

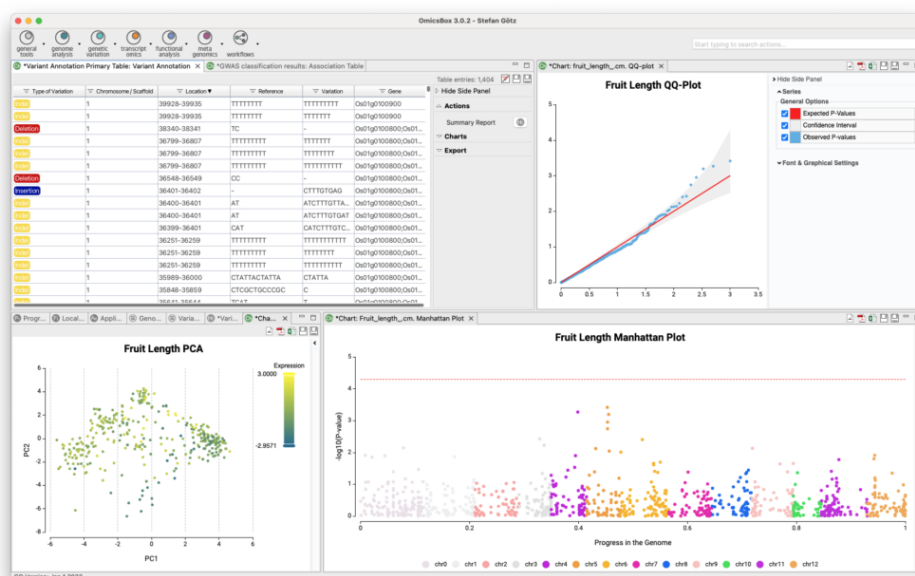
OMICSBOX を使用した GWAS



ゲノムワイド関連研究 (GWAS) は、遺伝的変異と表現型形質の関係を調査するための貴重なアプローチです。ただし、分析オプションが多数あるため、特に統計学や生物情報学に関する強力な背景がない研究者にとって、GWASの実行は困難な場合があります。

OmicsBoxのGenetic Variation Moduleは、強力な分析ツールをユーザーフレンドリーなインターフェイスで表示することで解決策を提供します。これには、GAPIT3 Rパッケージを利用するGWASオプションが含まれており、解析をより分かりやすく簡単に行うことができます。

この投稿では、GWAS分析にOmicsBoxを使用する利点について説明し、ステップバイステップのガイドを提供します。また、正確かつ効果的な分析を確保するための重要な考慮事項とベストプラクティスも強調しています。



主な入力ファイル

GWAS分析を実行するには、バリエーション情報を含む VCFファイルと、表現型の情報を含む形質ファイルの2ファイルが必要です。

File #1. VCF File

VCFファイルには、各サンプルの現在のバリエーションに関する情報が含まれています。これは、OmicsBoxの超高速[Variant Calling Tools](#)のいずれかを使用して取得できます。その後、[Variant Filtering Tool](#)により、最も信頼性の高いバリエーションを選択し、低品質の変異を破棄することができます。

File #2. Traits File

表現型ファイルを作成するには、バリエーションに関連付ける必要があるさまざまな特性を記述した、対応する値を含むサンプル名のプレーンリストを提供します。このファイルは TSV ファイル形式である必要があります。

Genome Wide Association Study

Input

Genome Wide Association Studies (GWAS) examine the association between SNPs and traits. Please select the parameters to run your GWAS.

To start, you will need to provide a VCF file and a TSV file containing phenotype data. The VCF file is a result of the variant calling process and may have been filtered, while the TSV file includes sample names and phenotype information.

VCF File ?

/home/enrique/gwas/melon.vcf.gz

Phenotypic Traits File ?

/home/enrique/gwas/phenotypes.tsv

Default < Back Next > Cancel Run

Figure 1. Main Input Files Wizard

適切なデータセットを作成する

フィルタリング：

特定のバリエーションフィルターは、重要な結果へ導くために GWAS で一般的に使用されており、詳細については、この[チュートリアル](#)をご参照ください。SNPをフィルタリングするだけでなく、[ハーディー・ワインベルグ平衡](#)に違反するバリエーションを排除するためのp値しきい値や[MAFしきい値](#)などの集団遺伝学パラメータに基づいてVCFファイルをフィルタリングすることもできます。表現型形質が正規分布していることを確認することも同様に重要です。データがこの分布に従わないと思われる場合は、正規化オプションを使用します。

GWAS オプション

GWAS を実行する場合、次の3つの側面が重要であり、よく理解する必要があります。

1. Kinship Matrix

親族関係、すなわち集団内の遺伝的近縁度は、GWAS において一般的な交絡因子です。交絡因子とは、遺伝的変異と研究対象の表現型の両方に関連しているが、観察された関連性の原因因子ではない変数のことです。親族関係の場合、サンプル間の遺伝的近縁性が、遺伝マーカーと表現型との間に偽陽性の関連を作り出す可能性があります。親族関係を考慮することで、個体間の遺伝的近縁性の影響を補正し、真の関連と偽陽性を区別することができます。

親族関係を調整するために、研究者は独自の親族関係マトリックスを解析に加えることができます。しかし、GWAS プロセスを簡素化するために、OmicsBoxは自動親族関係計算機能を提供しています。

親族関係を考慮することで、研究者は遺伝的変異と表現型との真の関連をよりよく同定することができ、GWASにおいてより正確で信頼できる結果を導くことができます。

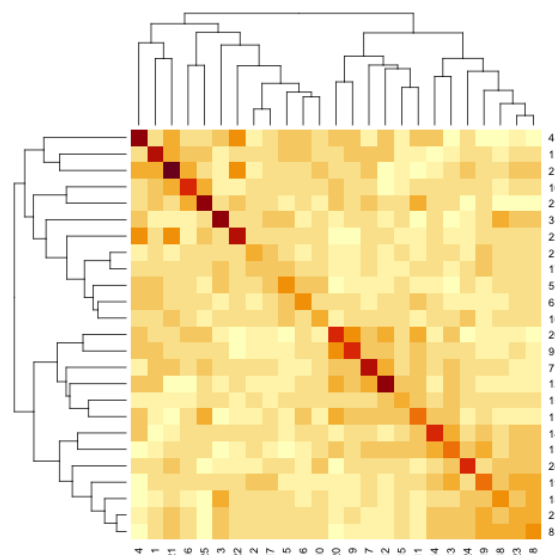


Figure 2. Kinship Matrix

2. Covariate Matrix

Kinship Matrixに加えて、Covariate MatrixもGWASにおいて交絡因子を説明するために使用されるツールです。これらの因子を解析に含めることで、それらの影響をコントロールすることができ、結果の精度を向上させることができます。

covariate matrixの求め方がわからない場合は、OmicsBoxでVCFファイルを用いて自動計算することができます。

3. Model

形質と集団構造の間の相関関係は、使用する分析方法に応じて変化する可能性があり、最適なモデルも異なります。しかし、OmicsBoxを使ったGWASの実行は簡単で、結果を比較するために複数のモデルで実験を行うことをお勧めします。さらに、[QQ-プロット](#)は、モデルが適切に調整されているかどうかを判断するための品質管理尺度として使用することができます。

さまざまなGWASモデルを知るためには、以下の論文が役に立つ場合があります。：
「[Comparing Different Statistical Models and Multiple Testing Corrections for Association Mapping in Soybean and Maize](#)」

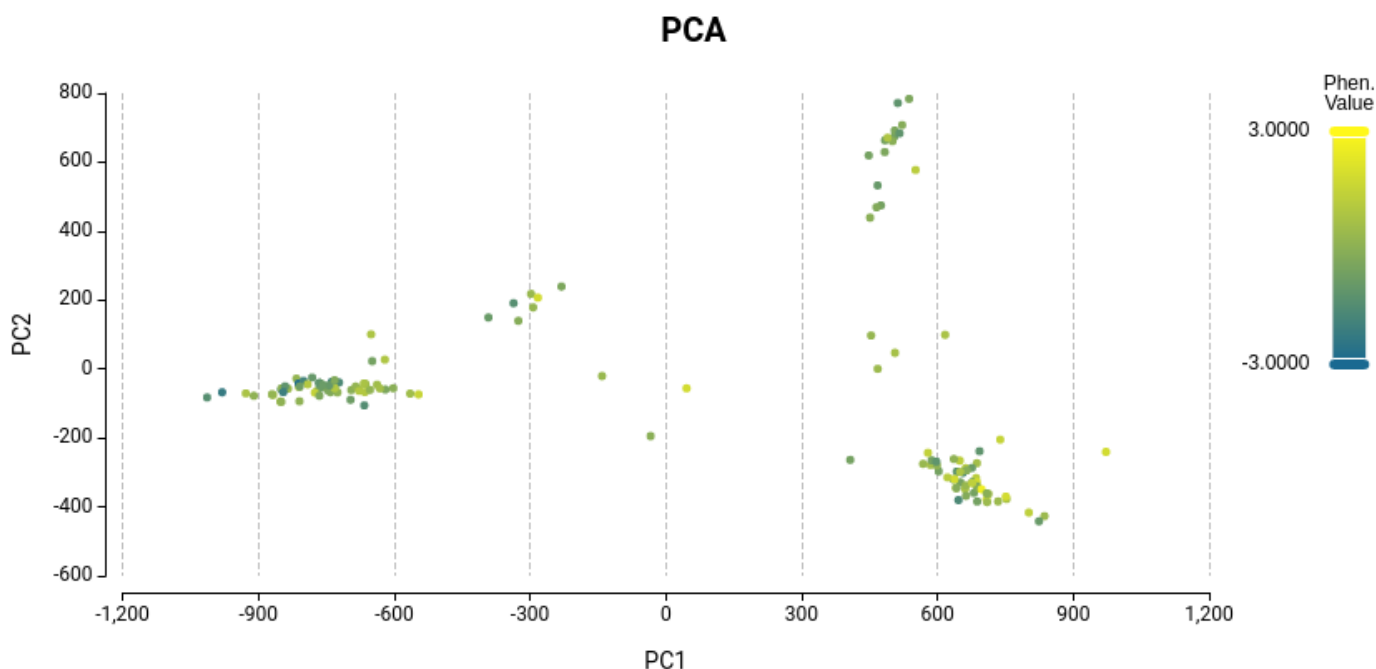


Figure 3. Covariate MatrixのOmicsBoxのPCA。サンプルは表現型の値によって色分けされています。

結果を分析する

適切なモデルでGWASを実行すると、異なる表現型シナリオごとに異なるバリエーションに関する情報を含む表が表示されます。さらに、使用したバリエーションとサンプルの数に関する情報と、表現型形質ごとの有意な変数の数を示す小さな表を含むサマリーレポートが表示されます。(Fig. 4を参照)。

このモデルがデータに適しているかどうかを確認するには、サイドバーの Charts > Phenotype Information > QQ-plots をクリックします。QQ-プロットは、GWASでp値の分布が期待される一様分布に従っているかどうかをチェックするために使用されます。良好なQQ-プロットは対角線を示し、p値の観察された分布が期待される分布に従っていることを示し、プロットの上部付近のわずかな偏差が有意な関連を示しています。観察された分布が期待された分布から大きくずれている場合は、データの潜在的なエラーやバイアスを示唆しており、さらなる調査が必要です。

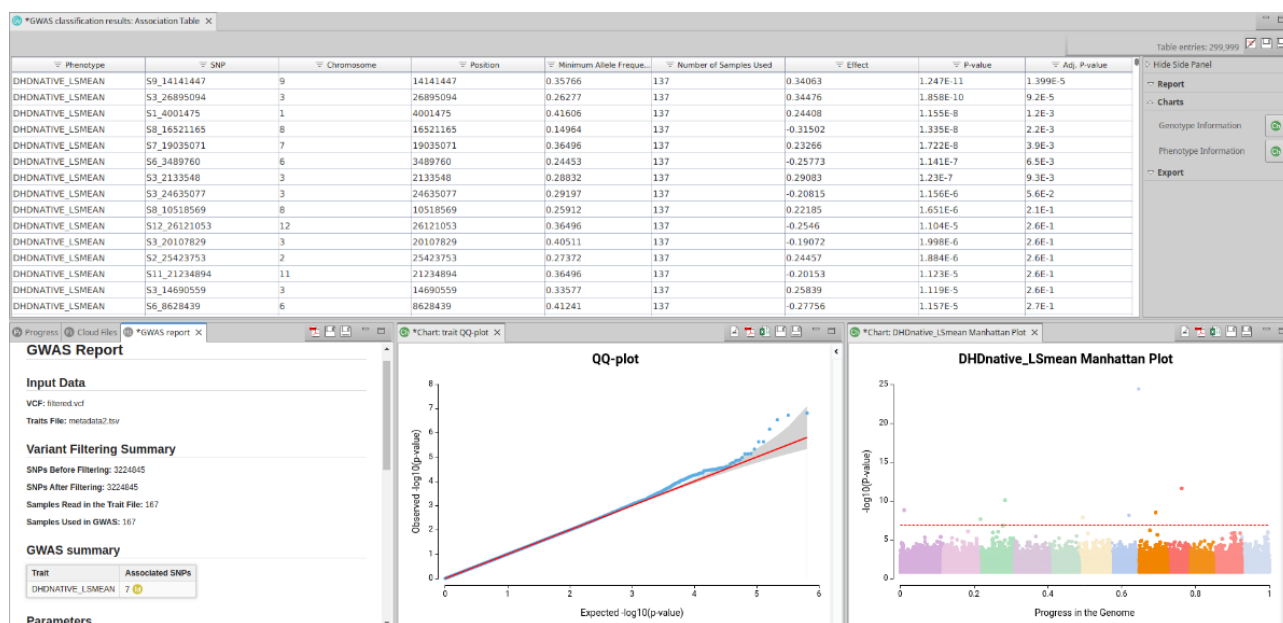


Figure 4. OmicsBoxにおけるGWAS結果の概要

モデルがデータに適切であることを確認したら、マンハッタンプロットをみてください(Charts > Phenotype Information > Manhattan Plot)。マンハッタンプロットは、遺伝的変異と対象の表現型の結果を表示するためのグラフィック表現ツールです。マンハッタンプロットの各点は表現型との関連性についてテストされた遺伝的変異を表し、点の高さは関連性の強さを示します (Fig. 5)。



Figure 5. OmicsBox のマンハッタンプロット

OmicsBox の GWAS

OmicsBoxを使用したGWASの実行は簡単で、必要なのはVCFファイルと表現型ファイルのみです。この使いやすさは、新規ユーザーにとって特に役立ちます。

経験豊富なユーザーには、親族関係やCovariate Matrixなどのカスタム情報を追加するオプションがあります。

OmicsBox は直感的なウィザードを提供し、GWAS分析のすべてのステップをユーザーにガイドします。

References

•Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., ... & Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics*, 28(18), 2397-2399.

•Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics*, 90(1), 7-24.



実績は高いがコマンドライン型であったりOSに制限があるオープンソフトウェアを多数搭載

それらの解析をマウス操作で簡単に解析できる

- 本稿で使用したGWASツールの他、このツールに使用するVCFファイル作成のためのバリアントコールツールを搭載しており、スムーズにGWAS解析に移行できます。
- OmicsBoxの紹介ページは[こちら](#)