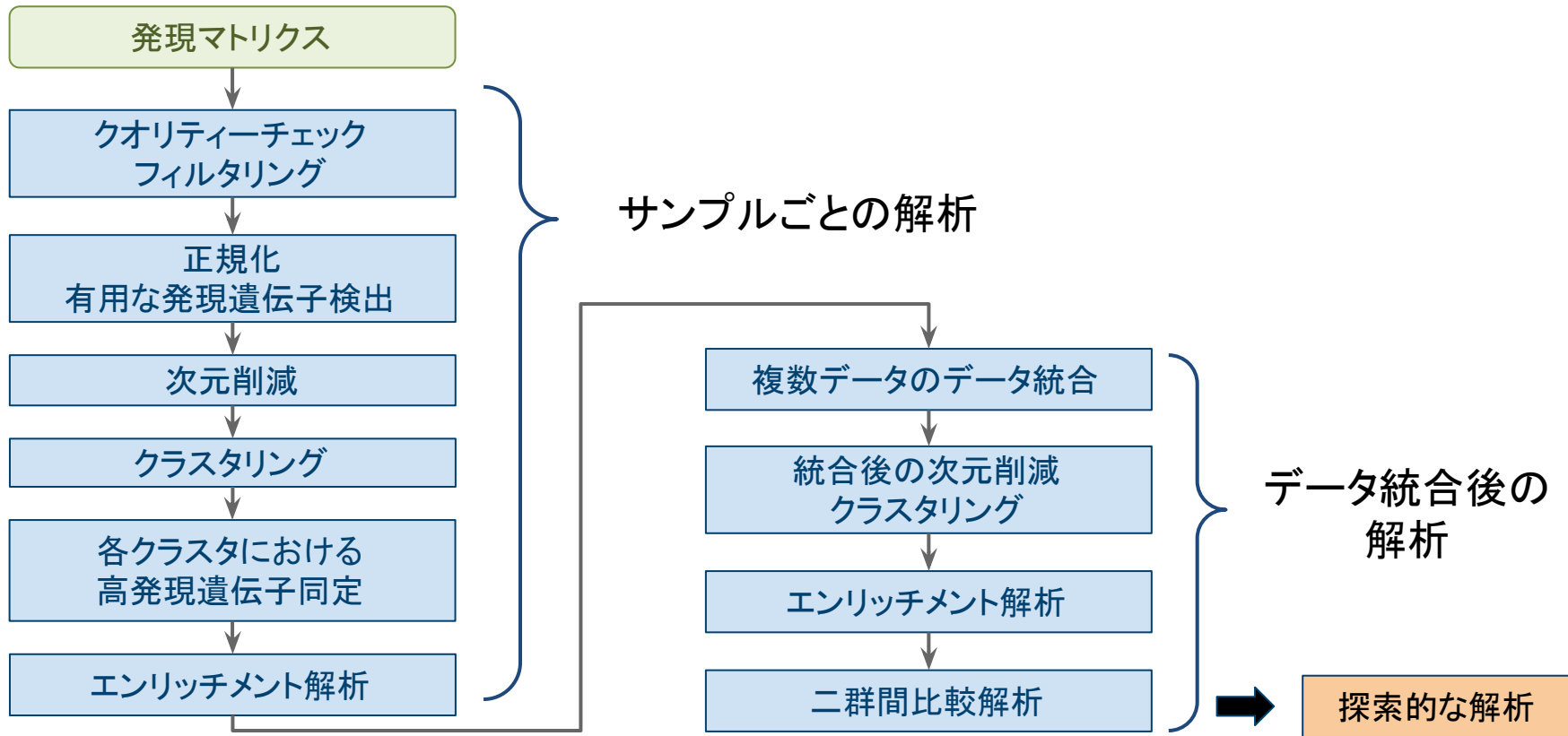
- ・参照データベースと変異データから変異を反映したたんぱく質配列データベースを構築



# RNAseq解析の流れ



# シングルセルRNAseq解析の流れ



# Nextflow パフォーマンス、実行時間のレポート

実行時に各ステップの実行時間や使用メモリ量などが記録されレポートされる

## Processes execution timeline

Launch time: 21 Jul 2023 20:30

Elapsed time: 6m 57s

Legend: job wall time / memory usage (RAM)

### プロセス一覧

### 実行時間 / 使用メモリ量

NFCORE\_RNASEQ:RNASEQ:PREPARE\_GENOME:UNTAR\_SALMON\_INDEX (salmon.tar.gz)

6.1s / 2.9 MB

NFCORE\_RNASEQ:RNASEQ:PREPARE\_GENOME:GUNZIP\_GTF (genes.gtf.gz)

6.1s / 2.8 MB

NFCORE\_RNASEQ:RNASEQ:INPUT\_CHECK:SAMPLESHEET\_CHECK (samplesheet\_test.csv)

NFCORE\_RNASEQ:RNASEQ:PREPARE\_GENOME:GUNZIP\_ADDITIONAL\_FASTA (gfp.fa.gz)

6.1s / 2.8 MB

NFCORE\_RNASEQ:RNASEQ:PREPARE\_GENOME:CAT\_ADDITIONAL\_FASTA (gfp.fa)

NFCORE\_RNASEQ:RNASEQ:PREPARE\_GENOME:GTF2BED (genome\_gfp.gtf)

NFCORE\_RNASEQ:RNASEQ:PREPARE\_GENOME:CUSTOM\_GETCHROMSIZES (genome\_gfp.fasta)

NFCORE\_RNASEQ:RNASEQ:PREPARE\_GENOME:STAR\_GENOMEGENERATE (genome\_gfp.fasta)

NFCORE\_RNASEQ:RNASEQ:PREPARE\_GENOME:BBMAP\_BBSPLIT

37.2s / 345.8 MB

NFCORE\_RNASEQ:RNASEQ:FASTQ\_FASTQC\_UMITOOLS\_TRIMGALORE:FASTQC (RAP1\_UNINDUCED\_REP1)

31s / 188.7 MB

NFCORE\_RNASEQ:RNASEQ:FASTQ\_FASTQC\_UMITOOLS\_TRIMGALORE:TRIMGALORE (RAP1\_UNINDUCED\_REP1)

30.9s / 198.9 MB

NFCORE\_RNASEQ:RNASEQ:CAT\_FASTQ (RAP1\_UNINDUCED\_REP2)

1.3s / 2.9 MB

NFCORE\_RNASEQ:RNASEQ:CAT\_FASTQ (WT\_REP1)

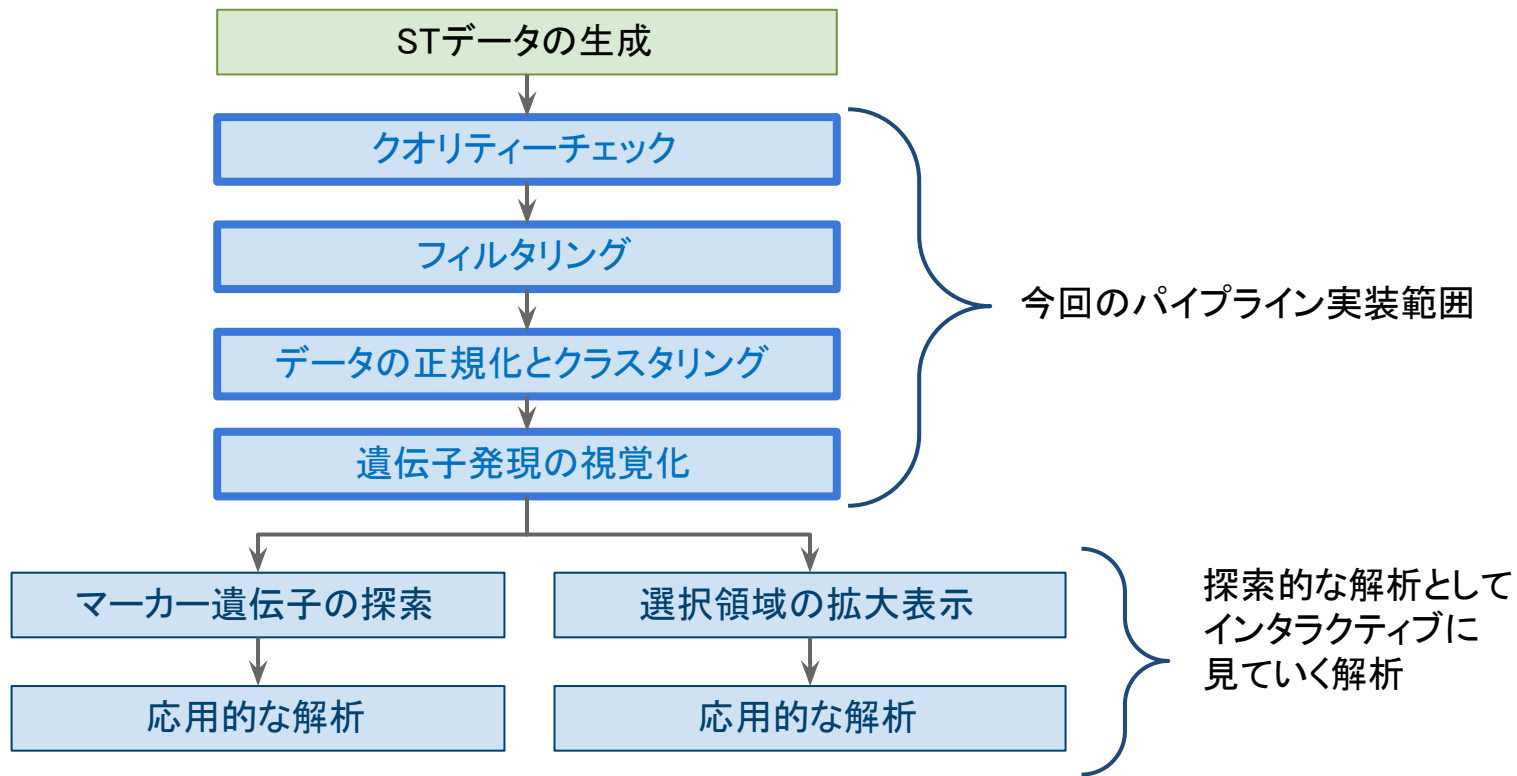
1.2s / 2.9 MB

24.2s / 181.8 MB

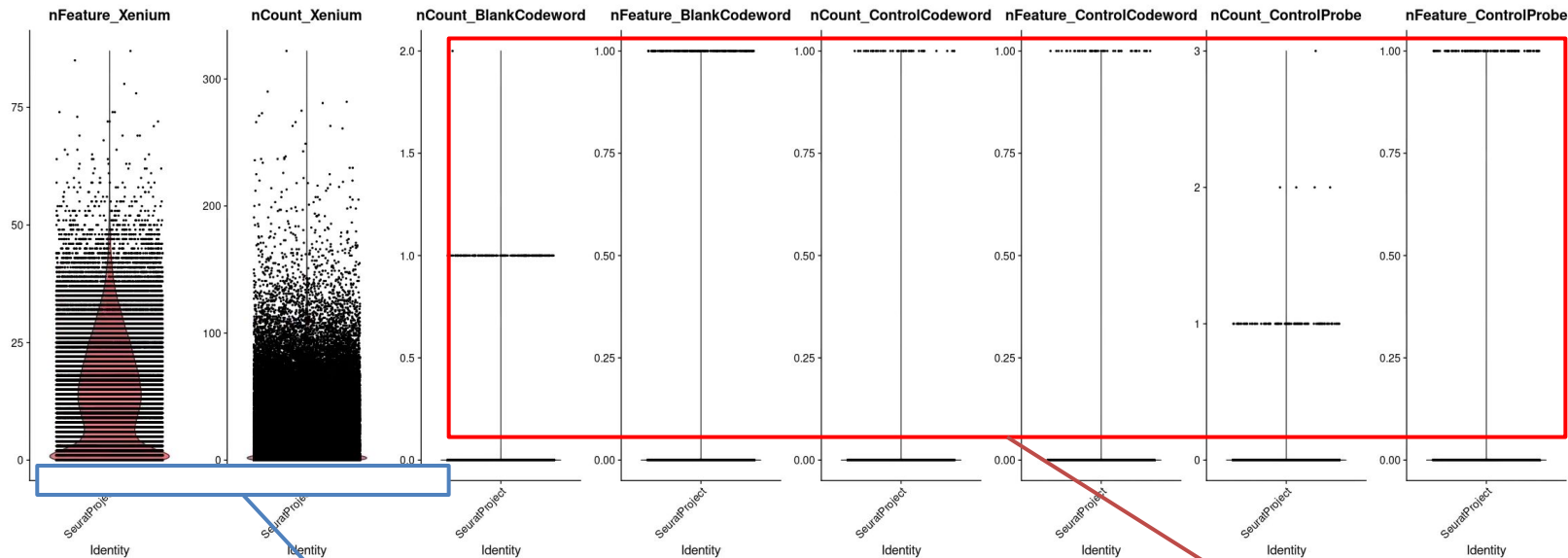
## 空間的遺伝子発現解析の実装例

---

# 空間的発現解析 Xenium In Situ のフロー



# クオリティコントロール



総UMIや総発現遺伝子数が0の場合、セルセグメンテーションがうまくできていない可能性

コントロールコードワードやコントロールプローブの数が多すぎる場合ノイズが高い可能性



# クラスタリング

細胞集団を亜集団(クラスタ/cluster)に分類し、特徴を抽出する

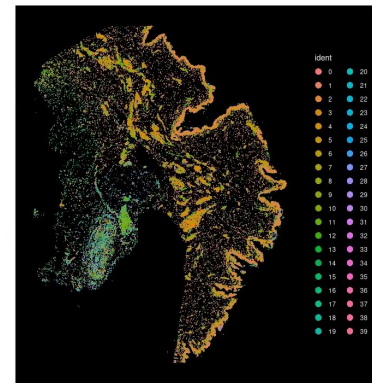
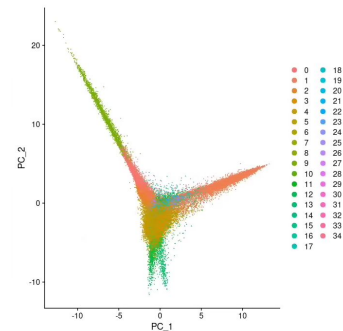
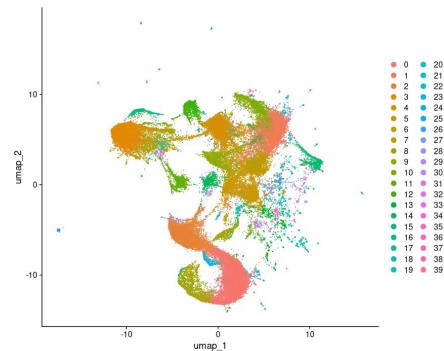
PCA Score Plotによる可視化



UMAP/tSNE Plot  
による可視化



ImagePlot  
による可視化

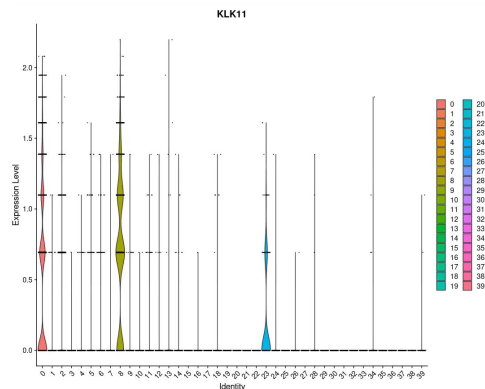


クラスタリングなどパラメータ調整などは複数条件の解析が必要  
Nextflowを用いることで環境に合わせてながら、複数条件の解析を並列化可能

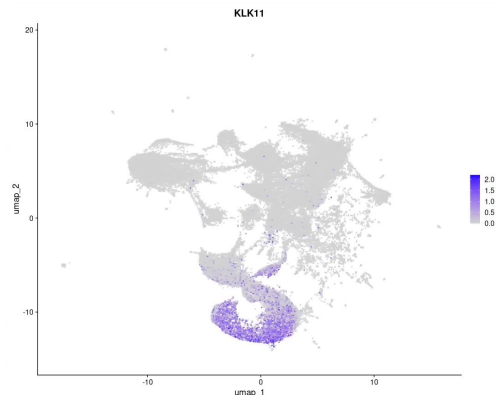
# マーカー遺伝子探索

ターゲットとしているクラスタと他クラスタとの判別に使用することができる遺伝子を選定する

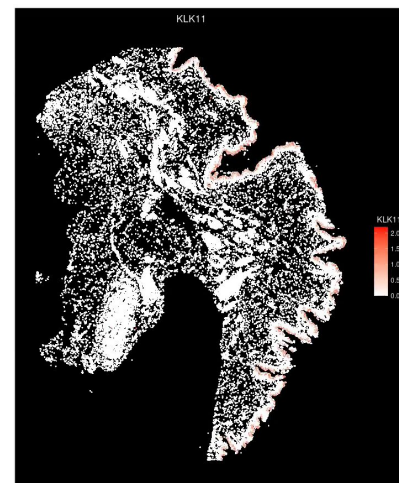
バイオリンプロットの描画



UMAP/ tSNEプロットへの  
特定遺伝子の発現量の表示



組織画像上への特定遺伝子の  
発現量表示



マーカー遺伝子の探索など探索領域、条件が広い解析ではインタラクティブに解析する環境を別途構築

# バイオインフォマティクス解析環境の選定と統合化

ラボでの解析処理を各プラットフォーム、各環境の特徴を理解しながら選定し、統合していくことが今後ますます重要になる

## クラウドプラットフォーム

必要な時に必要な解析環境を利用可能  
並列分散処理による大量の解析が可能  
GUI操作による簡便な実行  
大規模データソースの活用

## 社内、ラボ内 サンドボックス環境

実験的な処理を安全に実行  
ラボ内機器との連携  
解析のサイジング、コスト評価

## ラボでの解析処理

各環境に対して  
理解・選定・統合  
する必要性

## インタラクティブ 解析環境

各環境の操作系  
試行錯誤が必要な探索処理の実行  
環境のポータビリティ

# おわりに

製品ではNextflowを利用した内製パイプラインの整備  
クラウドプラットフォームへの適用など、システム構築の支援をしています。

ご興味のある方は弊社ホームページよりお問い合わせください。

[https://amelieff.jp/server\\_system/](https://amelieff.jp/server_system/)

# アメリエフのバイオインフォマティクスサービス

バイオインフォマティクス専門家が解決する  
「研究データ解析お悩み相談サービス」です

## コンサルティング

- ・手法の検討や結果解釈の支援
- ・メールサポートで疑問を解決
- ・コマンドレベルでノウハウ蓄積
- ・スモールスタートで最新技術試行
- ・研究目的の達成にコミット
- ・解析環境の準備不要

## データ解析サーバ導入 システム開発

- ・GPUなど解析を高速化
- ・研究テーマに最適化
- ・ソフトウェア導入サポート
- ・操作方法チュートリアル
- ・オンプレミスとクラウドに対応
- ・実績のある構成を導入



## バイオインフォマティクス トレーニング

- ・確かな解析ノウハウ
- ・経験豊富な講師陣
- ・実践環境で学習
- ・双方向オンライン学習
- ・5名までの少人数授業
- ・結果の解釈方法も学習

## 受託解析

- ・最新解析技術を研究に活用
- ・WEB会議で解析結果を解説
- ・中間報告で方向性のすり合わせ
- ・解析環境の整備運用が不要
- ・解析コストを変動化