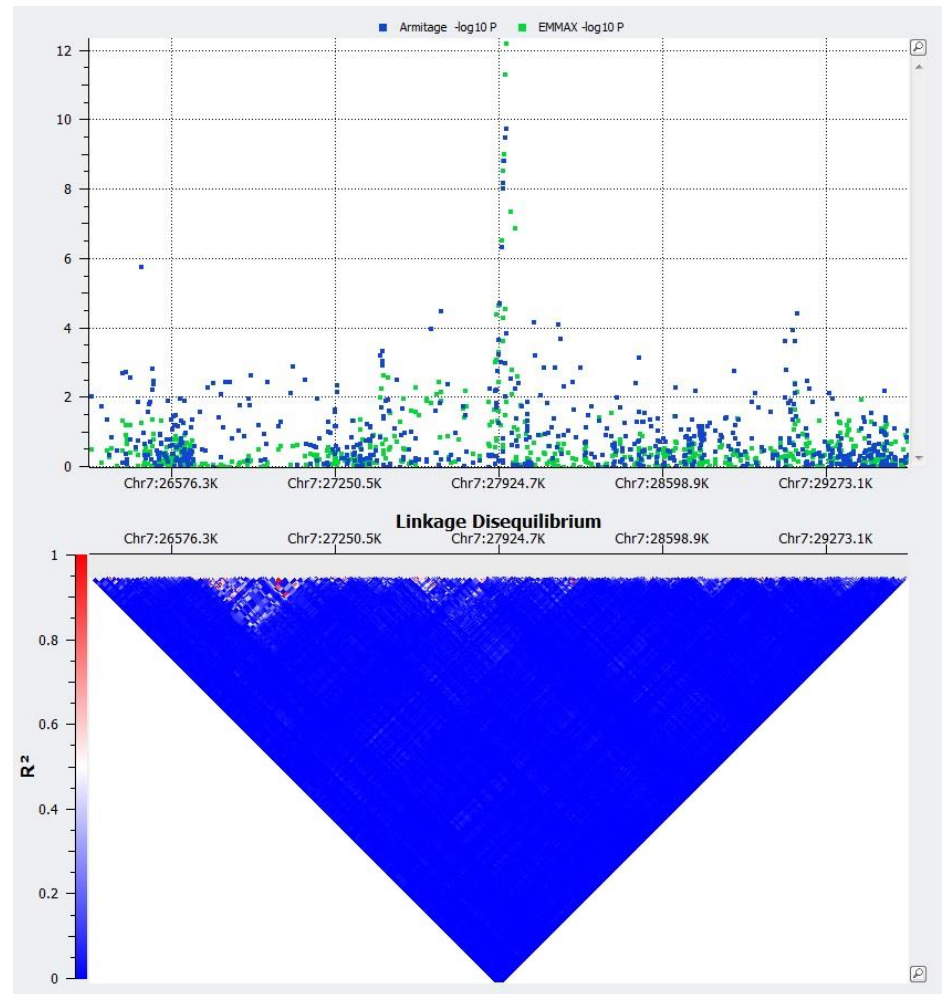


# SVSを用いた大規模ゲノムデータの 遺伝統計解析

フィルジエン株式会社 バイオサイエンス部  
(biosupport@filgen.jp)

- マイクロアレイや次世代シーケンサーの普及により、多数のサンプルについて膨大な量のゲノムデータを得られるようになったが、この大量のデータの中から有用な情報を抽出するためには、統計学を駆使したデータ解析手法を利用する必要がある
- Golden Helix社SNP & Variation Suite (SVS)では、ゲノム全域にわたってタイピングを行ったSNPデータを用いて、アレルの疾患発症リスクなどの検定を行うゲノムワイド関連解析 (GWAS) をはじめ、コピー数多型 (CNV) や稀少変異 (レアバリエント) のデータ解析用アルゴリズムが搭載されており、表現型との関連の評価など、様々な研究に応用されている



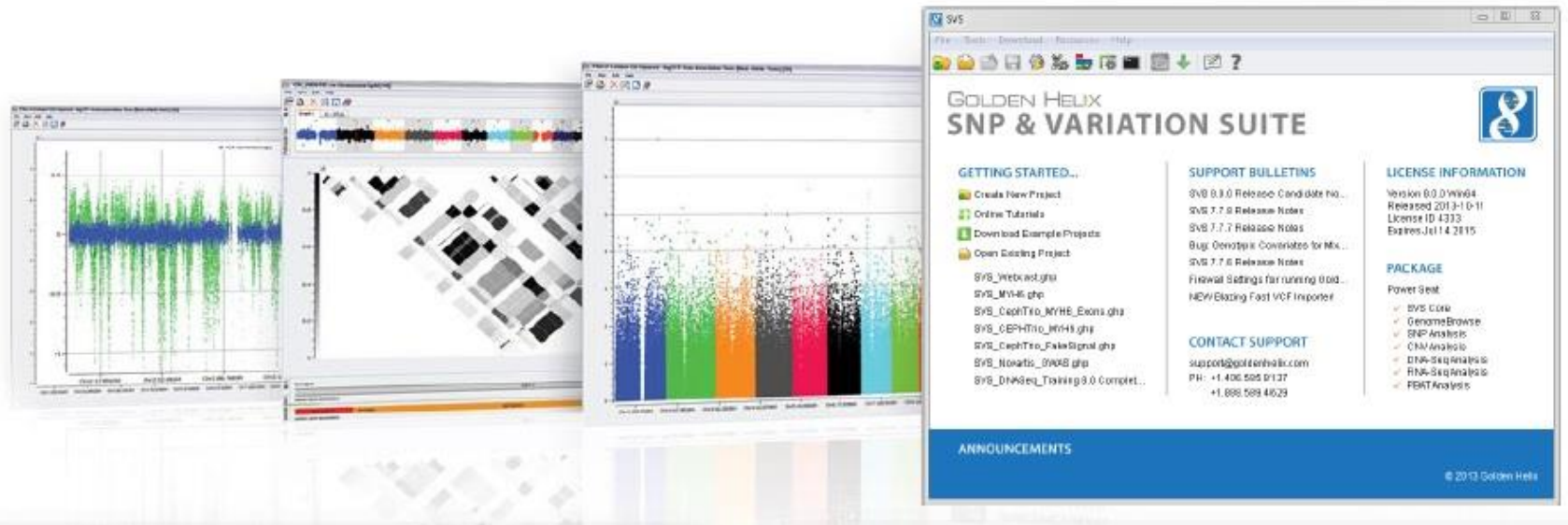


- GWAS & SNP Analysis
- Large-N DNA-Seq Analysis
- Genomic Prediction
- Copy Number Analysis
- RNA-Seq Analysis



- Cancer Diagnostics
- ACMG Guidelines Workflow Support
- CNV Calling
- Clinical Reporting
- High-throughput NGS Testing

- Golden Helix社では、遺伝統計解析ソフトウェア「SNP & Variation Suite」と、遺伝子診断支援ソフトウェア「VarSeq®」の2種類のソフトウェアパッケージを販売
- 医学・生物学研究や、家畜や作物の品種改良などの農学研究、さらに疾患の診断や最適な治療オプションの決定における医療分野、遺伝学的解析などの教育現場などで利用される

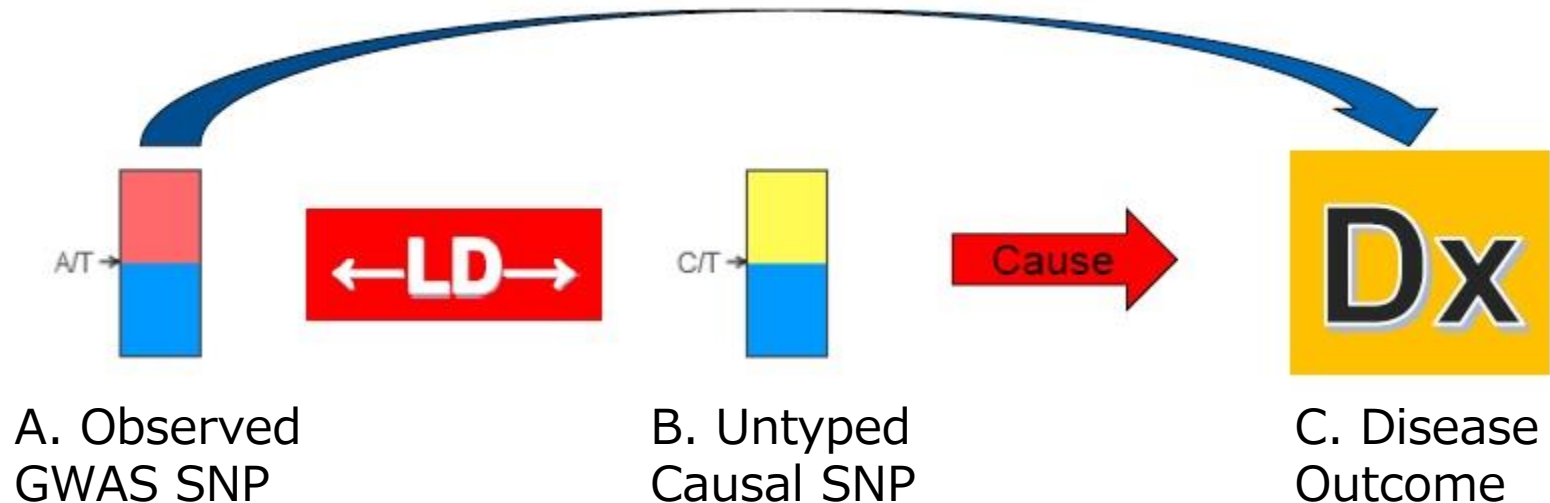


## Core Features

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

## Applications

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



- ゲノムワイド関連解析（Genome Wide Association Study: GWAS）では、サンプルの表現型（疾患の有無など）の原因となるマーカー（SNPなど）を見つけることを目的とする
- SNPマイクロアレイに搭載されていないSNPが疾患の原因となっているような場合は、そのSNPと連鎖不平衡（Linkage Disequilibrium: LD）の状態にあるSNPにより、間接的な関連を調べることもできる
- 一般的に多数サンプルのゲノムデータを解析に用いるため、解析前のデータのクオリティチェックや複雑な統計学的アルゴリズム、解析結果の精査やビジュアライゼーションが重要となる

## Import

- Microarray & NGS Data
- Phenotype Data
- Genotype Imputation

## QC

- Sample QC
- Marker QC
- Population Structure

## Test

- Association Test
- Test Correction Techniques

## Review

- Visualization
- Lambda / LD Score Regression
- Meta-Analysis

SVS - SVS8 Simulated GWAS

File Tools Import Download Resources Window Help

Project Navigator Window

Navigator Window Nodes

	ID	(Linked To)
SVS7_Project1 Project	1	
Study Genotypes	5	
Study Genotypes - Sheet 1	7	110
Statistics by Sample	45	
Call Rate - Histogram	47	
Het Rate from All Columns (Autosomes) - Histogram	49	
Study Genotypes + Study Phenotypes - Sheet 1	110	7, 42
Association Tests (Additive Model)	116	
Plot of Column Armitage -log10 P from Association Tests (Additive Model)	117	122
Plots from Association Tests (Additive Model) against Armitage expected -log10 P	119	
P-Values from Single-Locus Mixed Model	122	117
Plots from P-Values from Single-Locus Mixed Model against -log10(Expected P)	125	
Study Genotypes - Sheet 2	50	
Filtering Results	53	
Study Genotypes - Sheet 3	54	
Reference Panel	12	
All Populations	31	
All Populations - Sheet 1	33	80, 83, 105
All Populations + Principal Components (Additive Model) - Sheet 1	80	33, 57
All Populations + Principal Components (Additive Model) - Sheet 1	83	33, 72
Plots from All Populations + Principal Components (Additive Model) - Sheet 1 against EV = 28.7354	85	
Study Phenotypes	40	
Appended Spreadsheet	63	
Appended Spreadsheet - Sheet 1	65	
Filtering Results	68	
All Data	69	
Principal Components (Additive Model)	72	83
Plots from Principal Components (Additive Model) against EV = 28.7354	77	
PC Eigenvalues (Additive Model)	75	
IBS Distance ((IBS2 + 0.5*IBS1) / # non-missing markers)	88	
IBD Estimate: Estimated PI	100	105
Heat Map of IBD Estimate: Estimated PI	102	
IBD Estimate: Estimated PI + All Populations - Sheet 1	105	33, 100
Heat Map of IBD Estimate: Estimated PI + All Populations - Sheet 1	107	

Import

- Microarray Data
  - Affymetrix (e.g. CHP, CEL)
  - Illumina (e.g. Final Report, Matrix Text)
  - Agilent
  - NimbleGen

QC

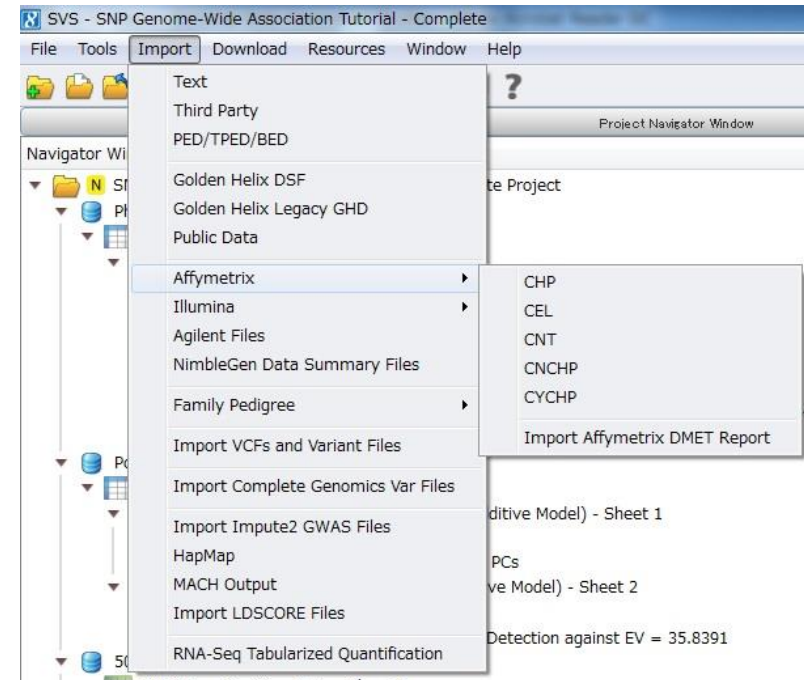
- NGS Data
  - VCF

Test

- Phenotype Data
  - Text
  - Third Party (e.g. Excel)

Review

- Others
  - PED/TPED/BED
  - Family Pedigree



Import

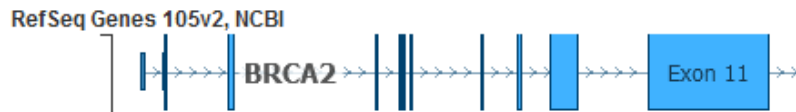
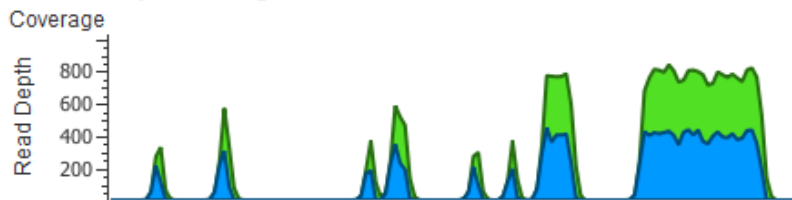
- エクソーム、遺伝子パネルなどのターゲットシーケンスで取得したBAMファイルと、ターゲット領域のBEDファイルなどを使用

QC

- ターゲット領域ごとのカバレッジを計算し、リファレンスサンプルとの比較による正規化を行って、コピー数データを取得する

\* 別途有償アドオンが必要

Test



Review

CVN on NGS Target Regions

Input and Outputs Parameters Advanced

Inputs

Select track for target regions

BAM Path Mappings

Associate from path using sample names:

No files associated

Use field in sample spreadsheet that specified BAM paths:

No field selected

SAMPLE1\_filtered - Variant Allele Freq(VAF)

Outputs

Sample summary table

Target Region CNV State

Categorical CNV state  Numeric CN (0, 1, 2, 3)

Filter out calls with QC flags:

High Controls Variation  Within Region IQR

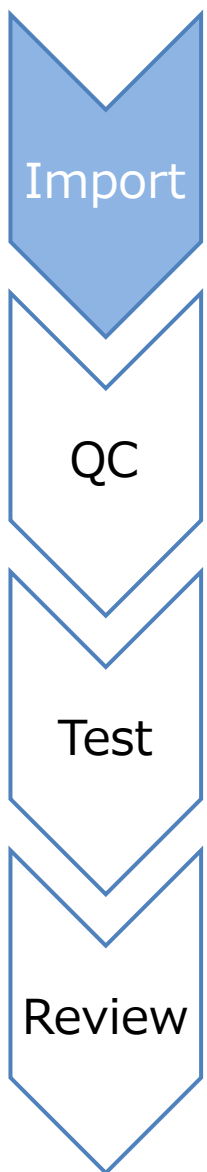
Low Controls Depth  Low Z Score

Target region Z-score table

Target region ratio table

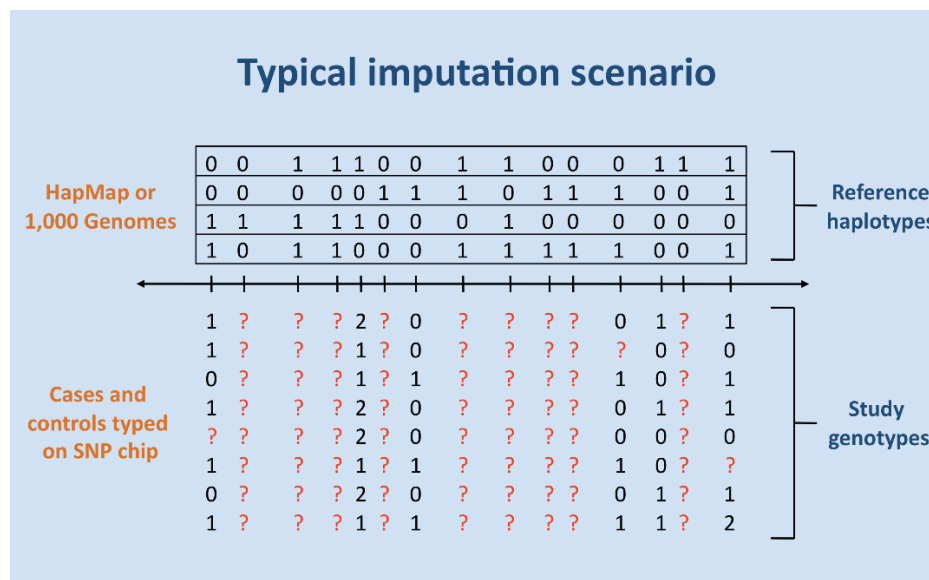
CNV event table for each sample, for:  samples





- 1000 Genome Projectなどのリファレンスパネルに含まれるハプロタイプ情報をもとに、サンプルデータには含まれていないゲノム領域上の遺伝子型の推定を行う
- 異なるプラットフォームのSNPマイクロアレイを用いている場合に、それぞれのアレイに含まれていないSNPの情報を補完することが可能
- ユーザー指定のVCFファイルをリファレンスパネルに使用でき、ヒト以外の生物種にも対応可能

\* サーバーライセンス限定で使用可能

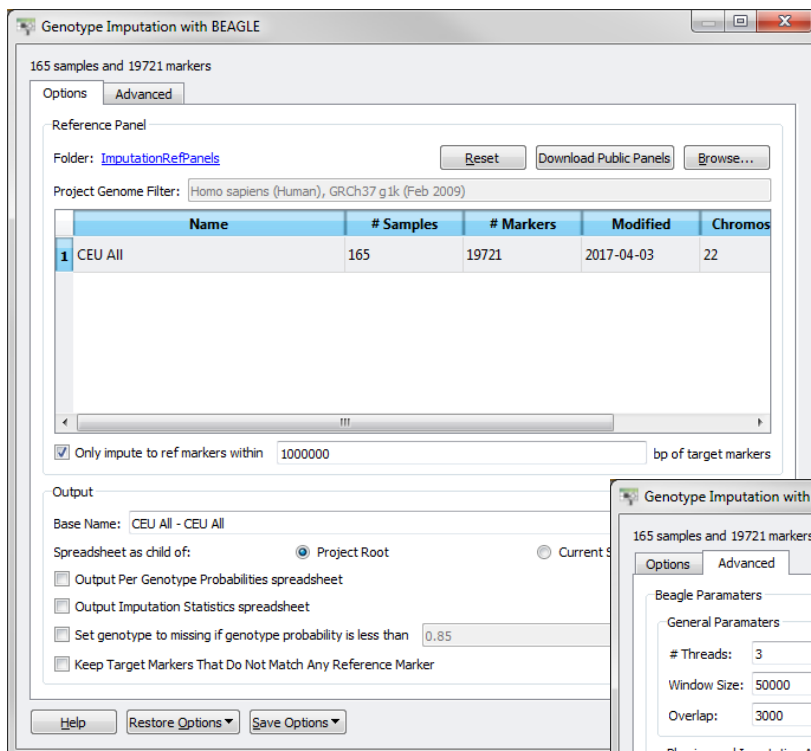


Import

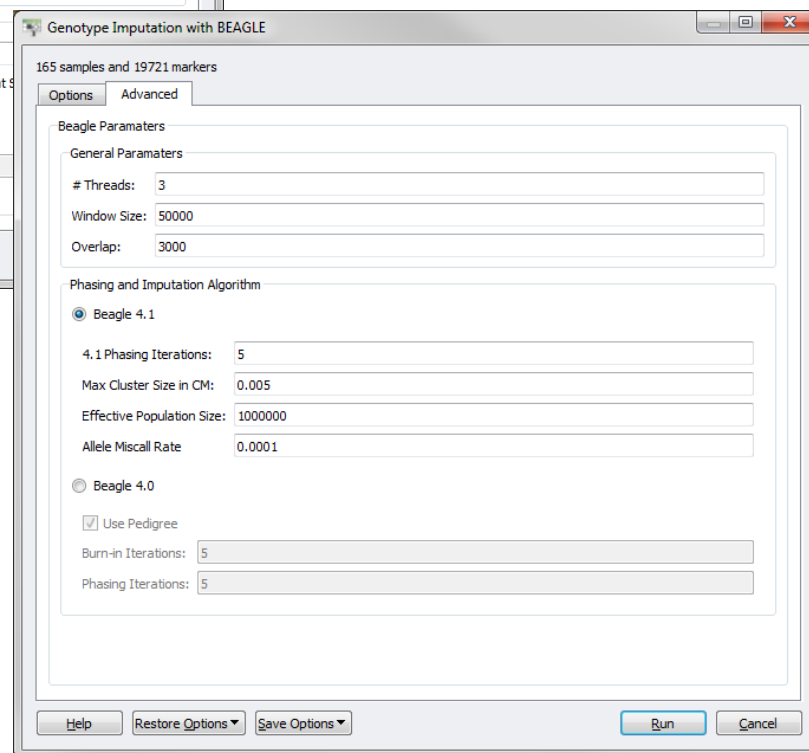
QC

Test

Review



- 標準搭載のダウンロード機能で1000 Genome ProjectのVCFファイルを取得するか、カスタムメイドのVCFファイルを所定のフォルダに保存
- Beagle 4.1または4.0の計算パラメータを設定し、計算を実行



# Import Result

Import

QC

Test

Review

Unsort		C 1	I 2	R 3	C 4	G 5	G 6	G 7	G 8	G 9	G 10
Map	sub	Phenotype 1	SBP	BMI	Gender	SNP_A-1909444	SNP_A-4303947	SNP_A-1886933	SNP_A-2116190	SNP_A-4291020	SNP_A-1902458
	Chromosome					1	1	1	1	1	1
	Position					752566	779322	785989	1003629	1097335	1130727
	dbSNP RS ID					rs3094315	rs4040617	rs2980300	rs4075116	rs9442385	rs10907175
	Associated Gene					?	?	?	?	?	TLL10
	Cytoband					p36.33	p36.33	p36.33	p36.33	p36.33	p36.33
	Reference Alleles A/B					[C/T]	[A/G]	[A/G]	[A/G]	[G/T]	[A/C]
	Top Alleles					[G/A]	[A/G]	[T/C]	[T/C]	[G/T]	[A/C]
	Bottom Alleles					[C/T]	[T/C]	[A/G]	[A/G]	[C/A]	[T/G]
	Strand					-	+	-	-	+	+
	Strand Versus dbSNP					same	same	reverse	same	same	same
1	GSM233256_GSM233257	Case	128	29.9	M	T_T	A_A	G_G	A_A	G_G	A_A
2	GSM233258_GSM233259	Control	137	32.8	M	C_T	A_A	G_G	A_A	G_G	A_A
3	GSM233262_GSM233263	Control	117	27.4	M	C_T	A_G	A_G	A_G	G_T	A_A
4	GSM233264_GSM233265	Control	140	29.7	M	T_T	A_A	G_G	A_A	G_G	A_C
5	GSM233266_GSM233267	Case	113	27.7	M	T_T	A_A	G_G	G_G	G_G	A_A
6	GSM233270_GSM233271	Case	125	29.2	M	T_T	A_A	G_G	A_A	G_T	A_A
7	GSM233276_GSM233277	Control	88	20	M	T_T	A_A	G_G	A_A	G_T	A_A
8	GSM233278_GSM233279	Case	131	31.2	M	T_T	A_A	G_G	A_A	G_G	A_A
9	GSM233280_GSM233281	Case	124	26.2	M	T_T	A_A	A_G	A_A	G_T	A_A
10	GSM233284_GSM233285	Control	111	22.8	M	C_T	A_G	A_G	A_G	G_G	A_A
11	GSM233286_GSM233287	Control	114	26.1	F	T_T	A_A	G_G	A_A	G_G	A_A
12	GSM233288_GSM233289	Control	107	23.5	M	C_T	A_G	A_G	A_G	G_G	A_C
13	GSM233290_GSM233291	Case	146	37.4	M	C_C	?_?	A_A	A_G	G_G	A_C
14	GSM233292_GSM233293	Case	122	27.6	M	T_T	A_A	G_G	A_G	G_G	A_A
15	GSM233298_GSM233299	Case	130	24.4	F	T_T	A_A	G_G	A_A	G_T	A_A
16	GSM233300_GSM233301	Control	116	24.4	M	T_T	A_A	G_G	A_A	G_G	A_A
17	GSM233302_GSM233303	Control	125	30.5	M	T_T	A_A	G_G	A_A	?_?	A_A
18	GSM233304_GSM233305	Case	112	22.2	M	T_T	A_A	G_G	A_G	G_G	A_C
19	GSM233306_GSM233307	Case	130	29.7	M	T_T	A_A	G_G	A_G	G_G	A_A
20	GSM233308_GSM233309	Control	136	30.7	F	C_T	A_G	A_G	A_A	G_G	A_A
21	GSM233310_GSM233311	Case	104	28	M	C_C	?_?	A_A	A_G	G_G	A_A
22	GSM233314_GSM233315	Case	115	26.1	M	T_T	A_A	G_G	A_G	G_G	A_A
23	GSM233316_GSM233317	Case	143	33.6	M	T_T	A_A	G_G	A_G	G_G	A_A
24	GSM233320_GSM233321	Case	103	27.7	M	T_T	A_A	?_?	A_A	G_G	A_C
25	GSM233322_GSM233323	Control	144	33.1	F	T_T	A_A	G_G	A_A	G_T	A_A
26	GSM233324_GSM233325	Case	131	32.3	M	T_T	A_A	G_G	A_A	G_G	A_A

- 全サンプルとマーカー（SNP, CNVなど）をまとめたスプレッドシートが作成される
- マイクロアレイのMarker Mapデータから、アノテーション付けも可能
- シートに表現型データを追加することも可能

Import

QC

Test

Review

- Sample QC
  - Call Rate / Het Rate
  - Gender Checks
  - IBD Testing
  - Principle Component Analysis
  - Mendelian Error

Genotype Statistics by Sample

(No variable is set as dependent.)

Genotype Count Statistics

NOTE: Call rate and heterozygosity are always output.

Number and fraction of genotypes with a minor allele (as determined from sample data)

Variant Statistics (Marker Map "Reference" Field Required)

Number of variant genotypes (non reference)

Number of singletons (variant genotype present only in given sample)

Mean  $T_i/T_v$  of variant genotypes

Autosomal Statistics

Hardy-Weinberg Thw P-Value (taken over all autosomal chromosomes and all samples)

Gender Chromosome Statistics

Gender inference:

Select chromosome to use for gender inference: Choose...

Threshold of heterozygosity for calling M/F 0.02

Additional Outputs (Verbose Output)

Output count and variant statistics for each autosomal chromosome

Help Restore Options Save Options Run Cancel

Import

QC

Test

Review

- Marker QC /Filtering
  - Call Rate / HWE
  - Minor Allele Frequency
  - LD Pruning
  - Genomic Annotations

Genotype Filtering by Marker

Case/Control dependent variable: Case? (47 cases and 650 controls)

Classify alleles by allele frequency       Classify alleles by reference/alternate  
(Reference field in map: "Reference")

Filter Genotype Columns

General Statistics Filtering

Drop if call rate      < 0.85

Drop if number of alleles      > 2

Drop if Minor Allele Frequency (MAF)      < 0.05

Drop if carrier count      < 10

Hardy Weinberg Equilibrium (HWE) Filtering

Perform HWE filtering based on:      All

Drop if Hardy Weinberg Equilibrium (HWE) P-Value      < 0.001

Drop if Fisher's exact test for HWE P-Value      < 0.001

Drop if signed HWE R (positive if more homozygous)      > 0.2

Actions

Inactivate genotype columns that meet above criteria for filtering

Output spreadsheet with marker statistics and "Drop?" columns

Additional Output

Output -log<sub>10</sub>(Value)

Help    Restore Options    Save Options    Run    Cancel

Import

QC

Test

Review

Genotype Principal Component Analysis

Classify alleles by allele frequency  Classify alleles by reference/alternate (Reference field in map: "Reference Alleles A/B")

Principal Components Source

Compute the principal components  
 Use precomputed principal components

PC sheet:

Principal Components

Find up to top 5 components

Output

Output corrected input data  
 Output principal components to spreadsheet  
 Output separate spreadsheet of eigenvalues

PCA Marker Normalization

Normalize each marker's data by its theoretical standard deviation under HWE  
 Normalize each marker's data by its actual standard deviation  
 Do not normalize marker data

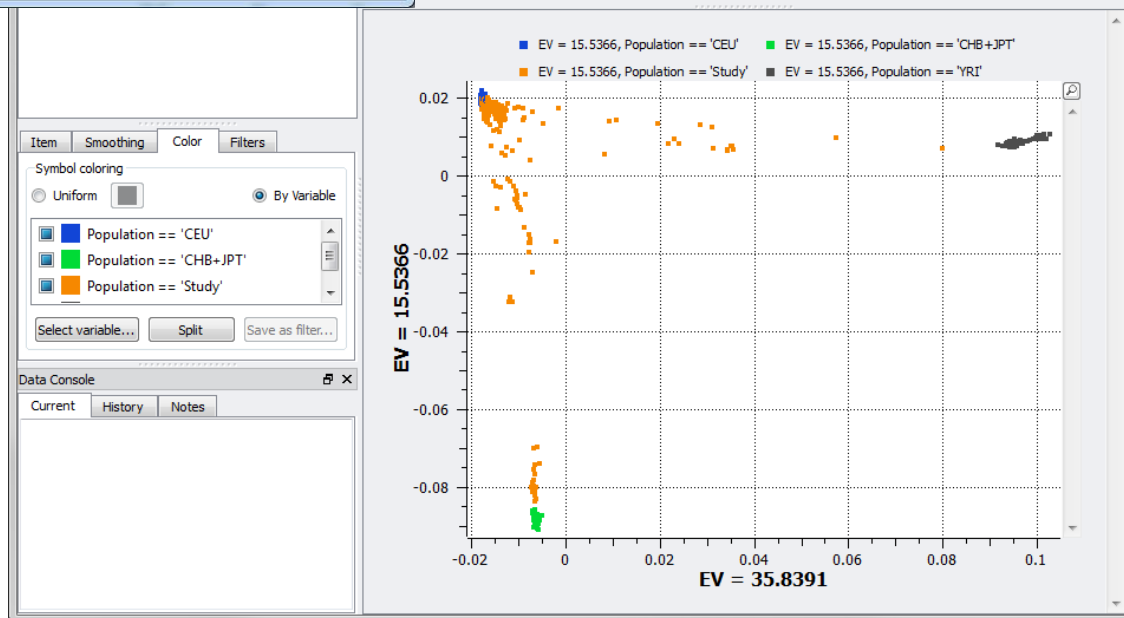
Genetic Model for PCA

Additive model: (dd) -> (Dd) -> (DD)  
 Dominant model: (DD, Dd) vs. (dd)  
 Recessive model: (DD) vs. (Dd, dd)

PCA Outlier Removal

Recompute components up to 5 times removing outliers (more than 6 standard deviations) from 5 components

- 遺伝子型データあるいはCNVなどの数値データを基に、主成分分析（PCA）を実行し、集団の階層化を確認可能



Import

## Genotype Association Tests

- SNPなどの遺伝子型データを利用した関連解析手法

QC

## Mixed Linear Model Analysis

- 遺伝子型データから計算したサンプル間相関データを用いて、サンプルの血縁関係に基づくバイアスを除外し、関連解析を行う手法

Test

## Haplotype Association Tests

- ハプロタイプデータを利用した関連解析手法

## Collapsing Methods

- 次世代シーケンサーなどで検出したレアバリエントを単一のユニット（遺伝子など）に統合して関連解析を行う手法

Review

## Numeric Association Tests

- 数値データを利用した関連解析手法

Import

QC

Test

Review

Filtered Data for Association Testing - Active Subset [97]

Unsort	C	1	B	2	R	3	I	4	R	5	G	6	G	7	G	8	G	9
Map	sub	Phenotype 1	Phenotype 1 - Binary	Phenotype 2	SBP	BMI	SNP_A-1909444	SNP_A-4303947	SNP_A-1886933	SNP_A-2116190								
1	GSM233256_GSM233257	Case	1	8.96	128	29.9	T_T	A_A	G_G	A_A								
2	GSM233258_GSM233259	Control	0	12.18	137	32.8	C_T	A_A	G_G	A_A								
3	GSM233262_GSM233263	Control	0	11.01	117	27.4	C_T	A_G	A_G	A_G								
4	GSM233264_GSM233265	Control	0	9.25	140	29.7	T_T	A_A	G_G	A_A								
5	GSM233266_GSM233267	Case	1	6.85	113	27.7	T_T	A_A	G_G	G_G								
6	GSM233270_GSM233271	Case	1	11.65	125	29.2	T_T	A_A	G_G	A_A								
7	GSM233276_GSM233277	Control	0	10.82	88	20	T_T	A_A	G_G	A_A								
8	GSM233278_GSM233279	Case	1	9.62	131	31.2	T_T	A_A	G_G	A_A								
9	GSM233280_GSM233281	Case	1	9.02	124	26.2	T_T	A_A	A_G	A_A								
10	GSM233284_GSM233285	Control	0	11.73	111	22.8	C_T	A_G	A_G	A_G								
11	GSM233286_GSM233287	Control	0	12.81	114	26.1	T_T	A_A	G_G	A_A								
12	GSM233288_GSM233289	Control	0	9.46	107	23.5	C_T	A_G	A_G	A_G								
13	GSM233290_GSM233291	Case	1	8.55	146	37.4	C_C	?_?	A_A	A_G								
14	GSM233292_GSM233293	Case	1	13.28	122	27.6	T_T	A_A	G_G	A_G								
15	GSM233298_GSM233299	Case	1	13.04	130	24.4	T_T	A_A	G_G	A_A								
16	GSM233300_GSM233301	Control	0	5.08	116	24.4	T_T	A_A	G_G	A_A								
17	GSM233302_GSM233303	Control	0	10.32	125	30.5	T_T	A_A	G_G	A_A								
18	GSM233304_GSM233305	Case	1	11.47	112	22.2	T_T	A_A	G_G	A_G								
19	GSM233306_GSM233307	Case	1	9.32	130	29.7	T_T	A_A	G_G	A_G								
20	GSM233308_GSM233309	Control	0	10.25	136	30.7	C_T	A_G	A_G	A_A								
21	GSM233310_GSM233311	Case	1	12.9	104	28	C_C	?_?	A_A	A_G								
22	GSM233314_GSM233315	Case	1	7.44	115	26.1	T_T	A_A	G_G	A_G								
23	GSM233316_GSM233317	Case	1	12.45	143	33.6	T_T	A_A	G_G	A_G								

- 解析実行時には、任意の表現型データを基準に指定できる
- Case/Controlのようなバイナリデータの他、身長・体重のような量的データも指定が可能



Case/Control dependent variable: Phenotype 1 - Binary (238 cases and 230 controls) Additive model: (dd) -> (Dd) -> (DD)

Classify alleles by allele frequency  Classify alleles by reference/alternate (Reference field in map: "Reference Alleles A/B")

Association Test Parameters PCA Parameters Overall Marker Statistics

Genetic Model or Tests Test Statistic or Method

Where D = minor allele, d = major allele

Basic allelic tests: D vs. d

Genotypic tests: (DD) vs. (dd) vs. (Dd)

Additive model: (dd) -> (Dd) -> (DD)

Dominant model: (DD, Dd) vs. (dd)

Recessive model: (DD) vs. (Dd, dd)

Case/Control additive model:

Correlation/Trend test

Cochran-Armitage test

Exact form of Cochran-Armitage test

Odds ratios: (Dd) vs. (dd) and (DD) vs. (Dd)

Logistic regression

Missing Values Multiple Testing Correction

Use missing values as predictors

Drop missing values

Bonferroni adjustment (on N SNPs)

False Discovery Rate (FDR)

Single value permutations

Full scan permutations

Additional Outputs

Output data for P-P/Q-Q plots

Output -log<sub>10</sub>(P)

Number of permutations:

Principal Components Analysis (PCA) Genomic Control of Output Data for Stratification

Correct for stratification with PCA

Show inflation factor (lambda), chi-squares, and corrected values

Correct using this inflation factor (lambda) instead:

Help Restore Options Save Options Run Cancel

Import

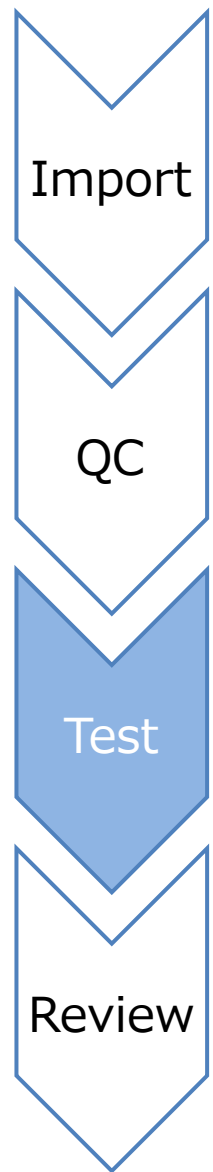
QC

Test

Review

- Genomic Model
  - Basic Allelic Tests
  - Genotypic Tests
  - Additive Model
  - Dominant Model
  - Recessive Model

- Test Statistics
  - Correlation/Trend Test
  - Armitage Trend Test
  - Exact Form of Armitage Test
  - Chi-Squared Test
  - Fisher's Exact Test
  - Odds Ratio
  - Analysis of Deviance
  - F-test
  - Logistic Regression
  - Linear Regression



Full Data

	DD	Dd	dd
Case	a	b	c
Control	d	e	f

(D:マイナーアレル、d: メジャーアレル、a~f: 遺伝型数)  
頻度の低いアレル (マイナーアレル) 数の、疾患への影響に合わせて、遺伝学モデルを選択する

Additive Model  
(dd -> Dd -> DD)

	DD	Dd	dd
Case	a	b	c
Control	d	e	f

Dominant Model  
((DD + Dd) vs (dd))

	DD + Dd	dd
Case	a + b	c
Control	d + e	f

Recessive Model  
((DD) vs (Dd + dd))

	DD	Dd + dd
Case	a	b + c
Control	d	e + f

Import

Case/Control, Additive model

Genetic Model or Tests

Where D = minor allele, d = major allele

- Basic allelic tests: D vs. d
- Genotypic tests: (DD) vs. (dd) vs. (Dd)
- Additive model: (dd) -> (Dd) -> (DD)
- Dominant model: (DD, Dd) vs. (dd)
- Recessive model: (DD) vs. (Dd, dd)

Test Statistic or Method

Case/Control additive model:

- Correlation/Trend test
- Cochran-Armitage test
- Exact form of Cochran-Armitage test
- Odds ratios: (Dd) vs. (dd) and (DD) vs. (Dd)
- Logistic regression

QC

Case/Control, Dominant model

Genetic Model or Tests

Where D = minor allele, d = major allele

- Basic allelic tests: D vs. d
- Genotypic tests: (DD) vs. (dd) vs. (Dd)
- Additive model: (dd) -> (Dd) -> (DD)
- Dominant model: (DD, Dd) vs. (dd)
- Recessive model: (DD) vs. (Dd, dd)

Test Statistic or Method

2 x 2 table:

- Correlation/Trend test
- Chi-squared test
- Chi-squared with Yates correction
- Fisher's exact test
- Odds ratio with confidence limits
- Analysis of deviance
- Logistic regression

Test

Quantitative, Additive model

Genetic Model or Tests

Where D = minor allele, d = major allele

- Basic allelic tests: D vs. d
- Genotypic tests: (DD) vs. (dd) vs. (Dd)
- Additive model: (dd) -> (Dd) -> (DD)
- Dominant model: (DD, Dd) vs. (dd)
- Recessive model: (DD) vs. (Dd, dd)

Test Statistic or Method

Quantitative additive model:

- Correlation/Trend test
- Linear regression

Review

- 指定した表現型データ、遺伝学モデルに合わせて、選択可能な統計手法のみが表示される

Import

QC

Test

Review

<b>Missing Values</b>	<b>Multiple Testing Correction</b>
<input type="radio"/> Use missing values as predictors	<input checked="" type="checkbox"/> Bonferroni adjustment (on N SNPs)
<input checked="" type="radio"/> Drop missing values	<input type="checkbox"/> False Discovery Rate (FDR)
<b>Additional Outputs</b>	<input type="checkbox"/> Single value permutations
<input checked="" type="checkbox"/> Output data for P-P/Q-Q plots	<input type="checkbox"/> Full scan permutations
<input checked="" type="checkbox"/> Output $-\log_{10}(P)$	Number of permutations: <input type="text"/>
<b>Principal Components Analysis (PCA)</b>	<b>Genomic Control of Output Data for Stratification</b>
<input type="checkbox"/> Correct for stratification with PCA	<input type="checkbox"/> Show inflation factor (lambda), chi-squares, and corrected values
	<input type="checkbox"/> Correct using this inflation factor (lambda) instead: <input type="text" value="1"/>

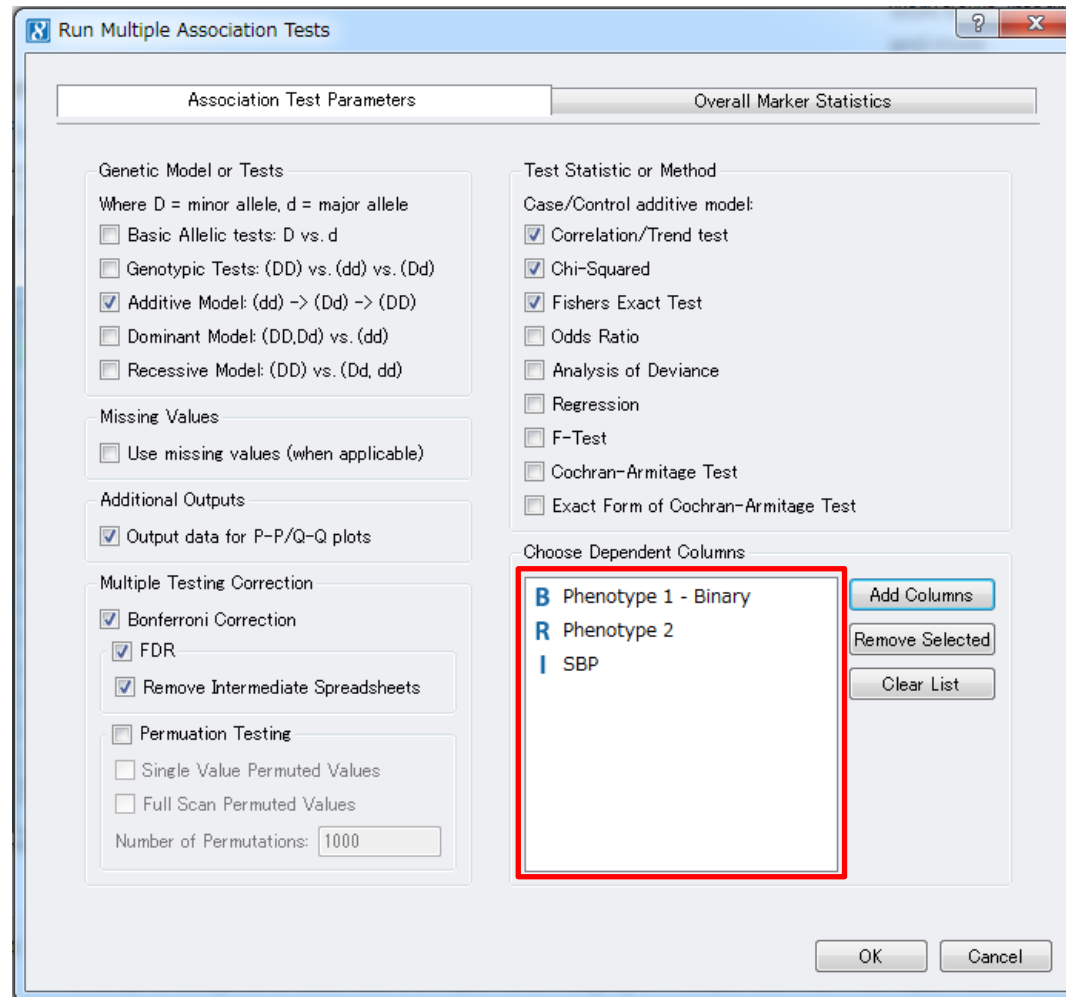
- 多重検定の実施やQ-Qプロット図の作成、さらに主成分分析（PCA）を利用した集団の階層化の補正を行うことも可能

Import

QC

Test

Review



- 無償のアドオンスクリプトを組み込むことで、複数の表現型データに対する計算をまとめて行うことが可能

Import

- おもにサンプルの血縁関係による偏りを除外し、関連解析を行う場合に用いられる手法
- SNPデータから計算したサンプル間の相関データを用いて、血縁関係の偏りを補正

QC

## Mixed Model GWAS using a single locus (EMMAX)

- 1か所のSNPごとに表現型との関連を計算

Test

## Multi-locus mixed model GWAS (MLMM)

- 複数か所のSNPをまとめて、表現型との関連を計算

Review

## Genomic Best Linear Unbiased Predictors (GBLUP)

- サンプルごとのランダム効果と、SNPごとのアレル代替効果を計算
- 農学分野における、ゲノム育種価の計算にも用いられる

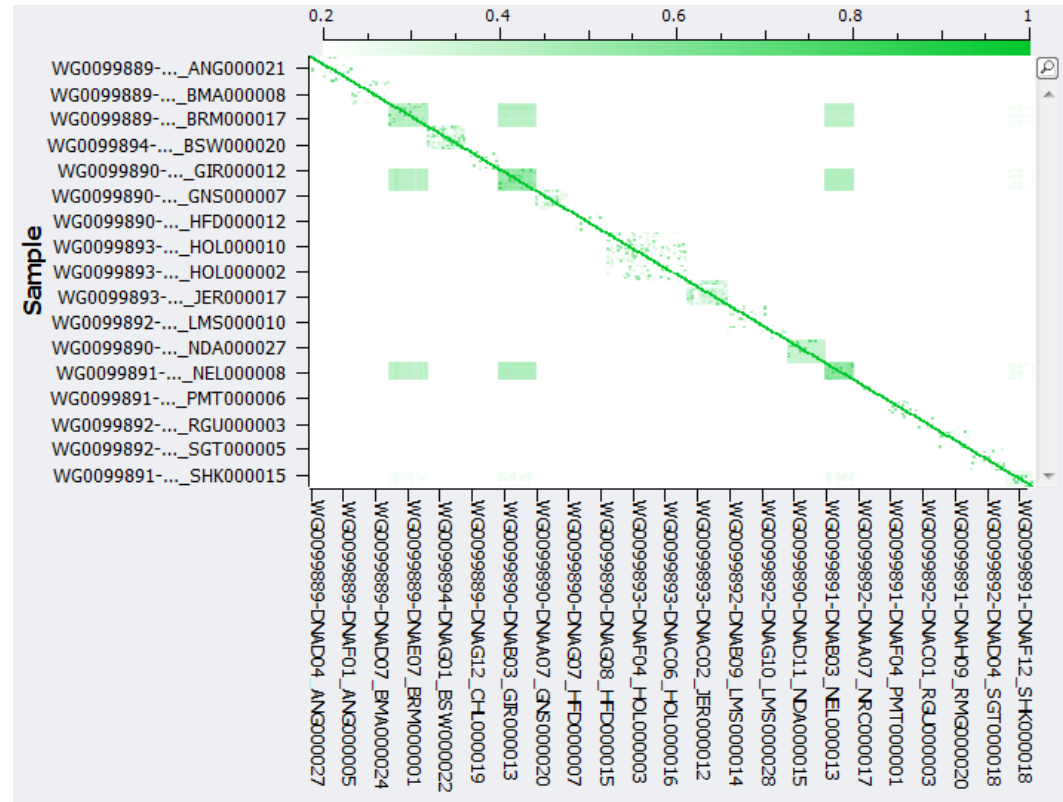
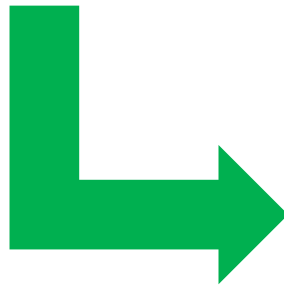
Import

QC

Test

Review

Identity by Descent Estimation
Fixation Index Fst
Fixation Index Fst (by Marker)
GBLUP Genomic Relationship Matrix
Compute Kinship A Matrix from Pedigree
Filter Samples by Call Rate
LD Pruning
SNP Density
Mendelian Error Check
Inbreeding Coefficients



- サンプル間相関データの計算には、Identical by State (IBS), Identical by Descent (IBD), GBLUP Genomic Relationship Matrixなど、様々な手法がある
- ここで計算したサンプル間相関データ、および表現型データと遺伝型データを使用して、関連解析を実行する

Import

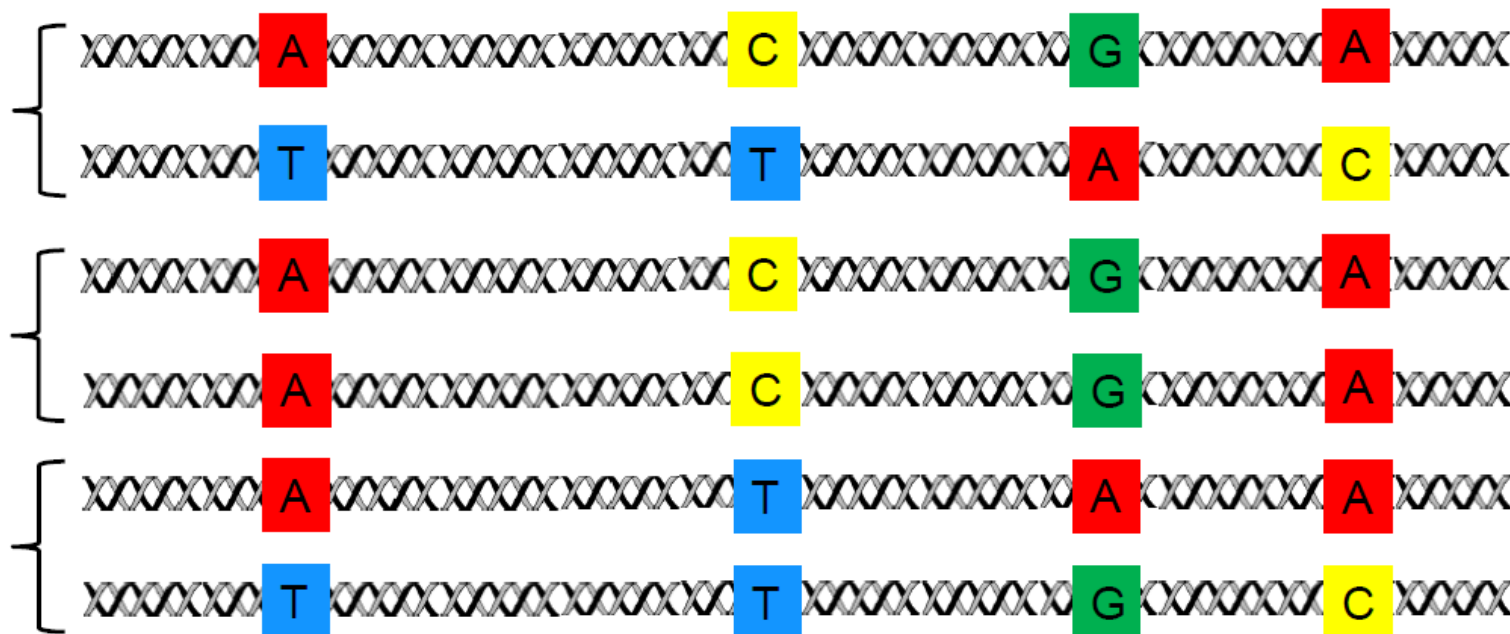
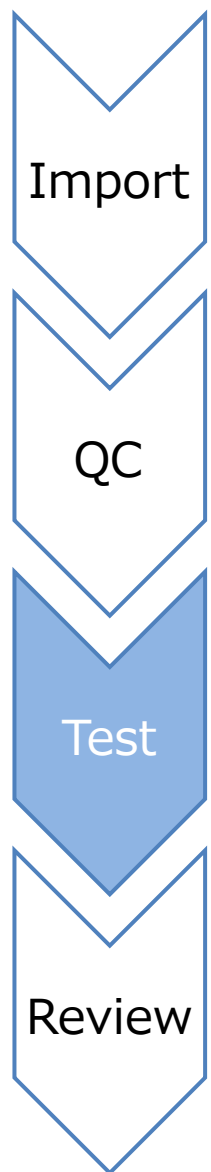
QC

Test

Review

- Mixed Linear Model Analysisのパラメータ設定時に、あらかじめ作成しておいたサンプル間相関データを選択する
- その他、解析に用いる手法や遺伝型モデルなども設定できる





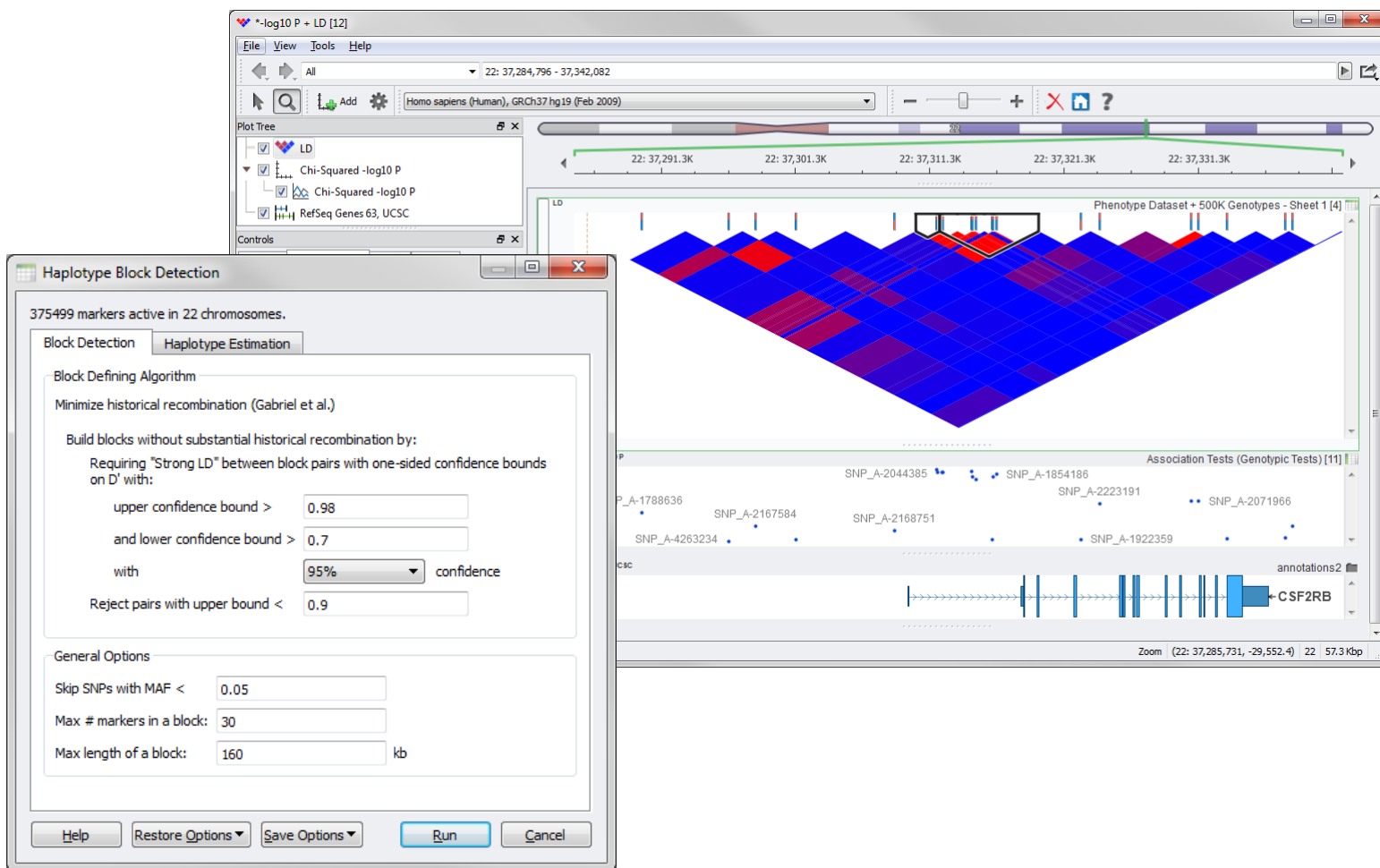
- マーカーとして、アレルの組み合わせであるハプロタイプを使用し、関連解析を行う手法
- 集団内で特定のハプロタイプが保存されているハプロタイプブロックを解析対象とする
- 1個体のSNPデータからハプロタイプを決定することはできないので、サンプル集団内の頻度からハプロタイプを推定し、解析に用いる

Import

QC

Test

Review



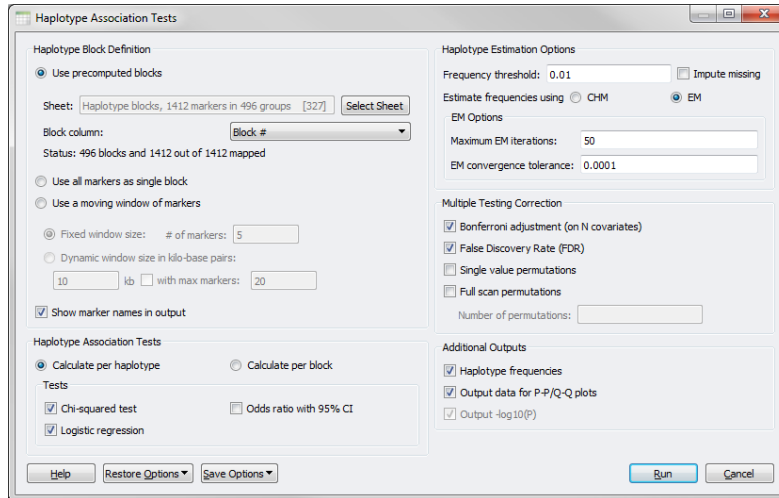
- SNP間の連鎖不平衡の度合いを表したLDプロット上から、あるいはSNPデータのスプレッドシート上から、ハプロタイプブロックを検出する

Import

QC

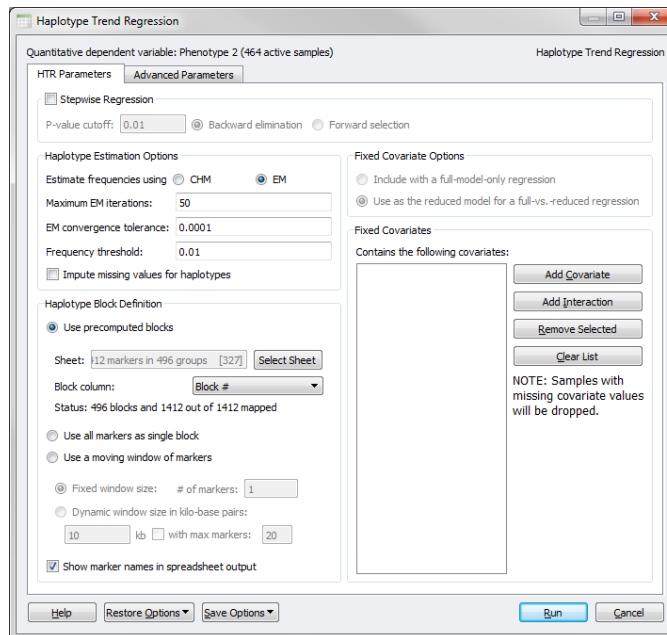
Test

Review



## Haplotype Association Tests

- 表現型データとして、Case/Controlのようなバイナリデータを使用
- Chi-Square test
- ハプロタイプごと、あるいはハプロタイプブロックごとの結果が出力される

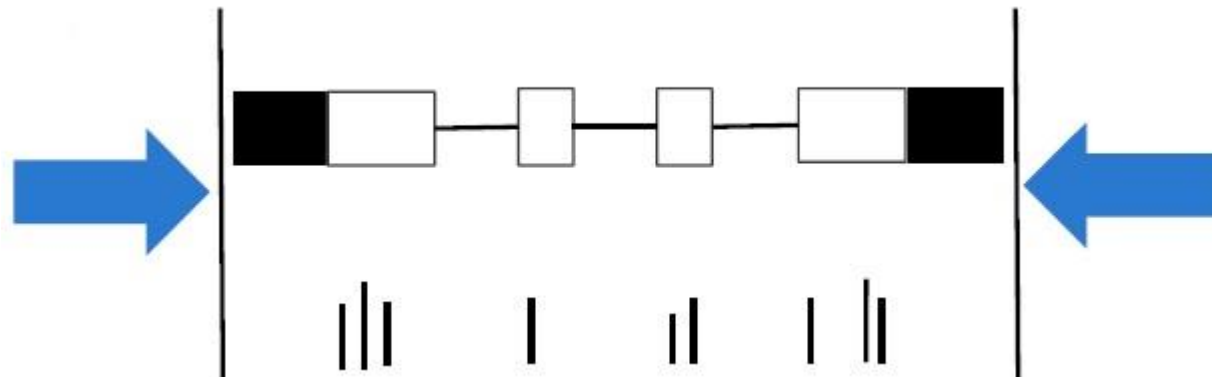


## Haplotype Trend Regression

- 表現型データとしてバイナリデータに加え、量的データも使用可能
- 共変量 (Covariate) による補正が可能
- ハプロタイプブロックごとの結果が出力される
- 個別のハプロタイプごとに、詳細情報が出力される

- マイクロアレイに基づくゲノムワイド関連解析では、基本的に遺伝型データとして、コモンバリエント(MAF  $\geq 0.05$ )を対象としてきた
- 次世代シーケンサーの利用により、レアバリエント(MAF  $< 0.01$ )も解析対象とすることが可能となった
- レアバリエントに対する関連解析は通常のゲノムワイド関連解析のアプローチでは困難

バリエントを集約し、単一ユニットとして解析



Import

## Burden Tests

- 複数のサイトのマイナーアレルを統合して検定を行う
- それぞれのアレルが疾患の原因であり、同方向の効果をもつことが前提となる
  - 重みづけなし : CMC
  - 重みづけあり : KBAC

QC

## Kernel Tests

- 複数のアレルの統計スコアの2乗を統合して検定を行う
- 各アレルの表現型への影響の方向が様々である場合に用いられる
  - SKAT
  - SKAT-O (Burden Testsとの組み合わせ)

Test

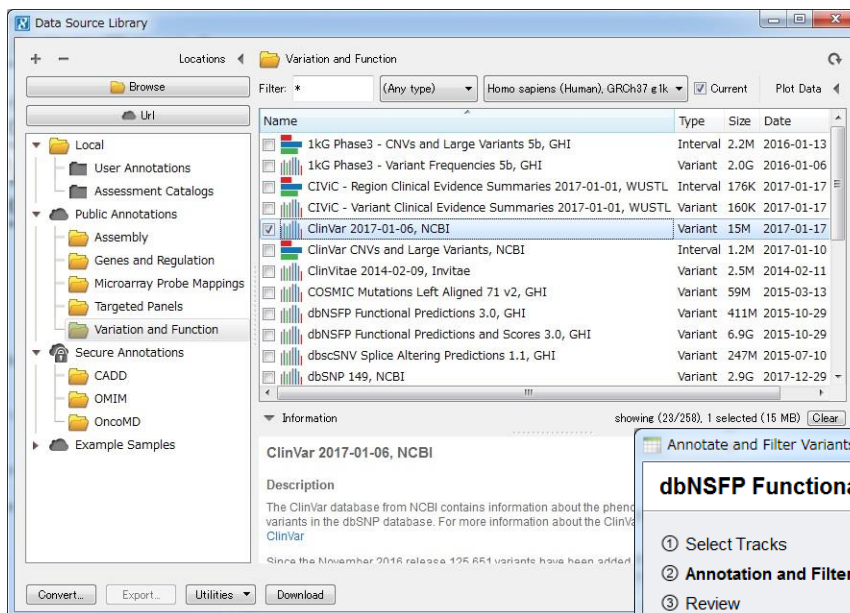
Review

Import

QC

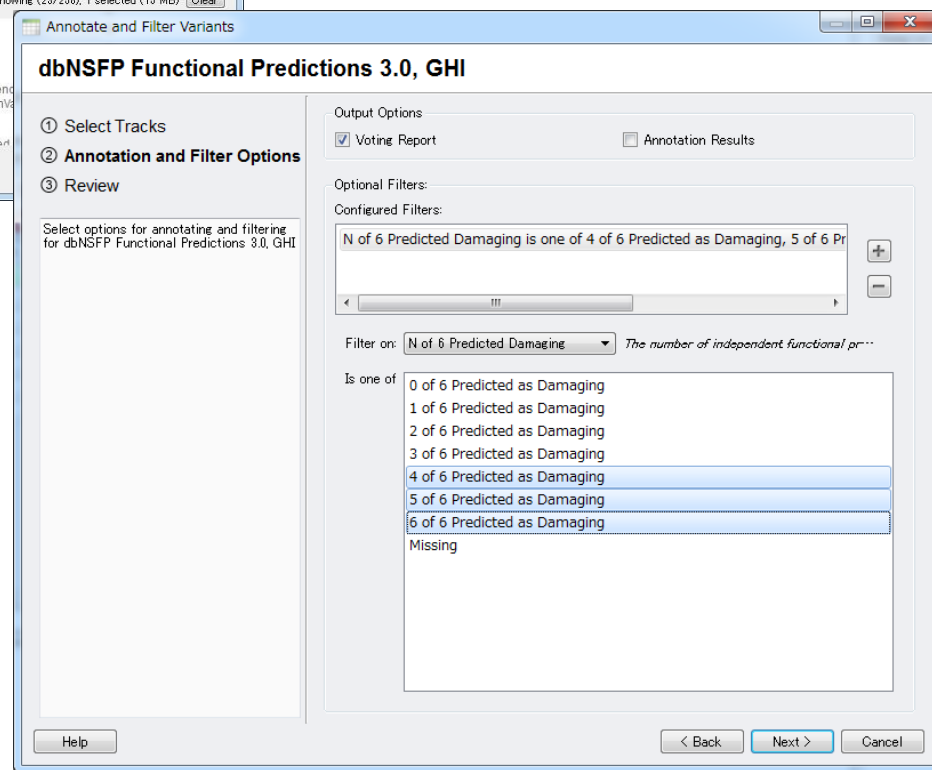
Test

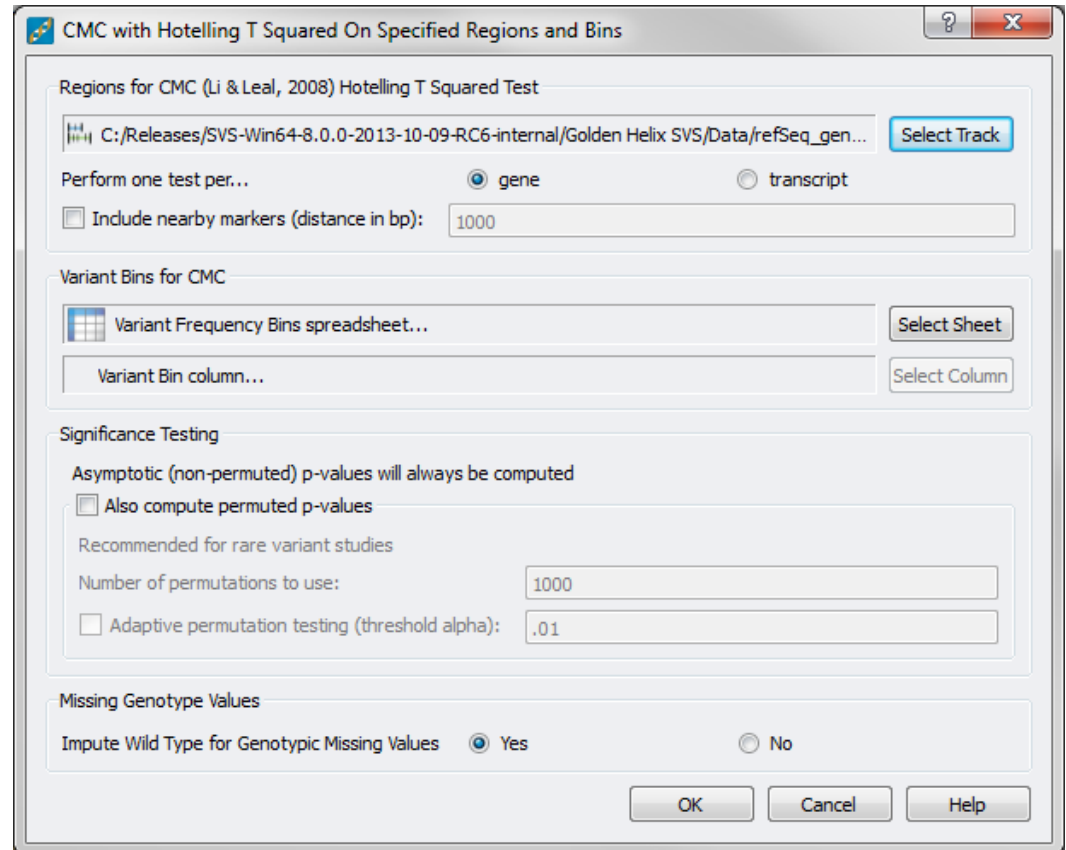
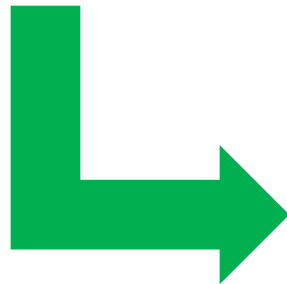
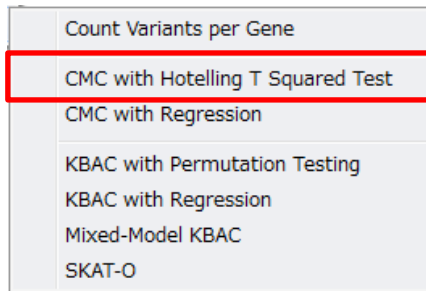
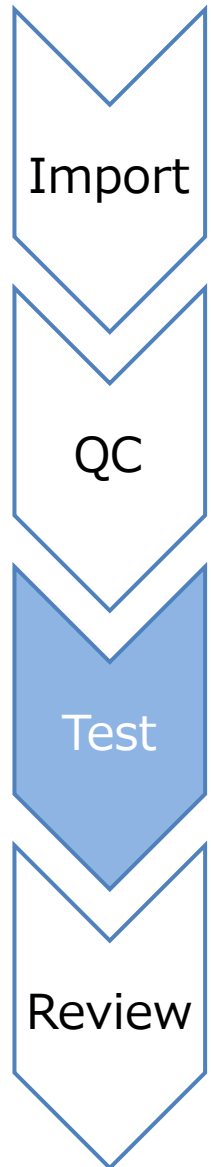
Review



■ Collapsing Methods実行前に、バリエーションの生体への影響を評価したデータベースの情報を用いて、サンプルのバリエーションデータのフィルタリングを行うことが可能

■ バリエーションの人種別のアレル頻度データベースもダウンロードでき、フィルタリングに用いることが可能





- オプションパラメータとして、ユニットとして使用する遺伝子アノテーションや、アリル頻度データベースなどを指定し、Collapsing Methodsを実行する

Import

QC

Test

Review

- Test Statistics
  - Correlation/Trend Test
  - T-test
  - Logistic Regression
  - Linear Regression



Import

QC

Test

Review

Association Tests (Additive Model) [112]

File Edit Select DNA-Seq Genotype Numeric RNA-Seq GenomeBrowse Plot Scripts Help

All: 384,255 x 31  
Active: 384,255 x 31

Unsort		R	1	R	2	R	3	R	4	R	5	
Map	Marker		Corr/Trend P		Corr/Trend -log10 P		Corr/Trend R		Corr/Trend X^2		Corr/Trend expected P	
1	SNP_A-1909444		0.823610992556257		0.0842778655816102		0.0103591936230651		0.0496858692368305		0.824214388882383	
2	SNP_A-4303947		0.607711955235808		0.216302219894464		-0.0241189244963382		0.263520301043608		0.609661552875044	
3	SNP_A-1886933		0.419039164827852		0.377745384494962		-0.0379673710741759		0.65300913362673		0.418345890099023	
4	SNP_A-2116190		0.390244651738654		0.408663040184332		-0.0399724317266052		0.738181427739809		0.389052842513435	
5	SNP_A-4291020		0.803995332242444		0.0947464726287053		-0.0117123528649632		0.0615934651254015		0.805960885349573	
6	SNP_A-1902458		0.204380733320902		0.68956004692674		-0.0591104809440866		1.6107565693804		0.202270627578041	
7	SNP_A-2109914		0.726161942202304		0.138966515991351		0.0162768295070365		0.122664987784925		0.726196145788604	
8	SNP_A-2291997		0.674846036031889		0.170795298767169		-0.019753807980215		0.175986031303362		0.677207843749593	
9	SNP_A-4277872		0.867355283244052		0.0618029717894659		-0.00779577000275364		0.0278952797405475		0.866165697258331	
10	SNP_A-4221087		0.229831562291807		0.638590330791662		0.056354897148033		1.44184699238468		0.228235156341492	
11	SNP_A-1866065		0.733342611279004		0.134693079349566		0.0164483094862814		0.116064613646336		0.733420515022576	
12	SNP_A-2288244		0.360313940856293		0.443318934325298		0.0425129440621263		0.83680324114005		0.358682385395115	
13	SNP_A-1884606		0.132259429616758		0.878573354570407		-0.070259028449073		2.26577596508091		0.129587123134377	
14	SNP_A-1783407		0.314539713220968		0.502324513513222		0.0471960009893667		1.01126797926229		0.312470885219451	
15	SNP_A-2082515		0.73530936325985		0.133529903787277		0.0157115144262479		0.114292330417148		0.735320295116524	
16	SNP_A-1910751		0.759533817513879		0.119452884825938		0.0145924136675373		0.093692956123593		0.761664779898765	
17	SNP_A-2235839		0.679716693228732		0.167672064066117		-0.0193547127390384		0.170445231870866		0.681915654968706	
18	SNP_A-2081399		0.830231187362503		0.0808009565211714		-0.0101184103144434		0.0459696200538541		0.830205202274531	
19	SNP_A-1919019		0.894433258677915		0.0484520603166122		0.00617360918199544		0.0176084140533925		0.893561567188456	

Association Tests (Additive Model)

- いずれの手法を用いた場合も、解析に用いたマーカーごとの統計値（P-valueなど）が出力される
- Marker Mapデータが存在する場合は、遺伝子名などのアノテーション情報も表示される

Import

QC

Test

Review

- データのビジュアライズ
  - Manhattan Plot
  - LD Plot
  - Q-Q Plot ...
- 集団の階層化の評価、遺伝率の計算など
  - Genomic Control
  - LD Score Regression
- メタ解析
  - Meta-Analysis

```
Markers analyzed:
  519308 markers with two alleles
  (35 markers with one allele were found.)

Analysis parameters:
Genetic model/test: Additive model
Use missing values: No

Correct Input Data for Stratification using PCA: No

Test Statistic or Method:
* Armitage Trend Test

Multiple Testing Correction:
Bonferroni adjustment for 519308 markers: Yes
False Discovery Rate: Yes
Single Value Permutations: No
Full Scan Permutations: No

Genomic Control of Output Data for Stratification: Yes
Output data for P-P/Q-Q plots: Yes
Output -log10 P: Yes

Markerwise Genotype Statistics:
  Call Rate: No
  Number of Alleles: No
  Allele Frequencies: Yes
  Carrier Counts: No
  HWE P-Value: No
  Fisher's Exact Test for HWE P-Value: No
  Signed HWE R: No

  Also output -log10(Value): No
  Also output data for P-P/Q-Q plots: No

  Genotype Counts: Yes
  Allele Counts: Yes

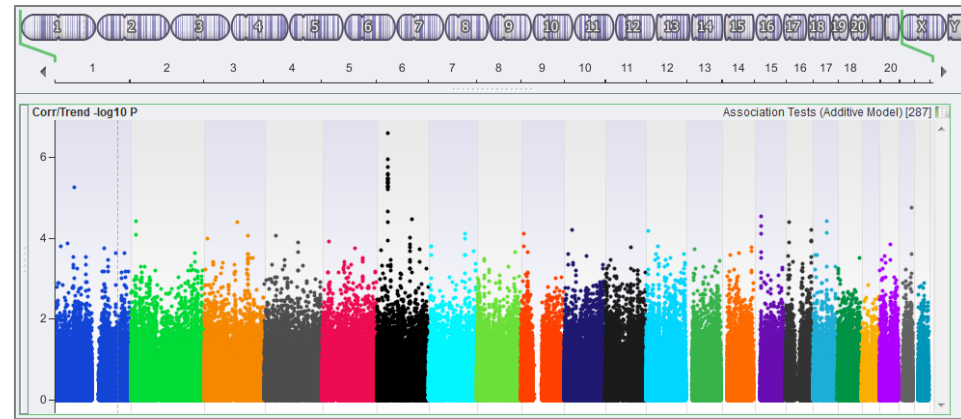
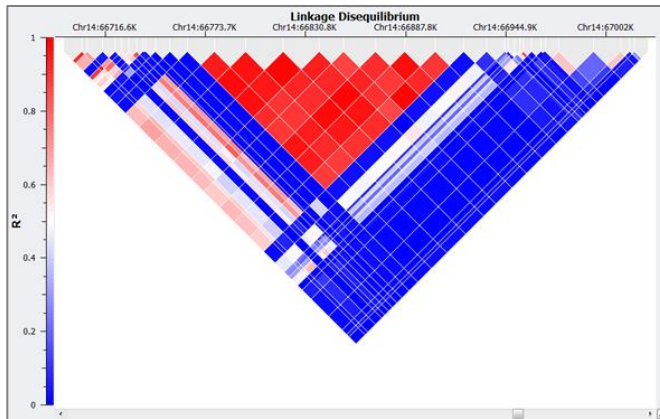
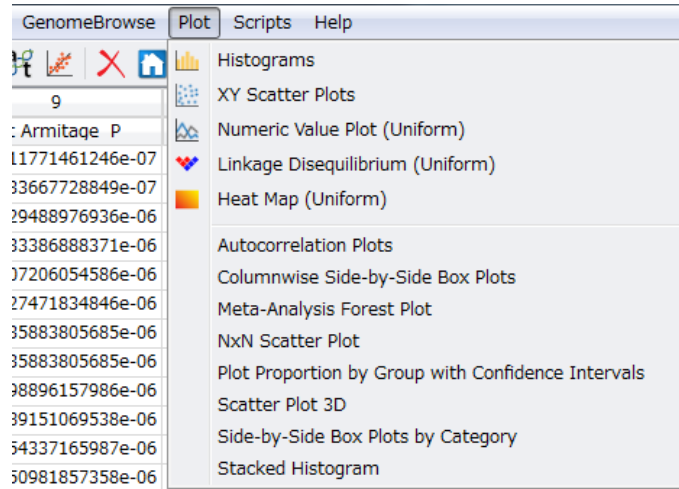
Inflation Factor (Lambda) Found for Armitage : 2.66031
```

Import

QC

Test

Review



- グラフ作成機能を利用し、解析結果の統計値などから簡単にプロット図を作成
- 2Dスキャッタープロットやヒートマップ、ゲノムブラウザーなどに各種データをプロット可能

Import

QC

Test

Review

- 関連解析におけるカイ二乗統計量の中央値の上昇を示すInflation factor (lambda)を指標に、この値が1より大きい場合は、サンプル集団の階層化の可能性がある
- 関連解析実行時に、補正值としてInflation factorの値を入力し、集団の階層化の補正を行うことが可能

Genomic Control of Output Data for Stratification \_\_\_\_\_

Show inflation factor (lambda), chi-squares, and corrected values

Correct using this inflation factor (lambda) instead:



```
Genomic Control of Output Data for Stratification: Yes
Output data for P-P/Q-Q plots: Yes
Output -log10 P: Yes

Markerwise Genotype Statistics:
Call Rate: No
Number of Alleles: No
Allele Frequencies: No
Carrier Counts: No
HWE P-Value: No
Fisher's Exact Test for HWE P-Value: No
Signed HWE R: No

Also output -log10(Value): No
Also output data for P-P/Q-Q plots: No

Genotype Counts: Yes
Allele Counts: No

Inflation Factor (Lambda) Found for Corr/Trend : 1
```

Import

QC

Test

Review

- ゲノム上の複数のSNPが表現型に影響を与えていること（ポリジェニックモデル）を仮定した計算手法
- あらかじめ計算しておいた各SNPごとの周辺SNPとの連鎖不平衡値の総和（LD score）と、関連解析で計算した統計量を用いて、両データ間の相関情報から、関連解析の結果を評価する様々な値を計算することが可能
- SVSでは遺伝率の計算と、他の表現型との遺伝学的相関の計算が可能

LD Regression Statistics [152]

### Summary Statistics

Heritability of GWAS + LD Scores - Mapped Sheet 1 - Mapped Sheet 1

Total Observed scale h2:	196.48381745 (6.40775007)
Lambda GC:	19.78388902
Mean Chi <sup>2</sup> :	37.16974346
Intercept:	33.24215882 (0.85318035)
Ratio:	0.89141243 (0.02358823)

Heritability of bip Dataset - Row Subset + LDSCORE Data

Total Observed scale h2:	0.14101099 (0.00299342)
Lambda GC:	1.13326885
Mean Chi <sup>2</sup> :	1.1401359
Intercept:	1.02380458 (0.01678472)
Ratio:	0.16986785 (0.11977457)

Genetic Covariance

Total Observed scale gencov:	-0.23000697 (0.02665641)
Mean z1*z2:	0.0009498
Intercept:	0.01916691 (0.0309698)

Genetic Correlation

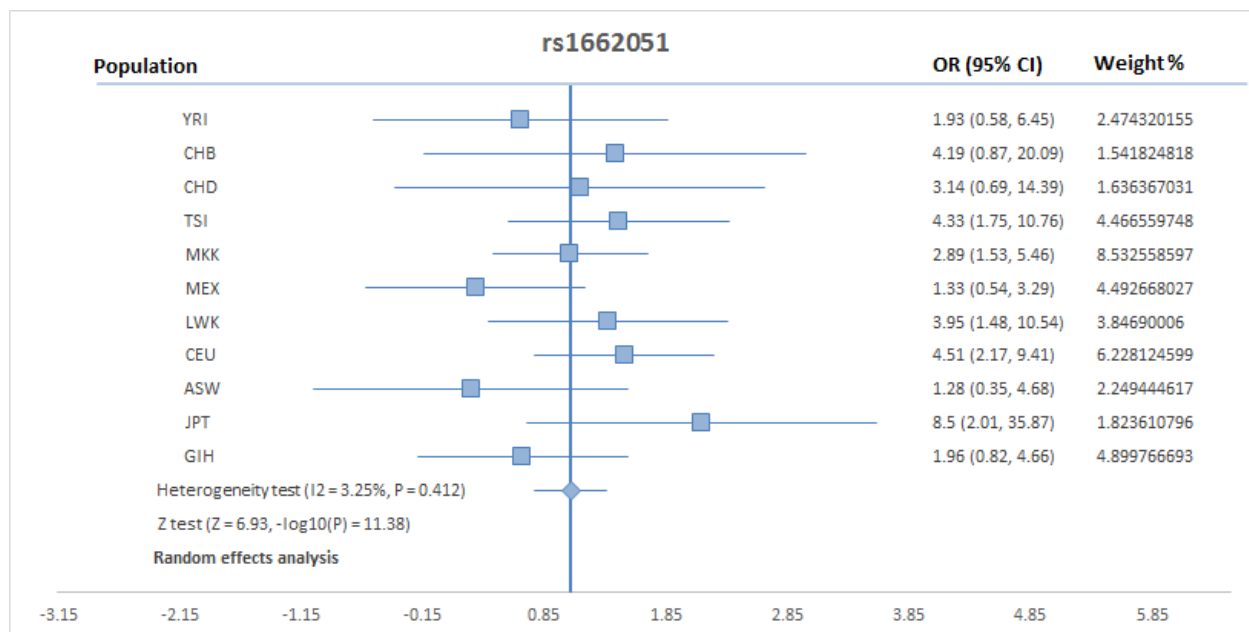
Genetic Correlation:	-0.04369695 (0.04141012)
Z-score:	-1.05522386
P:	0.29132295

Import

QC

Test

Review



- 同一の表現型に対して実施された、複数の研究プロジェクトの関連解析の結果を統合し、新たな知見を導き出す手法
- 固定効果モデルとランダム効果モデルの2種類のアルゴリズムをサポートし、研究プロジェクト間の不均一性の評価も可能
- 解析結果を比較したフォレストプロットの作成が可能

お問い合わせ先：フィルジエン株式会社  
TEL: 052-624-4388 (9:00～17:00)  
FAX: 052-624-4389  
E-mail: [biosupport@filgen.jp](mailto:biosupport@filgen.jp)