博淼生物
BIOMIAO BIOLOGICAL
-SINCE2009-
Your own Laboratory
您的专属实验室

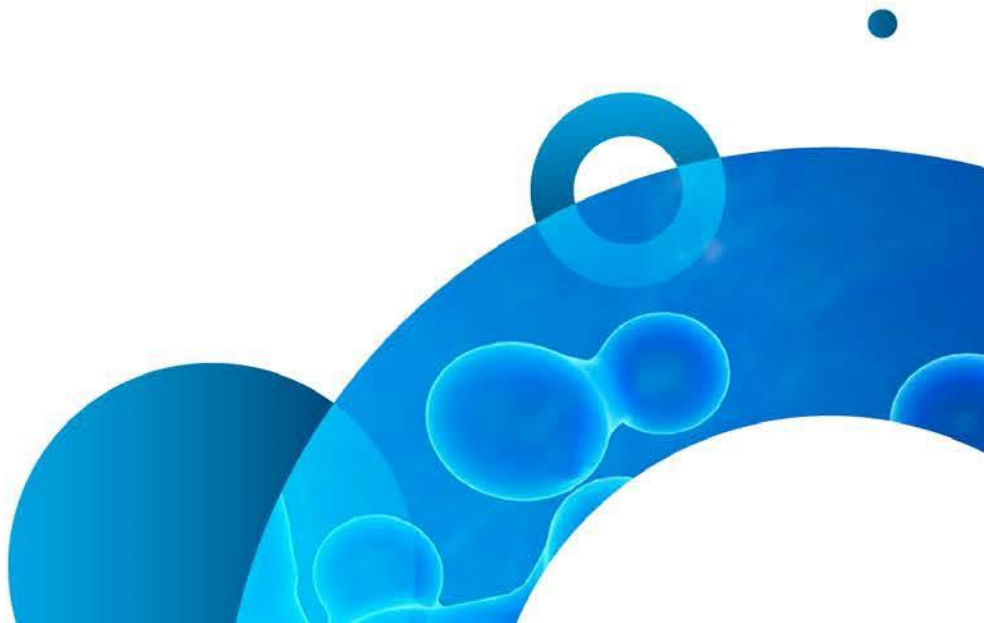
Post-GWAS分子流行病学研究创新策略

全国统一服务电话：400-6506-908

网址：www.biomiao.com

邮箱：marketing@biomiao.com

地址：北京市丰台区丰管路优橙创新中心3012-3015



目录
CONTENTS

GWAS经典研究策略

GWAS经典方案设计

博淼生物 BIO MIAO BIOLOGICAL SINCE 2009

01

Stage I—Discovery Sample GWAS

- ✓ Illumina–中华GWAS芯片
- ✓ Illumina–ASAMD GWAS芯片
- ✓ Illumina–CGA GWAS芯片

02

GWAS生物信息学分析

- ✓ Imputation PCA等
- ✓ Association analysisi等
- ✓ LD block and haplotype analysis等

03

Stage II—Replication Sample Target SNP Genotype

- ✓ 大样本、多中心
- ✓ Multi–PCR NGS、Massarray 、KASP Genotype

04

生物信息学分析

- ✓ Combine association
- ✓ Covariates、Various phenotype、Quantitative traits、Model based、Logistic等

经典单—GWAS研究的突破

突破——队列样本设计

- 精细化样本亚分组
- 样本数量的跨越式提升
- 多中心样本的设置

突破——因果数据分析

- Finemapping、Colocalization等分析策略的提升推动association结论往causal结论的跨越
- Imputation分析数据量的扩展
- 公开数据库多组学数据的完善，多维度诠释in silico analysis的价值

突破——遗传模型构建

- 多基因PRS/GRS模型算法的优化
- 跨中心/跨种族超大队列样本数据的整合

适合中国人群GWAS芯片介绍

- 中华芯片v1.4

✓ 位点数量：117万个

✓ 中国人群，MAF>1%，覆盖率最高

- GSA（Global screening array芯片

✓ 位点数量：>74万个

✓ 全球泛种族人群

✓ 价格低廉

✓ 第一款百万人群队列研究芯片

- ASAMD（Asian screening array multi-disease）芯片

✓ 位点数量：>74万个

✓ 东亚人群，MAF>1%

✓ 价格低廉

✓ 兼具科研及转化应用双重价值

- CGA(Chinese genotyping array)芯片

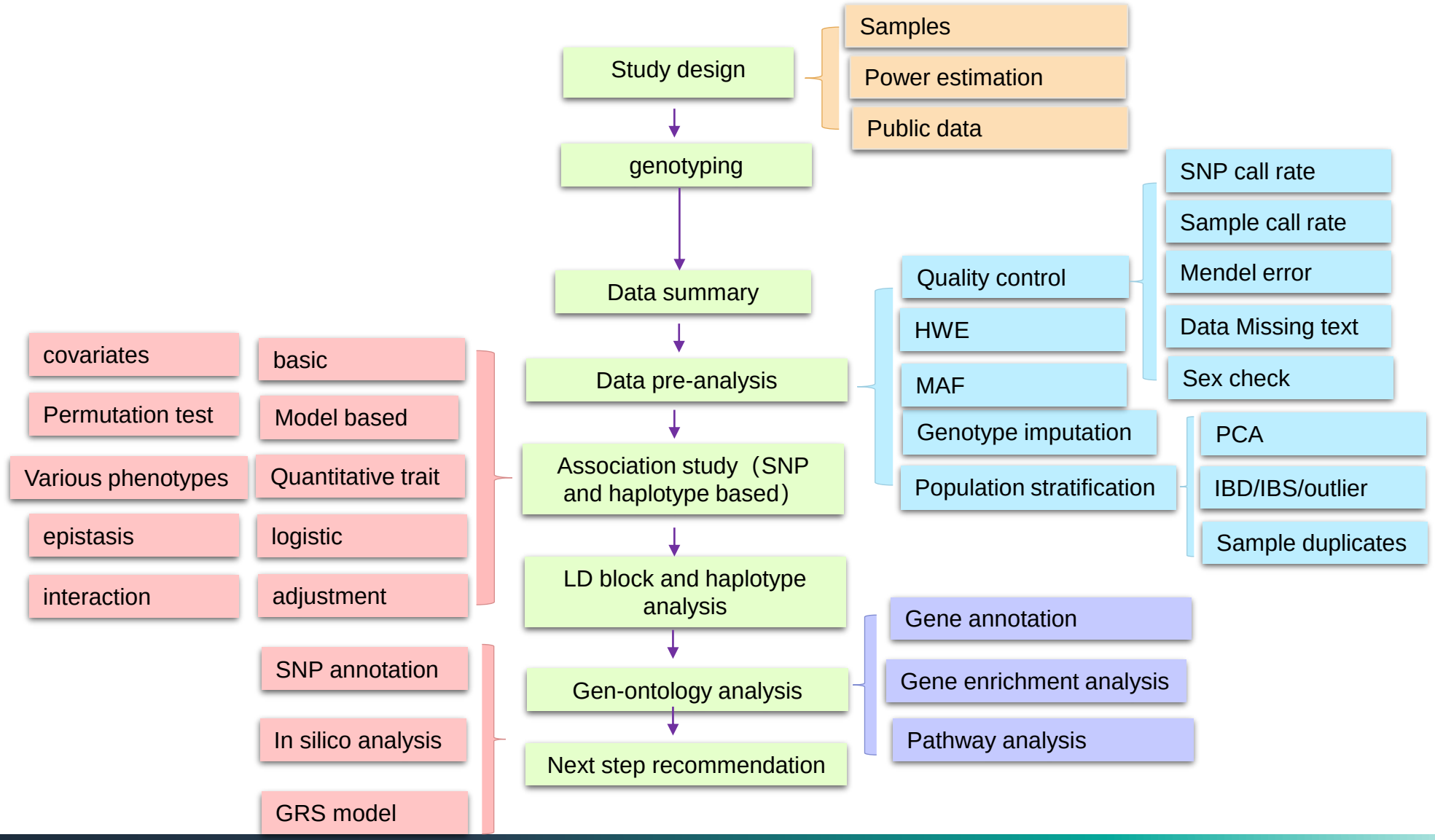
✓ 位点数量：>74万个

✓ 增加中国人群专属位点近20万个，MAF>1%

✓ 价格低廉

✓ 兼具科研及转化应用双重价值

GWAS生物信息学分析流程

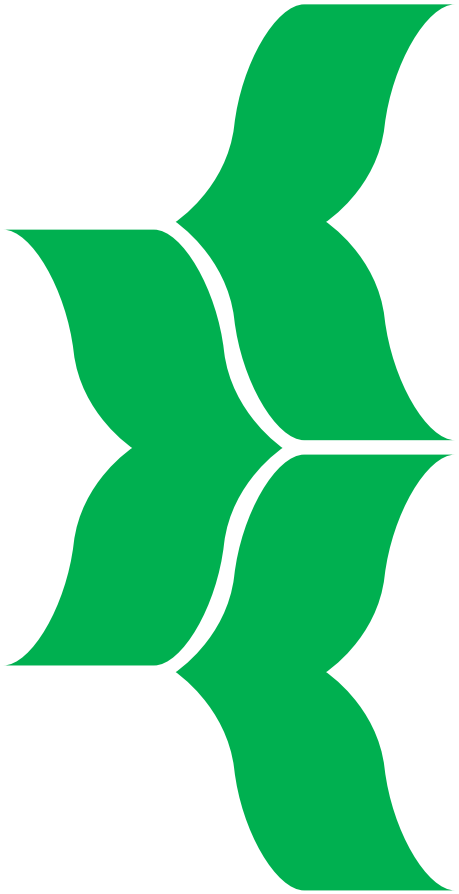


GWAS imputation analysis

博淼生物
BIO MIAO BIOLOGICAL
SINCE 2009

Imputation analysis

根据芯片位点数据，基于单体型进行未检测位点分型数据填充，提高全基因组SNP位点覆盖度



1000Genomes Phase3

优势：经典参考数据集，文章发表数量众多

劣势：跨种族人群，遗传异质性，填充质量较低

ChinaMAP/CKB Panel

优势：中国人群自主参考数据集，高质量填充

劣势：无法下载，online分析

Haplotype Reference Consortium/TopMed

优势：最大规模的单倍型数据库，高达64976个单倍型的reference panel

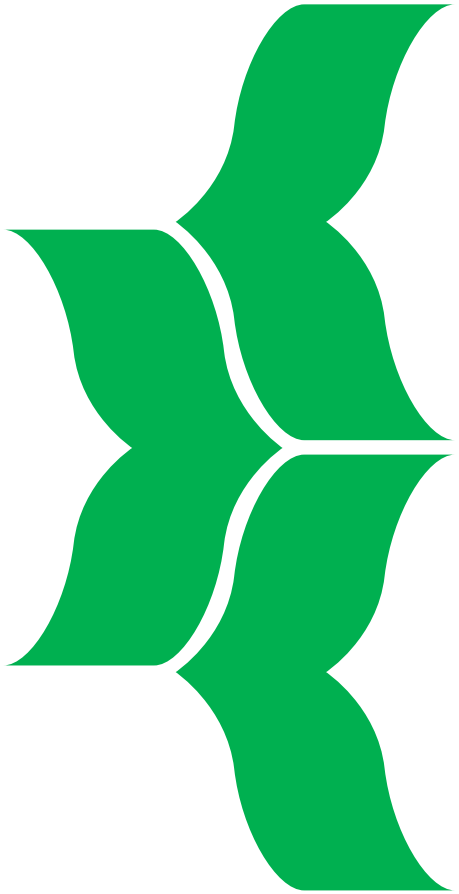
劣势：无法下载，在线填充准确性较低

GWAS causal snp analysis

博淼生物 BIO MIAO BIOLOGICAL SINCE 2009

In silico function annotation

GWAS数据与特定组织/细胞类型中的基因组/表观基因组数据联合分析，为精准定位因果调控变异提供了数据理论支撑.



GTEx Portal/TCGA/GEO数据库

基础的最全面不同组织中基因表达与SNP的eQTL\mQTL\sceQTL分析.

FUMA数据库

Functional consequences on genes , CADD score, RegulomeDB score, 15 chromatine state (127 tissue/cell types),eQTL, 3D chromation interactions (Hi–C),GWAScatalog

GWAS4D数据库

调控变异优先排序，筛选lead SNP，包括GTEx的QTL分析，1480个转录调控因子，Hi–C数据，非编码变异功能注释；SNP–靶基因相互作用调控预测；**Disease–causal SNP prediction**

GWAS经典策略文章亮点解析

Identification of Novel T1D Risk Loci and Their Association With AgeandIsletFunctionatDiagnosis in Autoantibody–Positive T1D Individuals:BasedonaTwo–Stage Genome–Wide Association Study

GWAS检测

- ✓ 1,045 T1D case + 1,308 control
- ✓ 中华GWAS芯片

Imputaiton 和 association分析

- ✓ Imputation of the non–MHC regionwas conducted using IMPUTE2 software + Imputation of the MHC region was performed with SNP2HLA
- ✓ 17 distinct genomic regions with P values<10E5

SNP Replication 检测

- ✓ 1,378 case +3,774 control
- ✓ 13个SNP位点，massarray分型技术+taqman分型技术
- ✓ 最终4个SNP位点与中国人群T1D显著相关

In Silico Bioinformatics Analysis

- ✓ define potentially causal variants
- ✓ assess Funciton such as eQTLs、 DNAse–I hypersensitive sites, enhancers, promoters, and histone peaks
- ✓ rs4320356 at 6p22.2 (PP = 0.14), rs9273471 at MHC (PP = 0.43), rs10905277 at 10p14 (PP = 0.21),and rs773125 at 12q13.2(PP=0.27) were the probable causal variants
- ✓ Further annotation revealed that rs4320356 and rs10905277 were involved in the regulation of T cells and associated with the expression of BTN3A1 and GATA3 in whole blood,

风险评估模型建立

- ✓ 模型一：The five independent variants of the MHC region，AUC=0.85
- ✓ 模型二： the identified eight variants in the Chinese population，AUC=0.86
- ✓ 模型三： additional validated variants from the Caucasian，AUC=0.87

02

GWAS多组学联合研究策略

GWAS多组学联合研究的创新

研究目的

研究手段

研究定位的转变

单一GWAS研究的局限性

- ✓侧重相关性结论的挖掘
- ✓遗传模型对于复杂表型的片面性
- ✓高质量队列样本的长周期
- ✓非功能区域易感SNP位点的分子调控功能机制挑战

GWAS多组学研究的广阔性

- ✓ 构建不同纬度的QTL调控network以及易感SNP causal结论的挖掘
- ✓ 不同多组学TWAS、MeWAS、PWAS等融合分析及MR随机化因果推断分析
- ✓ 筛选与表型有关的重要生物学通路及其中关键调控功能分子

Multi-Omics roles

博淼生物 BIO MIAO BIOLOGICAL — SINCE 2009 —

基因组学（Genomics）

—What is possible happen

- ✓ 胚系细胞SNP、CNV、InDel等遗传变异
- ✓ 体细胞mutation、InDel等环境突变变异

转录组学（Transcriptomics）

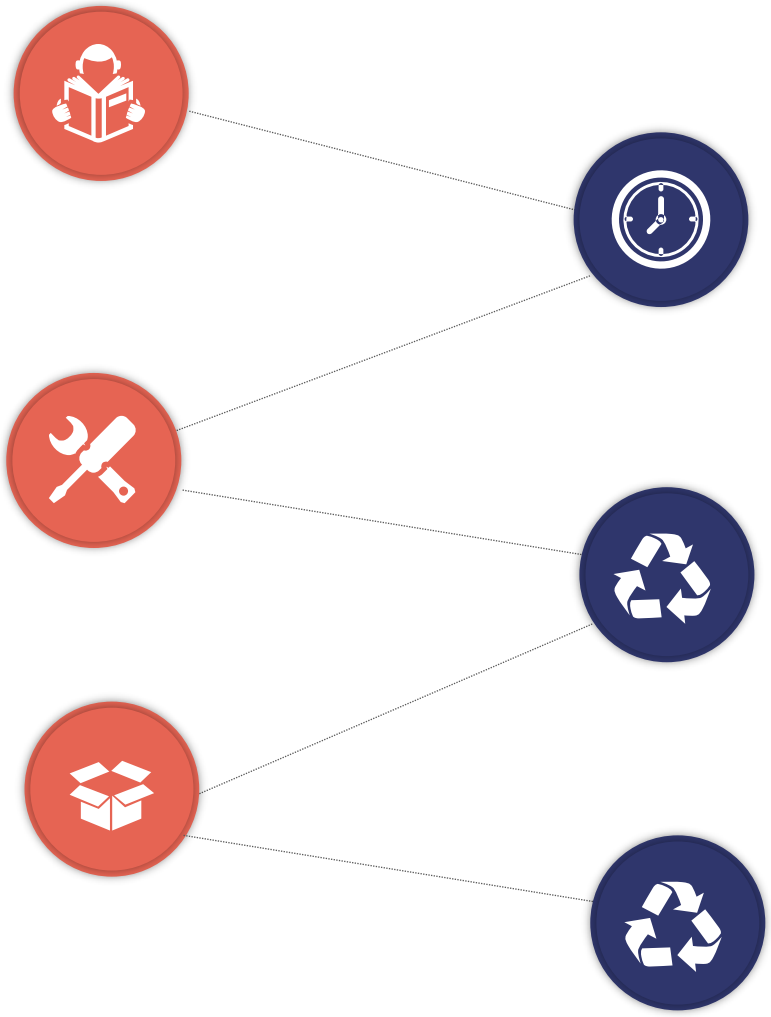
—What appears to be happening

- ✓ 转录本类型：mRNA、miRNA、lncRNA、cirRNA、piRNA等
- ✓ 转录定位：表达、融合、剪切、ceRNA等

代谢组学（Metabolomics）

—What has happened

- ✓ 脂质组、氨基酸类、胆汁酸类、神经递质类等等
- ✓ 宿主代谢物、微生物代谢物



表观基因组学（Epigenomics）

—What appears to be happening

- ✓ 基因组：ATAC、HiC
- ✓ DNA甲基化：5mC、5hmC、6mA等
- ✓ RNA甲基化：m6A等

蛋白质组学（Proteinomics）

—What is happening

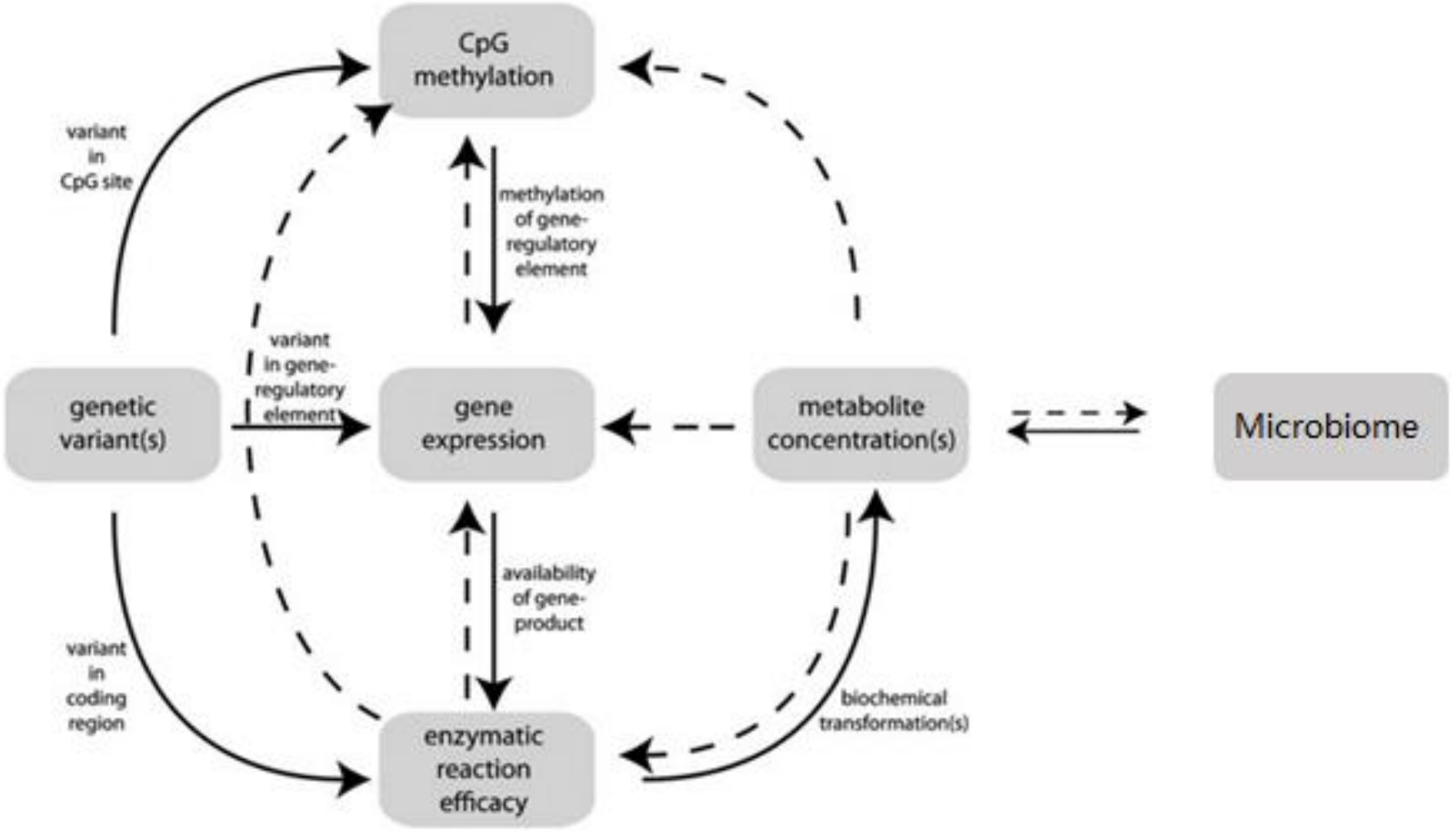
- ✓ 功能蛋白质：酶、转录因子等
- ✓ 修饰蛋白质：糖基化、乙酰化等

微生物组学（Microbiomics）

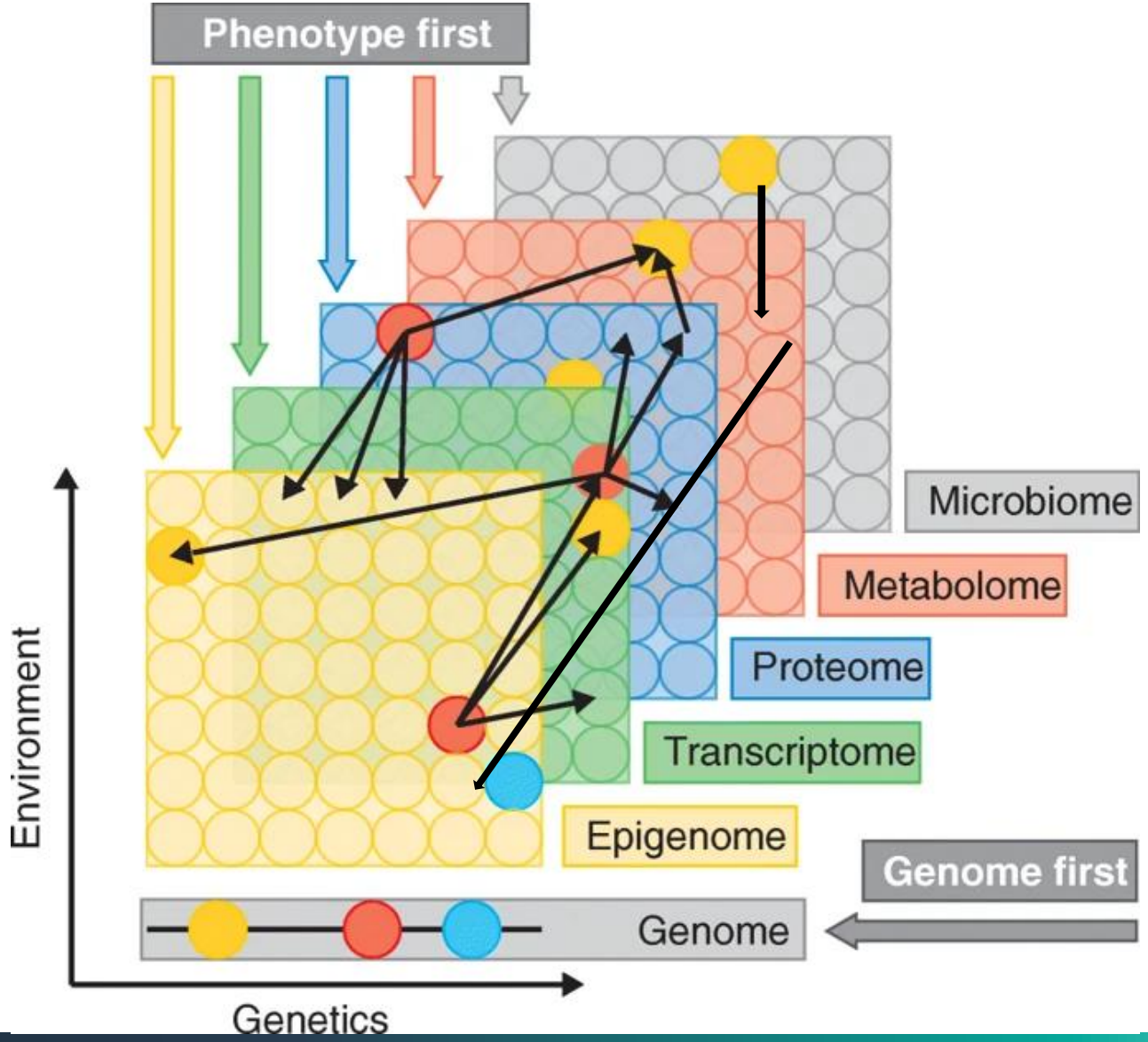
—The second omics：interact

- ✓ 宏基因组、宏转录组、宏蛋白质组、宏代谢组

Multi-Omics联合研究意义



Multi-omics approaches to disease research



GWAS创新理论体系

An Expanded View of Complex Traits: From Polygenic to Omnigenic Cell , 2017

颠 覆 一

多基因调控转向泛基因调控理论

颠 覆 二

核心基因与次要基因协同遗传效应

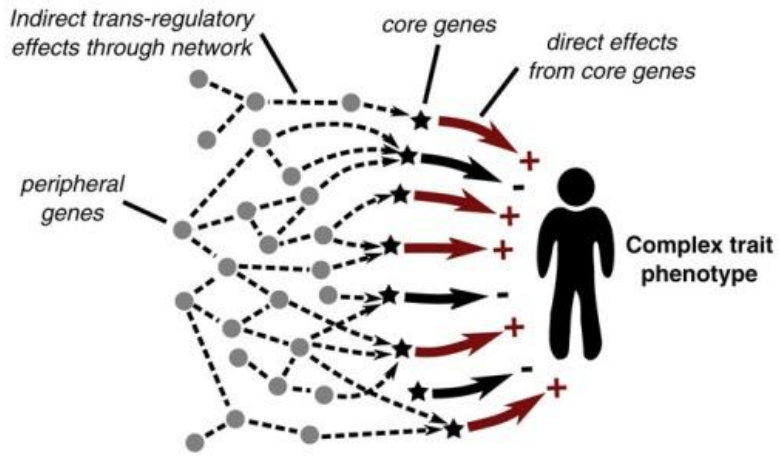
颠 覆 三

泛基因调控模型建立

GWAS创新理论体系

Question: Why is the architecture of complex traits dominated by huge numbers of small effect variants?

Approach: We built a quantitative phenotype model based on core gene expression



$$\mathrm{Var}(Y_i) = \underbrace{\sum_{j=1}^{M} \gamma_j^2 V_{j,\mathrm{cis}}}_{M \text{ core terms}} + \underbrace{\sum_{j=1}^{M} \gamma_j^2 V_{j,\mathrm{trans}}}_{M \text{ trans terms}} + \underbrace{\sum_{j=1}^{M} \sum_{k=1}^{j-1} 2\gamma_j \gamma_k C_{j,k}}_{M^2-M \text{ covariance terms}}$$

Conclusion: Most of the trait heritability is explained by many small *trans*-regulatory effects from peripheral genes

Highlights

✓

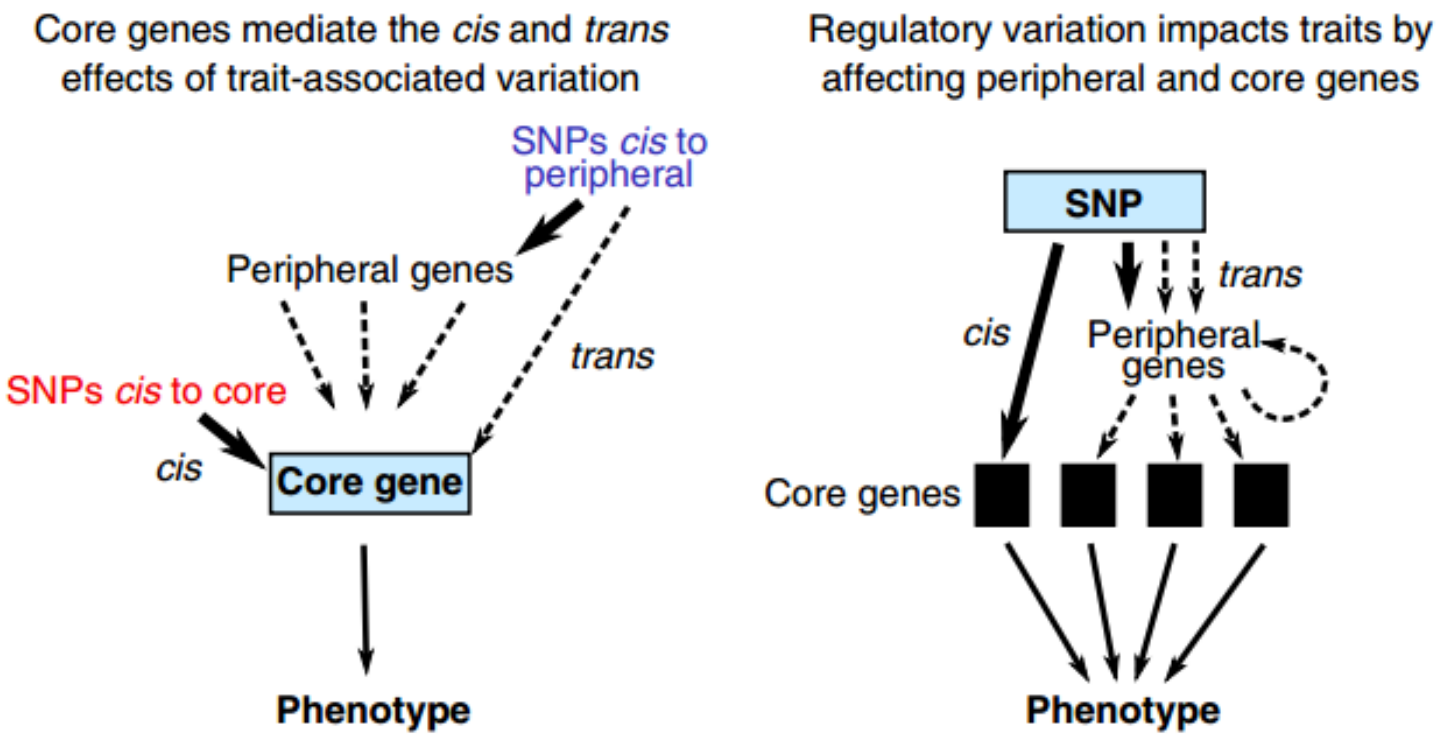
✓

✓

✓

Trans Effects on Gene Expression Can Drive Omnigenic Inheritance Cell , 2019

GWAS创新理论体系



Causal Pathways for Variants Affecting a Trait through Core Genes

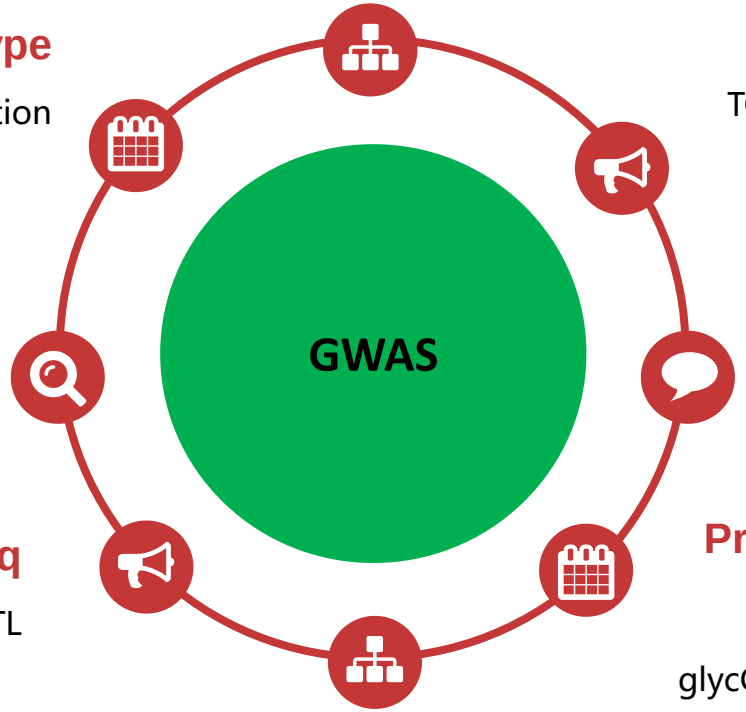
GWAS&多组学研究策略—*样本遗传基线数据库*

Methylation

meGWAS-meQTL

Phenotype

GWAS-asscoiation



RNA-Seq

TGWAS-eQTL、sQTL、roQTL

mRNA/miRNA

/LncRNA/CirRNA/Splice

Single cell RNA-seq

Sc-eQTL

Protein&Glycomics

pGWAS-pQTL

glycGWAS-glycQTL

Metabolome&Microbiomics

mGWAS-mQTL、miGWAS-miQTL

GWAS&多组学研究意义

科研主旨

关联分析研究

✓QTL分析Cis_与trans_分析遗传调控网络network构建

✓精准筛选重要驱动功能基因

✓以PRS为基础的多组学标记物联合模型优化构建

因果效应分析研究

✓MR分析、COLOC分析、Mediation分析、Condition分析等

✓探讨不同组学与表型组、暴露因素与表型组、不同表型组间因果关系

GWAS&Proteomics approach研究策略

pQTL图谱构建及分子调控机制分析

✓基于Cis_与trans_的pQTL关联分析，构建遗传与蛋白质的调控network网络，筛选关键分子遗传调控通路

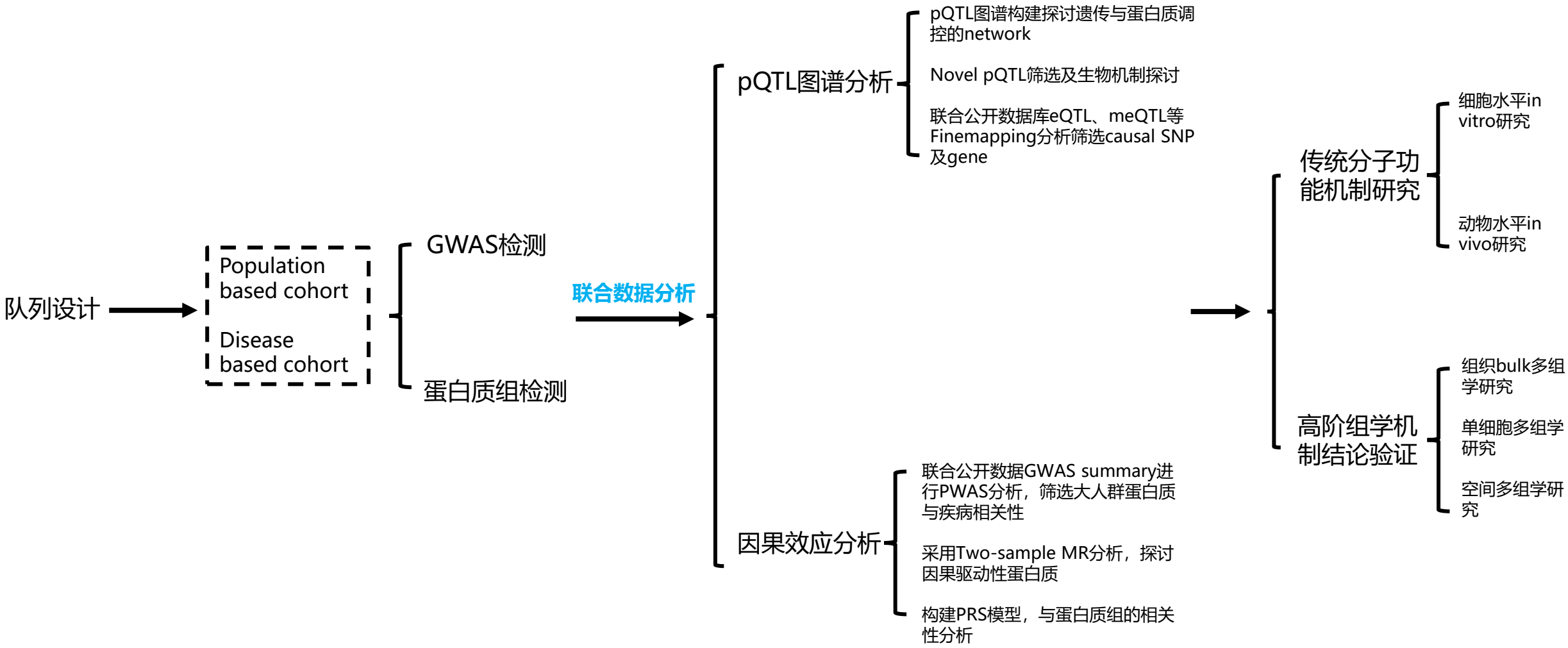
✓对比公开数据库的pQTL,筛选novel pQTL及对应的潜在生物学机制

因果效应推断分析

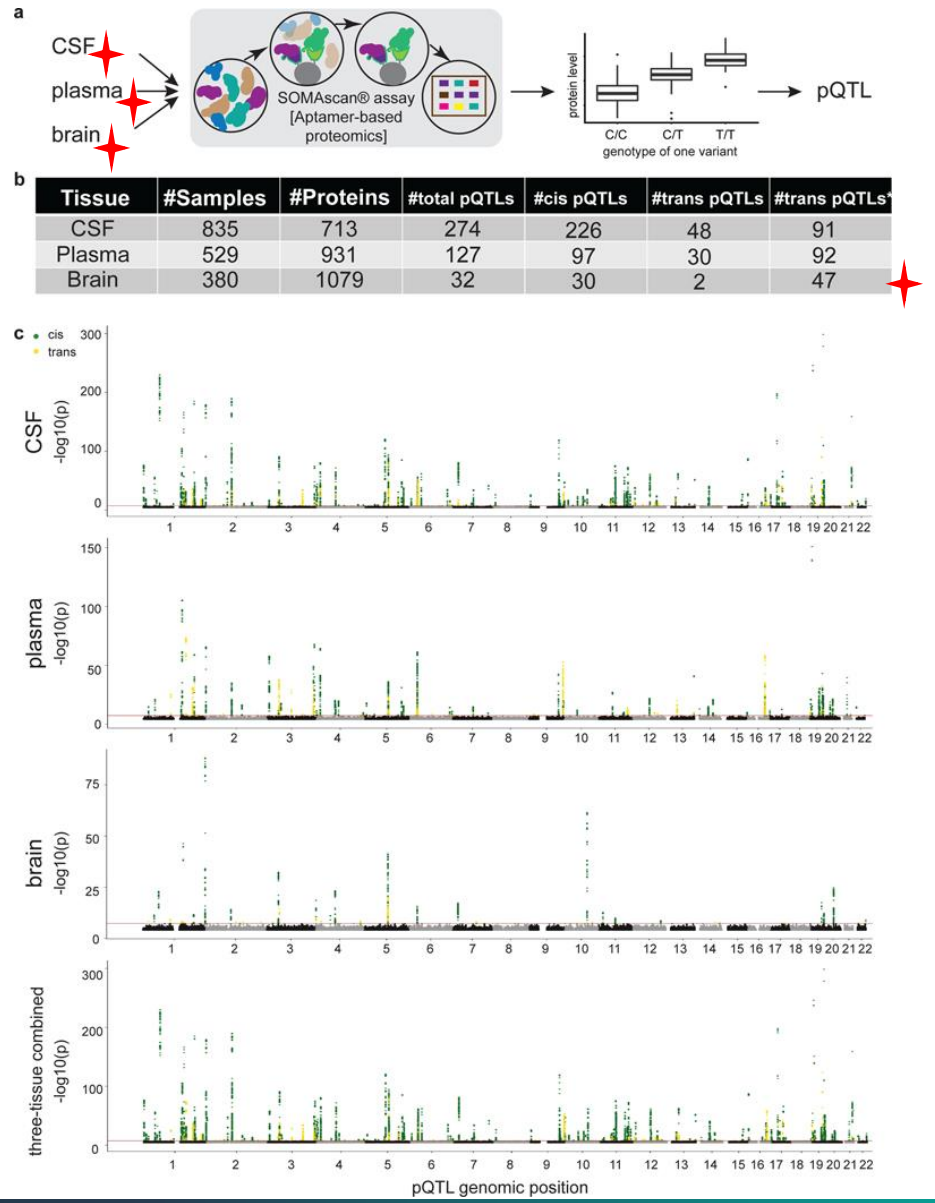
✓通过融合公开表型GWAS大数据库的PWAS、MR、COLOC分析策略，筛选关键蛋白质与表型的因果关系

✓基于公开数据库TCGA、GEO、GTEx等进行eQTL、meQTL等组学数据进行finemapping分析，筛选causal gene与variants

GWAS&Proteomics approach研究策略—研究路径概述



GWAS&Proteomics approach研究策略—代表文献解析



Tissue	#Samples	#Proteins	#total pQTLs	#cis pQTLs	#trans pQTLs	#trans pQTLs*
CSF	835	713	274	226	48	91
Plasma	529	931	127	97	30	92
Brain	380	1079	32	30	2	47

Nature Neurosci. 2021

GWAS&Proteomics approach研究策略—代表文献解析

Abstract

Understanding the tissue-specific genetic controls of protein levels is essential to uncover mechanisms of post-transcriptional gene regulation. We generated a genomic atlas of protein levels in three tissues relevant to neurological disorders (brain, cerebrospinal fluid (CSF), and plasma), by profiling thousands of proteins from participants with and without Alzheimer disease (AD). We identified 274, 127, and 32 protein quantitative trait loci (pQTLs) for CSF, plasma, and brain, respectively. Cis-pQTLs were more likely to be tissue-shared, but trans-pQTLs tended to be tissue-specific. Between 48.0 to 76.6% of pQTLs did not colocalize with expression, splicing, DNA-methylation, or histone-acetylation QTLs. Using Mendelian randomization (MR), we nominated proteins implicated in neurological diseases, including AD, Parkinson's disease or stroke. This first multi-tissue study will be instrumental to map signals from genome-wide association studies (GWAS) onto functional genes, to discover pathways, and to identify drug targets for neurological diseases.

GWAS&Proteomics approach研究策略—代表文献解析

Disease specific analyses: To investigate a disease-specific effect on pQTLs, we performed linear regression on the same protein-loci pairs (before conditioning on top variants) identified from the above default model using three additional models: 1) joint analysis including disease status as another covariate (CO vs non-CO); 2) AD case (CA) only using the same covariates as the default model; 3) cognitive unimpaired (CO) only using the same covariates as the default model. Using scatterplots, we visualized the correlation between each of the additional models and our default model. Using model 1 for comparison, we observed a Pearson correlation coefficient of 0.999, 0.999, 0.999 for CSF, plasma, and brain, respectively. Using model 2 for comparison, we observed a Pearson correlation coefficient of 0.991, 0.989, 0.998 for CSF, plasma, and brain, respectively. Using model 3 for comparison, we observed a Pearson correlation coefficient of 0.999, 0.998, 0.602 (p-value = 0.002) for CSF, plasma, and brain, respectively. The relatively low correlation seen when using model 3 for comparison with controls only in brain samples was due to a much smaller sample size.

GWAS&Proteomics approach研究策略—代表文献解析

Performing MR using TwoSampleMR R package

Mendelian randomization is a method of using measured variation in genes of known function to examine the causal effect of a modifiable exposure on disease. This method obtains unbiased estimates of the effects of a putative causal variable without conducting a traditional randomized trial. We used the R package TwoSampleMR v0.4.22[44]. For single SNP remained after clumping, the most basic method, the Wald ratio, was used. This package also implements the harmonization steps before performing MR, and these steps are: 1) Correcting the wrong effect/non-effect alleles; 2) Correcting the strand issues; 3) Fixing the palindromic SNPs; 4) Removing the SNPs with incompatible alleles. The SNPs selected for the analysis were the based on a suggestive threshold of 1×10^{-5}. The beta-coefficients and standard errors (SEs) for the selected variants (pQTL) from this study were used as input of instrumental variables. These instrumental variables were also extracted from the summary statistics from the latest GWAS for the outcome on neurological disease related traits. (Details see Table S26; Briefly, AD-risk GWAS was published in 2019[3]; AD-progression GWAS in 2018[45]; AD-age at onset GWAS in 2017[46]; PD-risk GWAS in 2019[38]; ALS-risk GWAS in 2016[47]; FTD-risk GWAS in 2014[48]; Stroke-risk GWAS in 2018[49]). To check the specificity of protein-neurological disease associations, we also chose asthma-risk GWAS[50] as an outcome of non-neurological disease. To test the directionality of exposure causing outcome is valid, we used the directionality_test function from the same R package. The method confirms whether the exposure (protein) and outcome (trait) directions are correct or not.

GWAS&Proteomics approach研究策略—代表文献解析

博森生物 BIO MIAO BIOLOGICAL SINCE 2009

Colocalization analyses

We performed Bayesian colocalization analysis using the coloc.abf function from the coloc R package[68,69] v3.1. We used the default priors with $p_1 = 1\times10^{-4}$, $p_2 = 1\times10^{-4}$, and $p_{12} = 1\times10^{-5}$. Evidence for colocalization was assessed using the posterior probability (PP) for hypothesis 4 (indicating that there is an association for both protein and disease and that they are driven by the same causal variant(s)). We used PP.H4 > 0.8 as a threshold to suggest that associations were highly likely to colocalize.

For colocalization of pQTLs with disease status: We downloaded and used the full GWAS summary statistics for each disease/trait from their original publications as the same for MR analysis.

For colocalization of cis-pQTLs with cis-eQTLs, cis-sQTLs from GTEx v8 release: We downloaded and used the significant cis-eQTLs and cis-sQTLs summary statistics for two single tissues, cortex and whole blood, from GTEx[5] (https://gtexportal.org/home/datasets). For cis-sQTLs we used gene-level sQTL results, rather than exon-level sQTLs.

For colocalization analysis of plasma pQTLs with eQTLs from eQTLgen: We downloaded and used the significant cis- and trans-eQTL summary statistics for blood, from eQTLgen[9] (https://www.eqtlgen.org/index.html). In both cases we analyzed cis- and trans-QTLs.

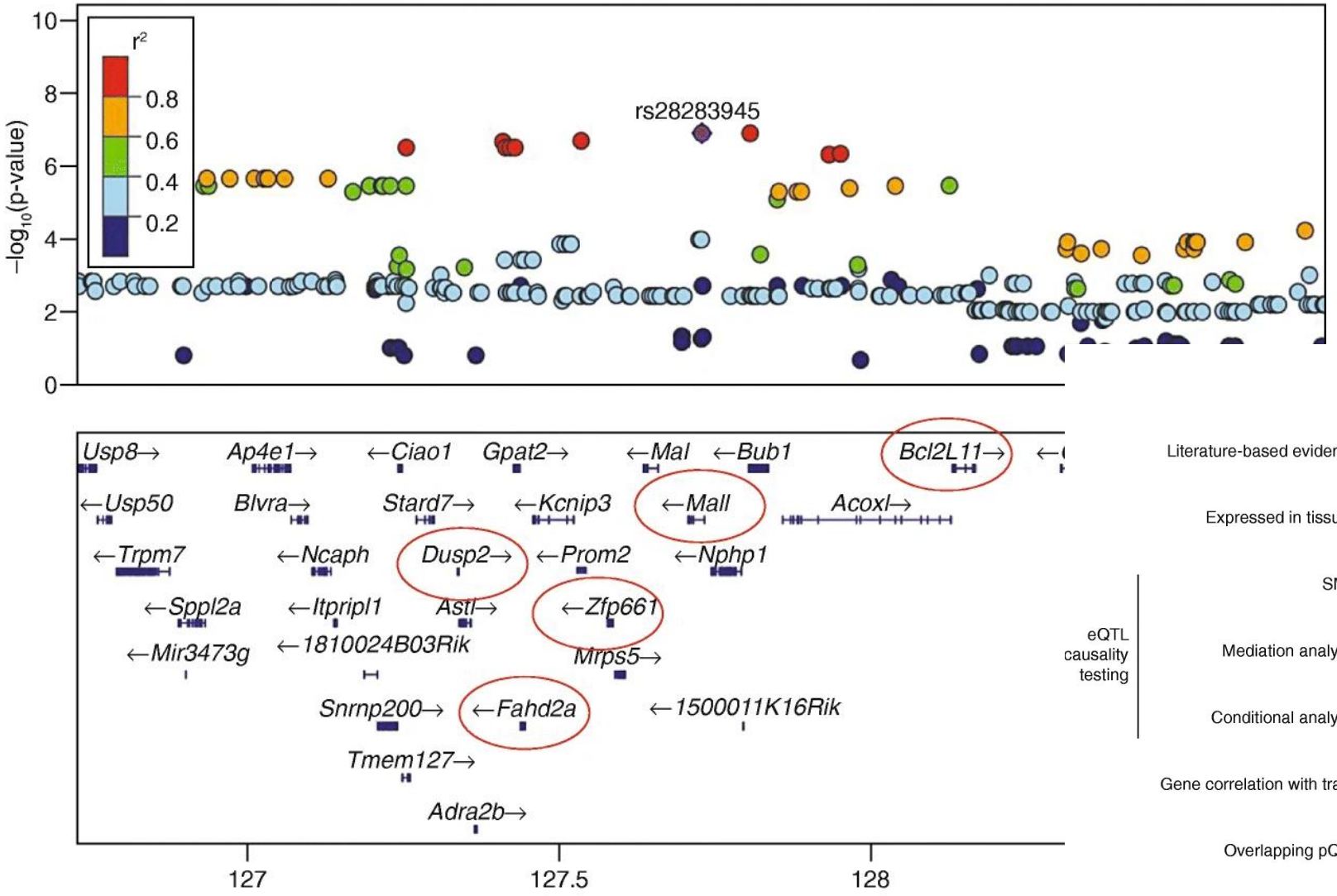
For colocalization of cis-pQTLs with cis-DNA-methylation-QTLs, cis-histone-acetylation-QTLs from ROSMAP: We downloaded and used the significant cis-DNA-methylation-QTL summary statistics for brain tissue, from ROSMAP[70] (http://mostafavilab.stat.ubc.ca/xQTLServe/). We downloaded the significant cis-histone-acetylation-QTL summary statistics (assigning to up to 10Mb upstream of the transcription start site given the same gene) for brain tissue, from ROSMAP[70] as well. To ensure that DNA-methylation-QTLs affecting pQTLs are mediated by eQTLs, we further subset the DNA-methylation-QTLs-pQTLs colocalization result with eQTLs-pQTLs colocalization result.

For colocalization of cis-pQTLs with cell-type-specific cis-eQTLs from ROSMAP: We identified the neuron-, oligodendrocyte-, microglia-, and astrocyte-eQTL data using a pseudo-bulk strategy on snRNA-seq (N=48) from ROSMAP data[71]. In total, we recreated the expression matrices on five cell-types (microglia, excitatory neurons, inhibitory neurons, oligodendrocytes, and astrocytes). We next identified cis-eQTLs for each cell type using fastqtlv2.0[72] after integrating with the whole-genome sequencing data from ROSMAP (N=39). Using both the nominal and permutation modes, we followed the significant eGene calling approach from the GTEx pipeline. We used different priors because the pseudo-bulk derived cell-type specific eQTLs were underpowered compared with bulk-level pQTLs with p1 as 1×10^{-4}, p2 as 1×10^{-2}, and p12 as 1×10^{-3}. The results can be found in Supplementary Fig 9 & Table S35).

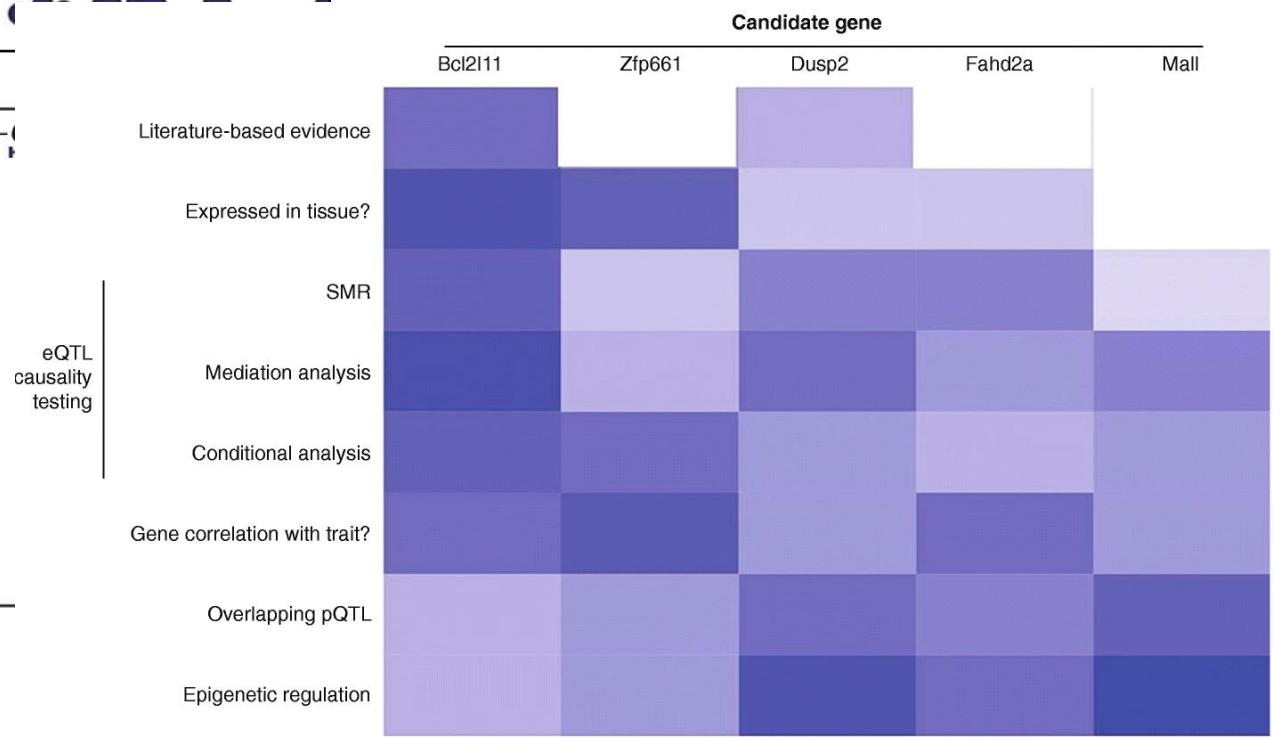
Overlap of proteins with pQTLs and drug targets

To obtain information on drugs that target proteins with pQTLs from this study, we used the DrugBank database (as of 1/3/2020)[73]. This is a manually curated database that maintains profiles for >15,000 drugs (including FDA-approved and experimental drugs). For our analysis, we focused on the protein target for each drug. For each protein assayed, we identified all drugs in the DrugBank with a matching protein target based on UniProt ID, annotated via https://www.uniprot.org/database/DB-0019. We further integrated the MR results on proteins as drug targets.

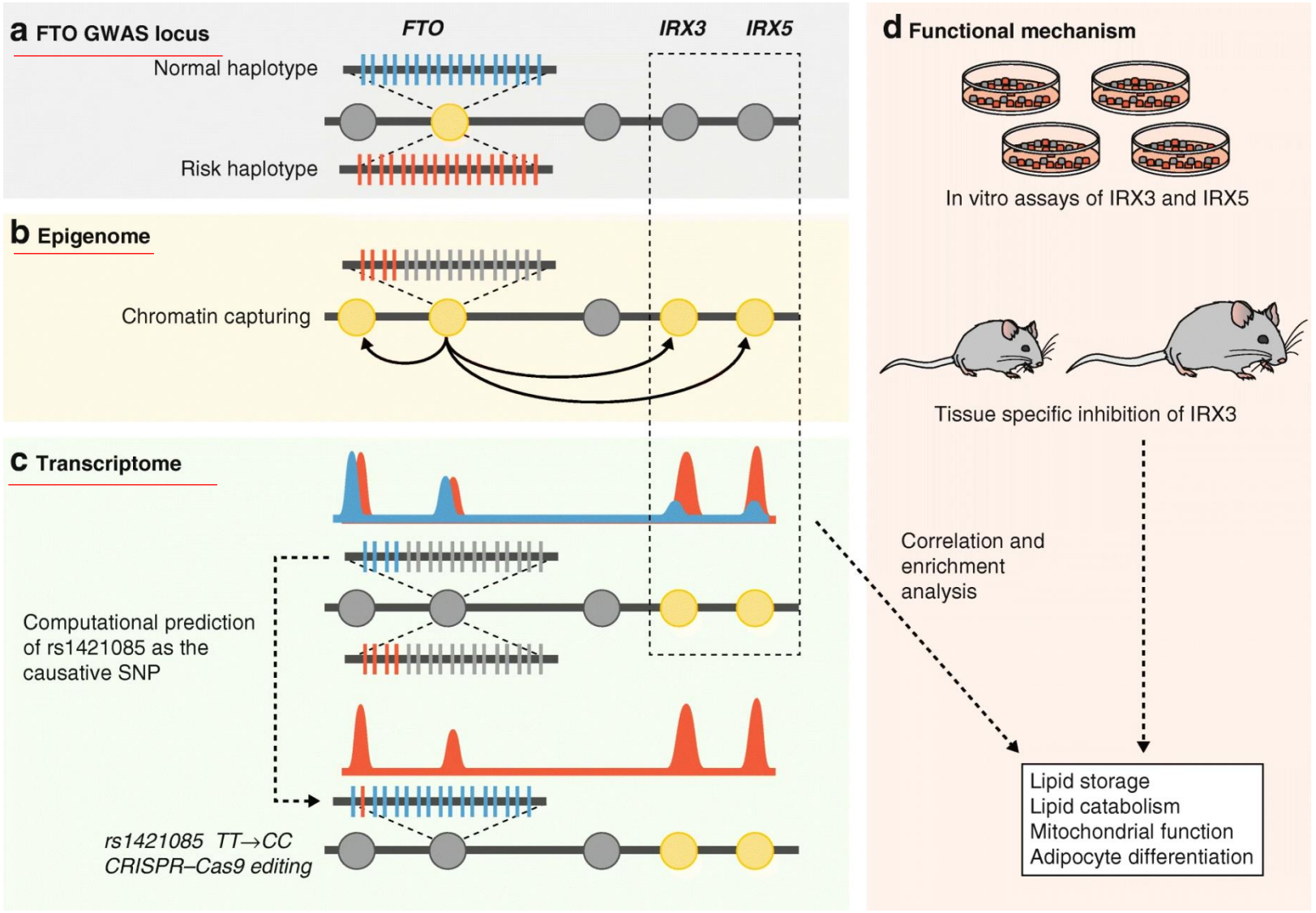
GWAS多组学联合研究——优选易感位点及因果推断分析



Narrowing causal mechanisms

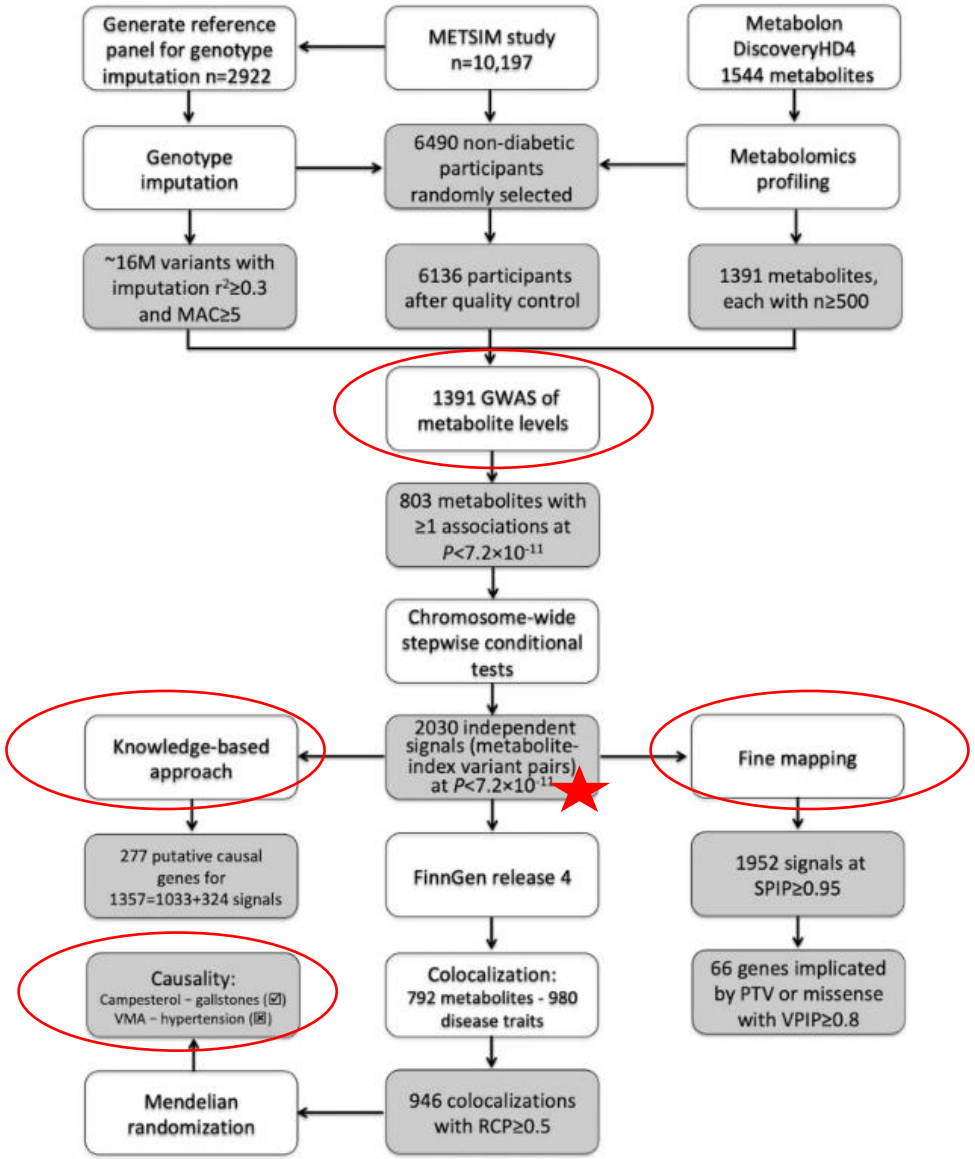


GWAS多组学联合研究——筛选驱动功能基因及分子调控机制验证



FTO obesity variant circuitry and adipocyte browning in humans. *N Engl J Med*

GWAS多组学联合研究——mQTL图谱构建及分子调控机制挖掘



Nature communications

GWAS多组学联合研究——联合蛋白质组PWAS分析及MR因果分析

nature genetics
LETTERS
https://doi.org/10.1038/s41588-020-00773-z
Check for updates

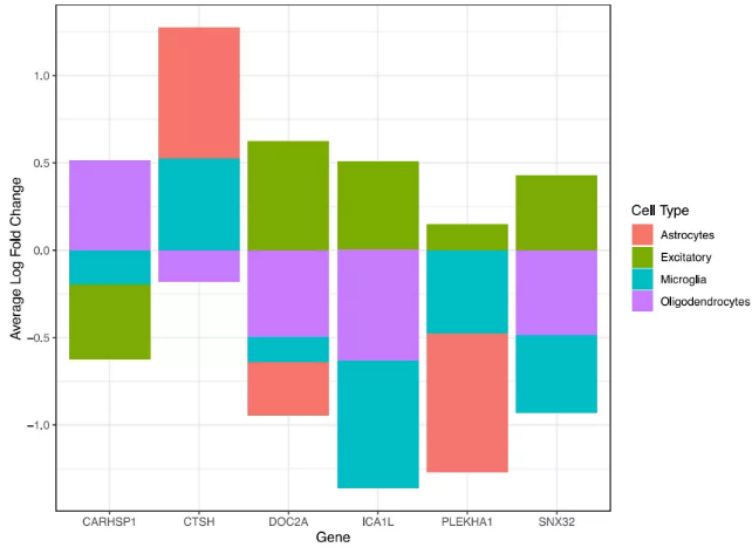
Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis

Genome-wide association studies (GWAS) have identified many risk loci for Alzheimer's disease (AD)[1,2], but how these loci confer AD risk is unclear. Here, we aimed to identify loci that confer AD risk through their effects on brain protein abundance to provide new insights into AD pathogenesis. To that end, we integrated AD GWAS results with human brain proteomes to perform a proteome-wide association study (PWAS) of AD, followed by Mendelian randomization and colocalization analysis. We identified 11 genes that are consistent with being causal in AD, acting via their *cis*-regulated brain protein abundance. Nine replicated in a confirmation PWAS and eight represent new AD risk genes not identified before by AD GWAS. Furthermore, we demonstrated that our results were independent of *APOE e4*. Together, our findings provide new insights into AD pathogenesis and promising targets for further mechanistic and therapeutic studies.

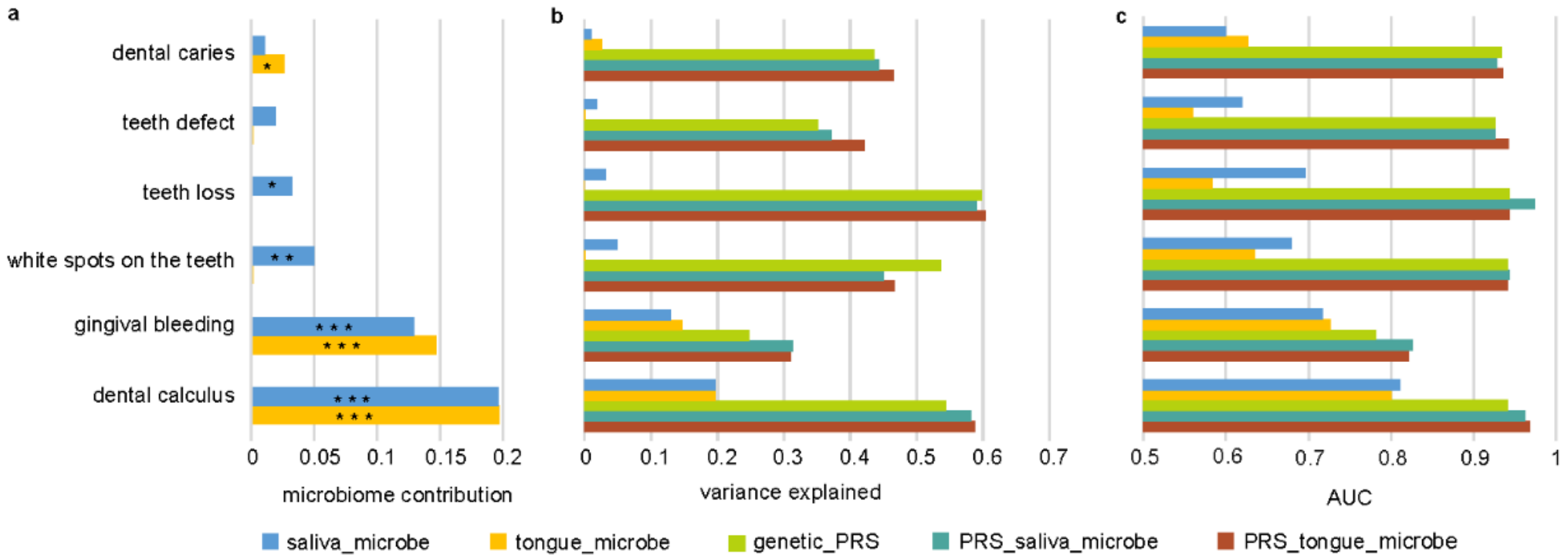
Table 3 | Summary of the 11 AD PWAS-significant genes with evidence for being consistent with a causal role in AD

			Discovery PWAS	Confirmation PWAS	Evidence for causality		TWAS significant	New gene
	Gene	Chromosome			COLOC	SMR		
1	*CTSH*	15	Significant	Replicated	Yes	Yes	Suggestive	Yes
2	*DOC2A*	16	Significant	Replicated	Yes	Yes	N/A	Yes
3	*ICA1L*	2	Significant	Replicated	Yes	Yes	No	Yes
4	*LACTB*	15	Significant	Replicated	Yes	Yes	Suggestive	No
5	*SNX32*	11	Significant	Replicated	Yes	Yes	Yes	Yes
6	*ACE*	17	Significant	Replicated	Yes	No	Yes	Yes
7	*RTFDC1*	20	Significant	Replicated	Yes	No	Suggestive	No
8	*CARHSP1*	16	Significant	Replicated	No	Yes	Yes	Yes
9	*STX6*	1	Significant	Replicated	No	Yes	Yes	Yes
10	*STX4*[a]	16	Significant	-	Yes	Yes	Yes	No
11	*PLEKHA1*[a]	10	Significant	-	Yes	Yes	N/A	Yes

[a]Proteins not found in the confirmation PWAS. N/A refers to genes that did not have significant heritability estimates to be included in the TWAS of AD. The full results for the TWAS are shown in Supplementary Tables 17 and 18. 'Suggestive' in the 'TWAS significant' column refers to genes with 0.05 < TWAS nominal $P < 0.1$. New gene refers to genes not within a 1-Mb window of SNPs with $P < 5 \times 10^{-8}$ identified in Jansen et al. AD GWAS[1].



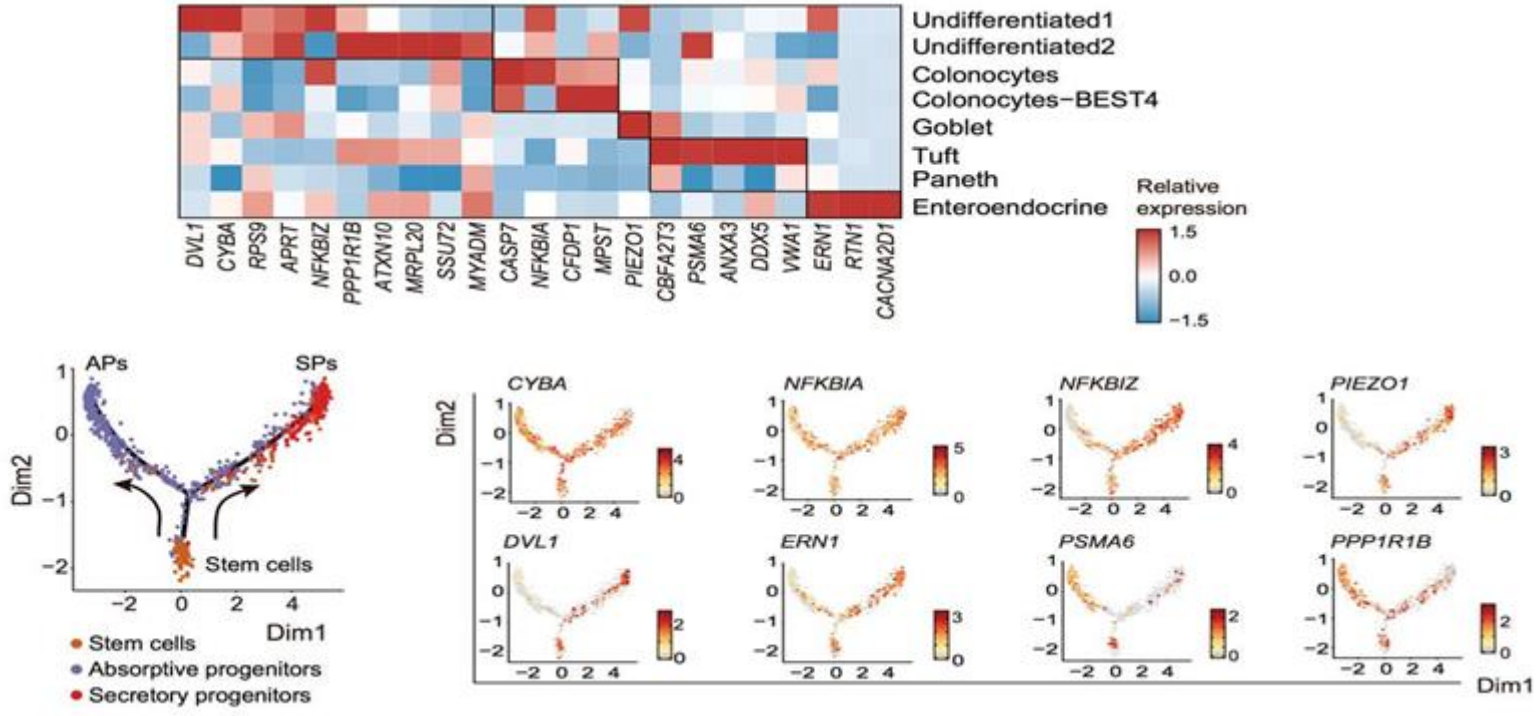
GWAS多组学联合研究——PRS联合多组学模型构建



图a：微生物对6种口腔问题的解释度；图b：微生物，基因PRS以及两者结合对6种口腔问题的解释度；图c：微生物，基因PRS以及两者结合对6种口腔问题风险的预测效能（AUC）

GWAS多组学联合研究——ScRNA-seq联合应用研究

探究遗传风险基因差异表达对疾病的促进作用



- 本研究用GWAS分析鉴定出244个PIBD风险相关的基因，在上皮效应细胞和干或祖细胞亚型中叠加候选PIBD风险基因，例如，*ERN1*与CD和UC风险相关，并介导未折叠蛋白反应，在干细胞、祖细胞和肠内分泌细胞中高度表达；*CASP7*，一种细胞消亡蛋白酶，介导感染诱导的上皮细胞死亡，在结肠细胞（ colonocytes ）和colonocytes-BEST4 细胞中表达。
- PIBD风险基因的差异表达可能导致上皮完整性受损和应激反应。

GWAS&多组学系统研究路径—纵向分子功能机制验证

多组学调控网络构建及因果效应分析

生物学功能通路挖掘及关键调控分子筛选

细胞体外及动物体内分子调控功能机制验证

GWAS&多组学系统研究路径—横向组学调控机制验证

动物模型验证性研究

研究优势：环境因素可控、组织容易获取、精准表型设置等等

研究劣势：无法准确反应人体生物学通路、疾病表型不一致等等

单细胞组学验证性研究

研究优势：在单细胞水平验证原位细胞图谱及对应的关键功能基因、调控通路等

研究劣势：无法系统验证对于细胞的分子功能调控机制

GWAS芯片&NGS技术联合研究策略

GWAS与NGS技术联合方案设计

博淼生物
BIO MIAO BIOLOGICAL
SINCE 2009

常见变异关联分析

- ✓ MAF>5%常见变异
- ✓ MAF>1%稀有变异
- ✓ GWAS芯片技术
- ✓ Plink关联分析

罕见变异关联分析

- ✓ MAF<1%罕见变异
- ✓ WGS、WES测序技术
- ✓ SKAT分析

GWAS与NGS技术联合方案设计

博淼生物 BIO MIAO BIOLOGICAL SINCE 2009

GWAS研究——家系样本

✓ 家系中各代的case样本

外显子组测序

✓ 至少两代直系遗传的case样本
✓ 家系中的control样本/1000Genomes及ESP数据作为对照

生物信息学分析

✓ Linkage analysis
✓ Consegregation analysis
✓ Mutation prediction analysis

细胞/动物模型分子功能验证

筛选功能意义的突变进行invitro及invivo分子功能机制研究

GWAS研究——散发样本

✓ 大样本GWAS检测，基于Common SNP发现易感区域及基因
✓ Meta analysis易感区域及基因

外显子组测序/靶基因捕获测序

✓ 全外显子组测序，与上述GWAS易感基因进行稀有/罕见致病突变挖掘
✓ 根据上述GWAS易感基因通过基因捕获进行NGS更深度测序，挖掘稀有/罕见致病突变

Massarray/Taqman/KASP突变验证

针对上述发现的高频、低频、罕见突变进行大人群、多中心验证

细胞/动物模型分子功能验证

筛选功能意义的突变进行invitro及invivo分子功能机制研究

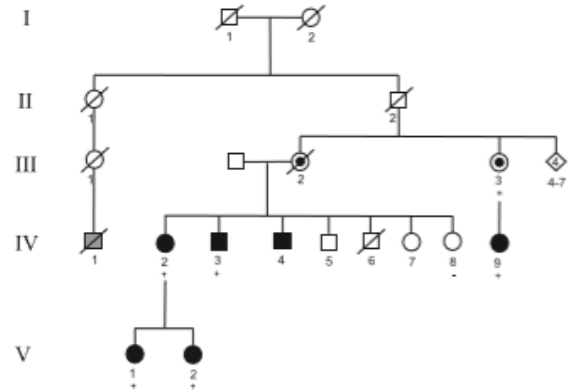
GWAS与NGS经典文章路线分享

博淼生物
BIO MIAO BIOLOGICAL
SINCE 2009

Exome Sequencing Identifies a Novel Variant in ACTC1 Associated With Familial Atrial Septal Defect

GWAS检测

- ✓ IV–2 V–1 V–2 IV–9 III–3五个case样本
- ✓ Illumina Omni 2.5M array



Exome sequence检测

- ✓ IV–2 V–1两个case样本
- ✓ Agilent SureSelect Human All Exon 50Mb

Validation 检测

- ✓ 84个散发case样本
- ✓ sanger测序技术验证上述筛选的潜在致病突变

结论

- ✓ p.M178L mutation was identified as benign by half of the mutation prediction algorithms tested. PolyPhen–2, MutationAssessor, and PROVEAN predicted that the p.M178L substitution will be benign or neutral in its effect on protein function

GWAS与NGS经典文章路线分享—重要结论展示

	p.M178L		p.M123V		p.E101K	
Algorithm	Score	Prediction	Score	Prediction	Score	Prediction
PolyPhen-2	0.035	Benign	0.011	Benign	0.788	Possibly damaging
PROVEAN	0.306	Neutral	−2.55	Deleterious	−2.919	Deleterious
MutationAssessor	−0.82	Neutral functional impact	3.785	High functional impact	2.84	Medium functional impact
SNPs&GO	7	Disease	9	Disease	9	Disease
Align-GVGD	14.3	Less likely to interfere with function	20.52	Less likely to interfere with function	56.87	More likely to interfere with function
MutationTaster	15	Disease-causing	21	Disease-causing	56	Disease-causing

Results of mutation prediction algorithms for 3 mutations in ACTC1

04

候选基因SNP研究策略

候选基因SNP方案设计

01

确定候选基因

- ✓ 既往GWAS\NGS\RNAseq筛选的关键基因
- ✓ 与疾病表型相关的生物功能基因
- ✓ GEO/TCGA公开数据库分析的关联基因
- ✓ 既往文献研究的靶向遗传易感基因

02

候选SNP位点筛选

- ✓ 1000Genomes/Hapmap数据库
- ✓ Functional Region SNP/tagSNP筛选
- ✓ Validated/Hot SNP筛选
- ✓ Function prediction analysis

03

样本分型检测

- ✓ KASP Genotype

04

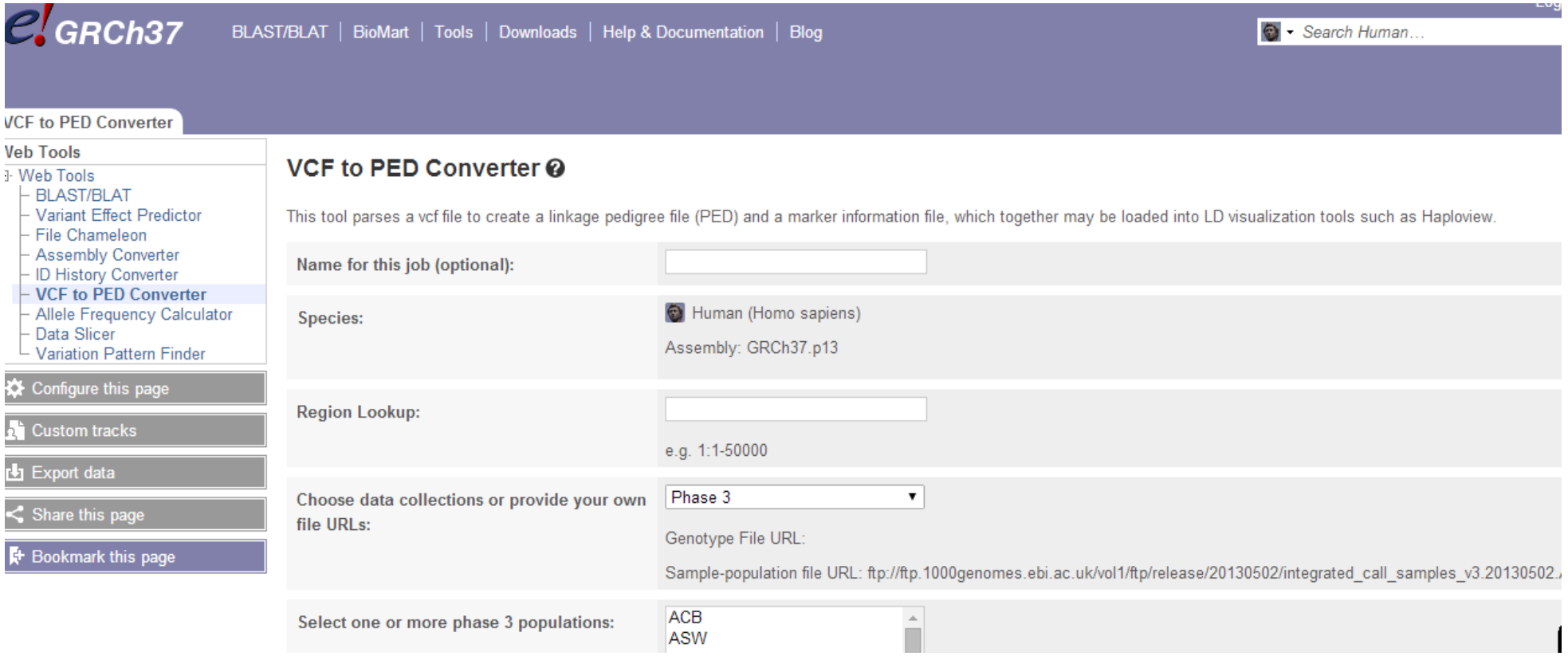
生物信息学分析

Covariates 、Various phenotype、Quantitative traits、Model based、Logistic等

候选基因SNP方案设计路线展示

博淼生物
BIO MIAO BIOLOGICAL
SINCE 2009

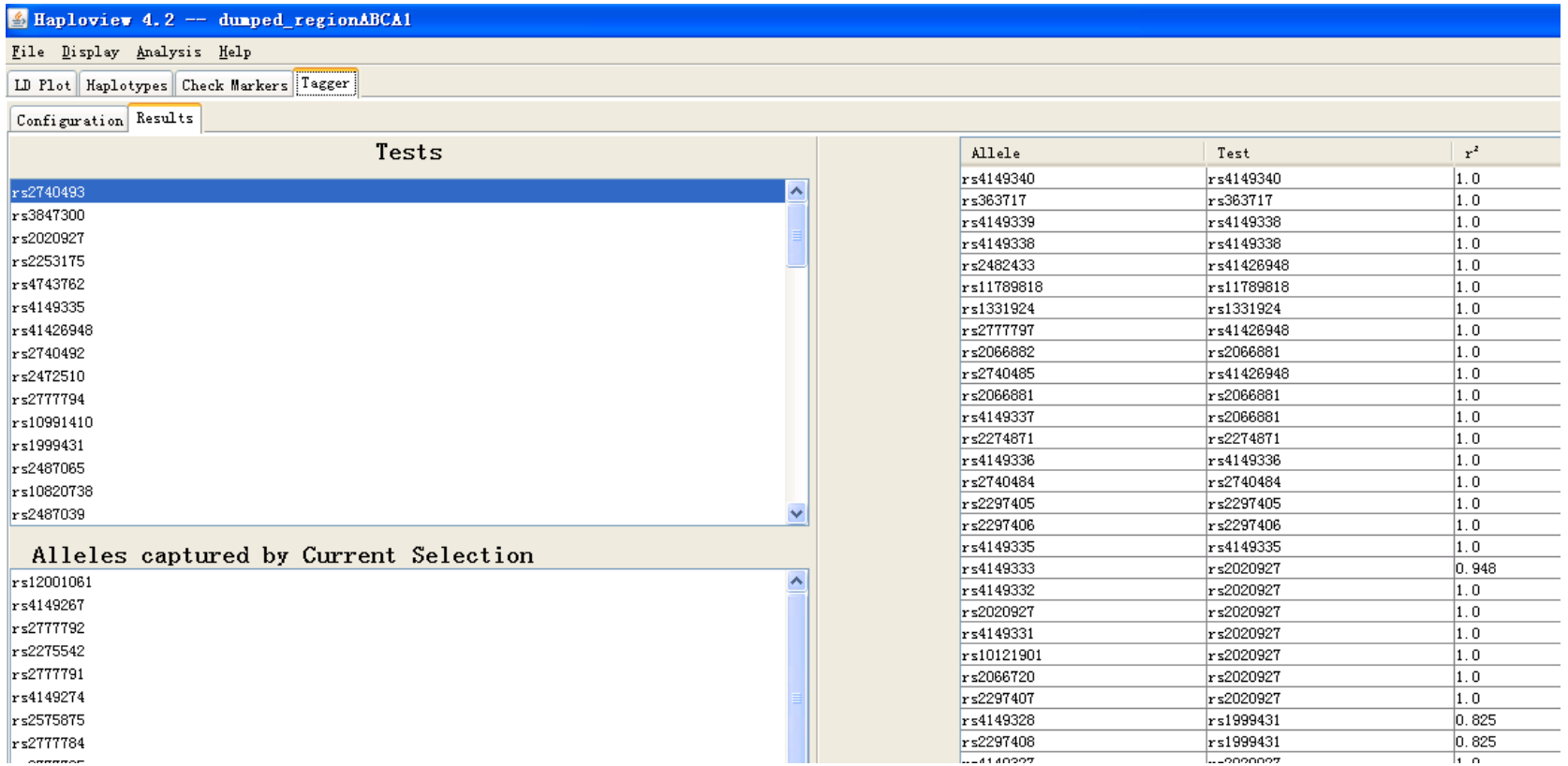
基于1000Genomes CHB/CHS数据的tagSNP筛选



候选基因SNP方案设计路线展示

博淼生物
BIO MIAO BIOLOGICAL
SINCE 2009

通过Haploview进行tagSNP分析



Allele	Test	r^2
rs4149340	rs4149340	1.0
rs363717	rs363717	1.0
rs4149339	rs4149338	1.0
rs4149338	rs4149338	1.0
rs2482433	rs41426948	1.0
rs11789818	rs11789818	1.0
rs1331924	rs1331924	1.0
rs2777797	rs41426948	1.0
rs2066882	rs2066881	1.0
rs2740485	rs41426948	1.0
rs2066881	rs2066881	1.0
rs4149337	rs2066881	1.0
rs2274871	rs2274871	1.0
rs4149336	rs4149336	1.0
rs2740484	rs2740484	1.0
rs2297405	rs2297405	1.0
rs2297406	rs2297406	1.0
rs4149335	rs4149335	1.0
rs4149333	rs2020927	0.948
rs4149332	rs2020927	1.0
rs2020927	rs2020927	1.0
rs4149331	rs2020927	1.0
rs10121901	rs2020927	1.0
rs2066720	rs2020927	1.0
rs2297407	rs2020927	1.0
rs4149328	rs1999431	0.825
rs2297408	rs1999431	0.825

候选基因SNP方案设计路线展示

博淼生物
BIO MIAO BIOLOGICAL
SINCE 2009

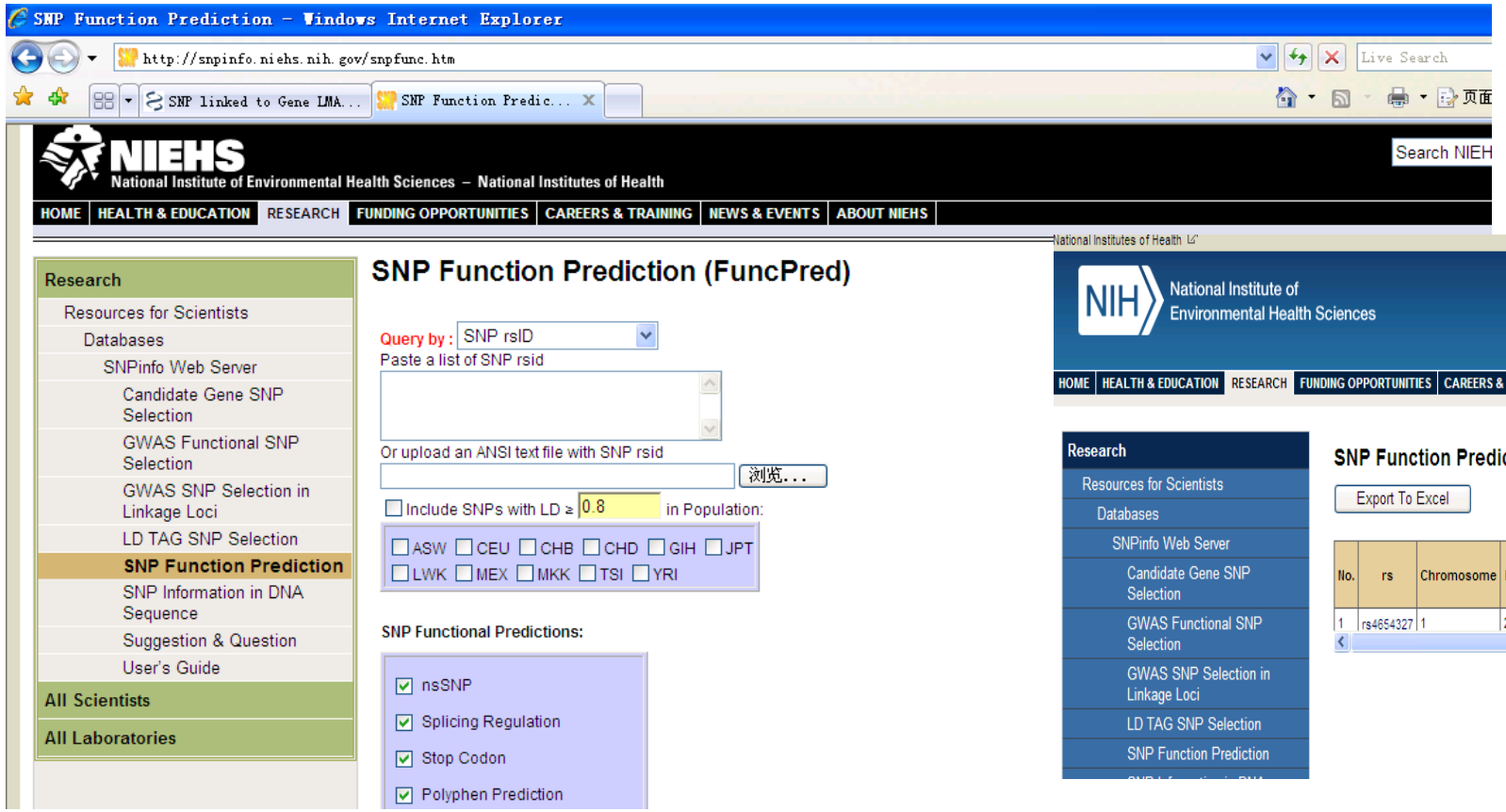
基于dbSNP数据库进行functional region SNP筛选

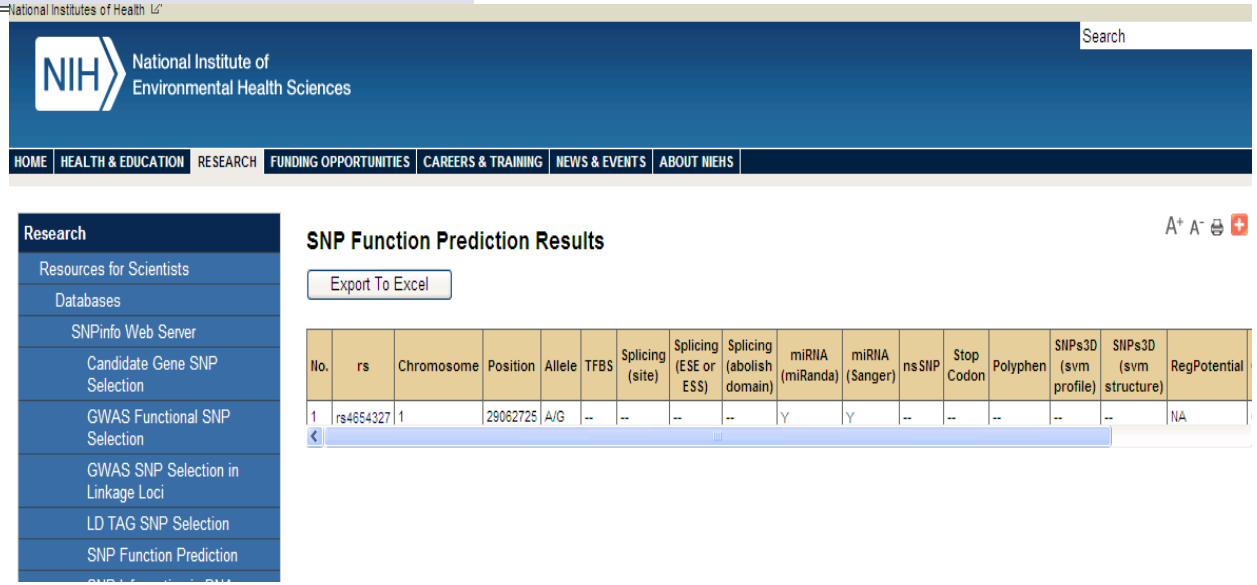
Number of variant consequences		Type	Description
0	-	Transcript ablation	A feature ablation whereby the deleted region includes a transcript feature (SO:0001893)
4	Show	Splice donor variant	A splice variant that changes the 2 base region at the 5' end of an intron (SO:0001575)
1	Show	Splice acceptor variant	A splice variant that changes the 2 base region at the 3' end of an intron (SO:0001574)
0	-	Stop gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript (SO:0001587)
0	-	Frameshift variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three (SO:0001589)
0	-	Stop lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript (SO:0001578)
0	-	Initiator codon variant	A codon variant that changes at least one base of the first codon of a transcript (SO:0001582)
0	-	Transcript amplification	A feature amplification of a region containing a transcript (SO:0001889)
0	-	Inframe insertion	An inframe non synonymous variant that inserts bases into in the coding sequence (SO:0001821)
0	-	Inframe deletion	An inframe non synonymous variant that deletes bases from the coding sequence (SO:0001822)
0	-	Missense variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved (SO:0001583)
7	Show	Splice region variant	A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron (SO:0001630)
0	-	Incomplete terminal codon variant	A sequence variant where at least one base of the final codon of an incompletely annotated transcript is changed
0	-	Synonymous variant	A sequence variant w
0	-	Stop retained variant	A sequence variant w remains (SO:0001567
0	-	Coding sequence variant	A sequence variant that changes the coding sequence (SO:0001580)
0	-	Mature miRNA variant	A transcript variant located with the sequence of the mature miRNA (SO:0001620)
0	-	5 prime UTR variant	A UTR variant of the 5' UTR (SO:0001623)
0	-	3 prime UTR variant	A UTR variant of the 3' UTR (SO:0001624)
142	Show	Non coding transcript exon variant	A sequence variant that changes non-coding exon sequence in a non-coding transcript (SO:0001792)
1228	Show	Intron variant	A transcript variant occurring within an intron (SO:0001627)
0	-	NMD transcript variant	A variant in a transcript that is the target of NMD (SO:0001621)
1375	Show	Non coding transcript variant	A transcript variant of a non coding RNA gene (SO:0001619)
451	Show	Upstream gene variant	A sequence variant located 5' of a gene (SO:0001631)
488	Show	Downstream gene variant	A sequence variant located 3' of a gene (SO:0001632)
2314	Show	ALL	All variations

候选基因SNP方案设计路线展示

博淼生物
BIO MIAO BIOLOGICAL
SINCE 2009

进行Function prediction analysis筛选





候选基因SNP方案设计路线展示

博淼生物
BIO MIAO BIOLOGICAL
SINCE 2009

通过google/pubmed进行validated/Hot SNP筛选

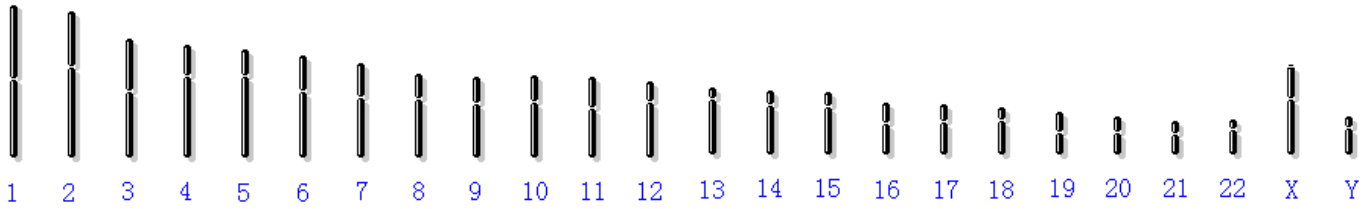
候选基因SNP方案设计路线展示

博淼生物
BIO MIAO BIOLOGICAL
SINCE 2009

基于ceRNA机制进行功能lncRNA-miRNA相关SNP筛选

lncRNASNP-human | SNP | lncRNA | miRNA | Download | Help | Contact | Go to Mouse | search | submit

In this module, we provide the information of 32,108 human lncRNA transcripts of 17,436 lncRNA genes. There are 2,717 human lncRNA transcripts of 1,543 lncRNA genes on **chromosome 1**.



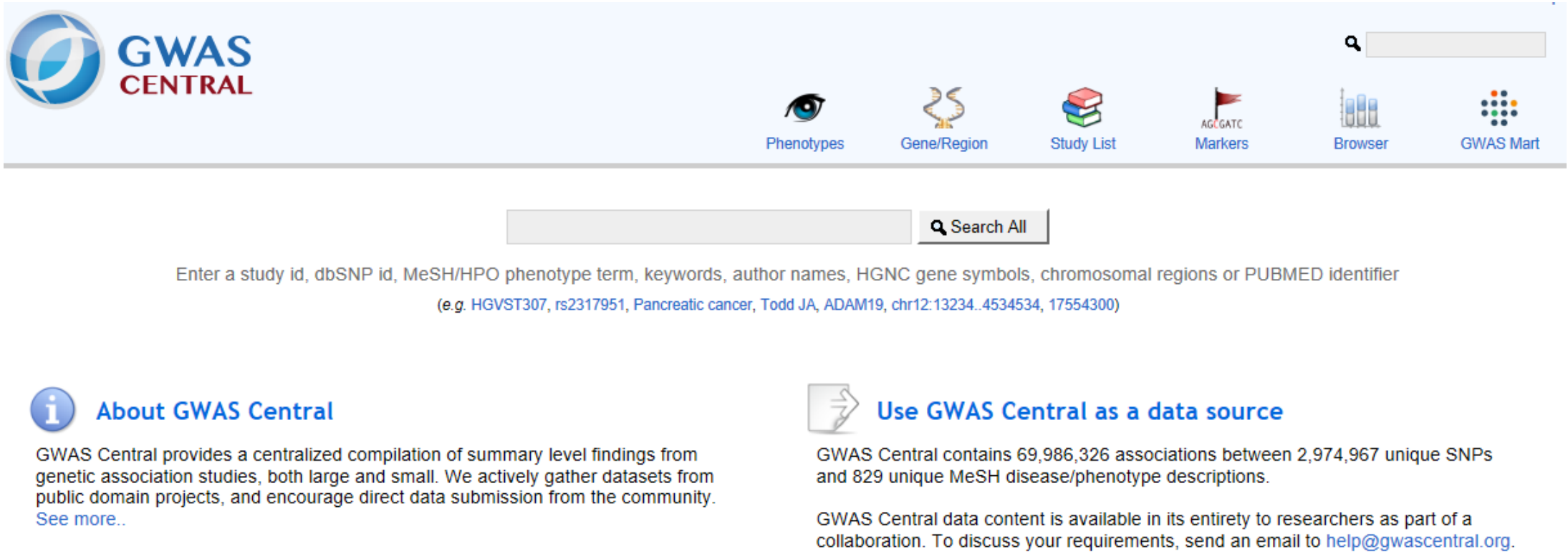
<< 1 2 3 4 5 6 7 8 9 10 ... 90 91 >>

lncRNA ID	lncRNA Gene	SNP	miRNA	Gain	Loss

候选基因SNP方案设计路线展示

博淼生物
BIO MIAO BIOLOGICAL
SINCE 2009

基于GWAS数据库筛选



候选基因SNP方案设计路线展示

博淼生物
BIO MIAO BIOLOGICAL
SINCE 2009

针对药物基因组SNP研究领域的数据库——CPIC

Genes-Drugs

CPIC assigns CPIC levels to genes/drugs with (1) PharmGKB Clinical Annotation Levels of Evidence of 1A, 1B, 2A and 2B, or (2) a PharmGKB PGx level for FDA-approved drug labels of "actionable pgx", "genetic testing recommended", or "genetic testing required", or (3) based on nomination to CPIC for consideration.

- View CPIC's process for assigning CPIC levels
- View CPIC's levels for genes/drugs
- View CPIC's process for prioritizing CPIC guidelines

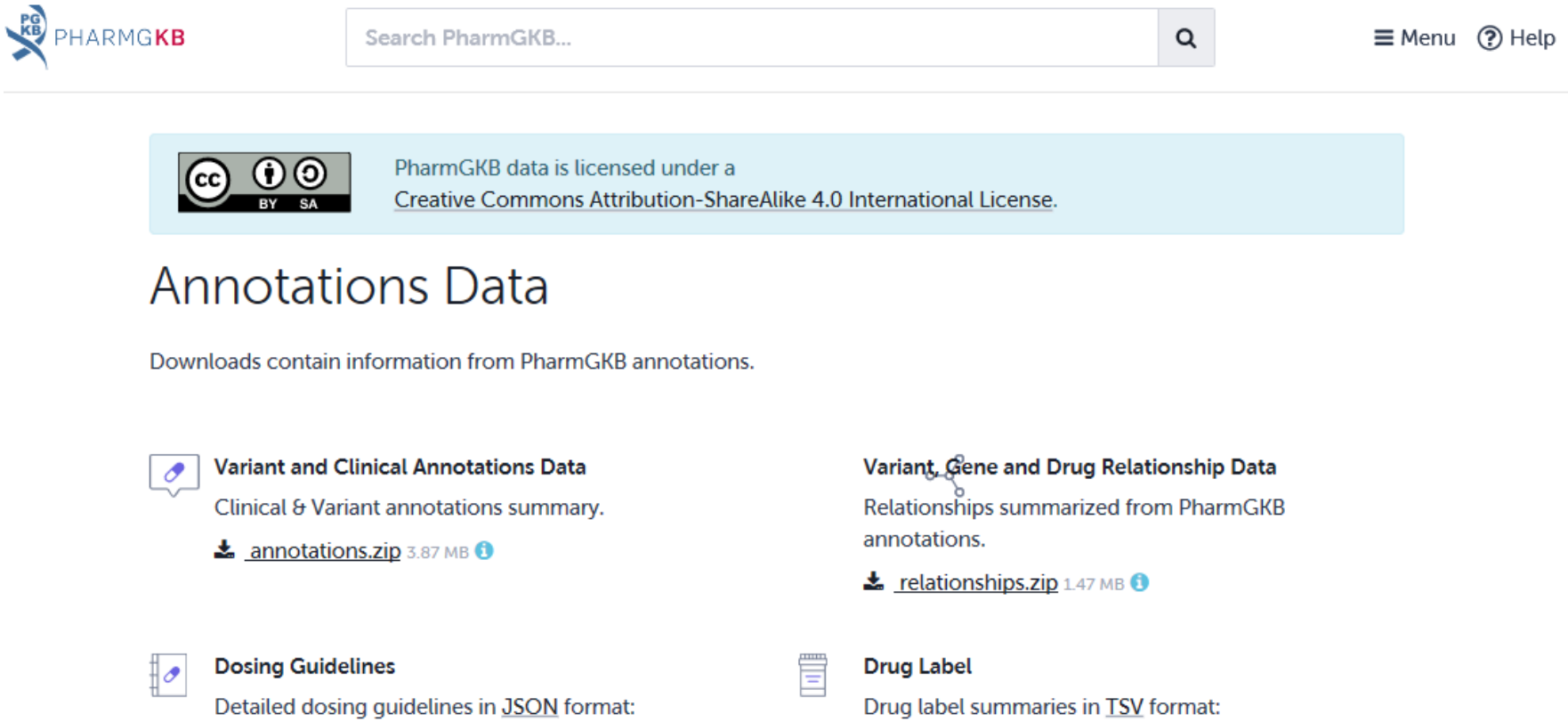
CPIC invites feedback on existing and planned gene/drug guidelines.

Download Table (CSV)

候选基因SNP方案设计路线展示

博淼生物
BIO MIAO BIOLOGICAL
SINCE 2009

针对药物基因组SNP研究领域的数据库——PharmGKB



SNP遗传调控功能注释数据库

候选SNP位点的eQTL注释数据库（外周血）：该数据库以外周血的SNP数据和RNA数据进行cis- or trans-eQTL的注释

Blood eQTL browser.

This web page accompanies the manuscript titled ***'Systematic identification of trans-eQTLs as putative drivers of known disease associations'***, by *Westra et al*, which has been published in Nature Genetics. If you want to use any of the *cis*- or *trans*-eQTL results displayed on this page in your publication, please cite this paper as indicated below. For further questions, contact the corresponding author: lude@ludesign.nl

Download eQTL Results

You can download the full *cis*- and *trans*-eQTLs, detected at a false-discovery rate of 0.50:
Cis-eQTLs (FDR 0.5)
Trans-eQTLs (FDR 0.5)

How to cite

If you use the eQTLs present on this website in your paper or research, please cite our work: Download citation directly from Nature Genetics

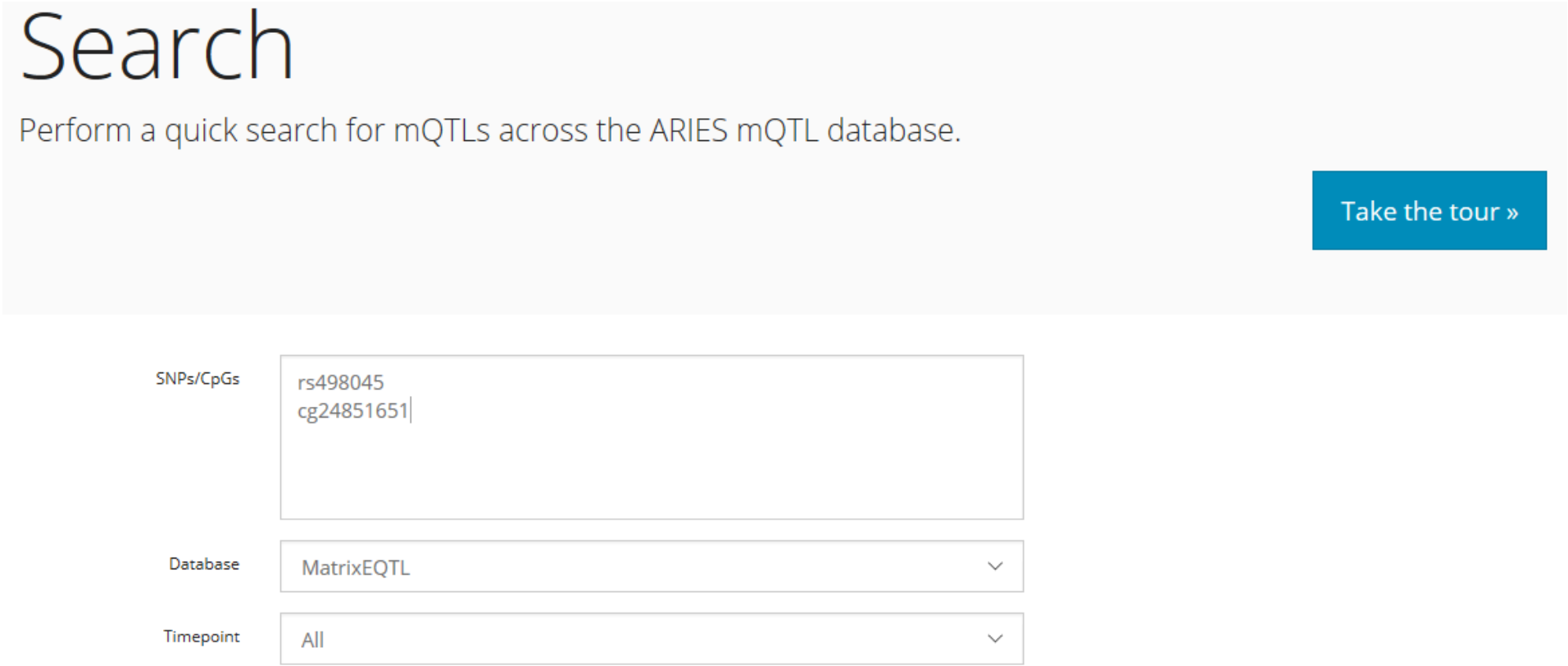
Query eQTL Results

Or, you can query the *cis*- and *trans*-eQTLs below (examples: rs7807018 or *VWCE*):

Gene or SNP name: Search

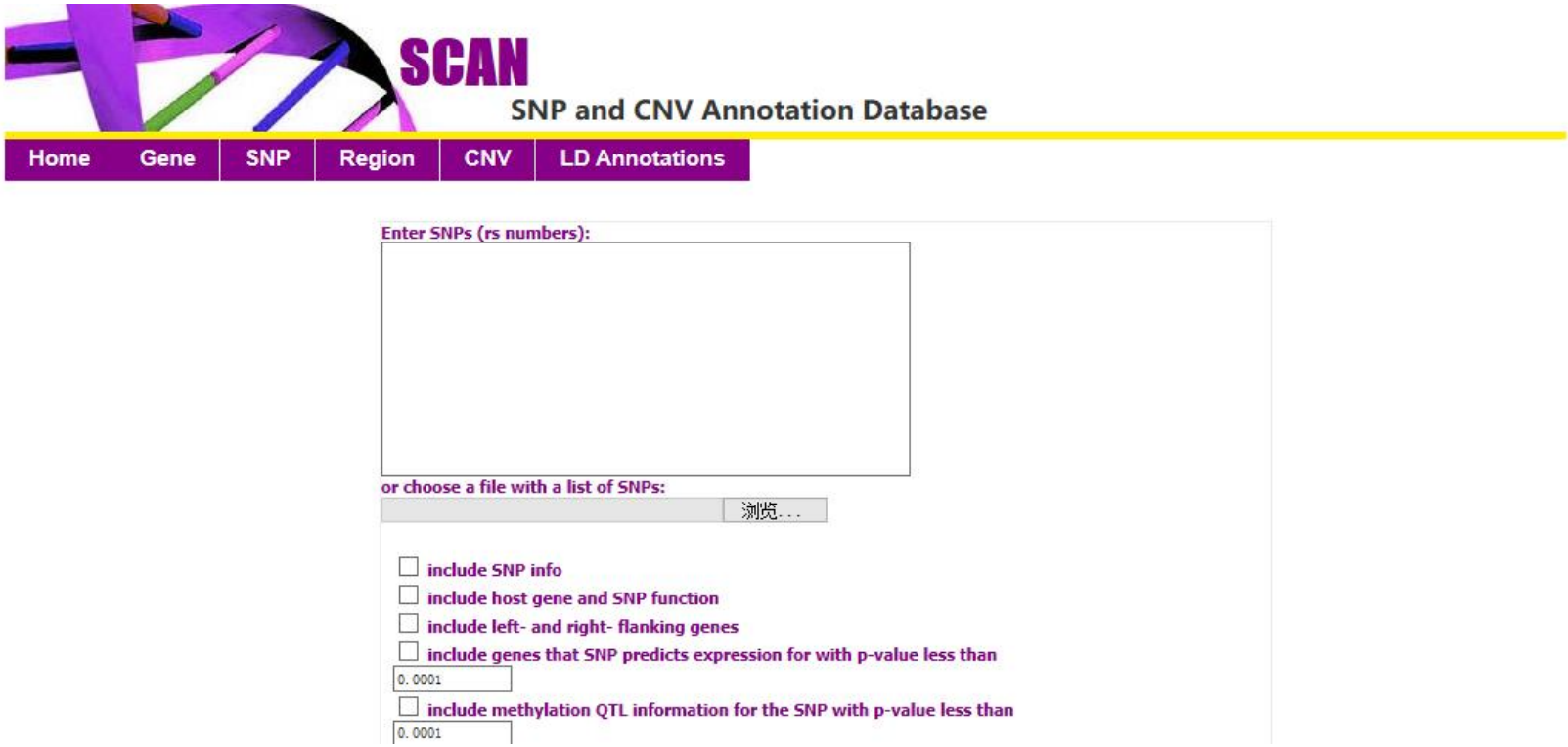
SNP遗传调控功能注释数据库

候选SNP位点的mQTL注释数据库：该数据库针对关注的SNP位点及甲基化位点，在五个生命周期阶段通过ARIES mQTL database进行mQTL注释。



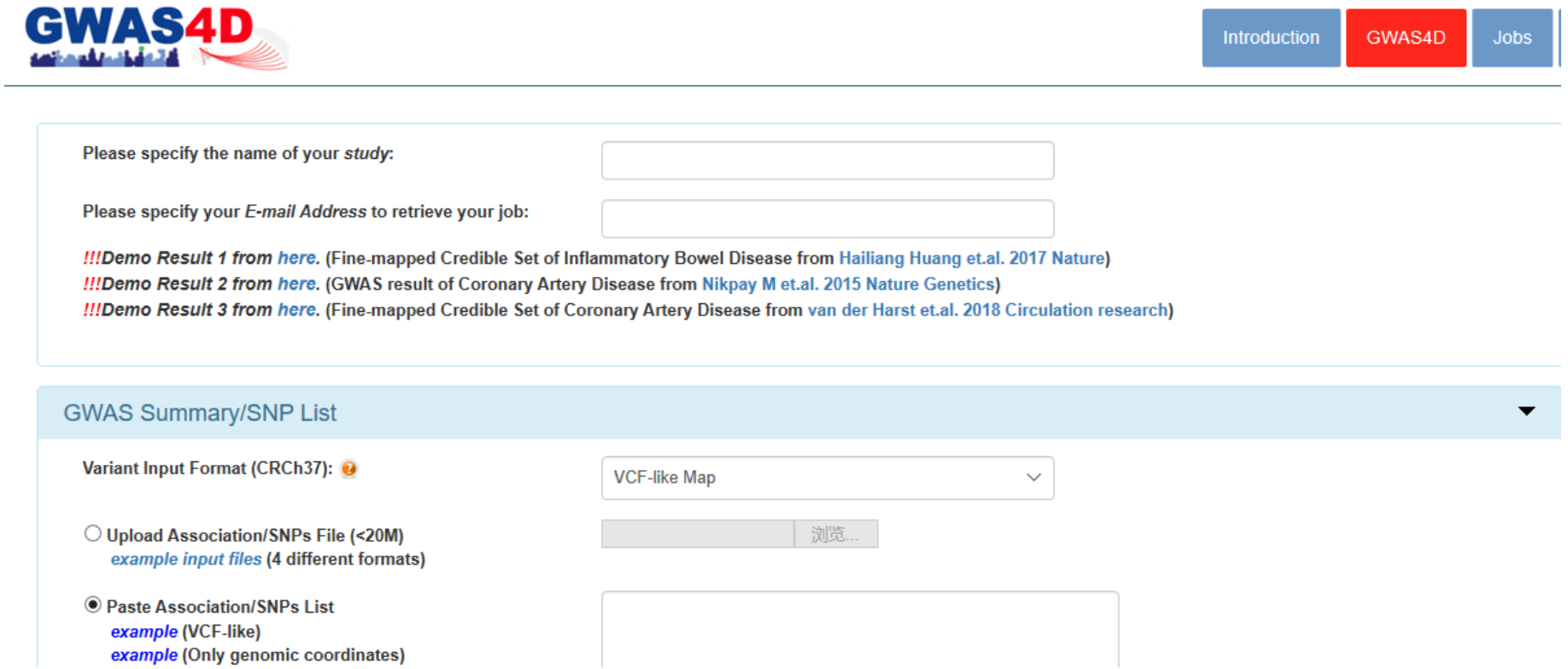
SNP遗传调控功能注释数据库

候选SNP位点的eQTL、mQTL注释数据库：该数据库以GTEx为源数据库，可以针对SNP位点或者基因进行不同组织中的eQTL及mQTL注释。



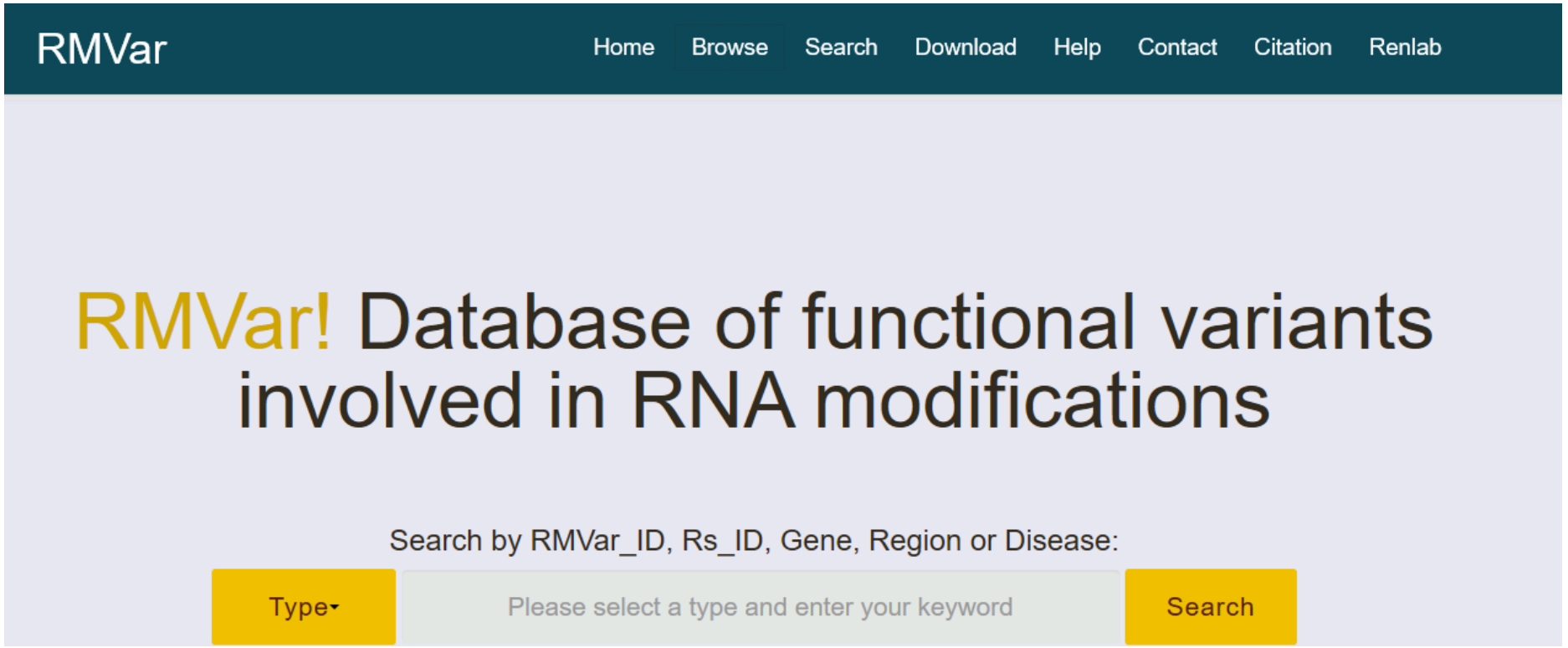
SNP遗传调控功能注释数据库

候选SNP位点的综合调控注释数据库：该数据库以TCGA为源数据，包括127中不同组织、细胞类型，及不同的疾病类型，进行TFBS、eQTL、HiC等丰富的遗传调控注释。



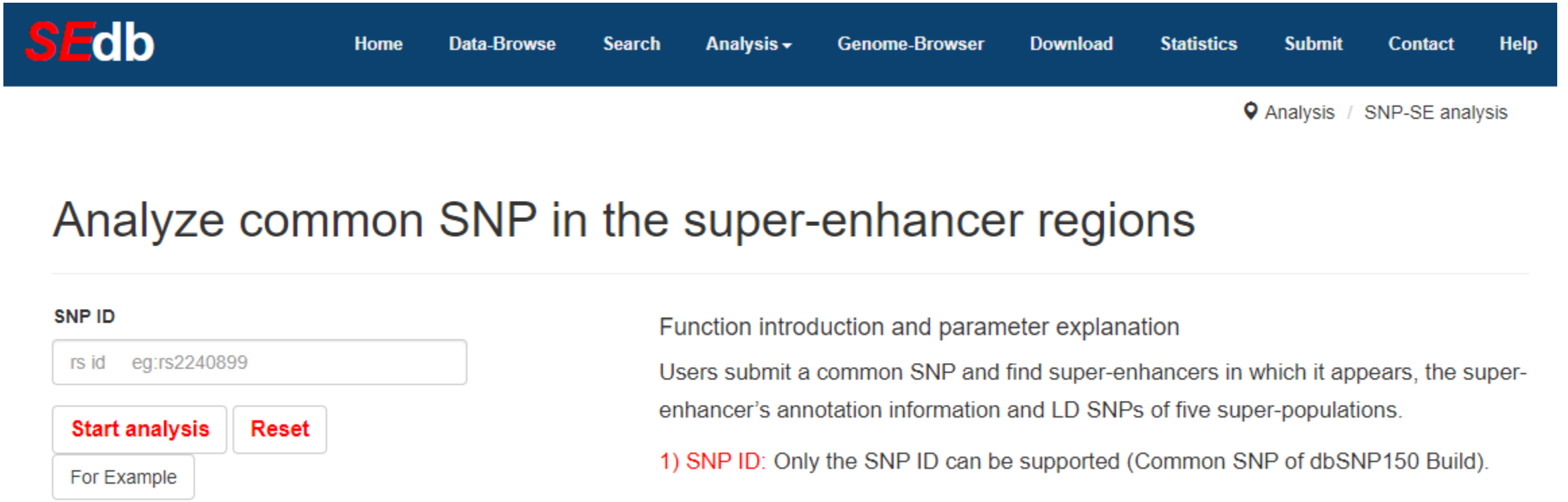
SNP遗传调控功能注释数据库

候选SNP位点的m6A注释数据库：通过该数据库可以覆盖如下不同层面的调控注释，m6A–associated genetic mutations (dbSNP)、m6A–associated cancer somatic mutations、Disease related m6A–associated variants、Splicing sites affected by m6A–associated variants、RNA binding protein affected by m6A–associated variants、miRNA targeting and processing affected by m6A–associated variants



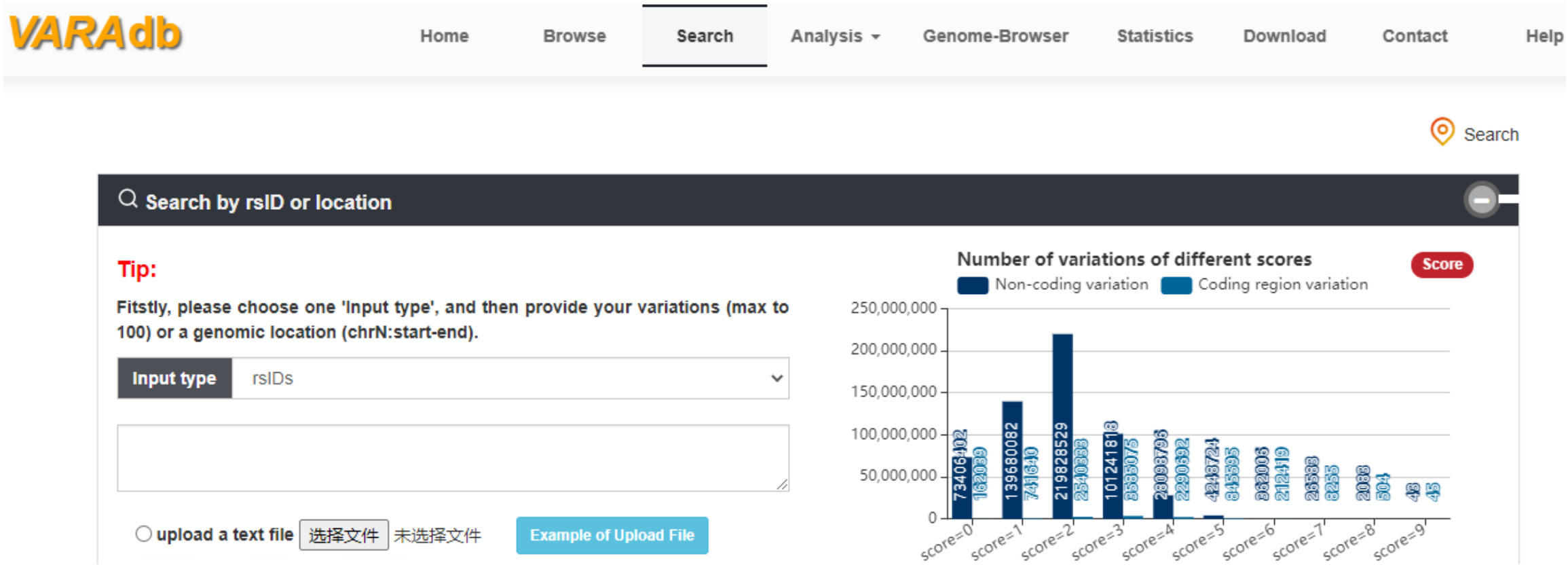
SNP遗传调控功能注释数据库

超强增强子遗传调控SNP位点注释：针对超级增强子区域注释常见的SNP，eQTL，风险SNP，TFBS，CRISPR/Cas9靶位点，DHS和增强子。



SNP遗传调控功能注释数据库

SNP位点综合调控功能注释：The information includes motif changes, risk SNPs, LD SNPs, eQTLs, clinical variant–drug–gene pairs, sequence conservation, somatic mutations, enhancers, super enhancers, promoters, TFs, ChromHMM states, histone modifications, ATAC accessible regions and chromatin interactions from Hi–C and ChIA–PET。



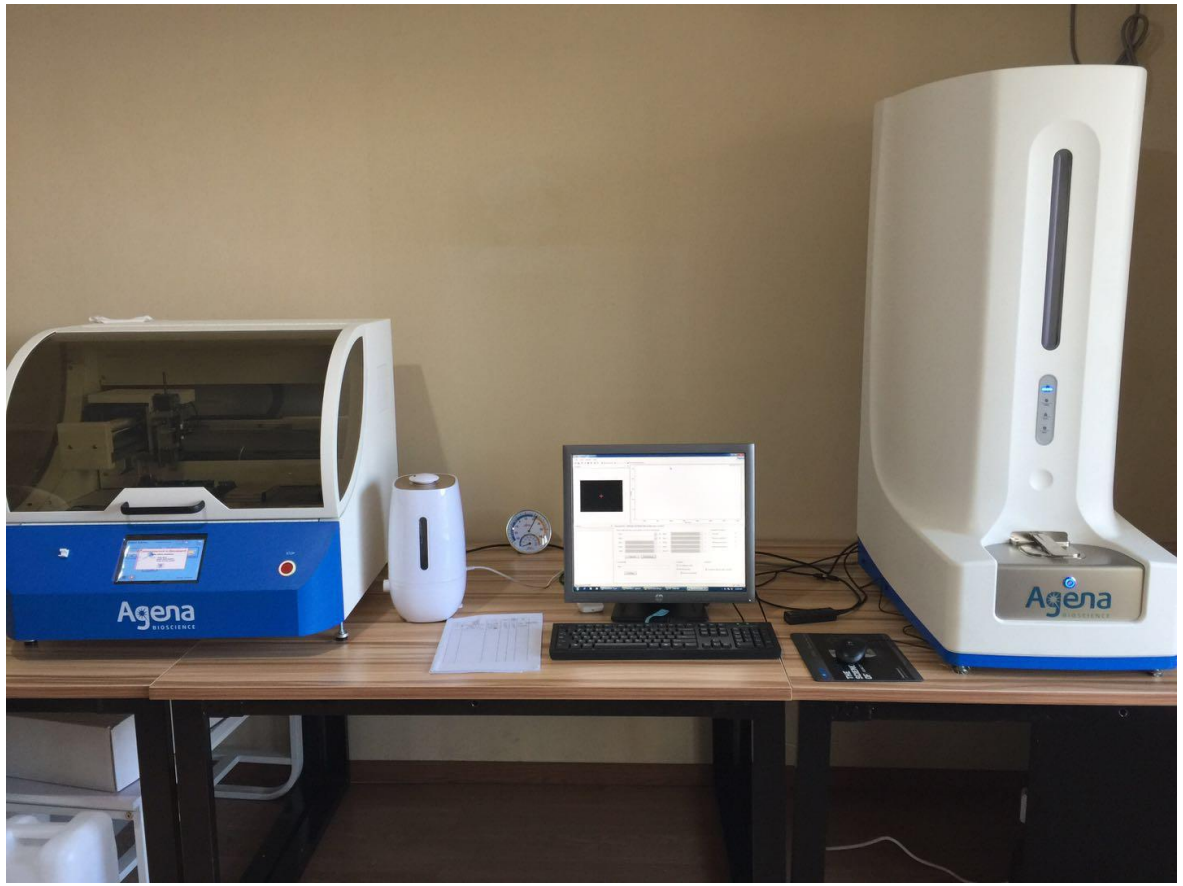
05

SNP分型主流技术

GWAS检测技术——iScan系统

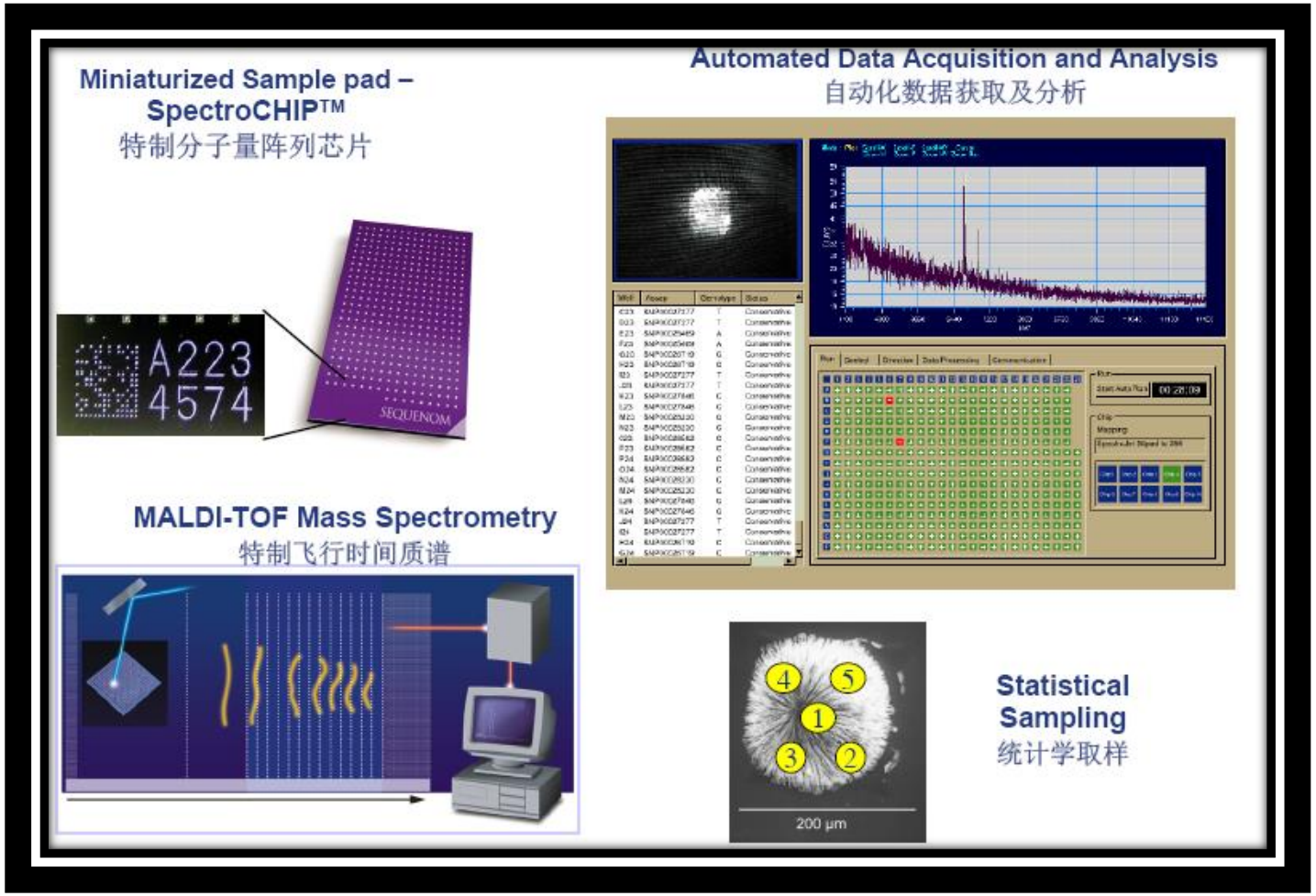
- GWAS芯片
  - 中华芯片
  - ASA芯片
  - CGA芯片
- 甲基化芯片
  - 850K芯片

靶向SNP分型检测技术——Massarray系统



- 候选SNP位点分型
- 候选甲基化位点定量检测

Massarray技术原理



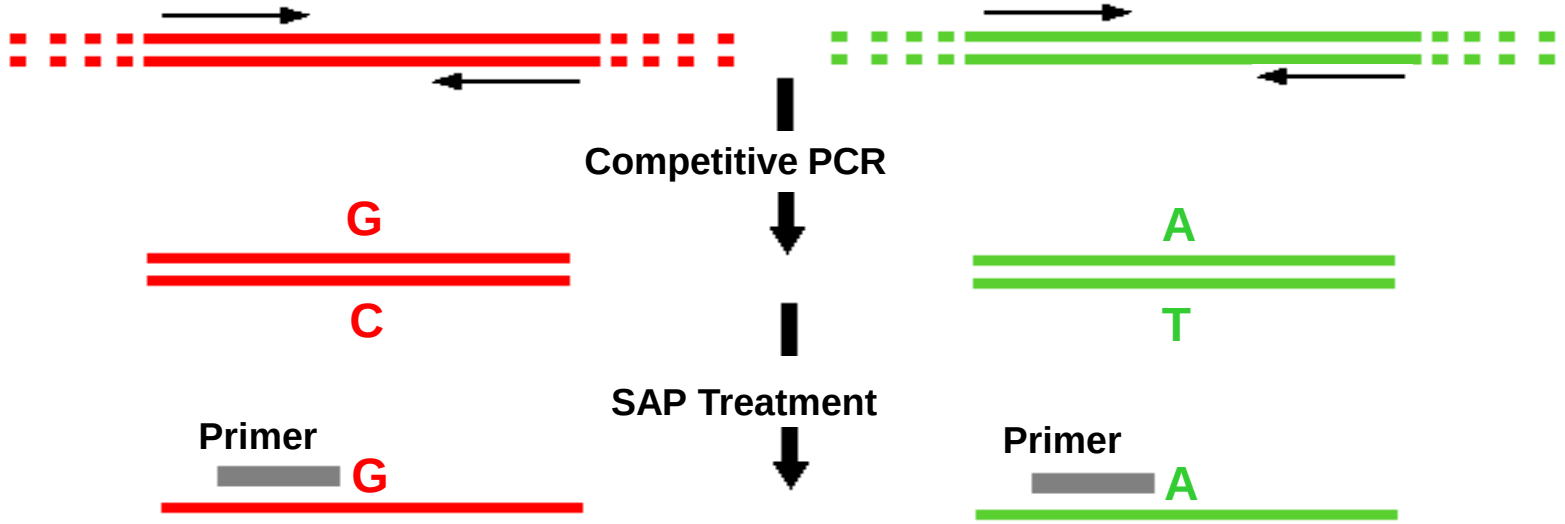
Massarray技术原理——SNP分型

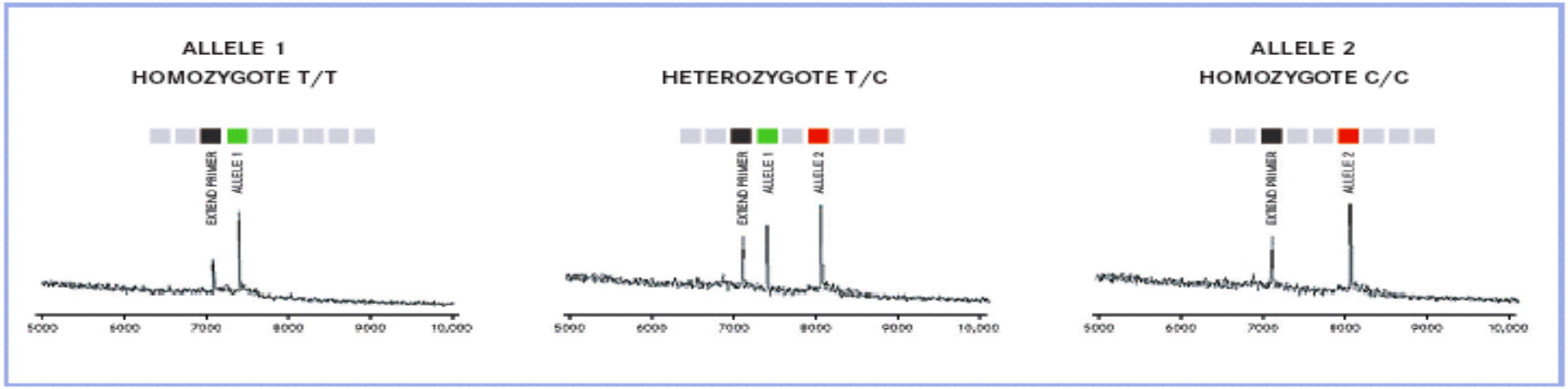
博淼生物 BIO MIAO BIOLOGICAL SINCE 2009

Genomic DNA

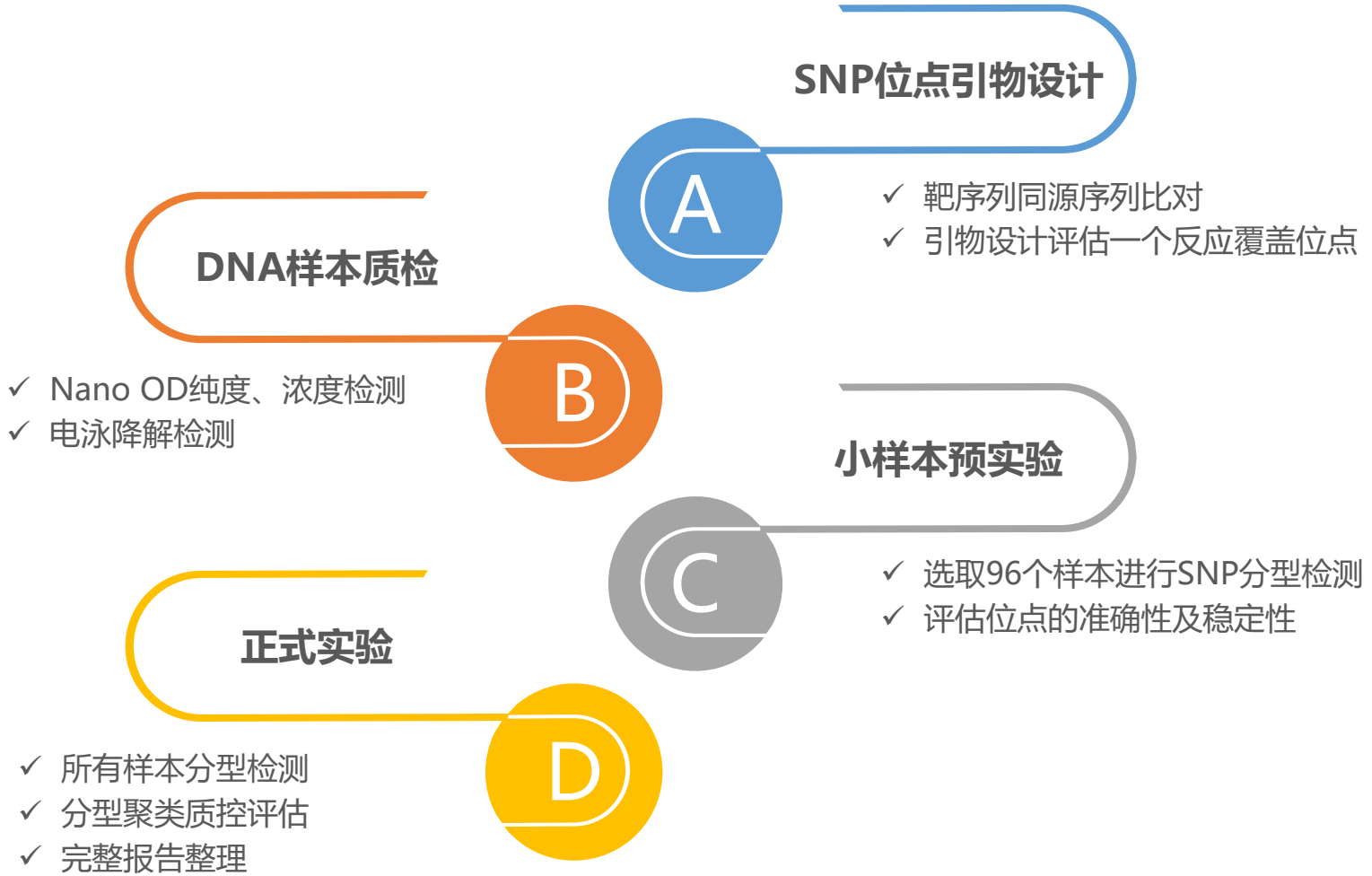
PCR Amplification

Primer Extension

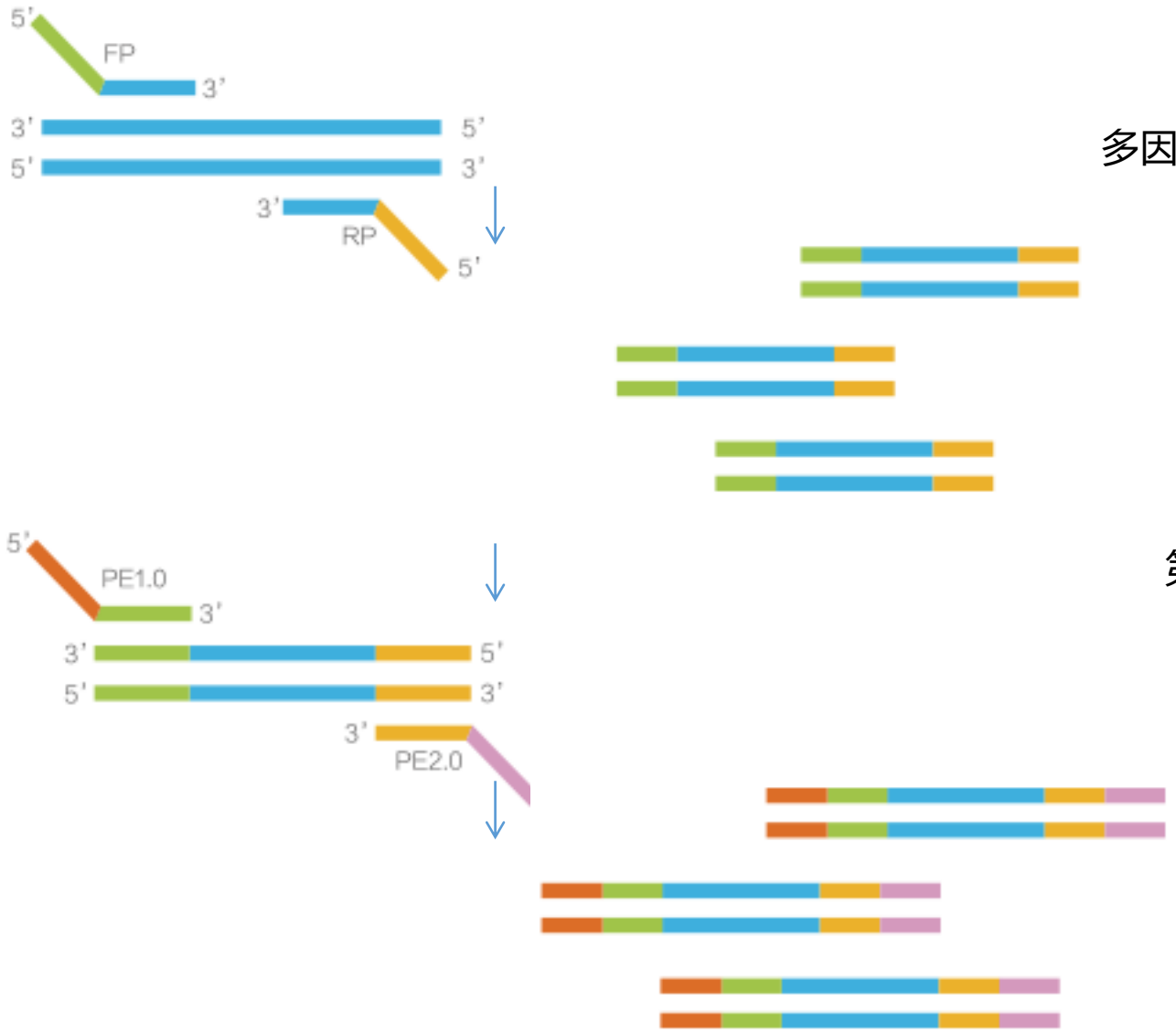




Massarray SNP分型 技术服务流程



Multi-PCR NGS SNP分型 技术原理



多因素算法设计多重PCR引物

↓

第1轮多重PCR反应

↓

磁珠纯化合并产物

↓

第2轮接头序列PCR反应

↓

磁珠纯化

↓

文库定量和质检

↓

NGS测序

Multi-PCR NGS SNP分型 技术特点

特点一：高通量验证技术，适用>300个SNP位点的分型

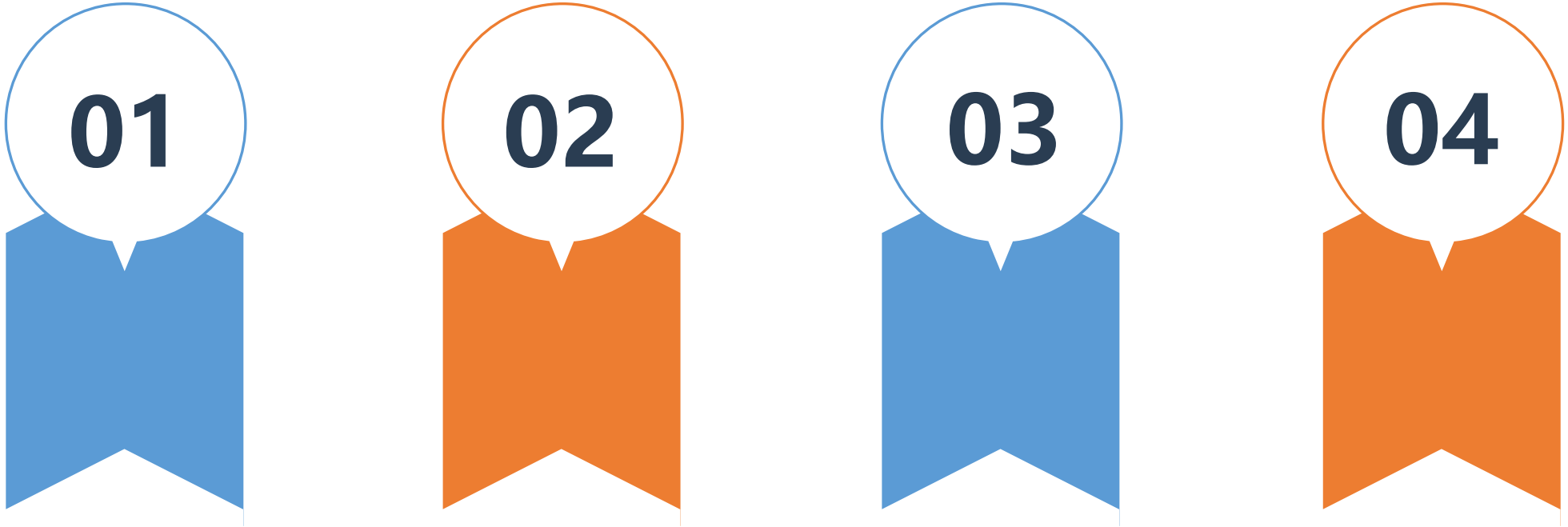
特点二：超高测序深度，平均测序深度>1000X

特点三：检测快速，价格低廉

特点四：多重PCR技术与NGS技术的理想搭配

Multi-PCR NGS SNP分型技术服务流程

博淼生物
BIO MIAO BIOLOGICAL
SINCE 2009



SNP位点引物设计
- ✓ 高度同源序列比对
- ✓ 多重PCR引物设计优化

DNA样本质检
- ✓ Nano OD浓度、纯度检测
- ✓ 电泳样本降解检测

预实验评估
- ✓ 48个样本预实验评估PCR扩增
- ✓ NGS测序的偏向性

正式实验
- ✓ 所有样本正式实验检测
- ✓ 完整实验报告

博淼技术服务项目体系

博淼生物
BIO MIAO BIOLOGICAL
SINCE 2009

基因组学服务

- GWAS芯片/WES-seq/Target NGS-seq
- Massarray /Multi-PCR NGS/Taqman /KASP SNP分型
- 16S扩增子测序/宏基因组测序
- HLA-seq/TCR&BCR-seq

表观基因组学服务

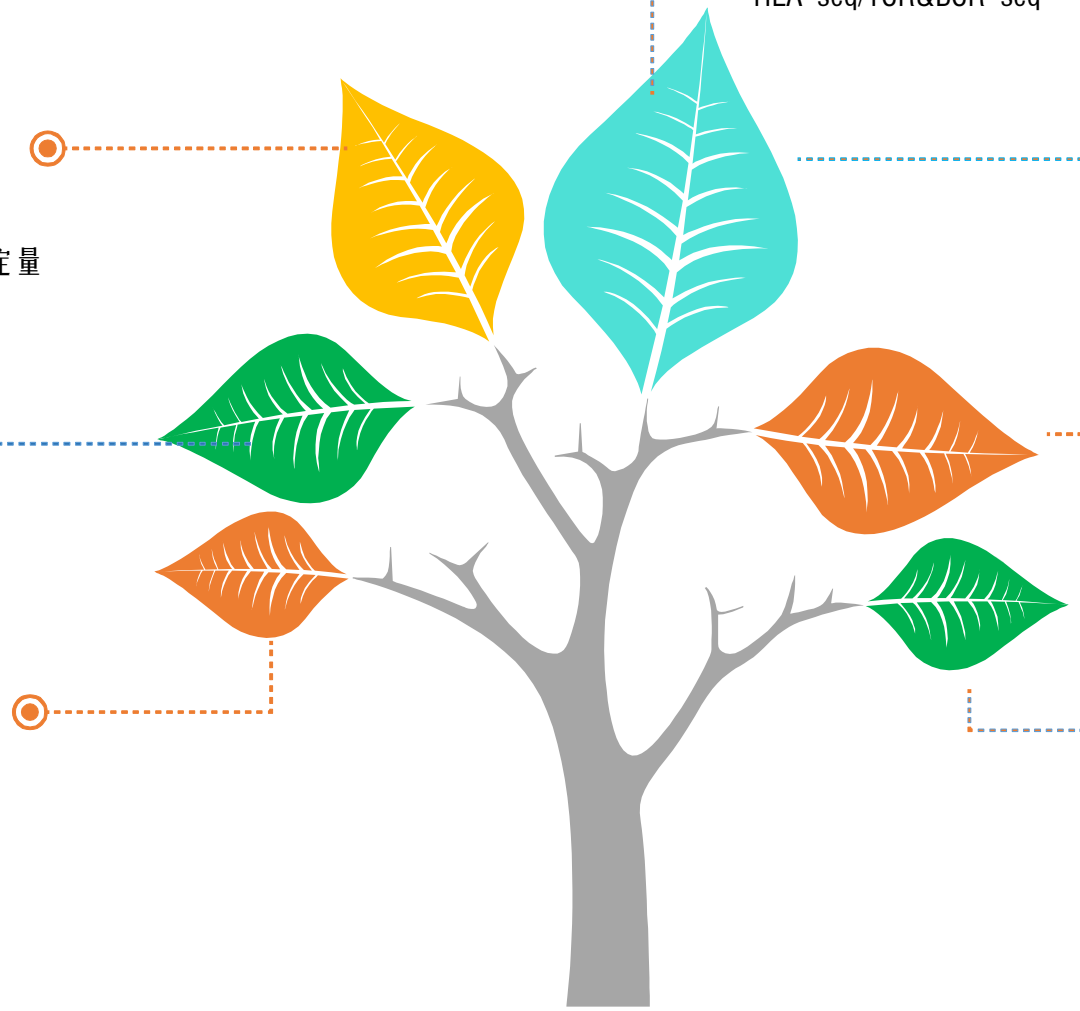
- EWAS芯片
- Multi-PCR NGS靶向DNA甲基化定量
- Massarray 靶向DNA甲基化定量

转录组学服务

- 转录组NGS/表达谱芯片
- RT-qPCR靶向转录本定量

单细胞组学服务

- 单细胞转录组测序
- 单细胞免疫组库测序
- 单细胞ATAC测序&转录组测序
- 空间转录组测序



代谢组学服务

- 非靶向代谢组
- 非靶向脂质组
- 靶向代谢组项目系列

蛋白质组学服务

- DIA定量蛋白质组/iTRAQ&TMT定量蛋白质组/Label free定量蛋白质组
- PRM靶向蛋白/Olink蛋白定量/ELISA
- 修饰蛋白质组
- 高密度自身抗体蛋白芯片

多组学联合研究服务

- GWAS&多组学技术服务
- EWAS&多组学技术服务
- 微生物基因组&代谢组技术服务
- 蛋白质组&代谢组技术服务

感谢各位的聆听

Your own Laboratory

——您的专属实验室