

◎小川誠司、油谷浩幸、上田宏生、浦上研一、岡田随象、片岡圭亮、柴田龍弘
白石友一 鈴木 穰、十時 泰、中村祐輔、南谷泰仁、藤本明洋、森下真一

1. シークエンスプラットフォームの検討 (進行中)

- 予備的解析 (進行中)
 - パイロットプロジェクト支援
 - 時系列解析
 - マルチオミックス解析(エピゲノム、免疫ゲノム、単一細胞)

2. 一次データ共有システムの構築(案)

3. 標準パイプラインの開発(案)

4. データ解析拠点の構築(案)

1. シークエンスプラットフォームの検討

1) シークエンスプラットフォームの検討

シークエンスデータはシークエンスプラットフォームに大きく依存する。従って、先行解析・本格解析は統一したシークエンスプラットフォームでデータを取得することが必要。

- 短鎖シークエンス
- 長鎖シークエンス

2) 解析を行うターゲットの検討

シークエンス解析の対象の評価

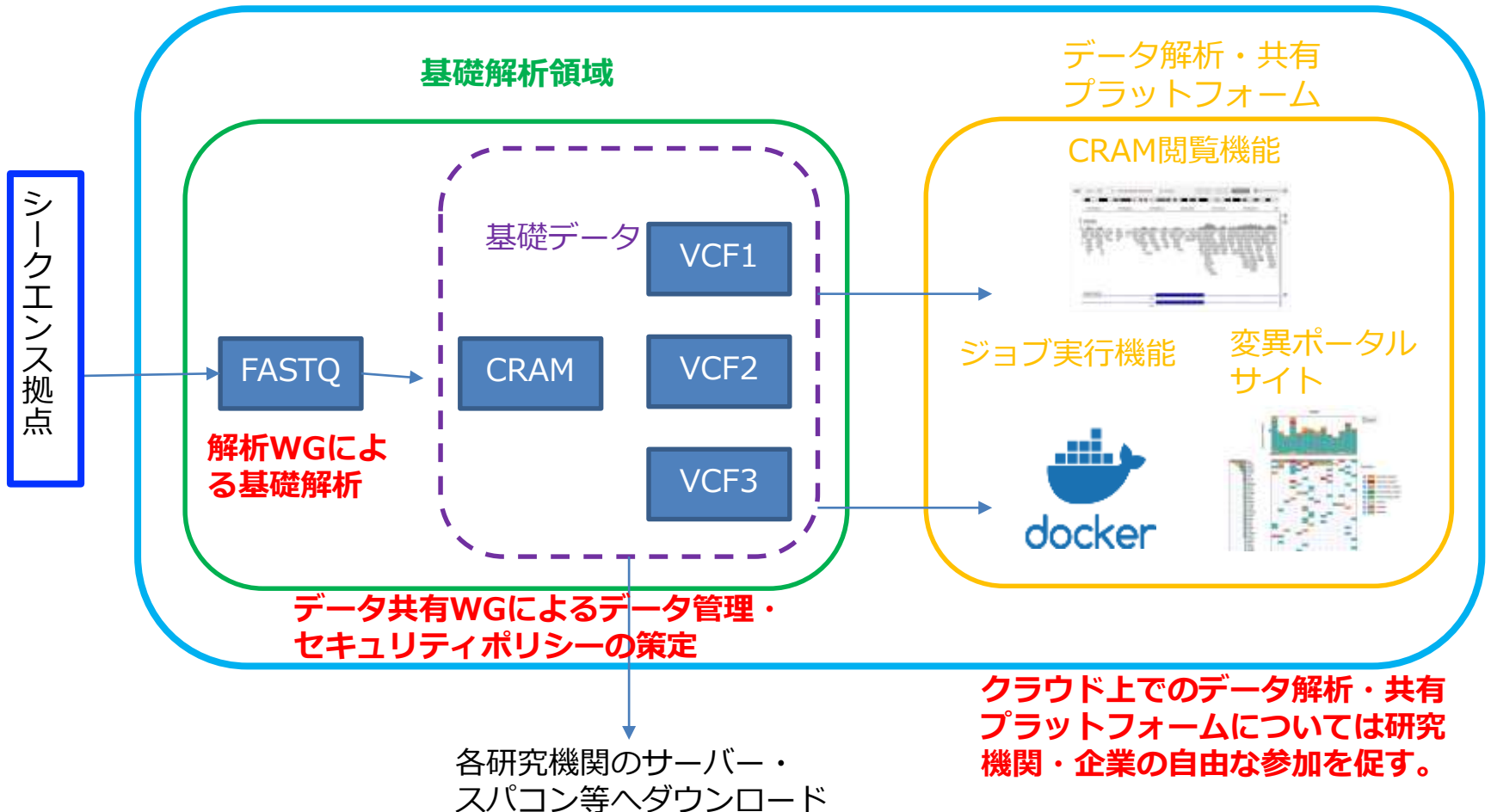
- 高深度の全ゲノムシークエンス (200x)
- RNAシークエンス
- 長鎖リードシークエンス
- エピゲノム
 - バイサルファイト WGS
 - (ATAC/deep HiC)
 - (ヒストン修飾)
- (単一細胞解析)

現在までの進捗

- **試料の調整:**
 - 大腸がん5検体（腫瘍-正常ペア）
 - 腫瘍：DNA抽出（カラム法）完了
（液体法）完了
 - 正常：DNA抽出（液体法）完了
 - 膵癌オルガノイド5検体
 - 培養中（9～10月に回収見込み）
- **シーケンス解析:**
 - 短鎖リードシーケンス
 - 契約終了
 - 一両日中に提出
 - バイサルファイト全ゲノムシーケンス
 - ライブラリー作成
 - PCRフリープロトコルの確認(済)
 - 9/14-の週でテスト検体受け取り
 - 長鎖リードシーケンス
 - 解析施設を計画書に加える必要あり（情報待ち、改訂に1ヶ月必要）
 - 腫瘍：DNA抽出（液体法）未（直ちに提出可能）
- **データ解析:**
 - 未
- **2020年12月までの終了を予定。**

2. 一次データ共有システムの構築(案)

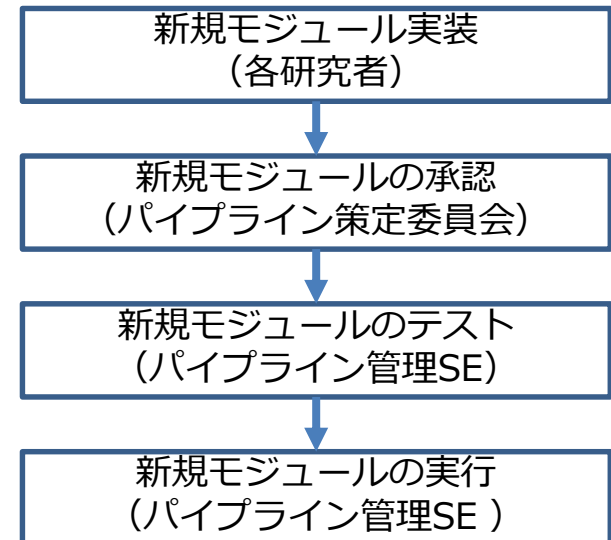
商用クラウド



3. 標準パイプラインの開発(案)

- **アラインメント方法の策定。**
 - リファレンスゲノム、アラインメントのオプションなど。
- **初期パイプラインの構築。**
 - 最初にプロトタイプを示して、システムを具体化するため。
- **パイプラインの実装方法の策定。**
 - パイプライン言語の利用の可否 (CWL, WDL)
 - パイプライン実行フレームワークの利用の可否 (snakemake, Nextflow)
 - 仮想コンテナの利用。
- **解析WGミーティングにおいて検討。**

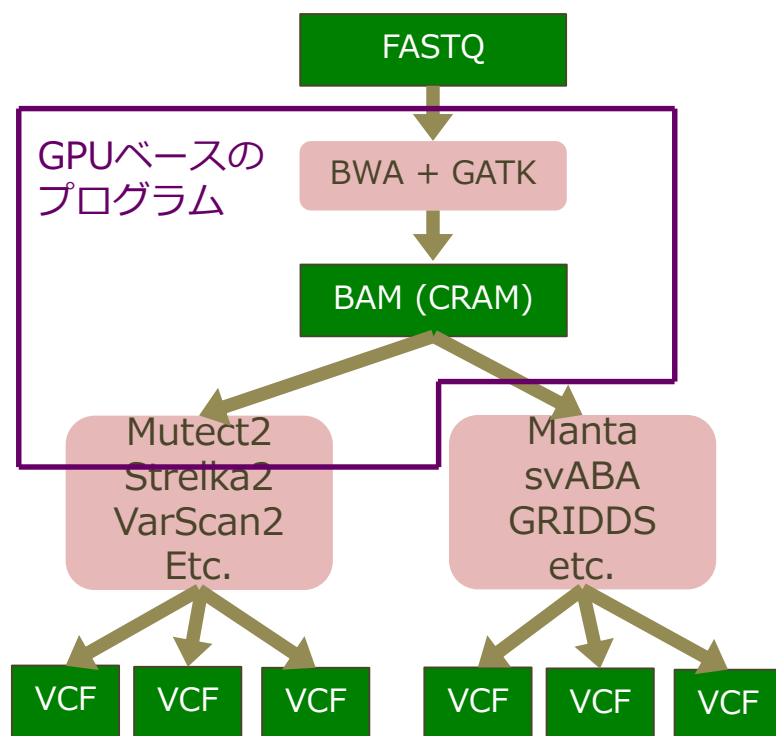
新規モジュールの追加の例



計算時間、有用性などを鑑みて
総合的に導入の可否を決める。

変異コール

- 変異検出については、ソフトウェアの優位性の決着がついていない。
 - 国際がんゲノムコンソーシアムなどでは、複数のソフトウェアの手法を統合的に利用することが一般的。
- 予め主要なソフトウェアについては解析結果を配備しておく。
 - ユーザーが独自にプログラムを実行させる手間を省く。
- アラインメントの方法はできるだけ世界の標準に合わせてクセのない形を選定すべき。
- CRAM以降の解析については、解析パイプラインWGなどで、アップデートを続ける体制を構築。



4. データ解析拠点の構築(案)

- **解析デザインの検討とパイプラインの開発**
 - 基本的な解析デザインの検討
 - そのためのパイプラインの整備
- **各研究グループの解析のサポート**
 - **解析WGがクラウドで行う解析**
 - 定型パイプラインによる基礎的な解析
 - 解析WGが承認したプログラムを全検体のCRAMファイルで実行する。
 - 小規模の仮想サーバー、VDIからCRAMを使った解析
 - **各研究グループの計算サーバー（スパコン含む）で行う解析**
 - VCF + 臨床情報を使った解析
 - 一部のCRAMファイルを使ったテスト的な解析
 - **サードパーティ共有プラットフォームを使った解析**
 - IGVなどのゲノムビューアを使ったCRAMファイルの閲覧。
 - 変異ポータルサイトの閲覧。
 - 任意のバッチジョブ実行。
- **人材の育成**
 - 解析に関わる人材の育成
 - Informatics