

# GA4GH Genomics API による マルチオミクスデータ統合検索

河野 信<sup>1</sup>、鈴木 稔<sup>2</sup>、菅野 純夫<sup>2</sup>

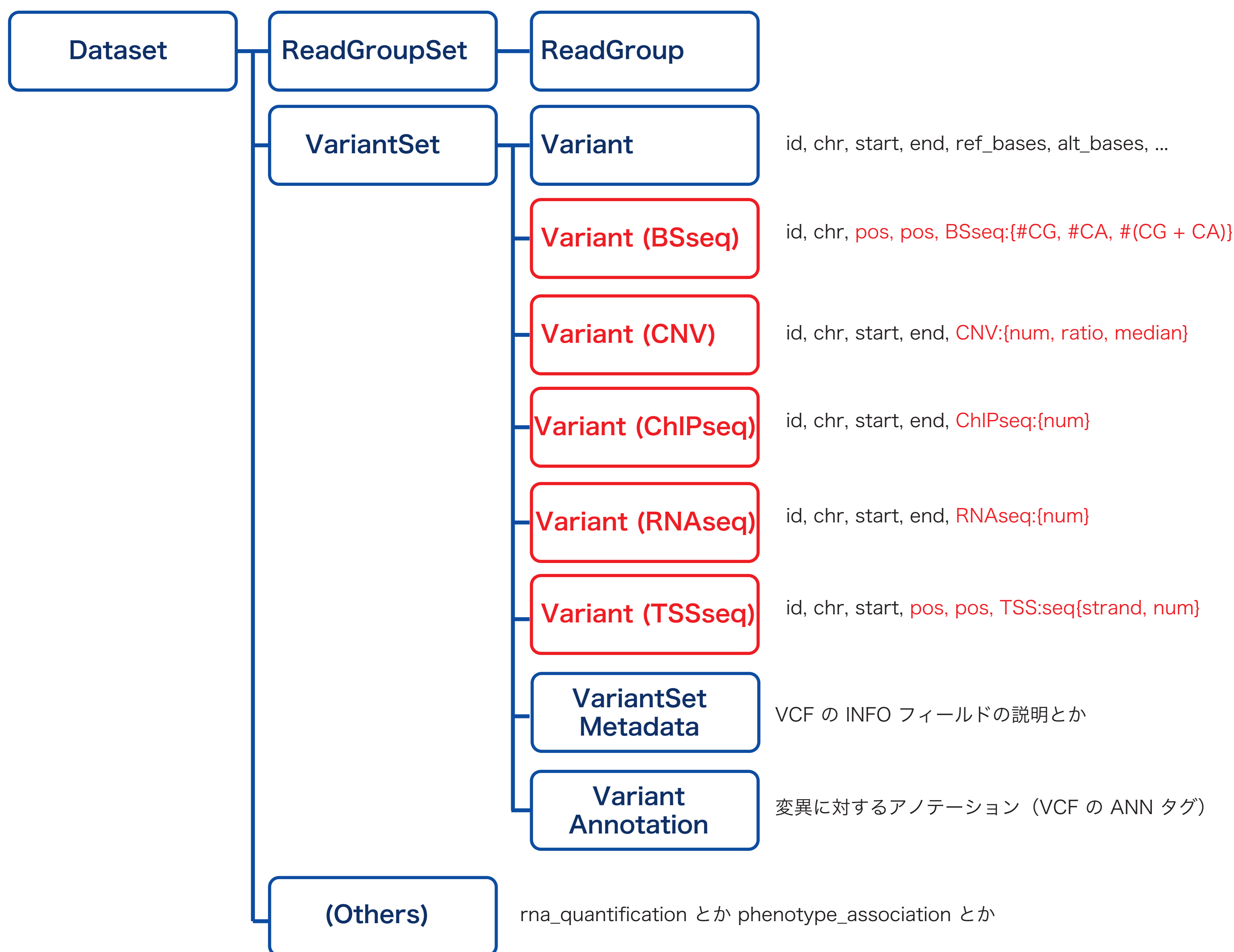
<sup>1</sup>情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター, <sup>2</sup>東京大学大学院 新領域創成科学研究科  
kawano@dbcls.rois.ac.jp

## 要旨

近年、希少疾患や癌の診断においてゲノム情報が利用されるようになってきた。これらを研究や治療、効率的な治療に応用するためには、データを共有することが重要である。ゲノム情報を国際的に共有するための仕組みとして Global Alliance for Genomics and Health (GA4GH) では GenomicsAPI を提供している。GenomicsAPI では変異の情報を共有することに主眼をおいているが、変異だけでなくエピゲノムや遺伝子発現 (以下、マルチオミクス) の情報も統合的に共有できれば、より粒度の細かいデータを研究・診断・治療に利用できる。そこで、ここでは GenomicsAPI のデータモデルを拡張し、マルチオミクスデータに対応させた。サンプルデータとして DBKERO (Database of Kashiwa Encyclopedia for human genome mutations in Regulatory regions and their Omics contexts, <http://kero.hgc.jp/>) に収録されている肺腺がん由来 26 細胞から得られたマルチオミクスデータを利用し、これらを VCF 形式に変換して GenomicsAPI リポジトリに登録した。この結果、変異を含むマルチオミクスデータの GA4GH GenomicsAPI を利用した統合検索が可能となった。

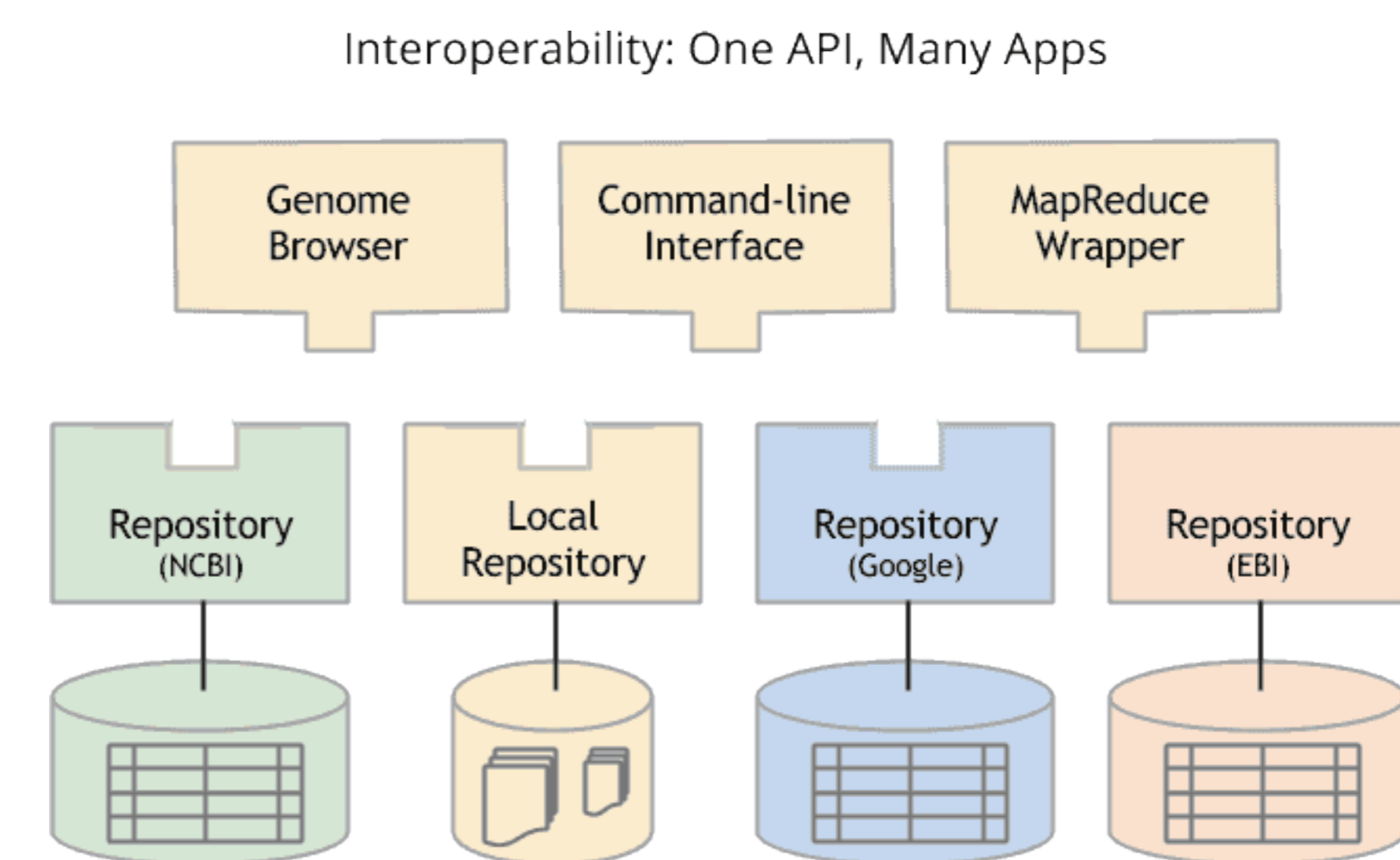
## マルチオミクスデータに対応した拡張スキーマ

既存の GA4GH Genomics API のスキーマ (青) の一部とそれを拡張した API スキーマ (赤)。各種オミクスデータを Variant の拡張として扱うことで、既存のしくみを変更することなくマルチオミクスデータを GA4GH Genomics API で取り扱うことが可能となった。



## GA4GH Genomics API

GA4GH Genomics API は、ゲノム関連情報 (特に変異情報と表現型情報) を国際的に交換するために開発されている、GA4GH によって標準化された API である。ゲノム変異情報に加えて、Genotype と Phenotype の関係や、リファレンス配列情報、メタデータ、アノテーション情報などを扱うことができる。ただし、DNA メチル化情報やクロマチン情報、転写開始地点情報など、変異情報以外のマルチオミクスデータには対応していないのが現状である。



<https://genomicsandhealth.org/work-products-demonstration-projects/genomics-api>

## 実装

マルチオミクスデータを変異情報と同等に扱うために、各種マルチオミクスデータを VCF 形式に変換するスクリプトを開発した。これを用いて DBKERO に登録されている肺腺がん由来 26 細胞から得られた DNA メチル化、コピー数、クロマチン情報、遺伝子発現、転写開始地点の各種データを VCF に変換した。これらと変異情報のデータを GA4GH Genomics API のデータレポジトリに登録し、API を通じてマルチオミクスデータの検索を可能にした。

### 検索例

```
$ ga4gh_client variants-search --referenceName 3 --start 0 --end 1000 http://...

aaaa [u' bs003' ] 3 100 101 A [u' T' ]
bbbb [u' bs004' ] 3 200 201 C [u' A' ]
cccc [u' -' ] 3 300 301 - [u' -' ] BSseq. CG=10;
dddd [u' -' ] 3 400 430 - [u' -' ] CNV. num=20;
eeee [u' -' ] 3 500 520 - [u' -' ] RNAseq. num=25

id gene name chr start end ref alt info
(dummy) (dummy)
```

## まとめ

- GA4GH Genomics API の仕様を拡張し、変異以外のデータについても統合的に検索可能とした
- 今後は本検索ツールを用いて、マルチオミクスの状態が類似したサンプル/患者を検索するツールを実装する予定である

## 謝辞

本システムはGenomedia株式会社様に実装していただきました。また、University of Toronto, SickKids Hospital の Orion Buske 氏からアドバイスをいただきました。本研究は科学技術振興機構バイオサイエンスデータベースセンターが支援する統合化推進プログラム「疾患ヒトゲノム変異の生物学的機能注釈を目指した多階層オミクスデータの統合」の一環として遂行されました。この場を借りて感謝いたします。