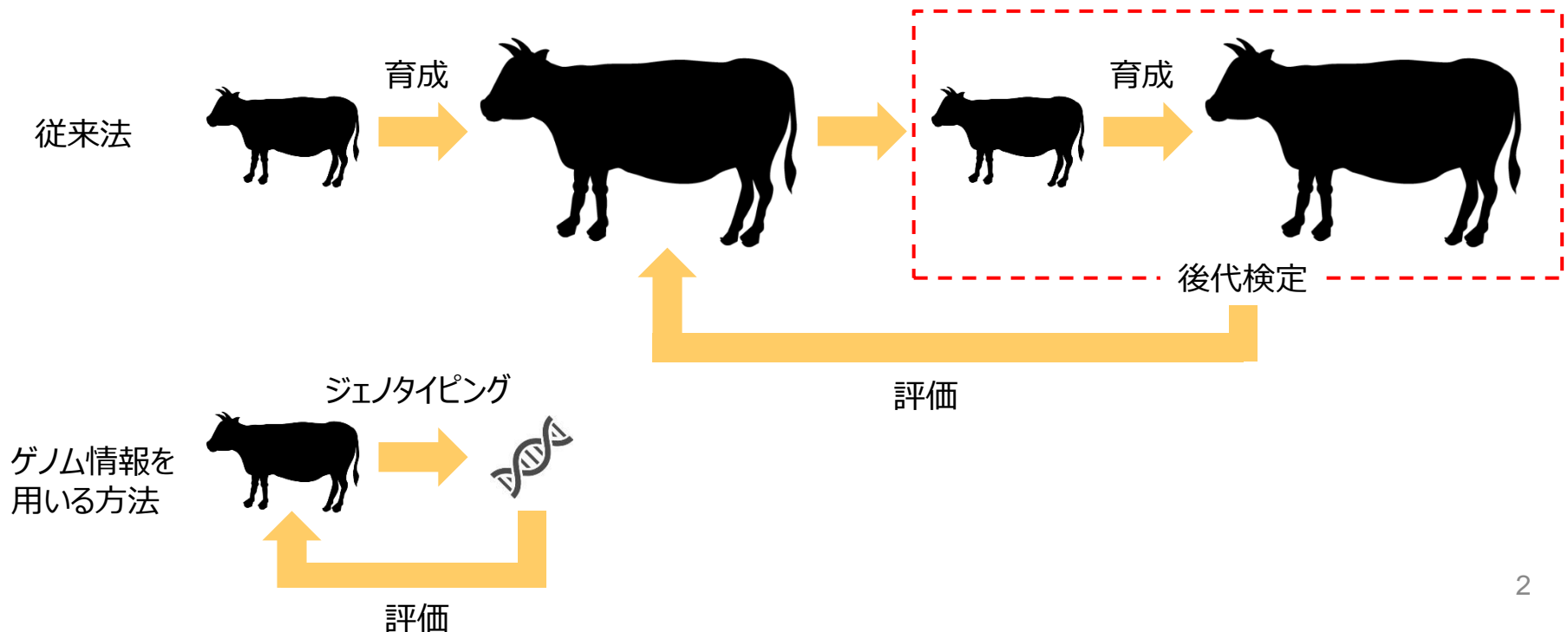
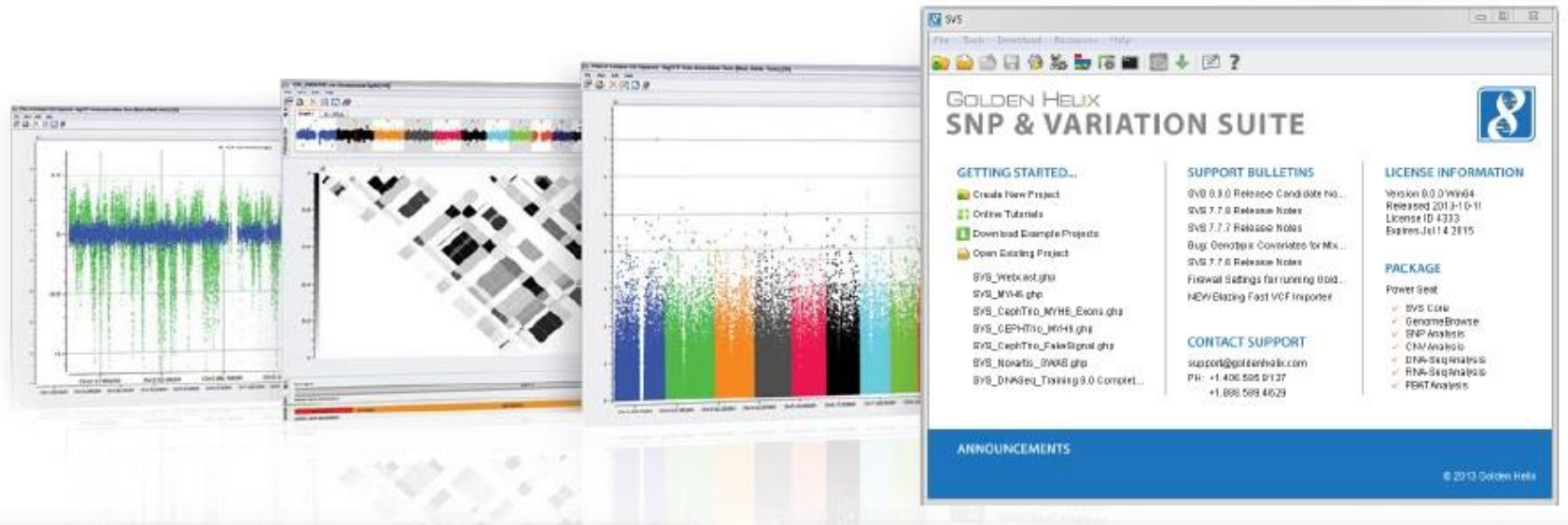


ゲノムデータを用いた表現型予測

フィルジェン株式会社 バイオインフォマティクス部
(biosupport@filgen.jp)

- ウシをはじめとした家畜の改良には、従来は選抜の候補となる個体の育成後、さらに後代となる個体の育成を行い、その成績より選抜を行うため、評価に時間とコストがかかる
- これに対して個体のゲノム情報に基づいた評価では、生後すぐに評価が可能のため、育成に必要な時間とコストを大幅に節約できるが、評価対象の表現型を、ゲノム情報から正しく予測するための特別なデータ処理を行う必要がある
- Golden Helix社SNP & Variation Suite (SVS)では、このようなゲノム情報から個体の表現型の予測を行うゲノミック予測の専用ツールが搭載されており、家畜や農作物などの改良に用いられている





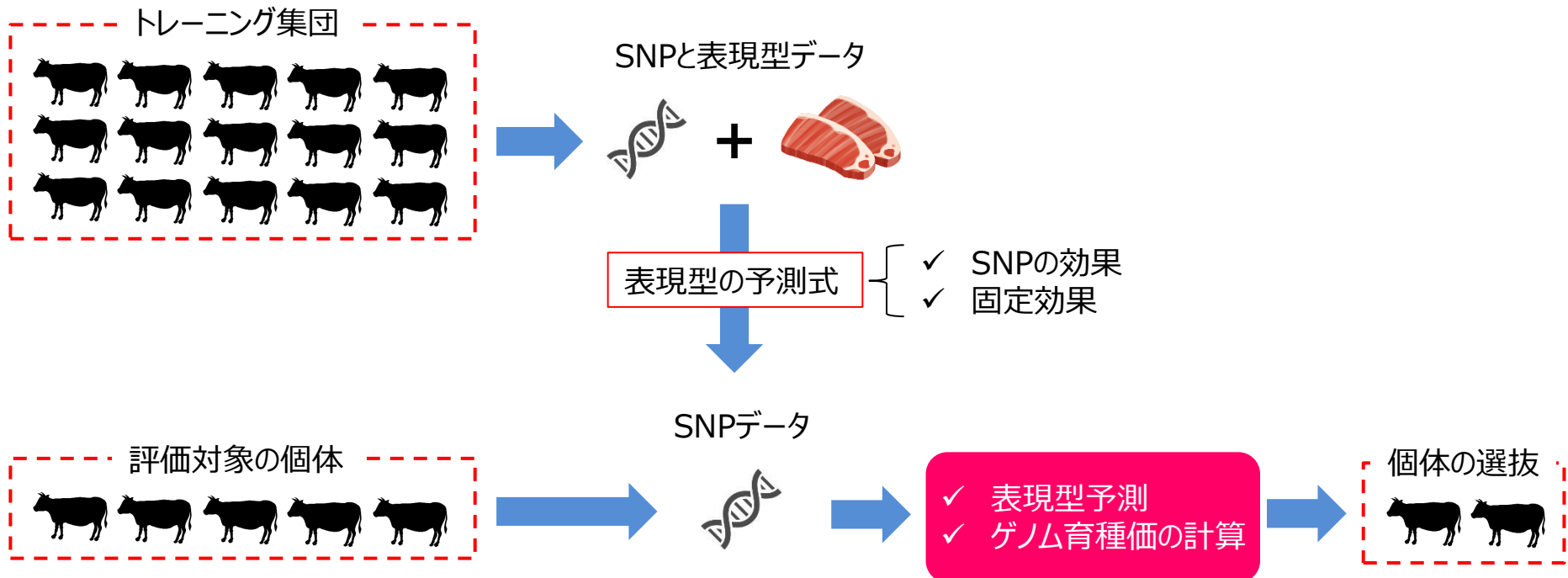
Core Features

- Powerful Data Management
- Rich Visualizations (GenomeBrowse)
- Robust Static
- Flexible

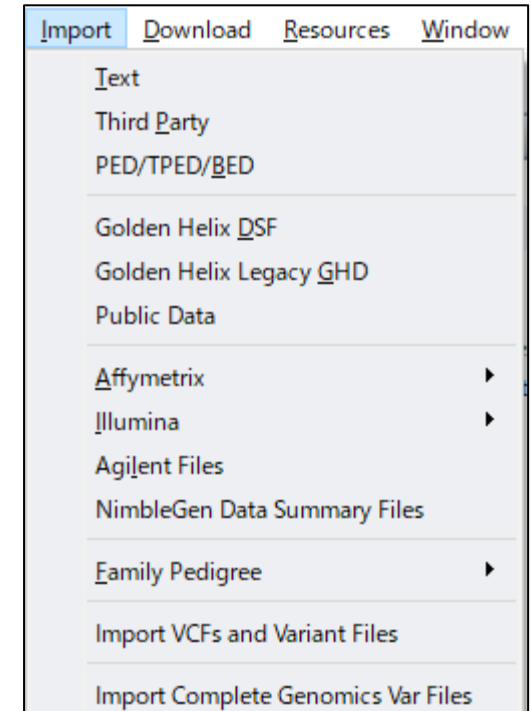
Applications

- Genotype Analysis
- Agrigenomics Analysis
- DNA/RNA Sequence Analysis
- CNV Analysis

- ゲノミック予測による予測モデルの構築には、評価対象とは別のトレーニング集団のデータを用いる
- トレーニング集団の個体は、個体ごとの表現型データと、全ゲノムを網羅したSNPデータを必要とし、これらに基づきSNPデータから表現型を予測する式を構築する
- 評価対象の個体はSNPデータのみを必要とし、これに予測式を適用することによって、個体ごとの表現型と、遺伝的能力の評価値であるゲノム育種価を算出する

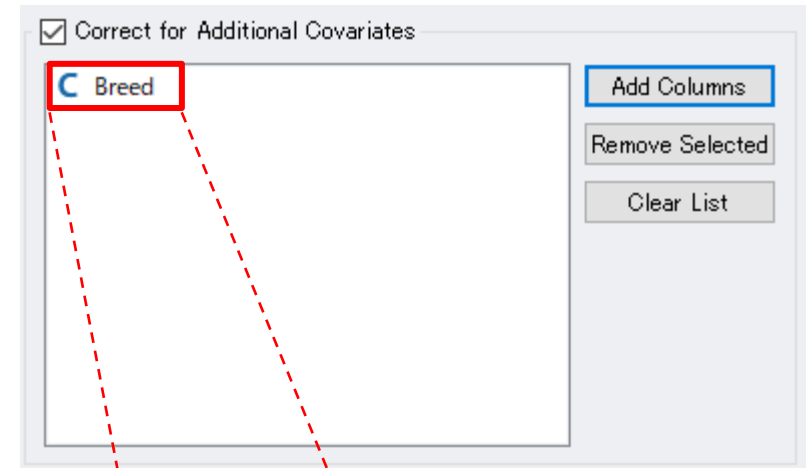


- トレーニング集団より表現型の予測式を作成する場合は、サンプルごとに SNPデータと表現型データの両方が必要
- それぞれのデータは、多様なフォーマットで作成されたファイルをソフトウェアに取り込み、解析に使用することが可能
 - ✓ SNPデータ
 - Affymetrix CHPファイル
 - Illumina Final Reportファイル
 - VCFファイル
 - ✓ 表現型データ
 - Excelファイル
 - タブ区切りテキストファイル
- 取り込まれたデータはスプレッドシートで表示され、データの追加などの加工や編集が可能

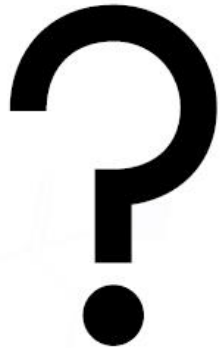


Unsort		C 1	C 2	R 3	G 4	G 5	G 6	
Map	Sample Label	SampleID	Breed	Phen5	Hapmap43437-BTA-101873	ARS-BFGL-NGS-16466	ARS-BFGL-NGS-19289	
1	WG0099889-DNAD04 ANG000027	ANG000027	Angus	24.8389532585512		G_G	C_C	A_A
2	WG0099889-DNAA02 ANG000008	ANG000008	Angus	21.7506929572886		G_G	C_T	A_A
3	WG0099889-DNAA03 ANG000016	ANG000016	Angus	17.3037717691138		G_G	C_T	A_A
4	WG0099889-DNAB03 ANG000017	ANG000017	Angus	26.6011234866098		G_G	C_C	A_A
5	WG0099889-DNAB04 ANG000025	ANG000025	Angus	23.4778765840422		G_G	C_C	A_A
6	WG0099889-DNAC02 ANG000010	ANG000010	Angus	27.0589945698839		G_G	C_C	A_A
7	WG0099889-DNAC03 ANG000018	ANG000018	Angus	15.7245152305321		G_G	C_C	A_A
8	WG0099889-DNAC04 ANG000026	ANG000026	Angus	6.34720842916488		G_G	C_C	A_A
9	WG0099889-DNAD02 ANG000011	ANG000011	Angus	23.8891657707582		G_G	C_C	A_A
10	WG0099889-DNAD03 ANG000019	ANG000019	Angus	16.2528640600238		G_G	C_T	A_A
11	WG0099889-DNAE01 ANG000004	ANG000004	Angus	25.5057310813833		G_G	C_T	A_A
12	WG0099889-DNAE02 ANG000012	ANG000012	Angus	29.8261600049517		A_G	C_C	A_A
13	WG0099889-DNAE03 ANG000020	ANG000020	Angus	25.3954670386242		G_G	C_T	A_A

- 計算実行時のパラメータオプションで、環境要因などの固定効果に用いる共変量を指定する
- 共変量データは、表現型データとともに、Excelにまとめたものを使用可能
- 指定された共変量では固定効果が計算され、予測式として利用が可能



	C 1	C 2	C 3	C 4	B 5
Sample Label	SampleID	Breed Code	Breed	Species	Phenotype1
WG0099889-DNAG01 ANG000006	ANG000006	ANG	Angus	Bos taurus	0
WG0099889-DNAG02 ANG000014	ANG000014	ANG	Angus	Bos taurus	1
WG0099889-DNAH02 ANG000015	ANG000015	ANG	Angus	Bos taurus	1
WG0099889-DNAH03 ANG000023	ANG000023	ANG	Angus	Bos taurus	1
WG0099889-DNAA05 BMA000005	BMA000005	BMA	BeefMaster	Bos taurus	0
WG0099889-DNAB05 BMA000006	BMA000006	BMA	BeefMaster	Bos taurus	0
WG0099889-DNAB07 BMA000022	BMA000022	BMA	BeefMaster	Bos taurus	0



OR



■ ゲノム育種価の計算

- SNPデータを使用し、個体の遺伝的能力（育種価）の推定を行う
- 表現型や血統情報のみの場合より、正確な遺伝的能力の評価が可能

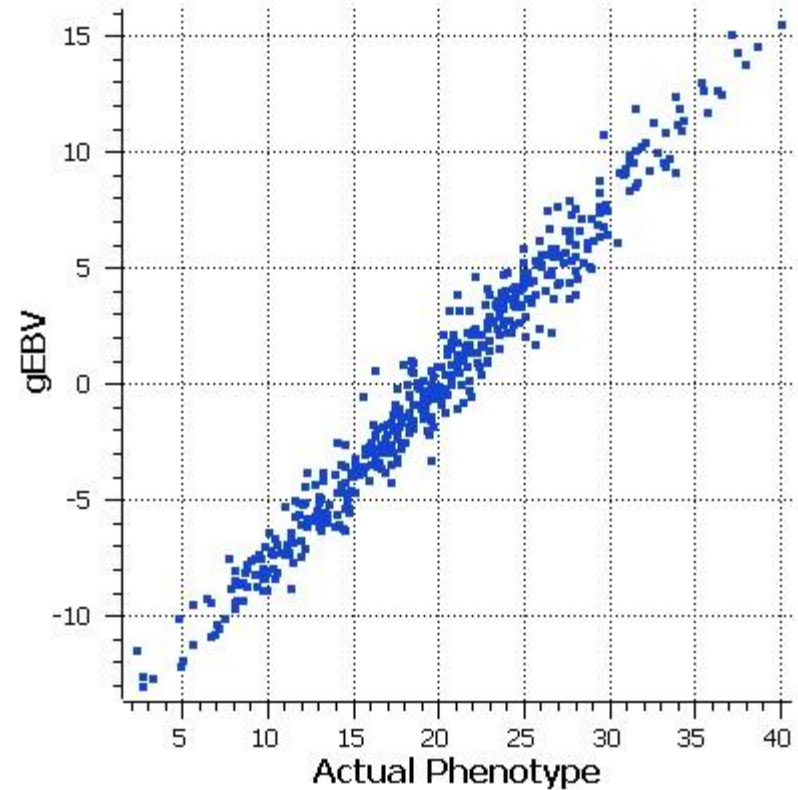
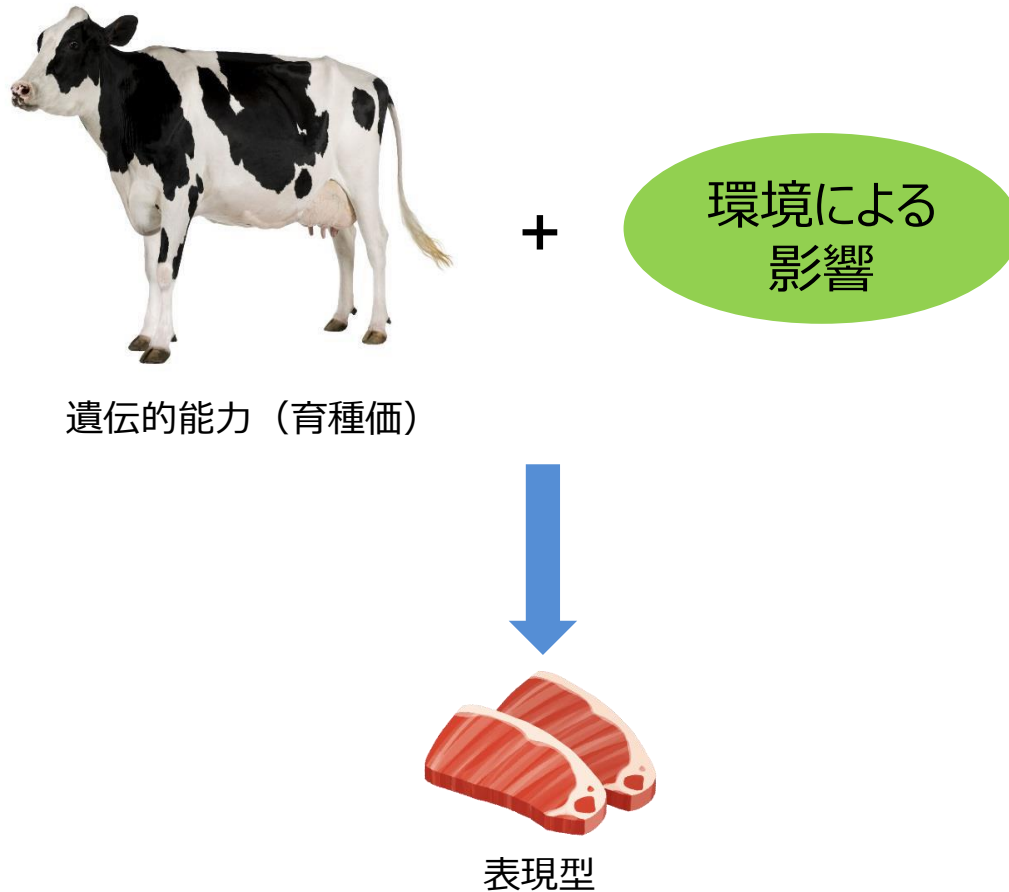
■ 表現型未知個体のゲノム育種価の推定

- トレーニング集団より算出した予測式を用いて、表現型未知個体のゲノム育種価と表現型を推定
- 表現型データが不要なため、飼育コストや期間の削減が可能

■ DNAマーカーの同定

- 表現型の発現に寄与するSNPを、網羅的に同定が可能

- 多数のSNPの表現型に対する累積効果より計算された、個体ごとの遺伝的能力のスコア
- 遺伝的能力に環境による影響を加えたものが、表現型となる
- 育種価と表現型は、高い相関を示す

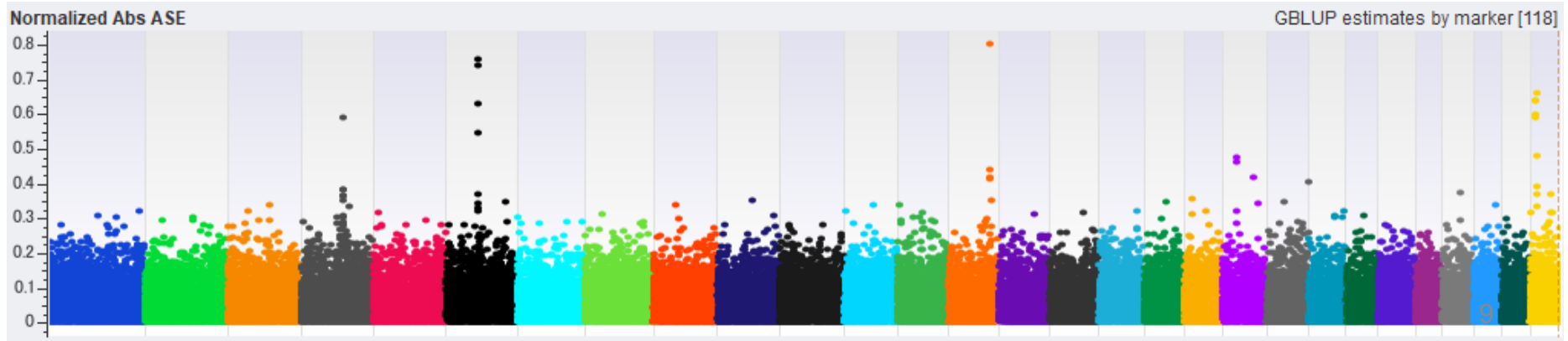


アリル置換効果 (ASE)

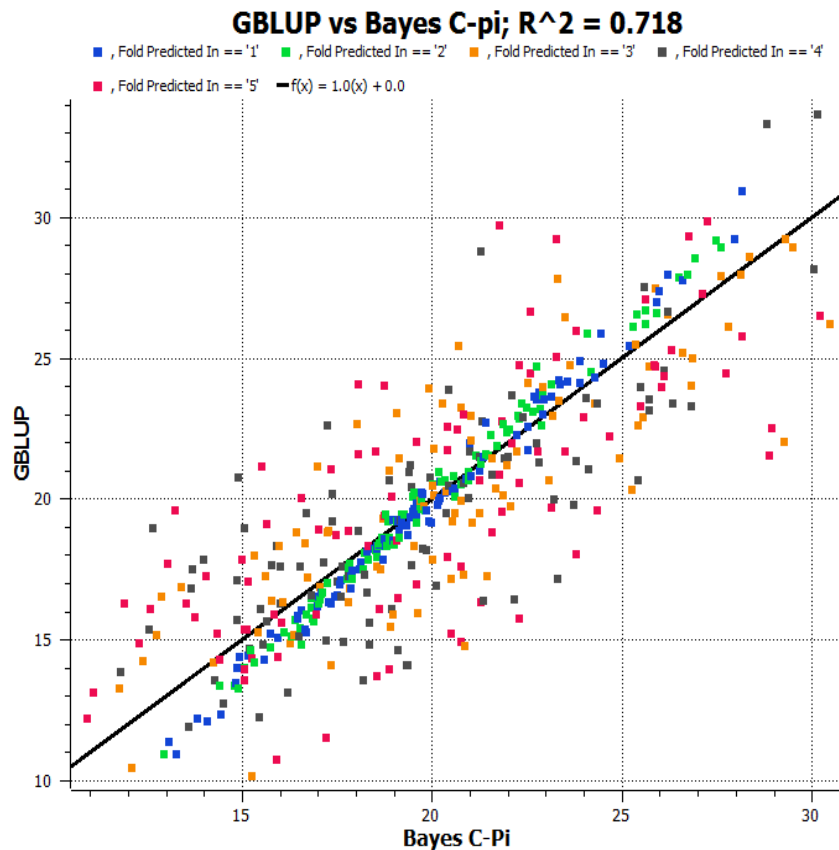
- ゲノミック予測では、ゲノム育種価に加え、SNPごとの表現型への寄与度を評価したアリル置換効果も算出される
- ゲノムブラウザーにデータを反映させることで、マンハッタンプロットの作成が可能
- トレーニング集団より算出したアリル置換効果データを、表現型の予測式として利用

Marker	Chromosome	Position	Design Strand	Allele substitution effect (ASE)
ARS-BFGL-BAC-33065	20	527561	[T/C]	0.00298217530987558
ARS-BFGL-NGS-104829	20	571472	[T/C]	-0.00588923650059982
ARS-BFGL-NGS-65452	20	593947	[A/G]	0.000775341189459952
ARS-BFGL-NGS-28987	20	654738	[C/G]	0.00150185299476017

ゲノムブラウザーにプロット



- ゲノミック予測の計算には、GBLUPとBayes C、およびBayes C-piの3つの手法が選択可能
- いずれの手法も、サンプルの血統情報を必要とせず、SNPデータを用いる
- SNPデータに加え、サンプルごとの表現型、あるいは表現型の予測式が計算時に必要



GBLUP

- ✓ 全SNPが表現型に影響を及ぼしていると仮定する

Bayes C

- ✓ 個々のSNPの表現型に及ぼす効果が、確率 π で0になる (π の値は固定)

Bayes C-pi

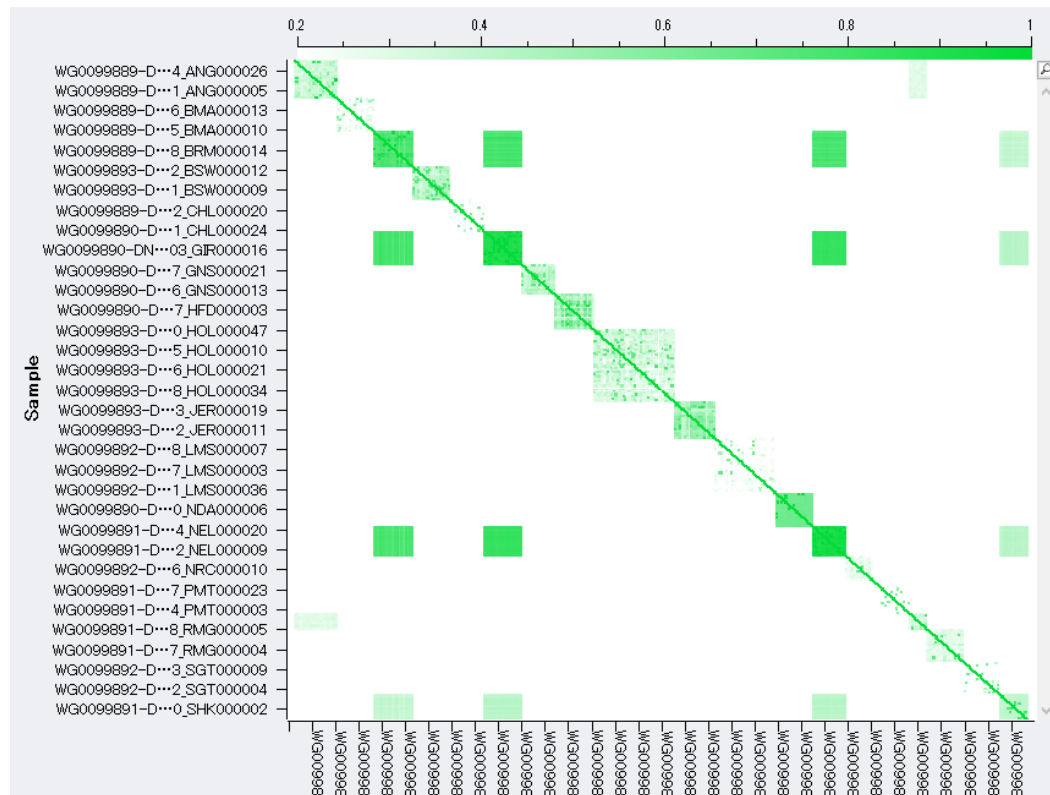
- ✓ 個々のSNPの表現型に及ぼす効果が、確率 π で0になる (π の値は調整可能)

ゲノム関係行列 (GRM)

Sample Label	WG0099889-DNAD04 ANG000027	WG0099889-DNAA02 ANG000008	WG0099889-DNAA03 ANG000016	WG0099889-DNAB03 ANG000017
WG0099889-DNAD04 ANG000027	1.2234246647787	0.268899125020094	0.272621627631719	0.266284395181538
WG0099889-DNAA02 ANG000008	0.268899125020094	1.25742844951306	0.289594298492229	0.345498875240711
WG0099889-DNAA03 ANG000016	0.272621627631719	0.289594298492229	1.25499658548318	0.328415665106116
WG0099889-DNAB03 ANG000017	0.266284395181538	0.345498875240711	0.328415665106116	1.22519831767652



ヒートマップでプロット



- GBLUPでは、サンプル間のゲノム関係行列データが同時に出力され、サンプル間のSNPの相関を確認できる
- 行列データは、ワンクリックでヒートマップ表示に切り替えが可能
- あらかじめゲノム関係行列データを作成しておき、GBLUP、Bayes C/C-piともに、計算時のパラメータオプションとして選択することで、サンプル間の補正に利用することが可能

GBLUP

- ✓ 全SNPが、表現型に弱く寄与していることを仮定
- ✓ 高速で計算が可能
- ✓ 分散推定にREML法を使用
 - EMMA
 - AI REML
- ✓ すべての座位の分散は同一を仮定

Bayes C-pi

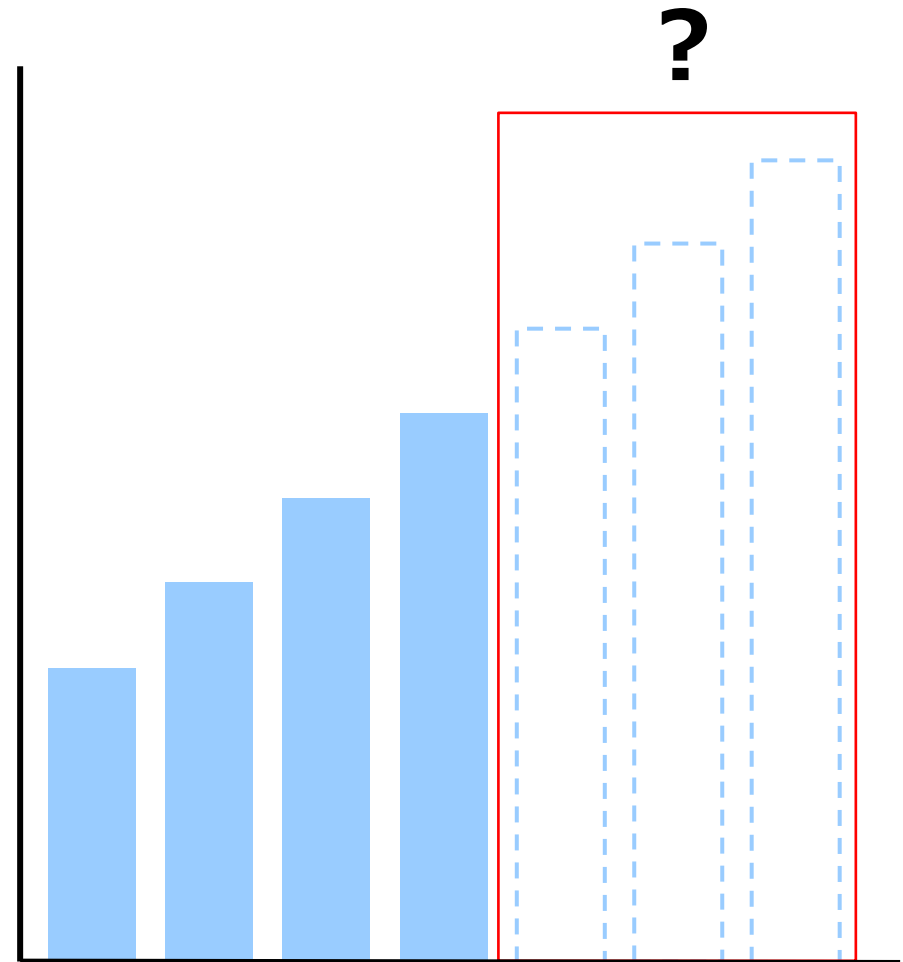
- ✓ ゲノム育種価とアレル置換効果、予測表現型を出力
- ✓ サンプル間のゲノム関係行列 (GRM) を計算に利用
- ✓ 少数の効果の強いSNPが、表現型に寄与していることを仮定
- ✓ 計算に時間がかかる
- ✓ パラメータ π により、SNP効果の事前確率を調整

算出した予測式の問題点

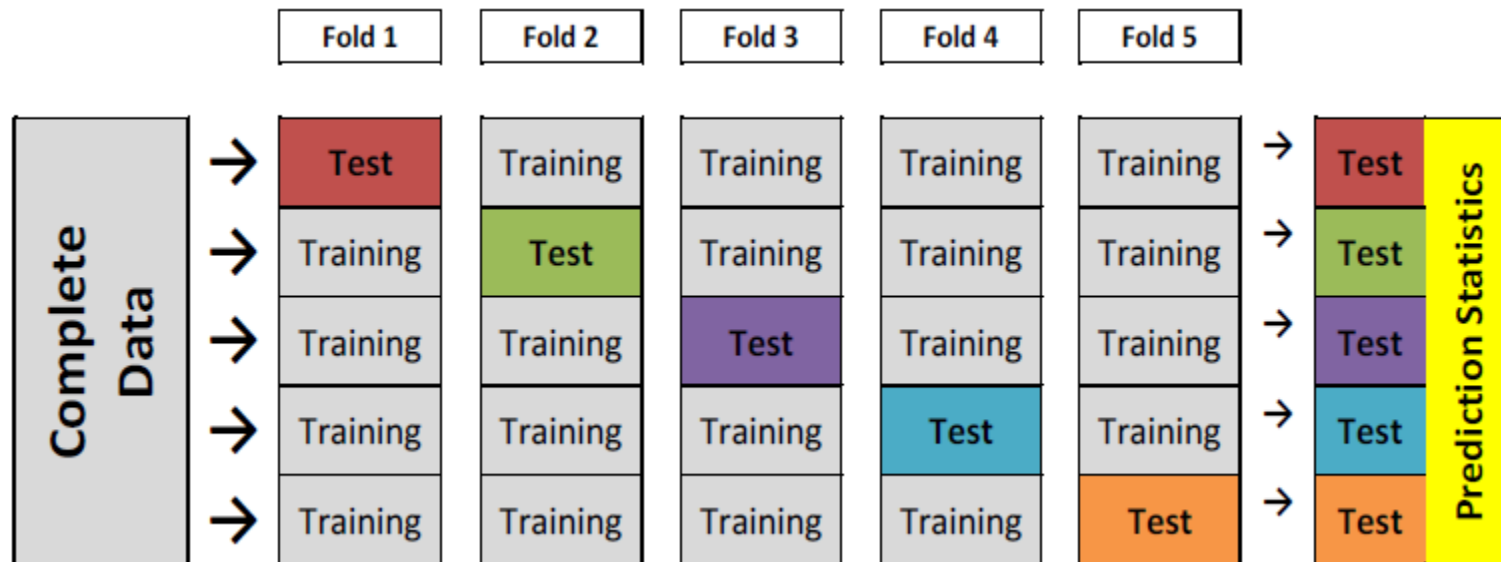
- ✓ 予測を正しく行えるのか？
- ✓ 採用した手法は妥当なのか？
- ✓ 選択したパラメータは妥当なのか？



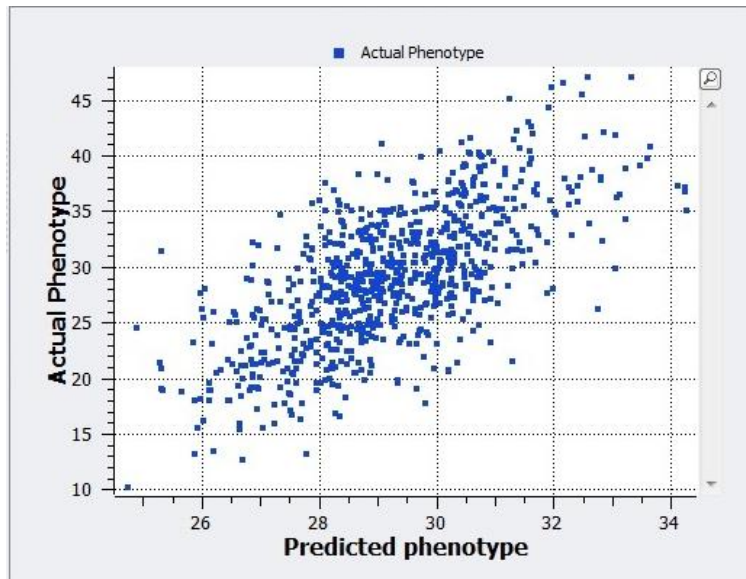
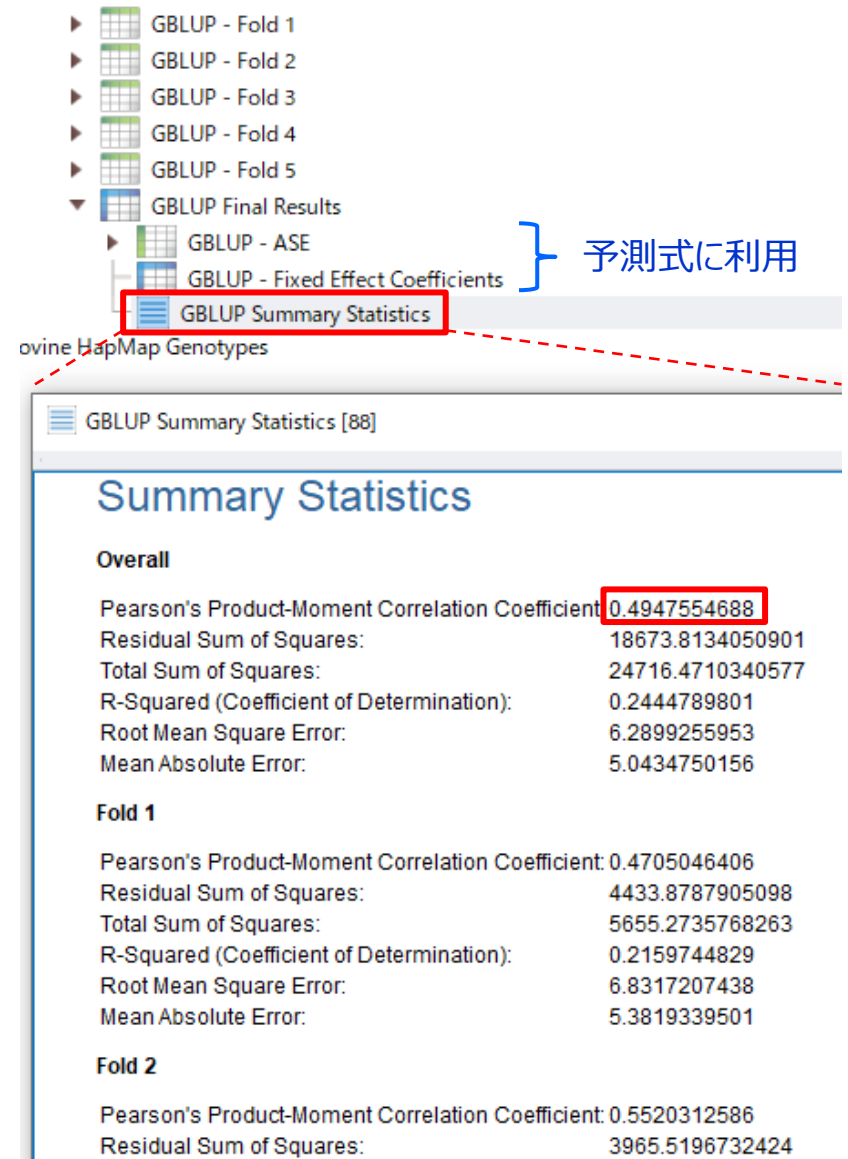
予測精度などの検証が必要



- データセットをK個に分割し、そのうち1個をテストセット、残りのK-1個をトレーニングセットとし、GBLUP、Bayes C/C-piをK回繰り返し検証を行う
- 計算に使用するサンプルデータは、表現型データとSNPデータの両方が必要
- 検証結果として、実際の表現型データと、作成した予測式によって予測された表現型データを比較した相関係数などをまとめたレポートが出力される
- 表現型予測式も同時に出力



- Foldごとに計算結果データ (gEBV, ASEなど) が出力される
- 最終出力データには、検証結果をまとめたサマリーレポート、さらに予測式として用いるASEと固定効果が含まれる
- サマリーレポートには、表現型の実データと、表現型の予測データを比較した相関係数が記載され、構築したモデルが、どの程度の精度で表現型を予測できるか確認が可能



- SNPデータをもとに、サンプルの表現型の予測を行うためのツール
- パラメータとして、トレーニング集団などより構築済みのモデル予測式（ASEと固定効果）が必要
- サンプルごとの表現型に加え、変量効果（ゲノム育種価）も出力

✓ 評価個体のSNPデータ

Unsort		G 1	G 2	G 3
Map	samples	ss715578788	ss715579193	ss715579265
1	PI605870A	G_G	C_C	C_C
2	PI594635C	G_G	C_C	A_A
3	PI605764B	G_G	C_T	C_C
4	PI567011A	A_A	C_C	C_C
5	PI171446	A_A	C_C	C_C

✓ トレーニング集団の予測式

- GBLUP - ASE
- GBLUP - Fixed Effect Coefficients

Predict Phenotypes From Existing Results

Unsort		R 1	R 2
Map	samples	Predicted Phenotypes	Random effect component
1	PI171446	0.154829218736358	0.280212873542408
2	PI240664	0.00476527551844433	0.130148930324495
3	PI307900	-0.0249905737481868	0.100393081057864
4	PI374157	0.183042488817044	0.308426143623094
5	PI567011A	0.0398180889204598	0.16520174372651
6	PI567773	-0.0400017959091541	0.0853818588968963
7	PI578324A	-0.316791518730737	-0.191407863924686
8	PI587678	-0.237937668548336	-0.112554013742286

- ✓ 評価個体の表現型とゲノム育種価

お問い合わせ先：フィルジエン株式会社
TEL: 052-624-4388 (9:00～18 : 00)
FAX: 052-624-4389
E-mail: biosupport@filgen.jp