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

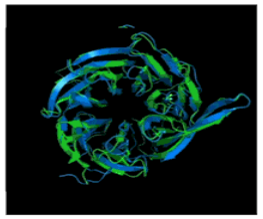
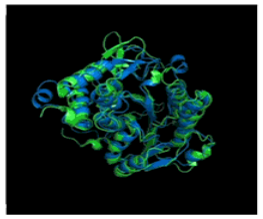
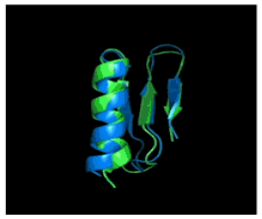
Proteomics-first approach在队列多组学研究中的应用体系

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

Proteomics-first approach研究体系

Multi-omics队列研究意义

生物世界

✓ 基因组、蛋白质组、代谢组…

✓ 单细胞组、空间组学

✓ 微生物组

因果调控功能机制

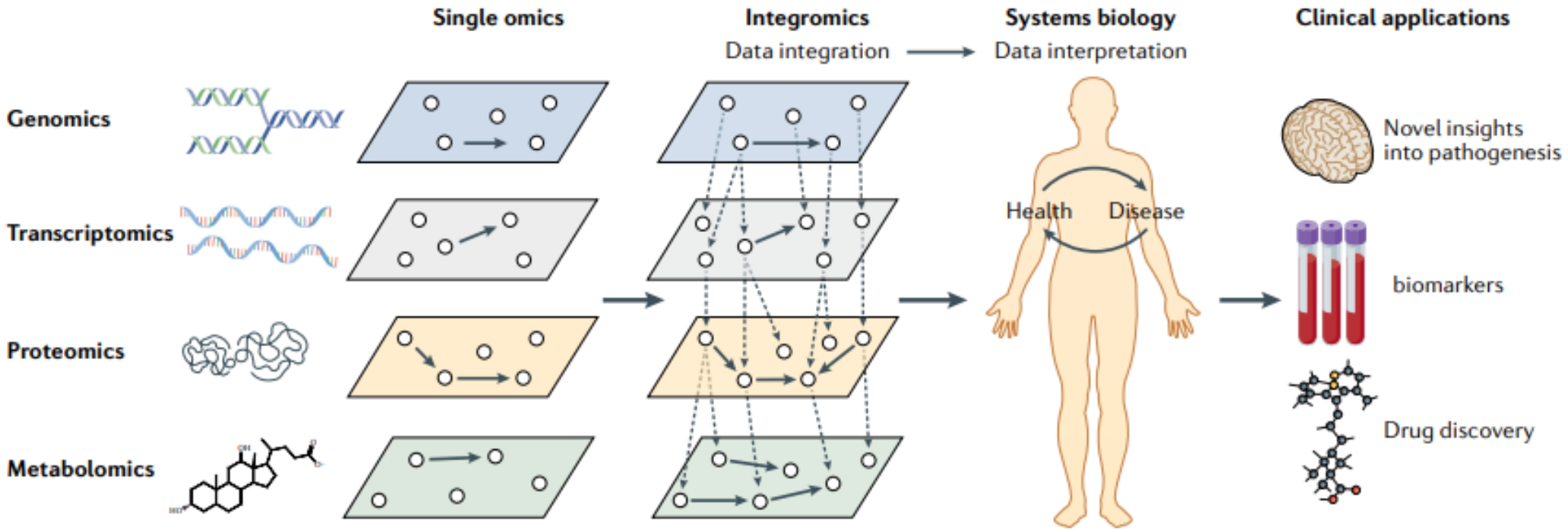
真实世界

✓ 环境暴露组

✓ 结局表型组

✓ 影像组…

Multi-omics队列研究意义



Multi-omics roles

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）

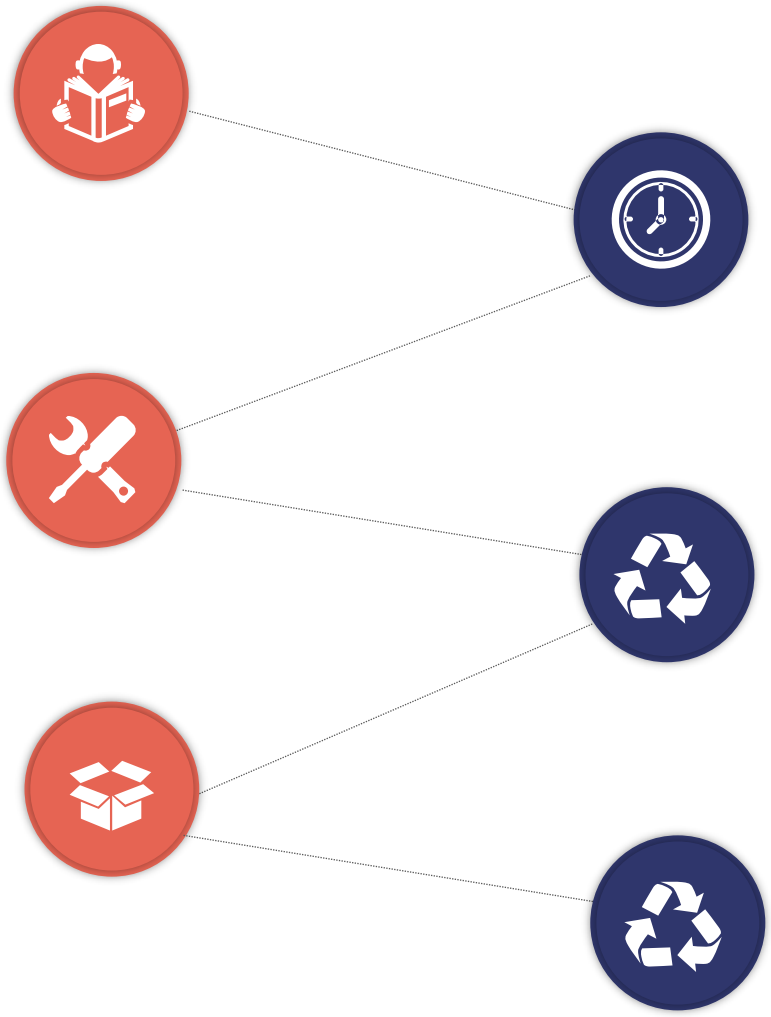
—What appears to be happening

- ✓ 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

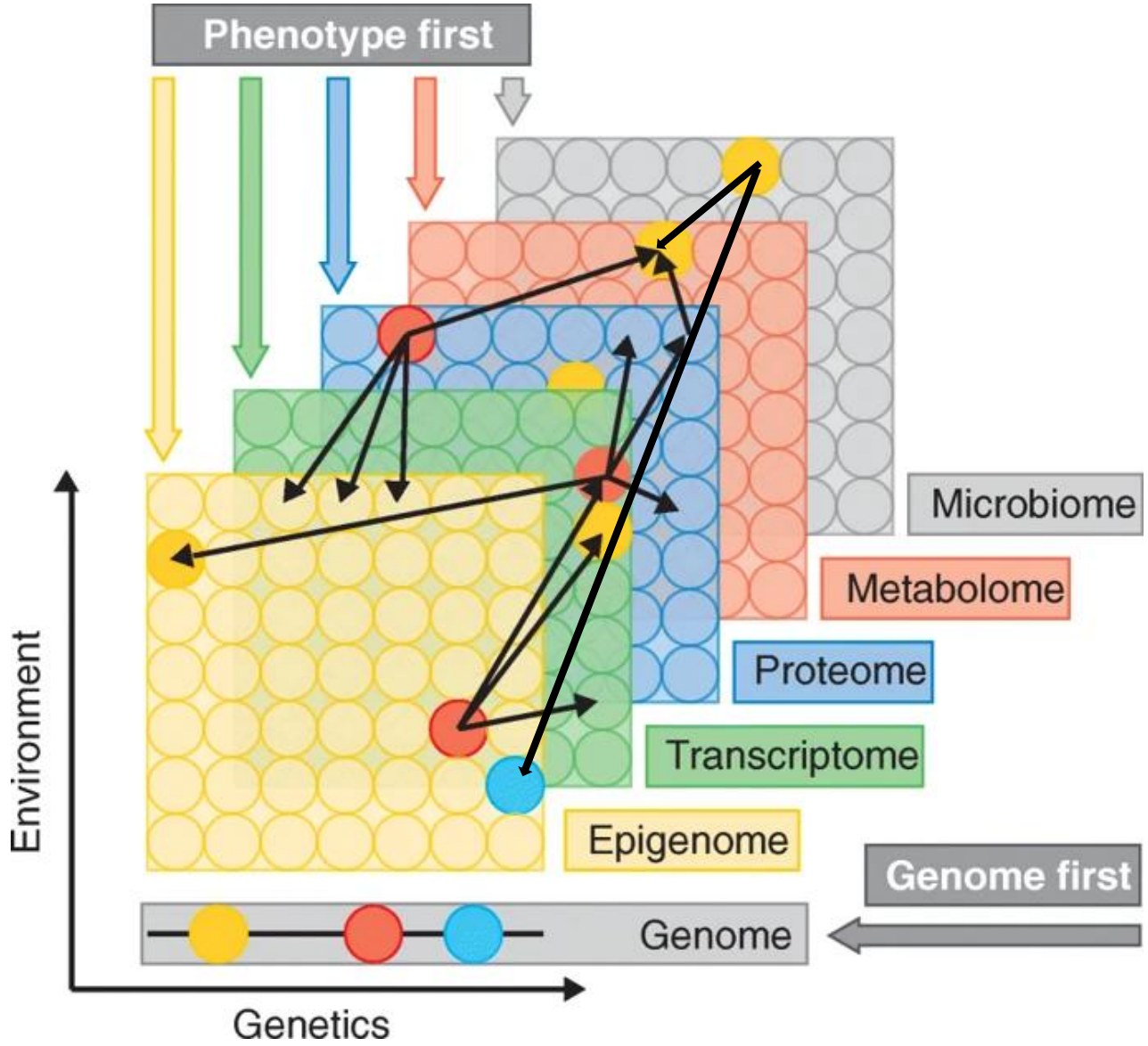
- ✓ 功能蛋白质：酶、转录因子等
- ✓ 修饰蛋白质：糖基化、乙酰化等

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 approaches to disease research



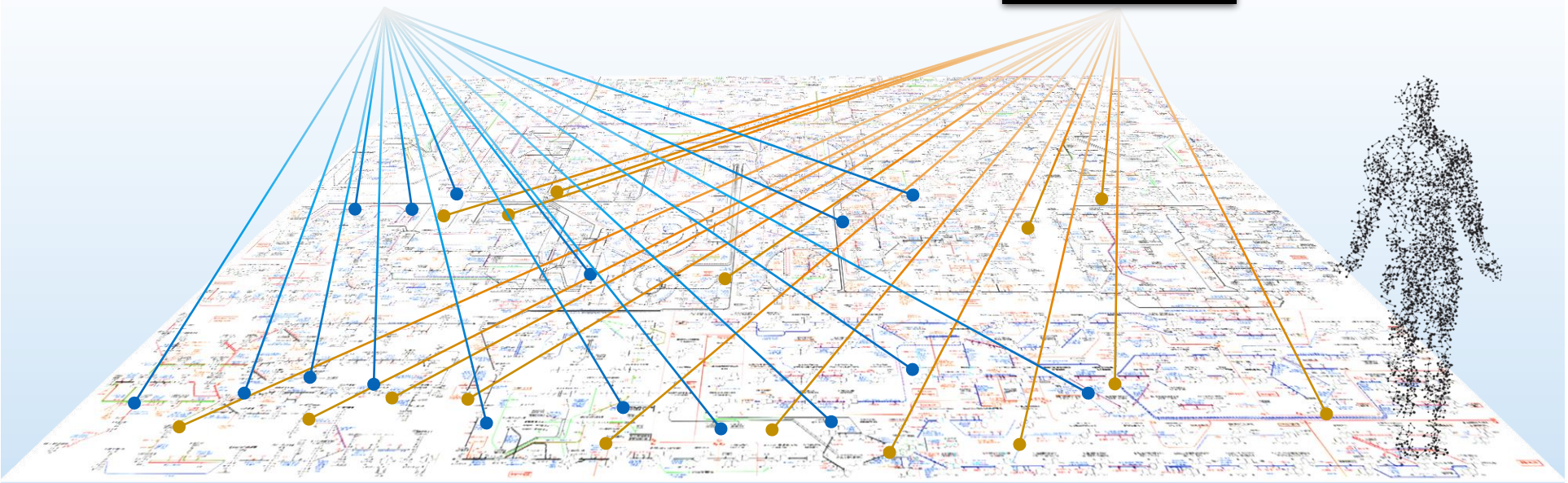
The Proteomics first approach —功能分子表型数据

遗传因素+环境暴露

↓

*Molecular Phenomics*分子表型组

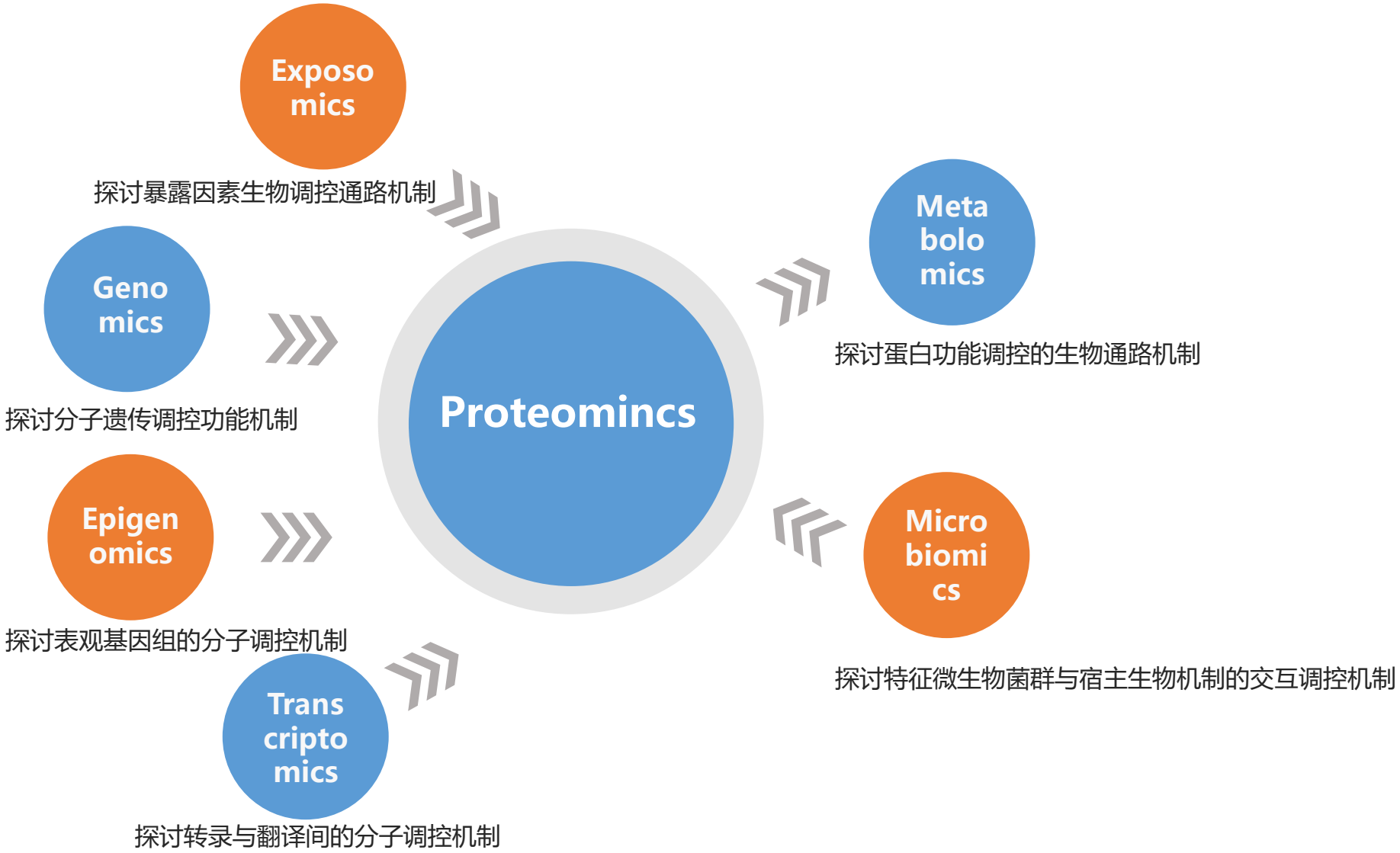
Metabolomics ***Proteomics***



↓

Marker panel, personalized medicine, bio-insight→ better health

The Proteomics first approach —功能分子表型数据



Multi-omics trajectory to disease research

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

Proteomics-first多组学研究策略解析

Proteomics&GWAS approach研究策略

研究目标

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

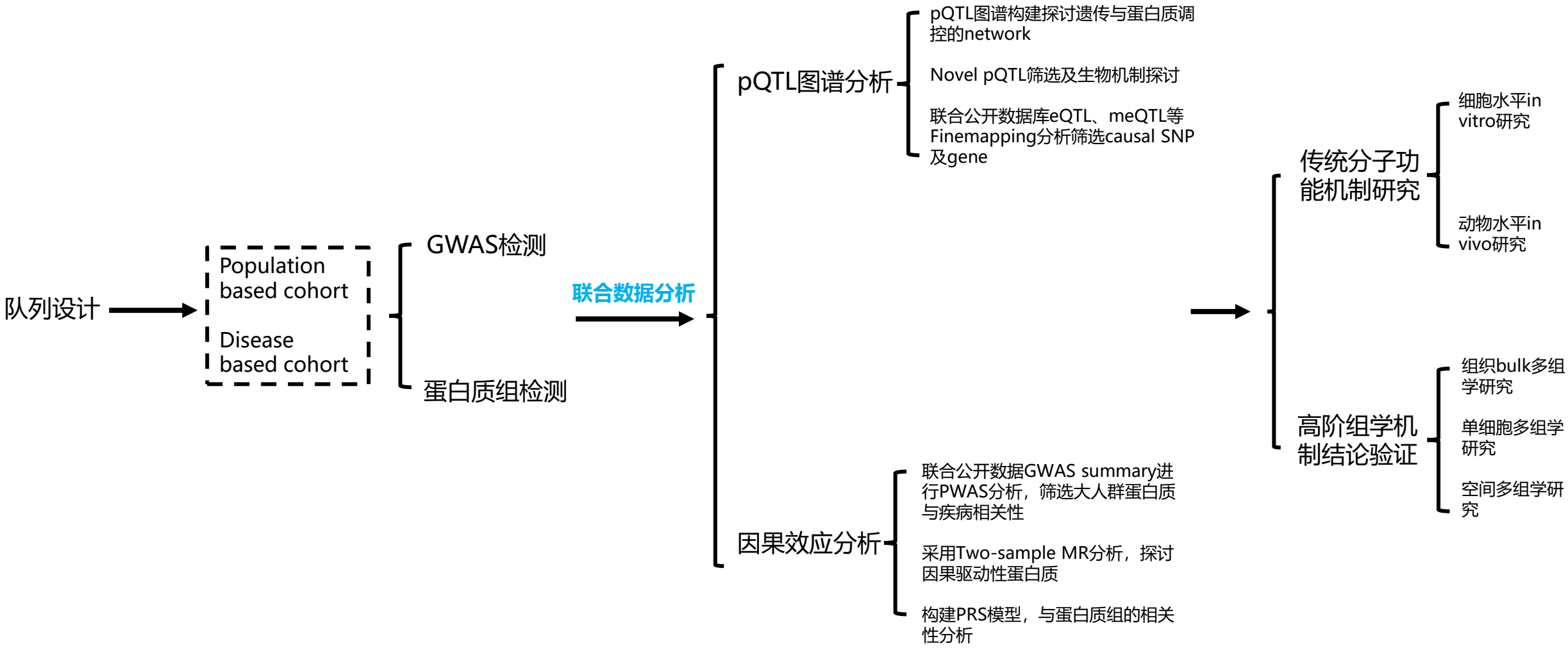
✓对比公开数据库的pQTL,筛选novel pQTL及对应的潜在生物学机制

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

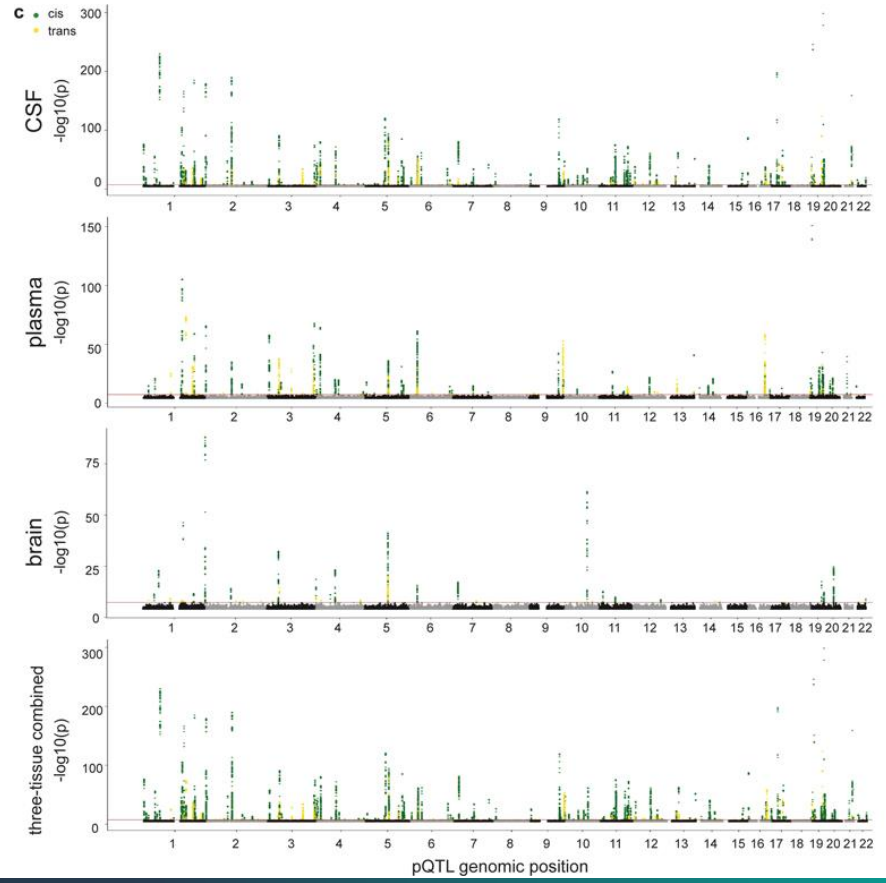
Proteomics&GWAS approach研究策略—研究路径概述



Proteomics&GWAS approach研究策略—代表文献解析

b

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



Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders *Nature Neurosci. 2021*

Proteomics&GWAS 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.

Proteomics&GWAS 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.

Proteomics&GWAS 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.

Proteomics&GWAS approach研究策略—代表文献解析

博淼生物 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.

Proteomics&EWAS approach研究策略

研究目标

pQTM图谱构建及分子调控机制分析

✓基于Cis_与trans_的pQTM关联分析，构建表观调控与蛋白质的调控network网络,筛选关键分子调控通路

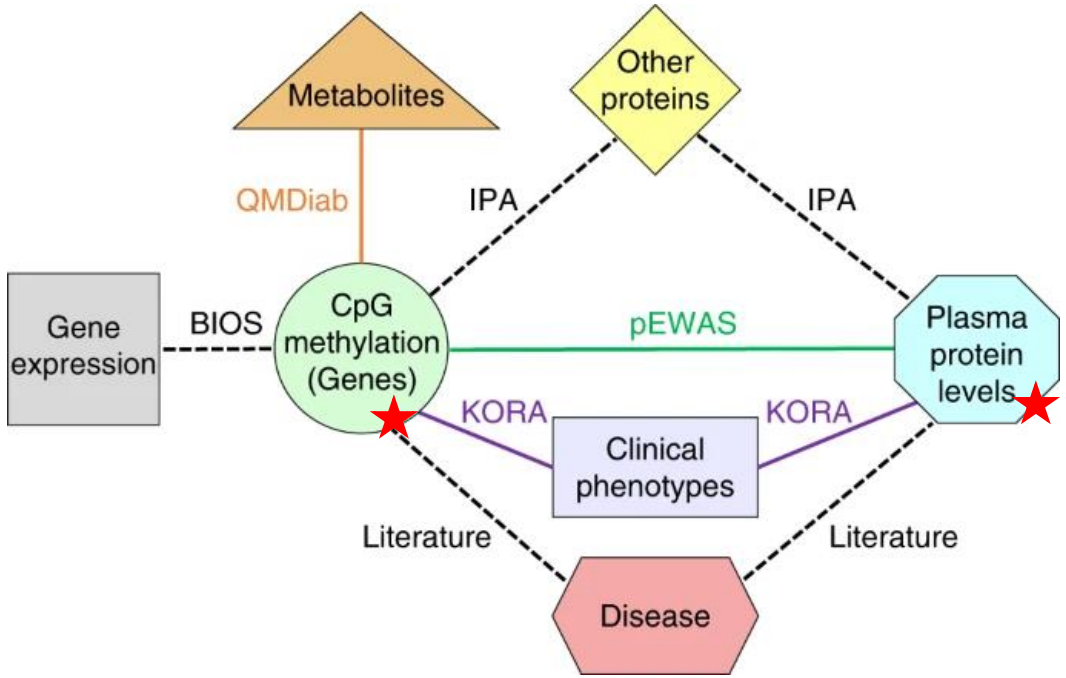
✓探讨甲基化对于诱发表型组的生物功能作用

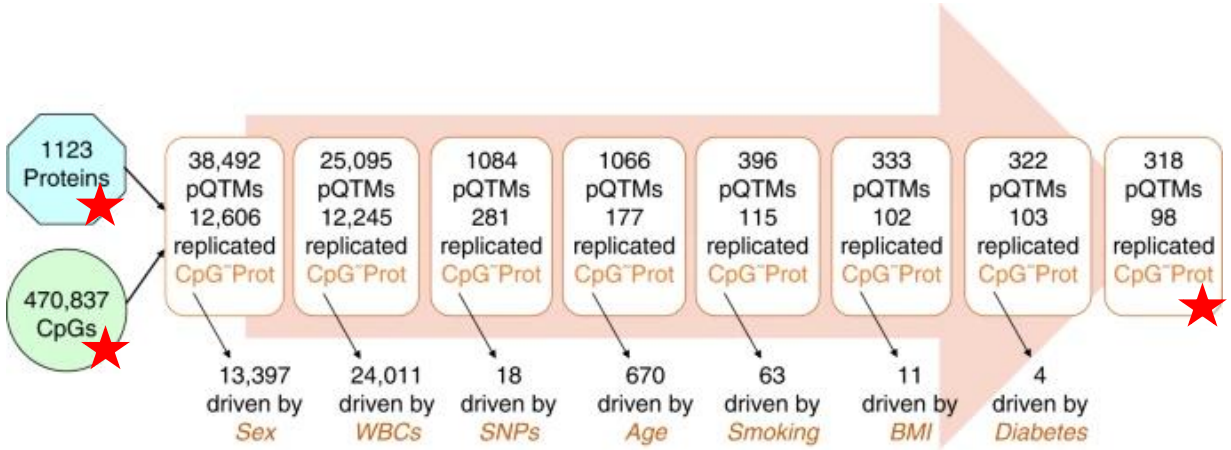
中介效应分析

✓通过Mediation分析，探讨甲基化—蛋白质—表型组的中介调控机制

✓基于公开数据库TCGA、GEO、GTEx等进行多组学数据 eQTM、mQTM等多维度注释，探讨causal 生物机制

Proteomics&EWAS approach研究策略—代表文献解析





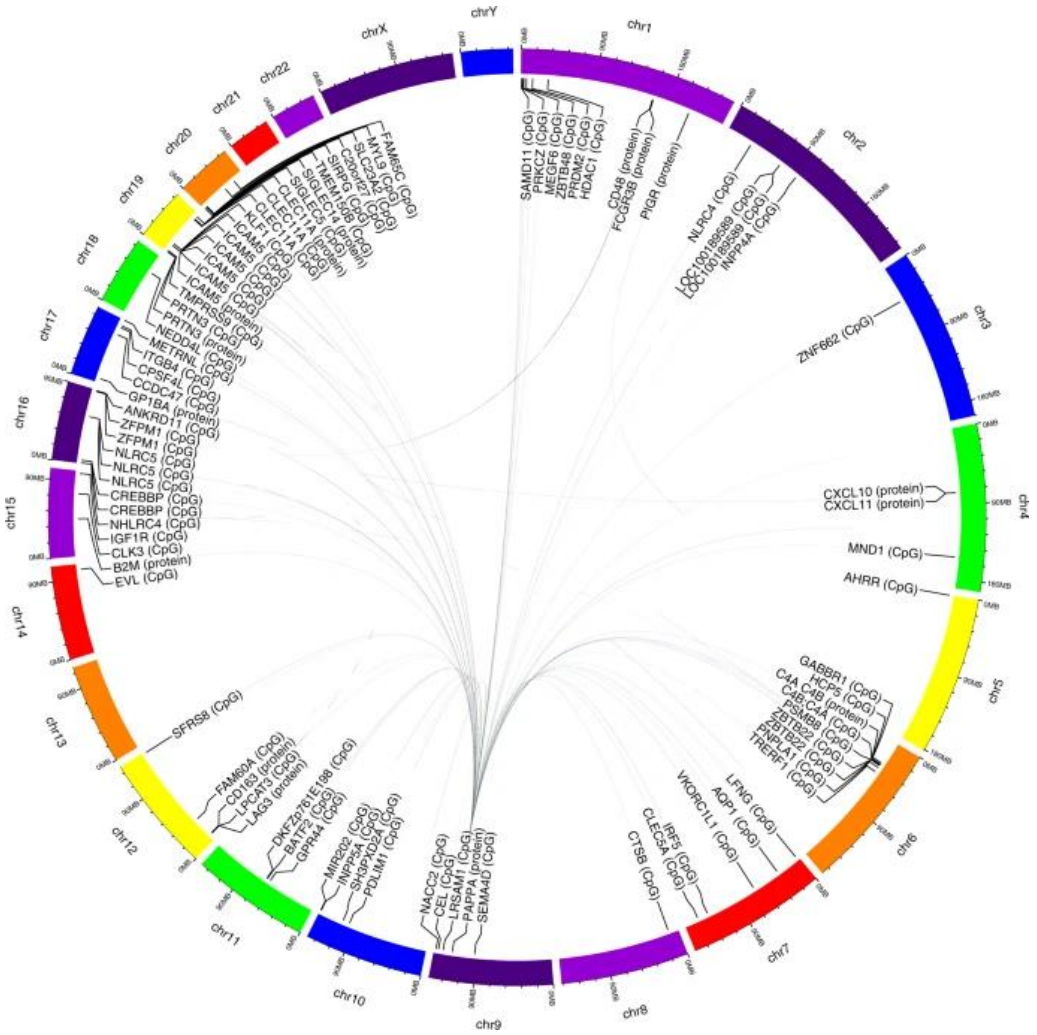
Epigenetics meets proteomics in an epigenome-wide association study with circulating blood plasma protein traits *Nature Communications 2020*

Proteomics&EWAS approach研究策略—代表文献解析

Abstract

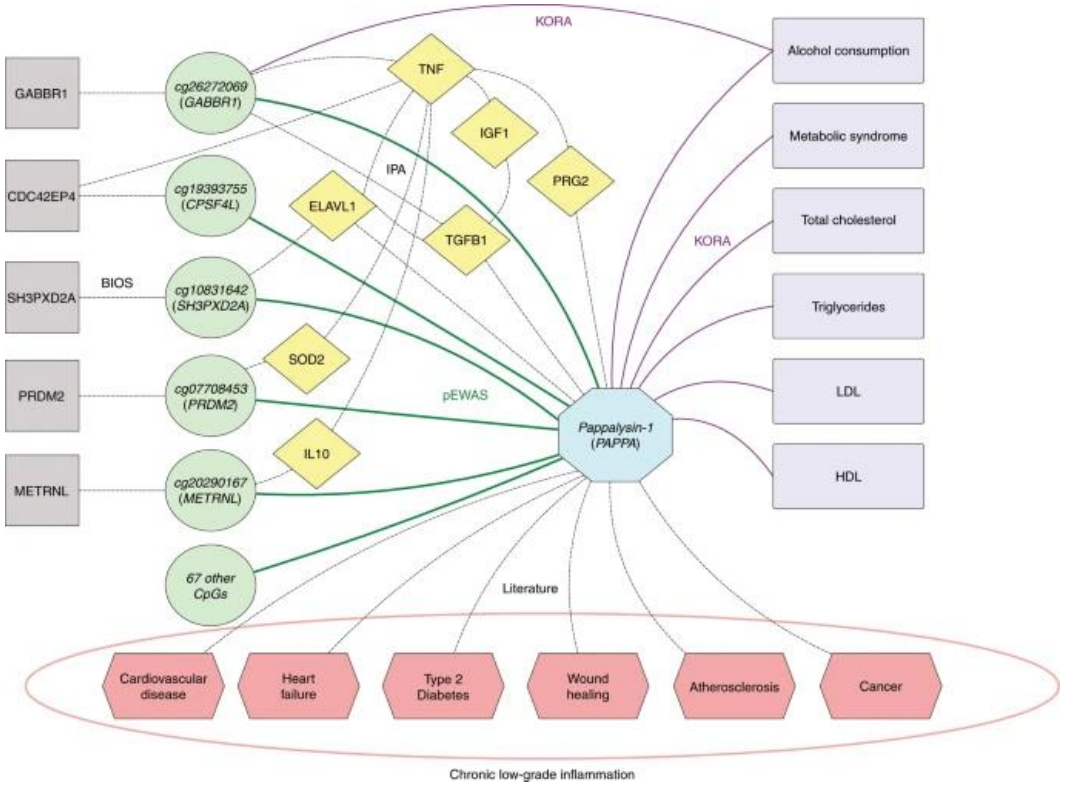
DNA methylation and blood circulating proteins have been associated with many complex disorders, but the underlying disease-causing mechanisms often remain unclear. Here, we report an epigenome-wide association study of 1123 proteins from 944 participants of the KORA population study and replication in a multi-ethnic cohort of 344 individuals. We identify 98 CpG-protein associations (pQTMs) at a stringent Bonferroni level of significance. Overlapping associations with transcriptomics, metabolomics, and clinical endpoints suggest implication of processes related to chronic low-grade inflammation, including a network involving methylation of *NLRC5*, a regulator of the inflammasome, and associated pQTMs implicating key proteins of the immune system, such as CD48, CD163, CXCL10, CXCL11, LAG3, FCGR3B, and B2M. Our study links DNA methylation to disease endpoints via intermediate proteomics phenotypes and identifies correlative networks that may eventually be targeted in a personalized approach of chronic low-grade inflammation.

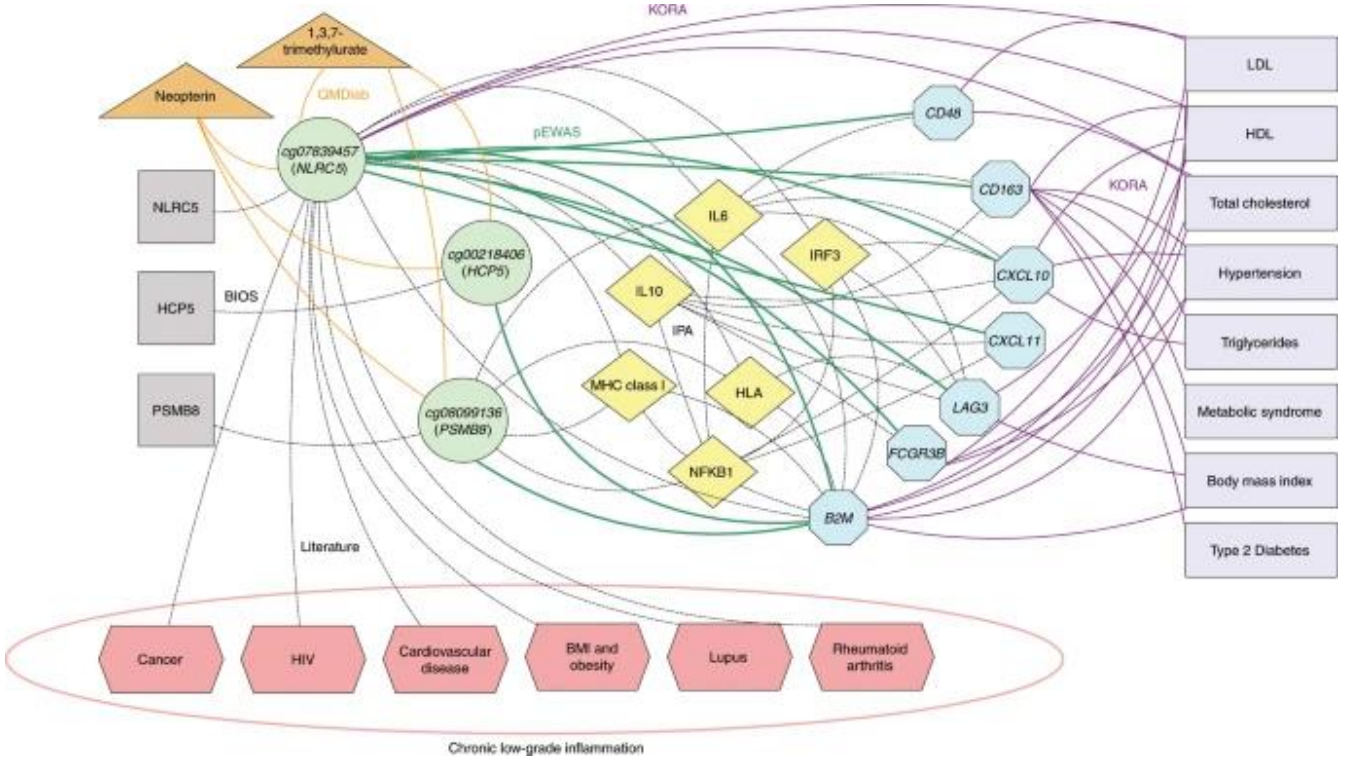
Proteomics&EWAS approach研究策略—代表文献解析



The methylation-proteome network: Circular plot of all 98 replicated cis- and trans- pEWAS associations

Proteomics&EWAS approach研究策略—代表文献解析





Pappalysin-1 protein network: This network comprises 72 CpG sites that associated with blood circulating levels of pappalysin-1 (PAPPA)

NLRC5 methylation network:*NLRC5* methylation is a hallmark of chronic inflammation and has been reported in association with several inflammation-related diseases

Proteomics&Metabolomics approach研究策略

研究目标

调控功能机制的验证

✓显著差异蛋白及代谢物的相关性分析、KEGG通路联合分析、PCA比较分析等

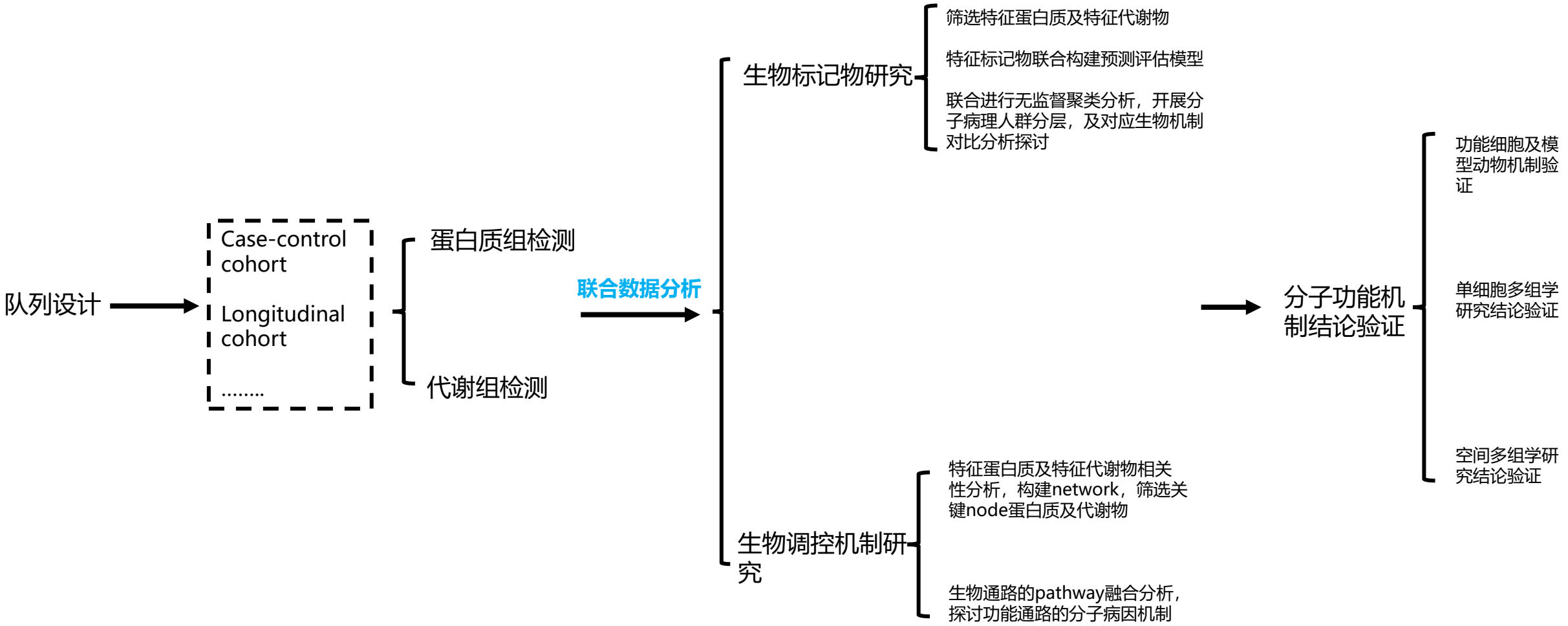
✓系统描绘蛋白至代谢的调控过程，挖掘关键蛋白与代谢物上下游调控通路

多组学联合预测模型构建

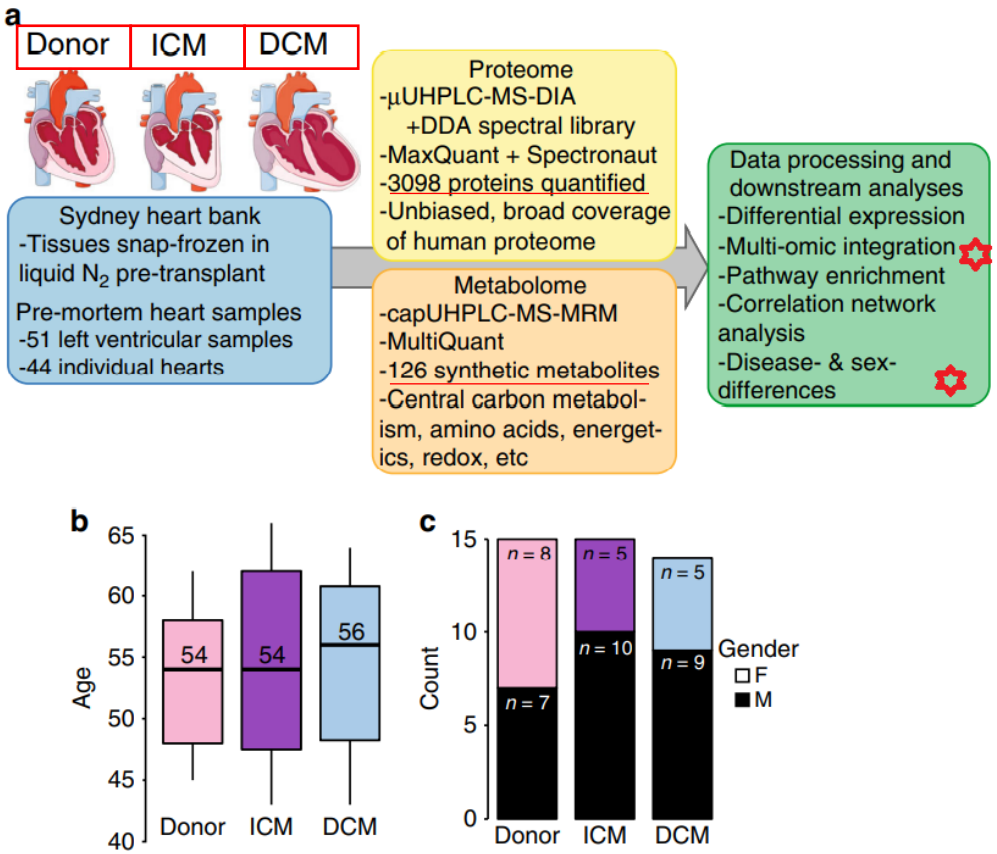
✓**单一组学标记物、多组学标记物联合，基于机器学习算法进行预测模型构建及AUC效率比较**

✓**分子表型聚类分析，构建不同的分子病理模型及对应的生物机制探讨**

Proteomics&Metabolomics approach研究策略—研究路径概述



Proteomics&Metabolomics approach研究策略—代表文献解析



研究设计

疾病表型：不同病因心肌病（缺血性心肌病、扩张性心肌病）的生物调控机制差异	特色：围绕不同性别对应疾病组织进行了蛋白质组学及代谢组差异分析

单一组学差异分析

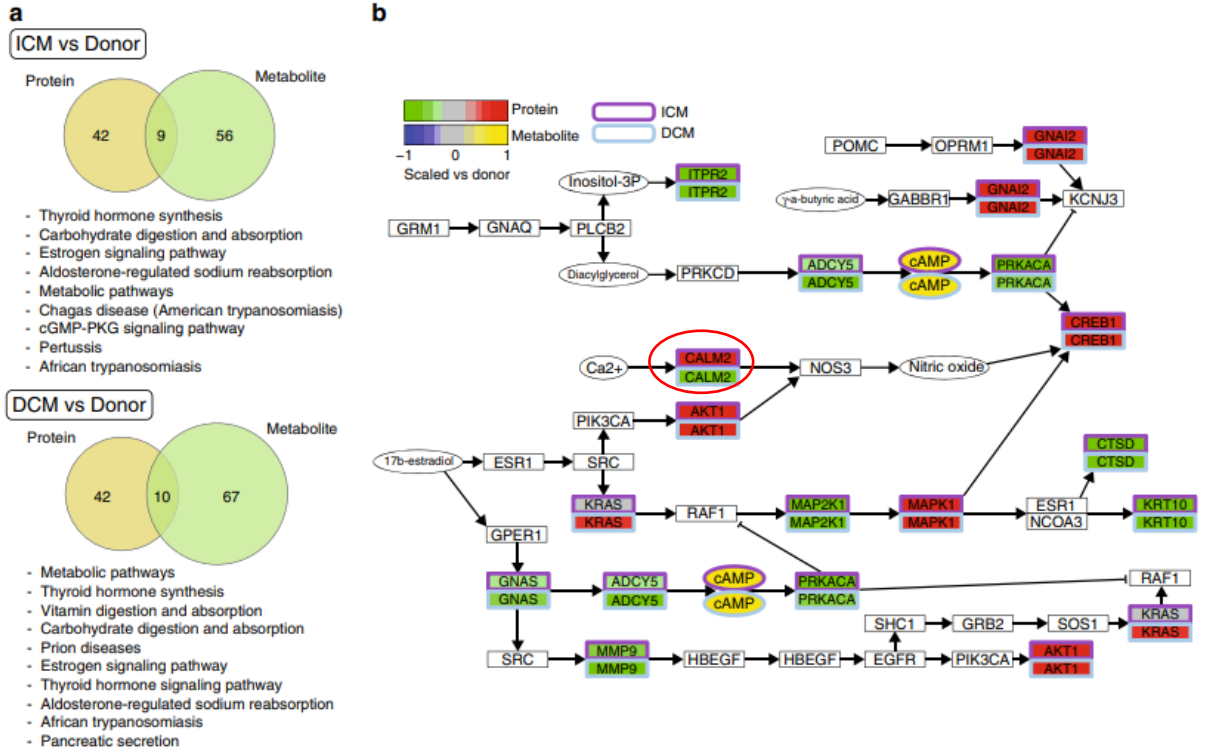
蛋白质组差异分析：根据ICM\DCM分别与Donor组间分析，筛选差异蛋白质及相关生物学特征	代谢组差异分析：根据ICM\DCM分别与Donor组间分析，筛选差异代谢物及相关生物学特征

蛋白质组&代谢组联合分析

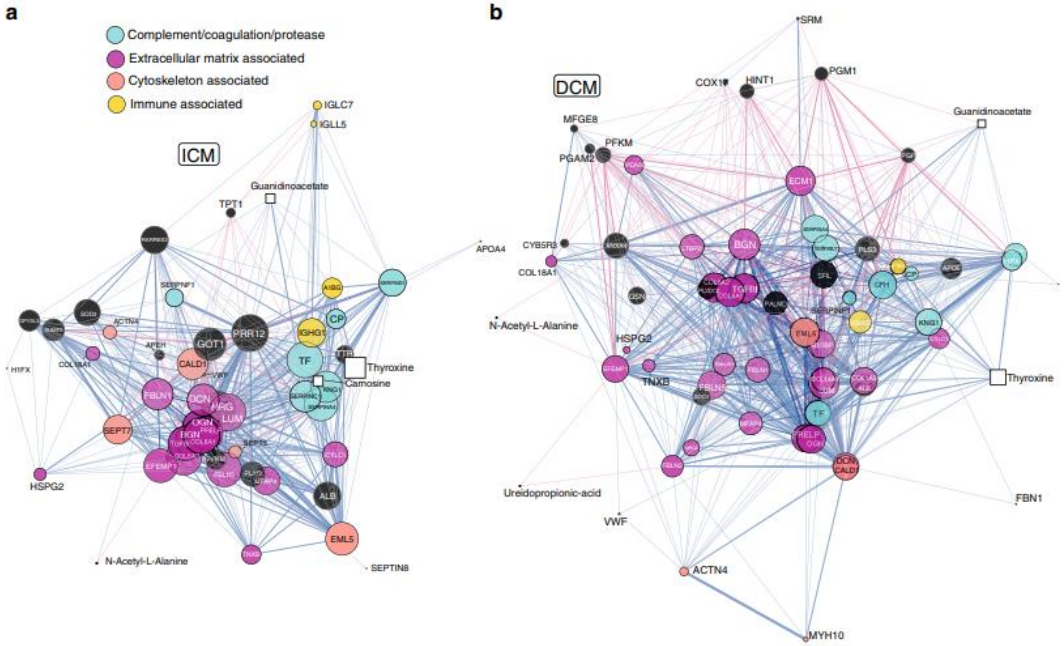
KEGG通路联合分析：对 ICM 和 DCM 的蛋白质和代谢物水平上进行KEGG通路分析,显示了两者在蛋白质和代谢物水平上调节的共同途径	显著差异蛋白质与代谢物相关性分析：在相关性分析基础上构建网络调控图，筛选关键节点蛋白质及代谢物

Core functional nodes and sex-specific pathways in human ischaemic and dilated cardiomyopathy
NATURE COMMUNICATIONS | (2020)

Proteomics&Metabolomics approach研究策略—代表文献解析



Pathways analysis



Network analysis

Proteomics&Microbiomics approach研究策略

研究目标

生物标记物筛选及联合模型优化

✓筛选与表型组显著相关的特征菌群及蛋白质

✓多组学标记物联合预测模型优化构建，显著提升评估效率

特征菌群及其生物调控机制挖掘

✓通过特征菌群与蛋白质的相关性分析，探讨其潜在的生物调控机制

✓采用粪便移植等技术干预手段，基于动物模型验证多组学统计性结论

Proteomics&Microbiomics approach研究策略—代表文献解析

ABSTRACT

METHODS

In this study, among 110 young children (mean age, 18 months) with linear growth stunting who were living in an urban slum in Dhaka, Bangladesh, and had not benefited from a nutritional intervention, we performed endoscopy in 80 children who had biopsy-confirmed EED and available plasma and duodenal samples. We quantified the levels of 4077 plasma proteins and 2619 proteins in duodenal biopsy samples obtained from these children. The levels of bacterial strains in microbiota recovered from duodenal aspirate from each child were determined with the use of culture-independent methods. In addition, we obtained 21 plasma samples and 27 fecal samples from age-matched healthy children living in the same area. Young germ-free mice that had been fed a Bangladeshi diet were colonized with bacterial strains cultured from the duodenal aspirates.

RESULTS

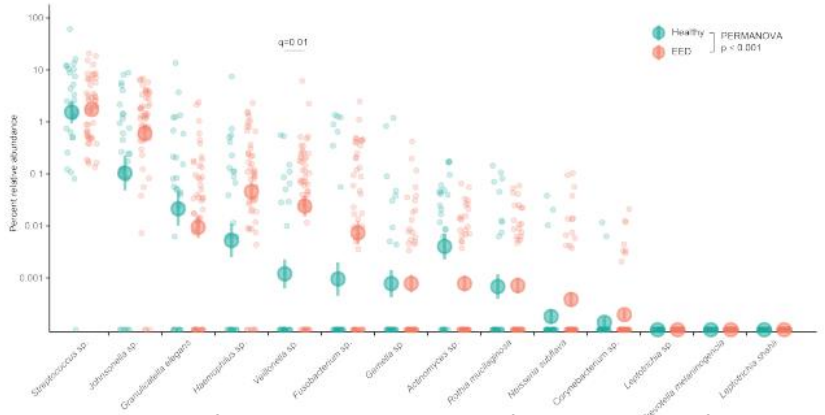
Of the bacterial strains that were obtained from the children, the absolute levels of a shared group of 14 taxa (which are not typically classified as enteropathogens) were negatively correlated with linear growth (length-for-age z score, $r=-0.49$; $P=0.003$) and positively correlated with duodenal proteins involved in immunoinflammatory responses. The representation of these 14 duodenal taxa in fecal microbiota was significantly different from that in samples obtained from healthy children ($P<0.001$ by permutational multivariate analysis of variance). Enteropathy of the small intestine developed in gnotobiotic mice that had been colonized with cultured duodenal strains obtained from children with EED.

CONCLUSIONS

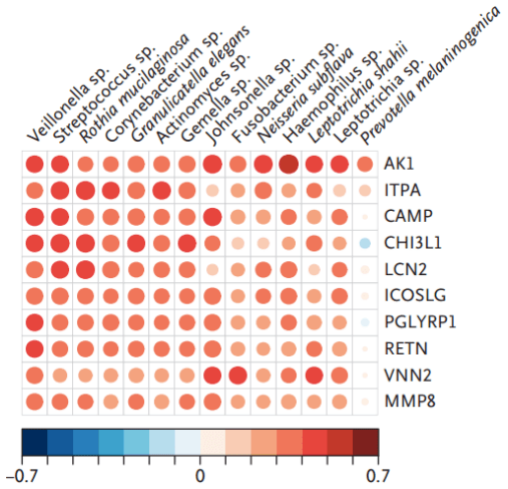
These results provide support for a causal relationship between growth stunting and components of the small intestinal microbiota and enteropathy and offer a rationale for developing therapies that target these microbial contributions to EED.

Duodenal microbiota in stunted undernourished children with enteropathy *NEJM* 2020

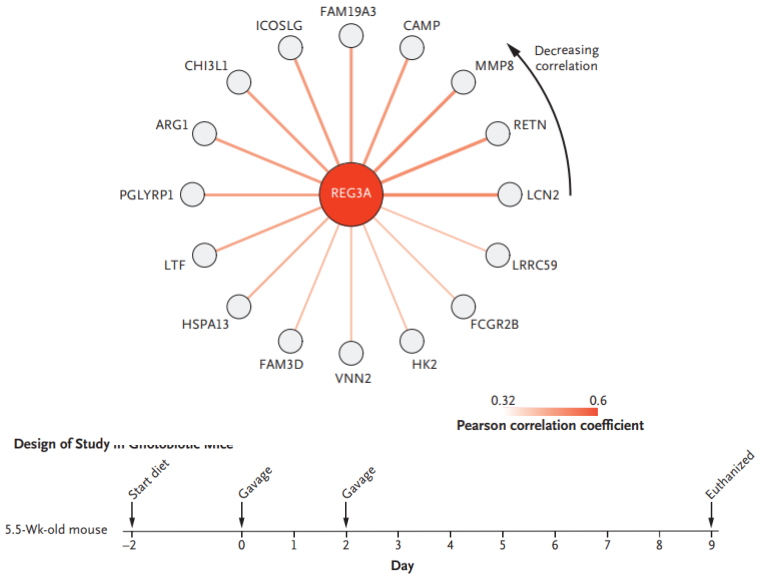
Proteomics&Microbiomics approach研究策略—代表文献解析



Relative abundances of core duodenal taxa in the fecal microbiota of children living in Mirpur who have healthy growth phenotypes (n=27) and those with EED (n=48)



The top 10 positive correlations between members of the 14 core taxa and duodenal proteins



A shows the design of the experiment in which gnotobiotic mice

分析 36 名环境肠功能障碍（EED）患儿的十二指肠微生物群，鉴定出 14 个不属于典型肠道病原体的 EED 核心细菌类群

特征菌群的绝对水平，与患儿生长负相关，与参与免疫炎症应答的十二指肠蛋白（如 LCN2）正相关，且在患儿粪便中的含量不同于健康儿童

和 EED 核心菌相关的十二指肠蛋白，与血浆 REG3A 等显著相关

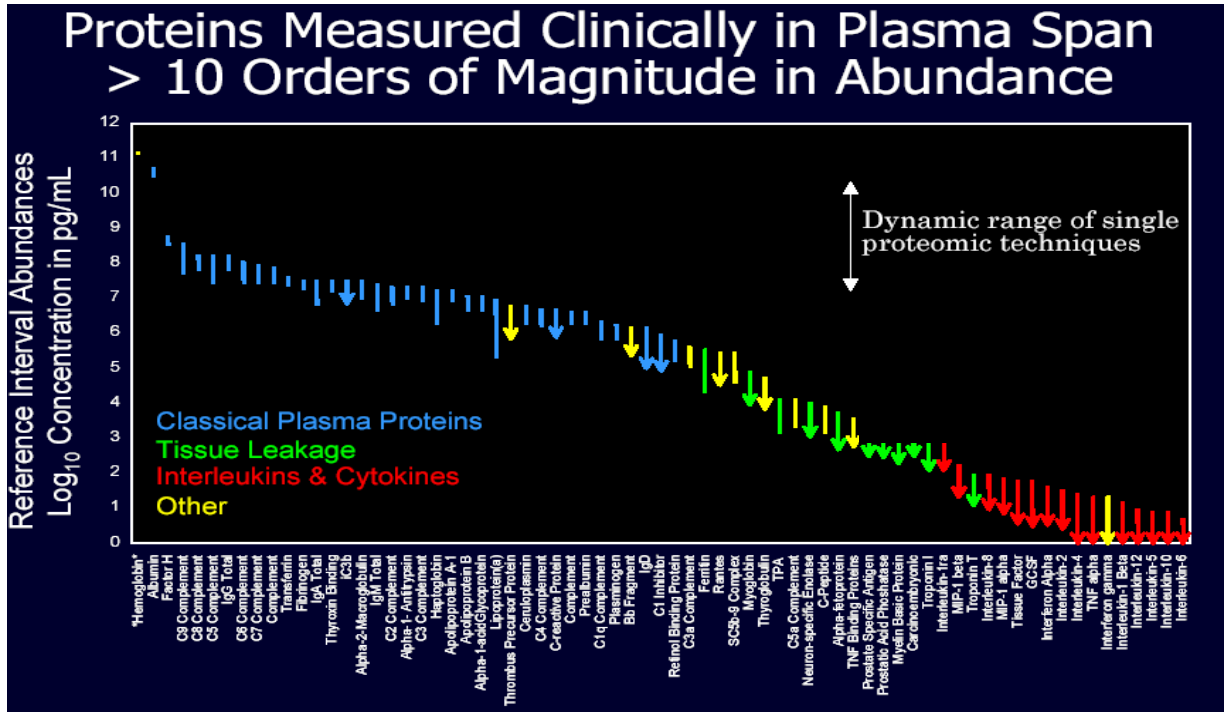
定植小鼠 EED 十二指肠菌群的部分分离菌（含大部分 EED 核心菌）可能通过促进十二指肠炎症应答，参与 EED 的发生发展，从而导致发育不良

Olink PEA技术在队列研究的创新优势

Olink创新技术与质谱蛋白质组学技术的互补

血清血浆蛋白质组学的复杂性

- 蛋白质在组成形式上非常复杂，至少有一万种蛋白的形式，并且大多数蛋白的丰度极低
- 蛋白的丰度差异非常大，远超过目前分离鉴定手段所能达到的灵敏度及检测的动态范围



质谱蛋白质组检测技术的限制

1. 前处理复杂，需要去除高丰度蛋白的干扰
2. 对于低丰度蛋白的检出率低
3. 数据的批次效应、缺失值较大
4. 后期需要通过Elisa或者Western其它平台进行验证

Olink蛋白质组&GWAS队列多组学研究代表文献汇总

Reference	Study population	Samples in study	Proteins assayed	pQTLs reported	Platform type
Zhernakova et al. (Nature Gen. 2018)	LifeLines Dutch population cohort	1264	92	214	Immuno-assay (Olink CVD II panel)
Hillary et al. (Nat. Comm., 2019)	Lothian Birth Cohort 1936	750	92	41	Immuno-assay (Olink neurology panel)
Klaric et al. (ASHG 2019 abstract)	genetically isolated ORCADES cohort (Scotland)	1059	1102	3,545	Immuno-assay (multiple Olink panels)
Höglund et al. (Scientific Rep., 2019)	Northern Swedish population health study (NSPHS)	1005	72	18 novel	Immuno-assay (from Olink panels Oncology I and CVD I)
Folkersen et al. (Nature Metabolism, 2020)	SCALLOP consortium	30,931	90	451 pQTLs for 85 proteins	Immuno-assay (Olink CVD-I panel)
Gilly et al. (Nat. Comm. 2020)	Hellenic Isolated Cohorts MANOLIS study	1328	257	131	Immuno-assay (Olink CVD-II, CVD-III, and MET panels)
Zhong et al. (BMC Genome Med, 2020)	Longitudinal wellness cohort from Sweden	101	794	144 pQTLs across 107 proteins	Immuno-assay (11 Olink panels)
Bretherick et al. (PLoS Genetics, 2020)	Isolated populations from the islands of Orkney (Scotland) and Vis (Croatia)	up to 1992	249	154	Immuno-assay (Olink CVD2, CVD3, and INF panels)
Png et al. (Nat Comm, 2021)	MANOLIS and Pomak, part of the Hellenic Isolated Cohorts (HELIC)	2,893	184	214 pQTLs for 107 proteins	Immuno-assay (Olink neurology and neuro-exploratory panels)
Chiou et al. (ASHG abstract, 2022)	UK Biobank Pharma Proteomics Project	35,306	1,424	28,867	Olink Explore 1536 platform

Olink蛋白质组&EWAS队列多组学研究代表文献汇总

Reference	Study Population/Cohort	Number of Samples	CpG-Trait Associations	Omics Phenotype
Hillary et al. (Nat. Comm., 2019)	Lothian Birth Cohort 1936	750 healthy older adults	26 sites with the levels of 9 proteins	92 protein levels from the Olink neurology panel
Hillary et al. (Genome Medicine 2020)	Lothian Birth Cohort 1936	876	3	70 blood circulating protein levels (Olink inflammation panel)

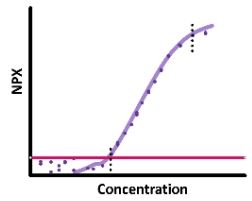
Olink PEA蛋白质组技术队列研究技术优势显著

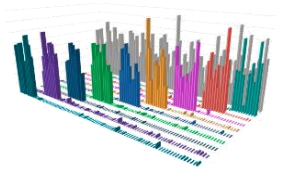
博淼生物 BIO MIAO BIOLOGICAL SINCE 2009

基于邻位延伸分析技术（PEA）

关键免疫分析指标验证

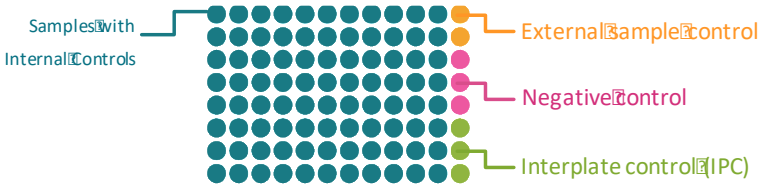
- 灵敏度，特异性，精密度，可扩展性，动态范围，干扰，分析前因素等



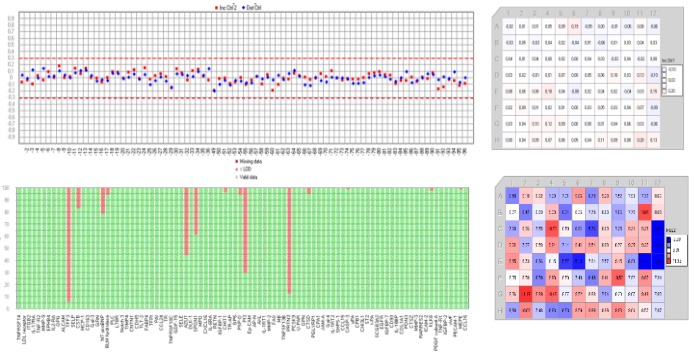


严格的内置质量控制

- 四个内部对照和八个外部对照



质控、标准化、可视化



Olink PEA蛋白质组精准医学应用场景多样性

博淼技术服务项目一览图

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

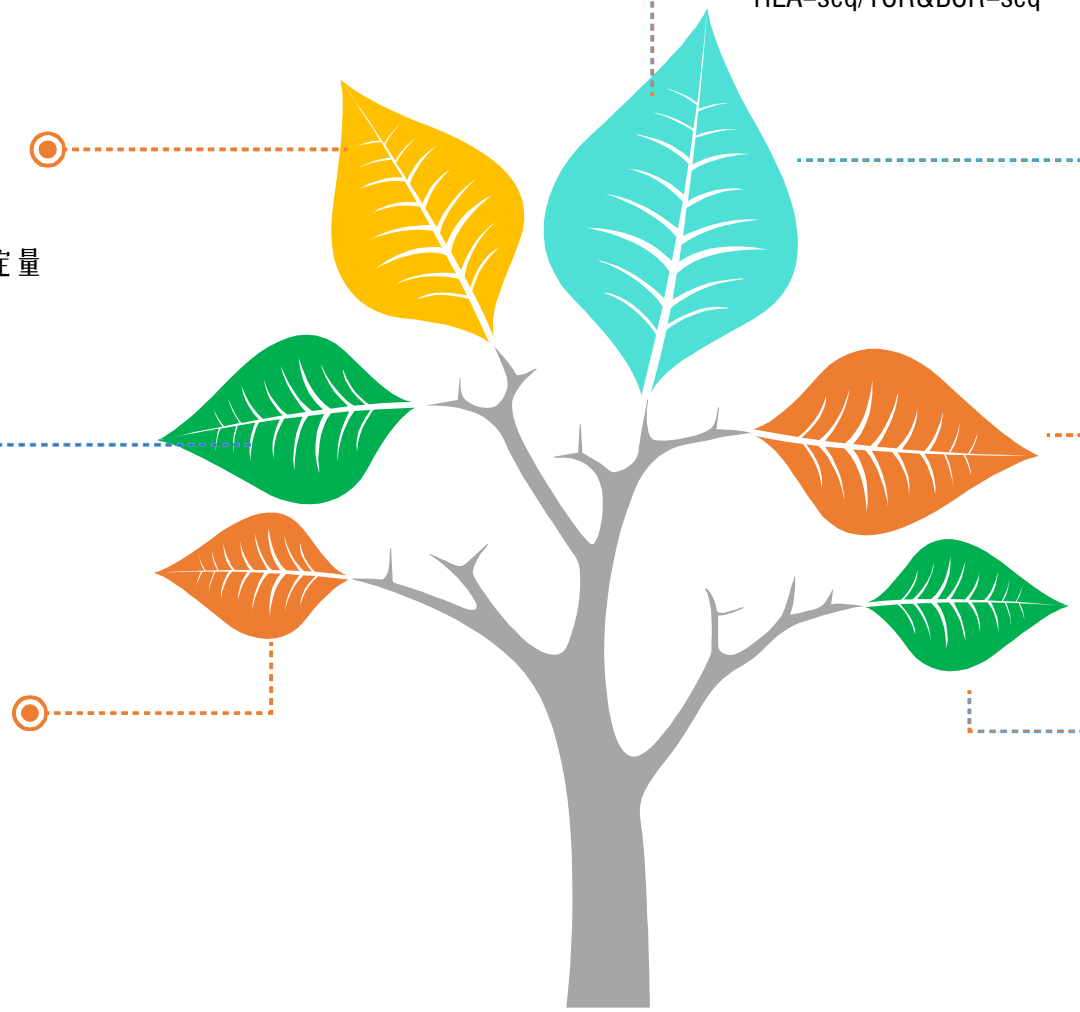
- EWAS芯片
- Multi-PCR NGS靶向DNA甲基化定量
- Massarray 靶向DNA甲基化定量

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