

No. 質問

回答

1 PRSのあるSNPの値はどのようにして求めますか？

(八谷講師・講義資料p33参照)

ご質問の意味を正しく理解できているかわかりませんが、リスクバリエントの集合(RV)とvariant j の重み(W_j)がモデルの部分になります。あるサブジェクトがどういったバリエントを持っているのか、 $(X_{i,j})$ は、測定した後にインピュテーションをして決まつてくるという、測定データ側のものになります。あるいは、リスクバリエントの集合をどのように決めているかというご質問だとしますと、モデルの作り方が何種類もあり、形質とのアソシエーションのp値でしきい値を決めるというものもあつたり、しきい値を使わずに、極力全部のバリエントを含むような形でやるアルゴリズムもあります。そうした形で、このリスクバリエントの集合は、あるものは數十個ぐらいまで絞られたりもしますが、逆に数百万バリエントをモデルに含むというものもありますので、ここでのリスクバリエントというのは、適用されるモデル、アルゴリズムによって全く定義が変わつてくるという形になっています。

2 使用経験がないのですが、マシンの必要スペックなどわからない場合は不明で申請可能なのでしょうか

わからぬものは、わからぬのでどうしたらしいですか、と言つていただければ対応できるかと思います。コア数とRAMは、推奨の値をとりあえず入れておいていただければ問題ないと考えております。送りづらいと思われてしまうフォーマットになつてゐるかも知れませんが、ためらわずにお送りいただければ幸いです。

3 ある特定のphenotypeに対するいくつかのモデルがある場合、どのようにして良いモデルを選ぶことができるでしょうか。

(八谷講師・講義資料p34参照)

これは、ボリュームニクスカのモデルについてかと思います。それについてには、まずPGSカタログ側で過去の評価結果を予測精度として示しているところがあつたり、PGSカタログを見いくと、トレーニングデータとして、どういったAncestryのデータが含まれているかなども書いてあります。その多くの場合は、ご参加いただいている方の研究ですと、東アジア人のデータがトレーニングデータに入っているかどうかですか、それがバリエーションとしてしっかり精度が出てるのか、そういうところを見つけていただくのがいいかと思います。もし、ご自身のデータで複数比較するというようなことでしたら、このTogoImputationの仕組みですと、かなり簡単に複数のモデルを選んで、その結果を一気に計算して、10個20個の計算結果を比較するようなことも可能になっていますので、ご自身のデータの中で、どれが一番オッズ比が高いか、というようなことを検討いただくにも手間が少なく、複数のモデルを試していただくことができるかと思います。

4 BBJ の 7472 人から構築したリファレンスパネル (JGAS000738) の実装予定はございますか？

具体的なリリース日などは申し上げられないのですが、ぜひやりたいと考えています。おそらくJGAにすでに登録されているパネルのことと推察いたします。具体的な人数は忘れたのですが、フルカバーレッジの4000人ぐらゐのパネルと、ローカバーレッジのものを足した7400が登録されていたかと思うので、それをTogoImputation用にコンバートして、こちらからTogoImputationで利用できるようにするということを、ぜひやりたいと考えているところであります。

5 PRSモデルの中身の説明ではGRCh37ベースとなっていましたが、GRCh38ベースもあるのでしょうか？

PGSカタログ側で37で登録されたモデルを38にコンバートするという、一律でコンバートするようにされていますので、38ベースのモデルもあります。しかし、コンバート自体がどこまで正しいのかというところは、リフトオーバーはどれだけ正しいのかという問題になるのですが、一応(38ベースのモデルも)あります。多くの場合は、そんなに大きな問題にはならないかと思います。注意点として、自動コンバートしたモデルは、パーエクトなコンバージョンになつてないかも知れないというのを念頭に置いてください。多くの場合は問題にならないかと思います。

6 TogoVarの検索で遺伝子名を用いて検索したときはエンハンサー領域はカバーされない結果になりますか？

はい、これはカバーされません。TogoVarではEnsembl GeneまたはEntrez Geneのいづれかで遺伝子とされているゲノム領域を遺伝子領域と定義しています。Ensembl GeneもEntrez Geneも転写産物ベースで遺伝子領域を決定していますので、多くの場合、その上流にあるエンハンサー領域はカバーされない結果になります。

7 SVの情報がそのうち入ることですが、いつ頃、どのくらいの規模の集団データについて予定されていますか？

これは準備をしております。1つはgnomADです。gnomADのv4.1 (<https://gnomad.broadinstitute.org/news/2023-11-v4-structural-variants/>)のSVを入れたいと思ってます。2つ目は、ToMMoのJSV1 (https://jmorp.megabank.tohoku.ac.jp/docs/guide-en/dataset_detail/tommo-jsv1/)です。この2つに関しては、すでに公開されているものになります。3番目は、先ほど紹介しましたJoGo (<https://jogo.csml.org/faq/>)です。JoGoでもSVがコードされていますので、JoGoのデータ、SV以外のものを含めて公開したいと思ってます。今交換準備を進めていますので、できれば3月には公開したいと考えております。

※AJACSの趣旨と異なる質問は、回答および掲載を控えさせていただきました。