整合组学数据的代谢网络模型研究进展

王雪亮^{1,2}, 张芸^{1,3*}, 温廷益^{1,3,4*}

1. 中国科学院微生物研究所, 中国科学院病原微生物与免疫学重点实验室, 北京 100101;

2. 中国科学院大学, 北京 100049;

3. 中国科学院绿色过程制造创新研究院, 北京 100190;

4. 中国科学院大学存济医学院, 北京 100049

* 联系人, E-mail: zhangyun@im.ac.cn; wenty@im.ac.cn

2020-12-28 收稿, 2021-02-01 修回, 2021-02-02 接受, 2021-02-03 网络版发表

中国科学院A类战略性先导科技专项(XDA17010503)、中国科学院绿色过程制造创新研究院A类重大示范工程项目(IAGM-2019-A02)和国家自然科学基金(31870070, 31870074, 32071460)资助

摘要 基因组规模代谢网络模型(genome-scale metabolic model, GEM)是根据化学计量平衡原理, 基于基因-蛋白-反应三者关联, 建立包含细胞生长必需的生化反应数学模型. 细胞代谢网络存在复杂的调控, 通过基因和蛋白表达量及代谢物浓度变化调节代谢反应通量, 而这些数据无法直接反映代谢反应的通量. GEM研究面临的一个挑战是如何整合不直接反映代谢通量的数据类型, 并将这些调控过程在代谢模型中进行准确的描述. 本文综述了将高通量组学数据整合到代谢化学计量模型中的方法, 包括根据基因和蛋白的表达判定代谢反应状态、筛选核心代谢反应集、代谢反应通量边界约束、转录调控网络整合和多时间尺度基因动态表达约束等; 比较分析了不同算法的优缺点和应用场景; 介绍了GEM与多组学数据整合在解析生物代谢特征、分析遗传和环境扰动、筛选癌症治疗潜在药物靶标和抗代谢药物等方面的应用; 展望了组学数据和代谢网络模型集成的发展趋势.

关键词 代谢网络模型, 转录组学, 蛋白质组学, 代谢组学, 组学数据整合

基因组规模代谢网络模型(genome-scale metabolic network model, GEM)是基于基因-蛋白-反应(gene-protein-reaction, GPR)三者关联, 模拟细胞内复杂代谢通路的一种数学模型, 可以预测细胞内的代谢通路和通量分布. 目前, 基于基因组测序注释和生化数据库已建立了许多微生物及动植物细胞的GEM, 其已经成为系统研究生物体内代谢的主要建模分析方法之一. 根据GPR关联数据, 利用数学模型, GEM系统描述了生物体内一整套基于化学计量学的质量平衡代谢反应, 可以通过通量平衡分析(flux balance analysis, FBA)^[1]、通量可变性分析(flux variability analysis, FVA)^[2]及代谢调节最小化分析(minimization of

metabolic adjustment, MOMA)^[3]等方法预测代谢反应的通量值. 随着测序和组学分析技术的不断发展, 大量组学数据的挖掘展示了更多的代谢调控信息, 而基因转录、蛋白表达及其相互调节作用不能直接反映代谢反应的通量, 这些数据无法直接整合到代谢网络模型. 因此, GEM研究面临的一个挑战是如何整合不直接反映代谢通量的数据类型, 并将这些调控过程在代谢模型中进行准确的描述, 提升模型的预测准确性. 本文综述了GEM与组学数据相结合的研究方法, 分析了不同方法在整合组学数据时所采取的策略, 阐述了GEM与多组学数据整合解析生物代谢特征的应用价值.

引用格式: 王雪亮, 张芸, 温廷益. 整合组学数据的代谢网络模型研究进展. 科学通报, 2021, 66: 2393–2404

Wang X L, Zhang Y, Wen T Y. Progress on genome-scale metabolic models integrated with multi-omics data (in Chinese). Chin Sci Bull, 2021, 66: 2393–2404, doi: 10.1360/TB-2020-1468

1 代谢网络模型的构建和发展

通过粗模型的构建和修正、数学模型的转换以及模型的评估调试等标准方法^[4,5],利用COBRA Toolbox工具可以手动构建GEM。但是,这一过程往往需要数月甚至到两年的时间,耗时耗力^[5,6]。现已开发出多种用于GEM自动重建的工具,如CarveMe^[7]、Path2Models^[8]、ModelSEED^[9]、AGORA^[10]、REVEN2.0^[11]和SuBliMi-naLToolbox^[12]等,具备自动注释基因组、生成GPR关联和预测反应可逆性等功能。自动构建的GEM可以通过Memote工具对自动建模进行基准测试^[13]。GEM发展至今已建立了多个模型数据库,如BiGG模型数据库(<http://bigg.ucsd.edu/>)^[14-16]和MetaNetX数据库(<http://www.metanetx.org/>)^[17]等。随着建模方法的不断发展,经过添加细胞分区、跨膜转运反应、电荷和质量平衡等一系列的精细化处理,模型的信息丰度(基因、代谢反应和代谢物数目)和质量逐步提升。大肠杆菌和谷氨酸棒杆菌作为重要工业微生物,其代谢网络模型经历了多代升级(表1),模型的质量、预测分析的准确度和应用范围不断提升,为工业微生物代谢网络重编程提供了一个研究平台和技术工具。

GEM是在基因组规模上用酶促反应的化学计量表示的代谢矩阵,其设定细胞大分子组成恒定,通过经验值确定总生物量各组成成分之间的关系,以生物量为目标函数,计算求解最优的代谢通量分布。细胞内的生

化过程还包括基因表达和功能蛋白质合成的反应,而酶的表达丰度直接影响代谢反应通量。由于GEM模型不包含对基因转录和蛋白翻译等的化学计量描述,无法定量对酶的合成进行准确模拟,无法模拟产生表型所需的全部代谢过程。将GEM和大分子表达(expression)相结合的综合模型(ME model),通过在GEM基础上增加酶的合成反应,包含转录过程与翻译过程的多种类型数据,增加了表型所需的代谢过程^[31](图1)。RNA和蛋白质组成在ME模型中设定为变量,不再是先验设定的生物量。这种整合表达数据的ME模型能够实现更大范围的约束模拟,准确计算胞内酶合成的成本,并将酶与催化的反应相偶联。大肠杆菌K12 MG1655的ME模型详细描述了近2000个不同基因产物的合成及相对应功能^[22,32],添加了线性约束将酶合成和催化反应偶联。相比大肠杆菌的代谢网络模型,这种多尺度模型对生长表型的预测准确性有所提高^[22]。由于ME模型包含了生命机制的生物合成成本,扩展了已建立的基因组规模的代谢网络模型,使人们对细胞生长有了新的理解。ME模型与GEM相比具有更高的预测能力和准确性,但其规模和复杂性也随之显著增加。

2 代谢网络模型与组学数据整合方法

细胞代谢是一个非常复杂的网络系统,存在多重调控机制,发生在多个不同的分子水平上,包括转录、转录后、翻译和翻译后水平。在不考虑表达调控的约

表1 重要工业微生物的GEM及ME模型

Table 1 The genome-scale metabolic models and ME-models of important industrial microorganisms

菌种	模型名称	模型信息	年份	文献
<i>E. coli</i> K-12 MG1655	iJE660	包含660基因, 627反应和438代谢物	2000	[18]
	iJR904	包含904基因, 931反应和625代谢物	2003	[19]
	iAF1260	包含1260基因, 2077反应和1668代谢物	2007	[20]
	iJO1366	包含1366基因, 2251反应和1136代谢物	2011	[21]
	iOL1650-ME	包含1650个基因和1295个蛋白复合物, 负责代谢、表达和大分子合成	2013	[22]
	EcoCyc-18.0-GEM	包含1445基因, 2286反应和1453代谢物	2014	[23]
	iJL1678-ME	基于iOL1650-ME整合了4个区室和膜约束等	2014	[24]
	iML1515	包含1515基因, 2719反应和1192代谢物	2017	[25]
	iJL1678b-ME	基于iJL1678-ME重新关约束	2018	[26]
<i>Corynebacterium glutamicum</i>	GMD-ATCC13032b	包含247基因, 446反应和411代谢物	2009	[27]
	GMD-ATCC13032a	包含277基因, 502反应和423代谢物	2009	[28]
	iJM658	包含658基因, 984反应和1065代谢物	2016	[29]
	iCW773	包含773基因, 950反应和1207代谢物	2017	[30]

图1 (网络版彩色)整合组学数据的代谢网络模型的构建。(a) GEM的构建; (b) ME模型的构建; (c) 整合模型的构建

Figure 1 (Color online) Workflow to construct a genome-scale metabolic model integrating omics data. (a) Construction of GEM; (b) construction of ME; (c) construction of integrated model

束条件下, GEM无法精确模拟计算基因表达和环境扰动对代谢通量分布的影响。随着高通量测序技术的发展,海量的组学数据提供了大量的基因表达调控信息,反映细胞在不同条件下的代谢状态。将组学数据整合到代谢网络模型,能够反映细胞在不同条件下的真实生理状态,进而在系统层面上研究细胞响应环境发生的代谢变化。至今已发展了多种将高通量组学数据整合到基因组规模的代谢网络模型中的方法^[33,34],包括从构建微生物的单细胞网络到构建人类不同组织的特异性网络、从定性约束网络到定量约束网络、从描述代谢的稳态过程到描述多时间尺度的动态代谢过程的整合方法。这些基于约束建模的方法,通过以组学数据为基础的数学边界或关系减小网络的解空间,从而使得代谢网络的解空间更贴近细胞真实的生理状态。本文总结了整合组学数据的不同方法,并分析了这些方

法在设定代谢反应约束时所采取的不同策略(表2)。

2.1 根据基因和蛋白的表达判定代谢反应状态的方法

细胞在不同环境条件下具有特异的基因表达谱,而基因表达谱与胞内代谢反应通量密切相关。早期开发的整合转录组数据的方法是基于基因为ON或OFF两种状态的设定,其作为“开关”决定基因所控制的代谢反应通量,直接将较低表达基因所控制的反应通量约束为0^[35]。这类方法主要包括GIMME^[36]、MADE^[43]、iMAT^[37,38]、INIT^[47]、tINIT^[57]和EXAMO^[55]等算法,其中GIMME方法基于一组基因表达数据和人为设定的阈值,根据基因调控的布尔逻辑和先验阈值确定各个基因是否表达,据此设定相应反应是否激活,重构代谢网络模型^[36]。MADE方法利用基因表达水平变化的差

表 2 组学数据与代谢网络模型整合的方法^{a)}

Table 2 The methods of metabolic model integrated with multi-omic data

方法名称	整合数据类型	方法特点	求解器	平台	文献
Akesson-04	T	首个整合转录组学数据至代谢网络模型的方法, 设定低表达基因对应的代谢反应通量为0	LP	Matlab	[35]
GIMME	T	代谢和表达调控的基因失活方法, 设定表达阈值, 移除低表达基因对应反应, 补充特定反应获得功能代谢模型	LP	Matlab (COBRA)	[36]
iMAT	T/P	综合代谢分析工具方法, 设定阈值, 将高、中和低表达基因分别用1、0和-1表示, 高表达基因对应反应通量最大化, 低表达基因对应反应通量最小化, 方法在 http://imat.cs.tau.ac.il/ 获取	MILP	Matlab	[37,38]
Moxley-09	T/M/F	使用“代谢物相互作用密度”参数判定基因表达值与代谢反应通量之间的相关性	QP	Metran	[39]
E-Flux	T	代谢通量和表达的组合方法, 根据基因表达数据设定代谢反应通量上下限, 不设反应上限为0或大常数	LP	Matlab	[40]
PROM	T	代谢概率调节方法, 整合转录调控网络与代谢网络, 添加调控作用影响产物生成, 方法在 http://www.igb.uiuc.edu/labs/price/downloads 获取	LP	Matlab (COBRA)	[41]
MBA	T/P/M	基于启发式修剪的模型构建算法, 根据置信度评分删减代谢反应	LP	Matlab (COBRA)	[42]
MADE	T/P	差异表达代谢调节方法, 分析多个条件下表达数据, 不设阈值, 利用t检验计算基因表达差异决定基因表达状态, 方法在 http://www.bme.virginia.edu/csbl/downloads/ 获取	MILP	Matlab	[43]
TIGER	T	整合基因组规模代谢表达和调控的工具箱, 提供算法构建整合调控网络和高通量数据的GEM, 方法在 http://csbl.bitbucket.org/tiger 获取	MILP	Matlab	[44]
tFBA	T	与MADE相似, 利用多个条件下表达数据	MILP	Matlab	[45]
MPA	T/P	代谢表型分析方法, 人类代谢模型整合表达数据推断代谢表型, 方法在 https://cs.tau.ac.il/~livnatje/MPA.zip 获取	MILP	Matlab	[46]
INIT	P/T/M	组织综合网络推断方法, 建立组织特异性模型	MILP	R	[47]
Lee-12	T/M	将绝对基因表达数据整合到约束方程, 方法在其补充材料获取	LP	Matlab (COBRA)	[48]
RELATCH	T/F	两步方法, 利用参考状态通量和表达数据预测扰动下的代谢状态, 方法在其补充材料获取	QP	Matlab (COBRA)	[49]
TEAM	T/P	基于时间表达的代谢分析方法, 结合dFBA与GIMME	LP	Matlab	[50]
AdaM	T/M	代谢适应方法, 利用EMF参数整合不同时间基因表达数据	MILP	Matlab	[51]
GX-FBA	T	基因表达FBA方法, 利用绝对基因表达数据, 方法在其补充材料获取	LP	Matlab (COBRA)	[52]
mCARDE	T/P	确定反应评分判断代谢特异性方法, 根据表达数据计算连接性分数确定移除反应, 方法在其补充材料获取	LP	Matlab	[53]
FCG	T/F	通量耦联基因方法, 利用基因表达与反应通量相关的基因集	LP	GAMS	[54]
EXAMO	T	选择性代谢优化探索方法, 扩展了iMAT方法, 构建特定模型, 方法在其补充材料获取	MILP	Python	[55]
GIM ³ E	T/M	GIMME方法的升级版, 使用代谢组和转录组数据, 方法在 https://open-cobra.sourceforge.net/ 获取	LP	Python	[56]
tINIT	P	任务驱动的组织综合网络推断方法, 是INIT的扩展, 方法在 http://www.sysbio.se/BioMet 获取	MILP	RAVEN	[57]
FASTCORE	T/P/M	重建特定环境代谢网络模型的通用算法, 定义模型核心代谢反应, 方法简单快速, 在 https://bio.uni.lu/systems_biology/software 获取	LP	Matlab (COBRA)	[58]
E-Fmin	T	表达数据调节的通量最小化方法, 改进GIMME代谢功能阈值要求, 方法在其补充材料获取	LP	Matlab (COBRA)	[59]
Aurich-15	P/M	适合细胞外代谢组数据的细胞代谢表征	LP	Matlab	[60]
FALCON	T/P	最小绝对偏差凸函数和归一化的通量分配方法, 分析量化酶复合物, 方法在 https://github.com/bbarker/FALCON 获取	LP	Matlab (COBRA), ATS	[61]
PROM2.0	T	PROM方法改进版, 扩展了框架, 可以预测转录因子过表达表型, 方法在其补充材料获取	LP	Matlab (COBRA)	[62]

(续表2)

方法名称	整合数据类型	方法特点	求解器	平台	文献
METRADE	T/P/M	代谢和转录组适配预测方法, 基于FBA的整合优化基因表达和密码子使用, 方法在其补充材料获取	LP	Matlab (COBRA)	[63]
FlexFlux	T	结合代谢和调控网络的分析程序, 方法在 http://lipm-bioinfo.toulouse.inra.fr/flexflux/ 获取	LP	Java	[64]
RegrEx	P/T	基于正则化最小二乘优化的方法, 自动分析数据获取通量分布, 方法在其补充材料获取	MIQP	Matlab (COBRA)	[65]
Lsei-FBA	T	等式和不等式的最小二乘FBA方法, 依据基因表达变化预测代谢反应通量变化	LP	R	[66]
FASTCOR-MICS	T	FASTCORE扩展版, 可以集成微阵列数据, 能快速从转录组数据创建代谢模型, 方法在 https://bio.uni.lu/systems_biology/software 获取	LP	Matlab	[67]
Tobalina-15	P/T/M	元网络方法, 利用宏蛋白质组等数据, 可以建立细菌群落环境特异性代谢网络模型	MILP	Matlab	[68]
CORDA	T/P	成本优化反应依赖评估法, 可以高效建立特定组织模型, 方法在其补充材料获取	LP	Matlab (COBRA)	[69]
TIMBR	T	转录推断代谢标志物反应方法, 将基因表达数据转化反应权重, 最小化反应通量加权计算目标产物生产所需代谢网络状态, 方法在 https://www.github.com/csl/ratcon1 获取	LP	R, Matlab (COBRA, RAVEN)	[70]
HPCOF	T/P	整合连续基因表达数据到GEM, 利用 Huber罚凸优化函数和通量最小化, 方法在 https://github.com/nwpu903/HPCOF 获取	LP	Matlab	[71]
TRFBA	T	转录调控FBA方法, 与PROM相似, 但所需基因表达数据较少, 改进了生长表型的定量预测, 方法在 https://sbme.modares.ac.ir 获取	LP	Matlab (COBRA)	[72]
CoRegFlux	T	将基因调控网络和代谢模型耦联, 提供更好的中值预测和更低的方差预测, 方法在 https://github.com/i3bionet/CoRegFlux 获取	LP	R	[73]
IDREAM	T	结合统计推断的环境和基因调控影响网络(EGRIN)与PROM框架, 方法在其补充材料获取	LP	Matlab	[74]
LBFBFA	T/P	线性结合FBA方法, 通过数据分析为每个代谢反应通量设定软约束, 方法在其补充材料获取	LP	Matlab	[75]
Dynamic ME	P/T/M	动态代谢与表达模型方法, 对代谢和蛋白质表达时间动态模拟的算法, 方法在 https://github.com/SBRG/dynamicme 获取	LP	Python	[76]
RIPTiDe	T	通过转录丰度和精简原则获得反应的方法, 方法在 https://github.com/mjenior/riptide 获取	LP	Python(COBR-Apy)	[77]
MOOMIN	T	组学数据在代谢网络的数学探索方法, 不需要生物目标函数, 不用先验反应, 使用差异表达的贝叶斯后验概率, 方法在 https://github.com/htpu-sa/moomin 获取	MILP	Matlab (COBRA)	[78]

a) T, transcriptomic data; P, proteomic data; M, metabolomic data; F, fluxome data; LP, linear programming; QP, quadratic programming; MILP, mixed integer linear programming; MIQP, mixed integer quadratic programming

异显著性判定基因的表达状态, 预测代谢反应激活或失活, 解决了任意选择阈值的问题^[43]。这两种方法的缺陷在于依赖阈值的选择, 可能将计算状态为0的代谢反应错误删除, 造成通量平衡分析与实验值之间的误差。

iMAT方法通过比较全部基因表达数据, 根据均值和方差将基因表达值离散化, 将基因表达数据分为高、中和低3种状态, 对3种状态所对应的代谢反应上下限进行约束, 利用混合整数线性优化算法, 确定最大化代谢反应数量的稳态通量分布, 其预测结果与基因表达数据以及模型的热力学约束一致^[37,38]。随后开发

的INIT、tINIT和EXAMO算法, 通过整合转录组与蛋白质组数据, 找到最符合表达数据的代谢反应网络。这几种方法的优点是无需人为设定细胞生长作为目标函数, 能够最大限度提高基因表达和网络通量状态之间的拟合程度; 使得高表达且携带通量的反应数量最大化, 低表达且携带通量的反应数量最小化, 从而获得与组学数据相一致的代谢网络模型。通过基因或蛋白表达数据判定代谢反应状态的方法缺点在于仅能够实现网络组成结构的约束, 无法根据基因或蛋白表达水平对代谢反应通量进行定量约束。

2.2 通过组学数据筛选核心代谢反应集的方法

细胞代谢在各种疾病的发病机制中起着关键作用,不同的细胞类型或组织代谢具有差异性,需要通过结合各种来源的组学和文献数据重建不同细胞类型或组织的特异代谢网络模型. MBA^[42]、mCADRE^[53]、CORDA^[69]、FASTCORE^[67]和FCGs^[54]等算法被应用于人组织特异性模型的构建,也被称为“剪枝”算法.通过组学数据获得一组核心反应,包含有功能的代谢反应,移除通用模型中的其余代谢反应,快速生成特定组织模型.这类算法中FASTCORE^[67]具有一定的通用性,由于其仅依赖一组预选的核心代谢反应作为输入,以通量一致性为计算目标,使每个反应有活跃的通量状态,搜索全局网络的通量一致子网络(包含来自核心集的所有反应和附加反应的最小集).每次迭代都通过两个线性程序进行,程序旨在最大化支持核心集内代谢反应模型,同时最小化核心集外的代谢反应数量. FASTCORE的优点在于简单易行,没有自由参数,比MBA算法快几个数量级,而且能够产生更紧凑的重构模型.这几种方法将整合组学数据的代谢网络模型的应用范围从单细胞生物扩大到具有复杂组织结构的高等真核生物.

2.3 通过基因表达值进行代谢反应通量约束的方法

组学数据反映了全细胞中基因和蛋白的表达水平,这些表达数据在一定程度上影响代谢反应通量.通过函数表达式对模型中的所有或大多数基因对应反应流量设定约束,开发了代谢反应通量约束的算法,其中具有代表性的算法是E-Flux^[40]和LBFBA^[75]. E-Flux算法使用“管道容量”类比法,根据基因与蛋白的表达水平对代谢反应的通量进行约束,通过函数表达式将特定状态下的基因表达量进行标准化处理,使用基因表达值相对调节模型中代谢反应可以携带的通量,实现对代谢反应的“通道”宽度作特定约束^[40]. E-Flux不仅能通过表达数据判定代谢反应的开和关,而且通过特定的函数关系将表达值与代谢通量相对应,将代谢反应通量约束在一定的范围内,求解最符合基因表达状态的代谢通量分布. LBFBA是一种线性结合FBA的整合方法,该方法基于转录组学与蛋白质组学数据的线性函数为每个代谢反应设定通量的上限和下限,而且这些通量界限是反应特异性的. LBFBA并非对所有反应使用单个结合函数,而是从训练数据集中的转录组学

(或蛋白质组学)和通量组学数据中学习,计算每个代谢反应特定的界限. LBFBA相对其他算法易于求解,需要4~5组转录组学、蛋白质组学和通量组学的实验测量数据,使用线性函数将代谢通量的子集约束在一个符合实际生理的范围内,而不是将表达和通量数据拟合到一个线性函数中,能够更准确预测代谢反应通量. LBFBA能够对真核和原核生物进行分析,可以预测野生型和基因敲除菌株的胞内通量分布.但由于通量组学测量限于中心代谢途径, LBFBA仅能够预测中心代谢通量^[75].随着基因组规模的通量测量方法的发展,有望实现对中心代谢之外其他代谢途径的通量预测.上述算法的优点在于突破了以往代谢网络模型定性约束的局限,利用表达数据进行代谢反应通量约束,实现了数据和代谢网络模型的定量整合,提高了代谢网络模型的模拟预测准确性.

2.4 转录调控网络与代谢网络整合的方法

基因之间存在着复杂的相互作用调控关系,将构建的转录调控网络(TRN)与代谢网络模型整合,开发了PROM^[41]、TIGER^[44]、FlexFlux^[64]、TRFBA^[72]、CoRegFlux^[73]等算法. TRN可以表示为顶点和边的二分图,顶点代表调控因子和靶基因,而边代表这些调控单位是如何连接和相互作用的.利用因果关系调节基因的表达状态,通过概率或者线性约束实现TRN与代谢网络模型的整合. PROM方法基于由转录因子及调控基因所构成的基因调控网络,生成概率模型判断代谢反应的状态,对代谢反应通量进行约束^[41]. PROM的缺点在于需要在各种环境和遗传干扰下测定的大量基因表达数据,以引入基因表达状态和基因与转录因子相互作用所需的概率,一定程度上制约了其应用范围. TRFBA算法将基因的表达水平视为一个连续变量,并引入了两个新的线性约束.一个约束利用常数参数限制代谢基因对应的反应速率,该参数将基因表达水平转换为代谢反应的上限.另一个约束基于约束建模概念将每个靶基因的表达水平与其调节基因的表达水平相关联^[72]. TRFBA可以实现持续约束代谢反应通量,仅需要转录因子及靶基因,通过自动添加调控基因的表达水平以约束靶基因的表达.与PROM方法相比, TRFBA不需要大量基因表达数据,也不需要调控子与其靶基因的关系,引入的约束条件更适用于改进FBA方法对生长的预测,是一种相对简单且准确的方法.与将基因表达数据整合到GEM模型中的方法相比,整合

TRN产生的模型提供了细胞内代谢和转录调节过程的全面描述。利用不同来源数据(如基因表达、转录因子结合位点或比较基因组学分析)重建TRN的方法较多,使得大多数已测序基因组的原核生物的TRN重建相对简单易行。但是,目前还缺少构建TRN的标准协议和易于使用且用户友好的计算工具。

2.5 基于时间的基因动态表达约束的方法

基因表达水平随环境变化发生动态变化,为了分析基因表达随时间发生的动态调节变化,开发了基于时间表达的代谢网络模型的分析方法。TEAM^[50]方法是基于时间表达的代谢分析,通过结合动态通量平衡分析(dFBA)和GIMME方法,利用时间次序基因表达数据预测瞬时代谢通量分布。使用特定基因阈值和成本最小化原则通过基因表达水平计算每个时间步长的通量分布,协调基因表达数据与代谢约束。AdaM^[51]方法是整合时间次序基因表达数据的另外一种方法,其在每个时间点寻找与差异基因表达模式一致的最小功能网络,不需要为每个时间点计算通量分布,而是为子网计算了一组基本通量模式(EFM),每个反应的时程模式由每个时间点包含该反应的EFM分数表示。这些方法的缺点是需要结合多个时间次序的表达数据,需要大量的基因表达数据,才能够获得动态的代谢网络信息,其优点在于不依赖于特定条件的目标函数,例如生物量和产量,突破了稳态分析的局限性,有利于对生物代谢过程进行动态分析。

2.6 其他整合组学数据的方法

为描述不同遗传和环境扰动等因素对细胞内代谢反应通量的影响,开发了利用不同参数和统计推断的方式预测环境和基因调节下代谢网络通量的算法,如RELATCH^[49]。RELATCH是一种两步方法,使用参考状态的代谢反应通量和基因表达数据,预测处于遗传扰动或环境扰动状态的代谢反应通量。第一步,利用转录组和通量组学数据预测参考状态的代谢反应通量分布和酶贡献(酶贡献指当考虑所有同工酶的贡献时,某一特定酶催化的代谢反应通量对总通量的贡献);第二步确定扰动状态的通量分布,使该状态通量分布对参考状态的代谢通量分布的调整最小,并且与预测的酶贡献相一致^[49]。该算法设定受干扰的菌株最初在有限调节适应时最小化相应代谢通量变化,适应干扰时可以增加之前活跃和不活跃的代谢反应的通量,根据扰动

的性质通过两个参数预测扰动对代谢反应通量的影响。RELATCH可以预测不同扰动条件下微生物系统的定量代谢变化,已被应用于预测遗传扰动或环境扰动下大肠杆菌、酿酒酵母和枯草芽孢杆菌菌株的通量分布,且在预测基因敲除菌株的细胞内通量分布方面具有优势。

上述这些方法通过不同方式整合组学数据,实现了对特定目的的研究。调控与代谢存在相关性,但不一定存在因果关系,未来需要开发更多新的方法,通过结合多种先验信息,如网络结构、成分相互作用和物理化学性质等,将数据作为约束与代谢网络模型进行深度结合,建立更准确的调控代谢关联。

3 整合组学数据的代谢网络模型的应用

3.1 整合模型在微生物研究中的应用

微生物具有代谢多样性,整合组学数据的代谢网络模型与GEM相比在转录水平、蛋白水平和代谢水平进行了升级,提高了模型预测的准确性。利用GEM模型和传统的通量平衡分析不能准确预测酿酒酵母部分表型,例如在以葡萄糖为底物的有氧分批培养中酿酒酵母指数生长期间降低的生物量和增加的副产物。Akeson等人^[35]将酿酒酵母恒化培养和分批培养的基因表达谱与代谢模型相结合,提高了对分批培养中代谢表型的预测准确度,可以获得以葡萄糖为底物发酵的酿酒酵母在指数生长期生物量和代谢副产物(乙醇和甘油等)的总产量系数的预测结果。计算的代谢通量分布与¹³C同位素标记的实验值相符,实现了细胞内通量变化的定性预测。应用ME模型整合转录组学数据分析极端嗜热菌胞内基因表达变化,发现了与阿拉伯糖和纤维二糖代谢相关的新调控因子,丰富了基因转录调控信息^[31]。利用MADE方法将酿酒酵母在含葡萄糖培养基到含甘油培养基的基因表达数据与酵母模型iND750进行整合,模拟了酿酒酵母从发酵呼吸向甘油呼吸转变过程中的代谢调节。整合模型分析表明,底物从葡萄糖转变为甘油后,脂肪酸和谷氨酸合成途径增强,膜离子转运活性在转变期结束时降低。当无葡萄糖时,一种全局转录调节因子重新激活了与其他营养物质运输和利用相关的途径。生长在YPD(yeast extract peptone dextrose medium)培养基的酿酒酵母显示出最小范围的通量可变性,转向甘油底物后,代谢多样性增加,代谢反应通量的可变性明显增加^[43]。通过AdaM方法将大

肠杆菌的代谢网络模型与冷应激和热应激的时间次序基因表达数据整合, 获得了不同时间次序下代谢反应具有持续活性和暂时活性这两种模式, 对比了维持基本细胞功能和适应温度所需的代谢反应, 预测出与时间和压力相关的代谢反应, 阐明了由于代谢网络结构和环境变化相互作用产生的鲁棒性和适应性之间的内在联系^[51]. 应用RELATCH算法预测了实验室适应性进化前后大肠杆菌MG1655敲除突变体(*pgi*、*ppc*、*pta*和*tpi*)的通量分布, 准确预测了4个敲除突变体的代谢反应通量分布, 有效地预测了对遗传扰动的最初反应, 包括潜在途径的有限使用和代谢缺陷反应通量的重新分布. RELATCH预测到3个突变体(*pgi*、*ppc*和*tpi*)的生长和葡萄糖吸收速率显著降低, *pta*突变体的葡萄糖吸收速率略有增加. 相比之下, 传统算法FBA、MOMA和ROOM高估了代谢反应通量值, 而RELATCH预测的葡萄糖吸收速率、乙酸分泌和生长率与亲本菌株生长曲线一致^[49]. RELATCH的预测相对准确、普适和数据要求量低, 使其能应用于更多的研究领域. 这些成功实例表明, 应用整合组学数据的代谢网络模型有助于研究不同微生物细胞的生长表型、代谢调控、代谢适应性等生理代谢特征以及遗传和环境扰动对代谢通量分布的影响.

3.2 整合模型在人类细胞研究中的应用

细胞代谢在各种疾病的发病过程中发挥关键作用, 利用组学数据建立不同人类细胞类型的组织特异性代谢网络模型, 有助于更好地理解复杂疾病发病过程中的代谢变化. 利用雷弗素姆病患者和正常成纤维细胞转录组、蛋白质组和代谢组数据, 使用FASTCORE算法创建的成纤维细胞特异性代谢网络模型, 发现了肌醇六磷酸摄取的通量显著减少, 表明肌醇六磷酸在体内的积累增加. 同时, 预测分析显示, 患者的二肽和三肽丰度与健康模型相比出现了明显的变化, 一些特定的肽参与了雷弗素姆病的发病过程, 可以作为诊断的生物标志物^[79]. 利用mCADRE方法在人类GEM模型Human Recon 1的基础上, 重建了26种基因组规模的肿瘤组织及其对应的正常组织模型和30种的脑组织模型, 通过对大量组织特异性模型的代谢路径分析, 发现了催化花生四烯酸产生白三烯的代谢反应是选择性影响肿瘤组织的潜在药物靶点^[53]. 使用蛋白质组学数据、HMR 2.0和tINIT算法, 为6名肝细胞性肝癌患者重建个性化的GEM用于识别抗癌药物(代谢物的结构类似物).

通过对83个健康细胞类型特异性GEM的评估, 分析预测了101种在全部肝细胞性肝癌患者中能够有效预防肿瘤生长的抗代谢物, 其中22种已经用于不同癌症的治疗, 而其余抗代谢物成为肝细胞性肝癌的新的潜在药物^[57]. 应用整合模型对癌症细胞和健康细胞的活性代谢网络之间进行差异分析, 能够识别癌症特异代谢特征、癌症治疗的潜在通用药物靶标和新型肿瘤抗代谢药物.

4 结论与展望

代谢过程是生命的基本过程, GEM将基因、蛋白和代谢物相关联, 提供了一种使用数学模型解码细胞代谢过程的方法. 大量的生物信息数据推动了GEM从最初的基于纯计量的通量平衡矩阵方程, 发展到“基于约束的代谢模型构建与分析”. 未来对GEM的研究将整合更多的生化信息, 包括采用更广泛的GPR关联; 添加蛋白质的生物化学性质, 例如酶与底物相互作用、蛋白质复合物的结构和翻译后修饰等; 开发更高效的数学建模技术等. 这些数据信息的扩充和建模方法的升级, 将进一步加快GEM模型重构的速度并大幅提升模型的质量.

细胞代谢反应的真实通量由代谢物浓度、酶动力学以及酶浓度决定, 而有效的酶浓度又取决于基因表达、转录和翻译、翻译后修饰和降解. GEM与组学数据相结合可以更准确地模拟生物的生理代谢过程, 对研究细胞代谢具有重要意义. 虽然现有的整合组学数据的方法提出了不同的数学公式, 用于解决基因蛋白表达水平对代谢反应通量约束的问题, 但是这些方法均有其特定的应用场景, 仍缺乏一种比较完善的代谢网络模型整合算法. 由于细胞代谢调控种类和方式复杂且多样, 仅考虑基因或蛋白的表达水平不能够完全反映代谢通量的真实水平, 因此需要集成更多的调控信息以提升模型模拟的准确性. 该领域未来的发展方向将是深度扩展组学数据与GEM的集成, 通过机器学习建立多组学数据源之间的关联; 结合多组学数据和动力学参数到基于约束的模型中, 充分考虑转录产物和其他代谢成分之间的相互作用; 整合更为完善的多层次调控关系网络信息, 扩增模型的模拟信息维度; 测定生理参数和细胞内通量, 将参数与通量分析相结合预测代谢过程. 更多生物信息与GEM的集成将推动代谢网络模型从单一维度向多维度、从静态向动态以及从定性向定量发展, 并为合成生物学研究中的“细胞

代谢重编程”提供一种高效设计分析工具, 指导微生物生产化学品、燃料和材料的细胞工厂, 推动绿色生物制造的发展。

参考文献

- Orth J D, Thiele I, Palsson B O. What is flux balance analysis? *Nat Biotechnol*, 2010, 28: 245–248
- Gudmundsson S, Thiele I. Computationally efficient flux variability analysis. *BMC Bioinformatics*, 2010, 11: 489
- Segrè D, Vitkup D, Church G M. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA*, 2002, 99: 15112–15117
- Ye C, Xu N, Chen X L, et al. Application of metabolic network model to analyze intracellular metabolism of industrial microorganisms (in Chinese). *Chin J Biotechnol*, 2019, 35: 1901–1913 [叶超, 徐楠, 陈修来, 等. 应用代谢网络模型解析工业微生物胞内代谢. *生物工程学报*, 2019, 35: 1901–1913]
- Thiele I, Palsson B O. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*, 2010, 5: 93–121
- Hao T, Ma H W, Zhao X M. Progress in automatic reconstruction and analysis tools of genome-scale metabolic network (in Chinese). *Chin J Biotechnol*, 2012, 28: 661–670 [郝彤, 马红武, 赵学明. 基因组尺度代谢网络自动重构及分析工具研究进展. *生物工程学报*, 2012, 28: 661–670]
- Machado D, Andrejev S, Tramontano M, et al. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucl Acids Res*, 2018, 46: 7542–7553
- Büchel F, Rodriguez N, Swainston N, et al. Path2Models: Large-scale generation of computational models from biochemical pathway maps. *BMC Syst Biol*, 2013, 7: 116
- Henry C S, DeJongh M, Best A A, et al. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol*, 2010, 28: 977–982
- Magnúsdóttir S, Heinken A, Kutt L, et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat Biotechnol*, 2017, 35: 81–89
- Wang H, Marcišauskas S, Sánchez B J, et al. RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLoS Comput Biol*, 2018, 14: e1006541
- Swainston N, Smallbone K, Mendes P, et al. The SuBliMinaL Toolbox: Automating steps in the reconstruction of metabolic networks. *J Integr Bioinf*, 2011, 8: 186
- Lieven C, Beber M E, Olivier B G, et al. Publisher Correction: MEMOTE for standardized genome-scale metabolic model testing. *Nat Biotechnol*, 2020, 38: 504
- Schellenberger J, Park J O, Conrad T M, et al. BiGG: A biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, 2010, 11: 213
- King Z A, Lu J, Drager A, et al. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucl Acids Res*, 2016, 44: 515–522
- Norsigian C J, Pusarla N, McConn J L, et al. BiGG Models 2020: Multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucl Acids Res*, 2020, 48: 402–406
- Ganter M, Bernard T, Moretti S, et al. MetaNetX.org: A website and repository for accessing, analysing and manipulating metabolic networks. *Bioinformatics*, 2013, 29: 815–816
- Edwards J S, Palsson B O. The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA*, 2000, 97: 5528–5533
- Reed J L, Vo T D, Schilling C H, et al. An expanded genome-scale model of *Escherichia coli* K-12 (*iJR904* GSM/GPR). *Genome Biol*, 2003, 4: 54
- Feist A M, Henry C S, Reed J L, et al. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol*, 2007, 3: 121
- Orth J D, Conrad T M, Na J, et al. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol Syst Biol*, 2011, 7: 535
- O’Brien E J, Lerman J A, Chang R L, et al. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol*, 2013, 9: 693
- Weaver D S, Keseler I M, Mackie A, et al. A genome-scale metabolic flux model of *Escherichia coli* K-12 derived from the EcoCyc database. *BMC Syst Biol*, 2014, 8: 79
- Liu J K, O’Brien E J, Lerman J A, et al. Reconstruction and modeling protein translocation and compartmentalization in *Escherichia coli* at the

- genome-scale. *BMC Syst Biol*, 2014, 8: 110
- 25 Monk J M, Lloyd C J, Brunk E, et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat Biotechnol*, 2017, 35: 904–908
- 26 Lloyd C J, Ebrahim A, Yang L, et al. COBRAme: A computational framework for genome-scale models of metabolism and gene expression. *PLoS Comput Biol*, 2018, 14: e1006302
- 27 Kjeldsen K R, Nielsen J. *In silico* genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. *Biotechnol Bioeng*, 2009, 102: 583–597
- 28 Shinfuku Y, Sorpitopon N, Sono M, et al. Development and experimental verification of a genome-scale metabolic model for *Corynebacterium glutamicum*. *Microb Cell Fact*, 2009, 8: 43
- 29 Mei J, Xu N, Ye C, et al. Reconstruction and analysis of a genome-scale metabolic network of *Corynebacterium glutamicum* S9114. *Gene*, 2016, 575: 615–622
- 30 Zhang Y, Cai J, Shang X, et al. A new genome-scale metabolic model of *Corynebacterium glutamicum* and its application. *Biotechnol Biofuels*, 2017, 10: 169
- 31 Lerman J A, Hyduke D R, Latif H, et al. *In silico* method for modelling metabolism and gene product expression at genome scale. *Nat Commun*, 2012, 3: 929
- 32 Thiele I, Fleming R M T, Que R, et al. Multiscale modeling of metabolism and macromolecular synthesis in *E. coli* and its application to the evolution of codon usage. *PLoS One*, 2012, 7: e45635
- 33 Ryu J Y, Kim H U, Lee S Y. Reconstruction of genome-scale human metabolic models using omics data. *Integr Biol*, 2015, 7: 859–868
- 34 Machado D, Herrgård M. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput Biol*, 2014, 10: e1003580
- 35 Akesson M, Förster J, Nielsen J. Integration of gene expression data into genome-scale metabolic models. *Metab Eng*, 2004, 6: 285–293
- 36 Becker S A, Palsson B O. Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol*, 2008, 4: e1000082
- 37 Shlomi T, Cabili M N, Herrgård M J, et al. Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol*, 2008, 26: 1003–1010
- 38 Zur H, Rupp E, Shlomi T. iMAT: An integrative metabolic analysis tool. *Bioinformatics*, 2010, 26: 3140–3142
- 39 Moxley J F, Jewett M C, Antoniewicz M R, et al. Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. *Proc Natl Acad Sci USA*, 2009, 106: 6477–6482
- 40 Colijn C, Brandes A, Zucker J, et al. Interpreting expression data with metabolic flux models: Predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS Comput Biol*, 2009, 5: e1000489
- 41 Chandrasekaran S, Price N D. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA*, 2010, 107: 17845–17850
- 42 Jerby L, Shlomi T, Rupp E. Computational reconstruction of tissue-specific metabolic models: Application to human liver metabolism. *Mol Syst Biol*, 2010, 6: 401
- 43 Jensen P A, Papin J A. Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics*, 2011, 27: 541–547
- 44 Jensen P A, Lutz K A, Papin J A. TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. *BMC Syst Biol*, 2011, 5: 147
- 45 van Berlo R J P, de Ridder D, Daran J M, et al. Predicting metabolic fluxes using gene expression differences as constraints. *IEEE/ACM Trans Comput Biol Bioinf*, 2011, 8: 206–216
- 46 Jerby L, Wolf L, Denkert C, et al. Metabolic associations of reduced proliferation and oxidative stress in advanced breast cancer. *Cancer Res*, 2012, 72: 5712–5720
- 47 Agren R, Bordel S, Mardinoglu A, et al. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput Biol*, 2012, 8: e1002518
- 48 Lee D, Smallbone K, Dunn W B, et al. Improving metabolic flux predictions using absolute gene expression data. *BMC Syst Biol*, 2012, 6: 73
- 49 Kim J, Reed J L. RELATCH: Relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. *Genome Biol*, 2012, 13: 78
- 50 Collins S B, Reznik E, Segrè D. Temporal expression-based analysis of metabolism. *PLoS Comput Biol*, 2012, 8: e1002781
- 51 Töpfer N, Jozefczuk S, Nikoloski Z. Integration of time-resolved transcriptomics data with flux-based methods reveals stress-induced metabolic adaptation in *Escherichia coli*. *BMC Syst Biol*, 2012, 6: 148
- 52 Navid A, Almaas E. Genome-level transcription data of *Yersinia pestis* analyzed with a new metabolic constraint-based approach. *BMC Syst Biol*, 2012, 6: 150
- 53 Wang Y, Eddy J A, Price N D. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst Biol*, 2012, 6: 153

- 54 Kim H U, Kim W J, Lee S Y. Flux-coupled genes and their use in metabolic flux analysis. *Biotechnol J*, 2013, 8: 1035–1042
- 55 Rossell S, Huynen M A, Notebaart R A. Inferring metabolic states in uncharacterized environments using gene-expression measurements. *PLoS Comput Biol*, 2013, 9: e1002988
- 56 Schmidt B J, Ebrahim A, Metz T O, et al. GIM3E: Condition-specific models of cellular metabolism developed from metabolomics and expression data. *Bioinformatics*, 2013, 29: 2900–2908
- 57 Agren R, Mardinoglu A, Asplund A, et al. Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol Syst Biol*, 2014, 10: 721
- 58 Vlassis N, Pacheco M P, Sauter T. Fast reconstruction of compact context-specific metabolic network models. *PLoS Comput Biol*, 2014, 10: e1003424
- 59 Song H S, Reifman J, Wallqvist A. Prediction of metabolic flux distribution from gene expression data based on the flux minimization principle. *PLoS One*, 2014, 9: e112524
- 60 Aurich M K, Paglia G, Rolfsson Ó, et al. Prediction of intracellular metabolic states from extracellular metabolomic data. *Metabolomics*, 2015, 11: 603–619
- 61 Barker B E, Sadagopan N, Wang Y, et al. A robust and efficient method for estimating enzyme complex abundance and metabolic flux from expression data. *Comput Biol Chem*, 2015, 59: 98–112
- 62 Ma S, Minch K J, Rustad T R, et al. Integrated modeling of gene regulatory and metabolic networks in *Mycobacterium tuberculosis*. *PLoS Comput Biol*, 2015, 11: e1004543
- 63 Angione C, Lió P. Predictive analytics of environmental adaptability in multi-omic network models. *Sci Rep*, 2015, 5: 15147
- 64 Marmiesse L, Peyraud R, Cottret L. FlexFlux: Combining metabolic flux and regulatory network analyses. *BMC Syst Biol*, 2015, 9: 93
- 65 Robaina Estévez S, Nikoloski Z. Context-specific metabolic model extraction based on regularized least squares optimization. *PLoS One*, 2015, 10: e0131875
- 66 Gavai A K, Supandi F, Hettling H, et al. Using bioconductor package BiGGR for metabolic flux estimation based on gene expression changes in brain. *PLoS One*, 2015, 10: e0119016
- 67 Pacheco M P, John E, Kaoma T, et al. Integrated metabolic modelling reveals cell-type specific epigenetic control points of the macrophage metabolic network. *BMC Genomics*, 2015, 16: 809
- 68 Tobalina L, Bargiela R, Pey J, et al. Context-specific metabolic network reconstruction of a naphthalene-degrading bacterial community guided by metaproteomic data. *Bioinformatics*, 2015, 31: 1771–1779
- 69 Schultz A, Qutub A A. Reconstruction of tissue-specific metabolic networks using CORDA. *PLoS Comput Biol*, 2016, 12: e1004808
- 70 Blais E M, Rawls K D, Dougherty B V, et al. Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nat Commun*, 2017, 8: 14250
- 71 Zhang S W, Gou W L, Li Y. Prediction of metabolic fluxes from gene expression data with Huber penalty convex optimization function. *Mol Biosyst*, 2017, 13: 901–909
- 72 Motamedian E, Mohammadi M, Shojaosadati S A, et al. TRFBA: An algorithm to integrate genome-scale metabolic and transcriptional regulatory networks with incorporation of expression data. *Bioinformatics*, 2017, 33: 1057–1063
- 73 Banos D T, Trébulle P, Elati M. Integrating transcriptional activity in genome-scale models of metabolism. *BMC Syst Biol*, 2017, 11: 134
- 74 Wang Z, Danziger S A, Heavner B D, et al. Combining inferred regulatory and reconstructed metabolic networks enhances phenotype prediction in yeast. *PLoS Comput Biol*, 2017, 13: e1005489
- 75 Tian M Y, Reed J L. Integrating proteomic or transcriptomic data into metabolic models using linear bound flux balance analysis. *Bioinformatics*, 2018, 34: 3882–3888
- 76 Yang L, Ebrahim A, Lloyd C J, et al. DynamicME: Dynamic simulation and refinement of integrated models of metabolism and protein expression. *BMC Syst Biol*, 2019, 13: 2
- 77 Jenior M L, Moutinho T J, Dougherty B V, et al. Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. *PLoS Comput Biol*, 2020, 16: e1007099
- 78 Pusa T, Ferrarini M G, Andrade R, et al. MOOMIN—Mathematical explORation of 'Omics data on a Metabolic Network. *Bioinformatics*, 2020, 36: 514–523
- 79 Wegrzyn A B, Herzog K, Gerding A, et al. Fibroblast-specific genome-scale modelling predicts an imbalance in amino acid metabolism in Refsum disease. *FEBS J*, 2020, 287: 5096–5113

Summary for “整合组学数据的代谢网络模型研究进展”

Progress on genome-scale metabolic models integrated with multi-omics data

Xueliang Wang^{1,2}, Yun Zhang^{1,3*} & Tingyi Wen^{1,3,4*}¹ Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China;² University of Chinese Academy of Sciences, Beijing 100049, China;³ Innovation Academy for Green Manufacture, Chinese Academy of Sciences, Beijing 100190, China;⁴ Savaid Medical School, University of Chinese Academy of Sciences, Beijing 100049, China* Corresponding authors, E-mail: zhangyun@im.ac.cn; wenty@im.ac.cn

The genome-scale metabolic network model (GEM) is a mathematical framework based on gene-protein-reaction associations combined with stoichiometric balance and is capable of facilitating the computation and prediction of multiscale phenotypes by optimizing the objective function of interest. It has been increasingly used as an important tool for understanding cellular metabolism and characterizing cell phenotypes. In cells, metabolism is tightly controlled by intricate regulatory mechanisms at the different system levels and is strictly regulated to ensure the dynamic adaptation of biochemical reaction fluxes for maintaining cell homeostasis to ultimately achieve optimal metabolic fitness. Advances in high-throughput screening and analysis technologies have generated massive amounts of genome sequences, along with transcriptomic, proteomic and metabolomic data, providing quantitative regulatory information to gain insights into cellular metabolism; however, integrating the available omics data into constraint-based metabolic models and quantitatively profiling genotype-phenotype relationships remains an outstanding challenge for computational biology. Here, we describe the recent developments in introducing macromolecular expression into GEMs and generating metabolic expression (ME) models, which increase the complexity and predictive capability of computational frameworks. Various algorithms employ different approaches to combine additional layers of omics data to limit the cone of allowable flux distributions in the metabolic model. In this review, we categorize all methods by five different grouping criteria and evaluate their practical perspectives. The first category of methods utilizes a threshold to distinguish active and inactive states of the corresponding reactions based on the gene expression measurement data. The second uses omics data to build cell- and tissue-specific models of human metabolism by removing unexpressed reactions from the global human metabolic network. The third category of methods involves modifying reaction bounds on the basis of mRNA and protein abundance, in which the width of the “flux cone” is adjusted via the maximum possible flux in the upper bound of the FBA optimization problem dependent on gene and protein expression levels. The imposition of constraints further defines the associated solution space of the model to improve the prediction accuracy. The fourth model incorporates transcriptional regulation networks (TRNs), which describe the phenomenological interactions between different biomolecules in response to genetic and environmental perturbation, into GEMs and avoids the obstacles of information formulation to achieve comprehensive knowledge regarding the metabolic and regulatory events occurring inside the cell. The last category integrates time-series transcriptomics data with flux-based bilevel optimization to comprehend the interplay between metabolism and regulation in time-dependent processes. We compare the advantages and limitations of different categories and explore the application areas of integrated models in analyzing metabolic characteristics, interpreting phenotypic states and the consequences of environmental and genetic perturbations while discovering potential drug targets and screening anti-metabolic drugs for cancer treatment. Finally, we also highlight the future perspectives and challenges for GEM-based reconstruction with omics data integration.

genome-scale metabolic model, transcriptomics, proteomics, metabolomics, omics data integrationdoi: [10.1360/TB-2020-1468](https://doi.org/10.1360/TB-2020-1468)