遗传 HEREDITAS (Beijing) 2013 年 1 月, 35(1): 17－26
ISSN 0253-9772 www.chinagene.cn

综 述

DOI: 10.3724/SP.J.1005.2013.00017

蛋白质泛素化修饰的生物信息学研究进展

卢亮, 李栋, 贺福初

军事医学科学院放射与辐射医学研究所, 北京蛋白质组研究中心, 蛋白质组学国家重点实验室, 北京 102206

摘要: 泛素-蛋白酶体系统(Ubiquitin-proteasome system, UPS)介导了真核生物 80%~85%的蛋白质降解, 该蛋白质降解途径具有依赖 ATP、高效、高度选择性的特点。除参与蛋白质降解之外, 泛素化修饰还可以直接影响蛋白质的活性和定位。由于泛素化修饰底物蛋白在细胞中的广泛存在, 泛素化修饰可以调控包括细胞周期、细胞凋亡、转录调控、DNA 损伤修复以及免疫应答等在内的多种细胞活动。近年来, 泛素-蛋白酶体系统相关的蛋白质组学数据不断产出, 有效地管理、组织并合理分析这些数据显得尤为必要。文章综述了当前世界范围内针对蛋白质泛素化修饰展开的生物信息学研究, 总结了前人的工作结果, 包括 UPS 相关蛋白质数据的收录、泛素化修饰网络的构建和分析、泛素化修饰位点的预测及泛素化修饰 motif 的研究等方面内容, 并对该领域未来的发展方向进行了讨论。

关键词: 泛素-蛋白酶体系统(UPS); 泛素化修饰; 生物信息学

Bioinformatics advances in protein ubiquitination

LU Liang, LI Dong, HE Fu-Chu

State Key Laboratory of Proteomics, *Beijing Proteome Research Center*, *Beijing Institute of Radiation Medicine*, *Beijing* 102206, *China*

Abstract: Ubiquitin-proteasome system (UPS) mediates 80% to 85% of the protein degradation in eukaryotic cells. The characteristics of UPS pathway are dependent on ATP, efficient and highly selective. Ubiquitination not only participates in protein degradation, but also directly affects protein activity and localization. Ubiquitination can regulate multiple cellular processes including cell cycle progression, apoptosis, transcriptional regulation, DNA damage repair and immune response. More and more datasets about UPS are published, and it is necessary to organize and analyze these data efficiently. We review the related bioinformatics studies in UPS datasets, such as collection of UPS related proteins, construction and analysis of ubiquitination networks, prediction of ubiquitination sites and motifs. Some potential perspectives are also discussed.

Keywords: Ubiquitin-proteasome system (UPS); ubiquitination; bioinformatics

收稿日期: 2012-04-28; 修回日期: 2012-06-20

基金项目: 国家重大科学研究计划项目(编号：2012CB910300, 2011CB910202), 国家高技术研究发展计划项目(“863”计划)(编号：2012AA020201)和蛋白质组学国家重点实验室开放课题基金项目(编号：SKLP-O201106)资助

作者简介: 卢亮, 硕士研究生, 专业方向：泛素化修饰网络的构建。Tel: 010-80727777-1125; E-mail: pikaliang@163.com

通讯作者: 贺福初, 研究员, 中国科学院院士, 研究方向：蛋白质组学。E-mail: hefc@nic.bmi.ac.cn

李栋, 副研究员, 研究方向：生物学网络。E-mail: lidong.bprc@foxmail.com

网络出版时间: 2012-8-9 10:44:53

URL: http://www.cnki.net/kcms/detail/11.1913.R.20120809.1044.001.html

泛素(Ubiquitin)是一种由 76 个氨基酸构成、在真核生物中广泛存在并具有高度保守性的多肽[1]。一个或多个泛素分子在一系列酶的作用下与底物蛋白质分子共价结合的翻译后修饰过程称为泛素化修饰(Ubiquitination/Ubiquitylation)。泛素化修饰最早被发现的功能是标记靶蛋白，使之被蛋白酶体识别并降解，整个过程涉及泛素分子、底物蛋白、多种酶系统(如泛素激活酶(Ubiquitin-activating enzyme, E1)、泛素结合酶(Ubiquitin-conjugating enzyme, E2)、泛素连接酶(Ubiquitin-protein ligase, E3)、去泛素化酶(Deubiquitinating enzyme, DUB))，以及蛋白酶体，它们共同构成了泛素-蛋白酶体系统(Ubiquitin-proteasome system, UPS)[2~5]。泛素-蛋白酶体系统介导了真核生物体内 80%~85%的蛋白质降解，该蛋白质降解途径具有依赖ATP、高效、高度选择性的特点。泛素化修饰是一种重要的翻译后修饰，其扮演的角色远比最初认识的要多得多，还参与调控诸如转录因子活性[6]、逆转录病毒的出芽[7]、受体的内吞作用和溶酶体的运输[8]、胰岛素水平[9]和TGF-β通路[10]等。研究表明泛素化修饰的作用已不仅仅局限于介导蛋白质的降解[11, 12]，泛素化修饰还可以直接影响蛋白质的活性和定位。蛋白质修饰及降解贯穿几乎所有细胞活动，包括基因表达、细胞增殖、分化、衰老、凋亡和自噬，也调控着个体的生长、发育和衰老。蛋白质的调控异常往往伴随着个体的发育异常或疾病的发生，与各种人类重大疾病的发生、发展、诊断和治疗密切相关。由于底物蛋白广泛存在于细胞中，泛素化修饰调控的细胞活动主要包括细胞周期、细胞凋亡、转录调控、DNA损伤修复及免疫应答等[13~17]。

随着泛素化修饰研究的深入，越来越多泛素化修饰相关的蛋白质组学数据不断产出，对这些数据进行合理地组织、存储、管理以及分析显得十分必要，现今泛素化修饰的生物信息学研究主要围绕泛素化修饰相关蛋白质数据库的建立、泛素化修饰网络的构建和分析以及泛素化修饰位点的预测等方面展开，随着实验数据的产出及生物信息学技术的发展，对泛素化相关蛋白质数据的解读将更为全面，帮助科研工作者更为深入地理解泛素化修饰的机制。

1 泛素-蛋白酶体系统相关蛋白质数据的收录

随着世界范围内泛素化修饰相关的研究报道越来越多，泛素-蛋白酶体系统相关蛋白质数据不断产出，合理地组织、储存、呈现以及更新这些数据显得很有必要，这也是进行后续数据分析工作的基础，针对这一问题，很多实验室予以关注，形成了一些较有特色的数据库，这里予以简要介绍。

1.1 UbiProt：泛素化修饰底物蛋白数据库

UbiProt(http://ubiprot.org.ru/)[18]是一个收录泛素化修饰底物蛋白的数据库，由俄罗斯下哥罗德国立医学院创建。该数据库旨在系统地组织收录泛素化底物蛋白数据，供深入研究使用，它的建立对泛素化领域的研究具有重要意义。UbiProt数据库中每个数据条目描述了特定的泛素化底物蛋白信息，包括蛋白的性质、物种来源、泛素化修饰特征、参考文献及相关链接等。该数据库主要包含了一些大规模组学实验数据，其余的底物蛋白数据是从针对特定蛋白的泛素化修饰实验研究中得到的，数据库所有的条目都是通过人工手动的方式进行提取并注释的。UbiProt数据库以泛素化底物数据为主要收录对象，但是该数据库数据信息并不完善，比如很多条目的底物并不包含相应的E1、E2、E3 信息，且数据较为陈旧，后续更新情况不理想，可能是由于其采用的是人工提取的方法并缺少及时维护。

该数据库收录的主要是酵母和人的泛素化底物蛋白数据，其中来源于酵母的数据有 948 个，来源于人的数据有 137 个，其中来源于人的泛素化底物数据规模很小，并且相应条目的注释信息不够完善，加之该数据库长久不更新，使用起来较为不便，但是该数据库的建立具有开创意义，它是首个专门针对泛素-蛋白酶体系统进行数据收集的数据库，首次尝试将分散的泛素化底物蛋白数据进行组织整理，且该数据库具备数据提交的功能，UbiProt 的发布对于泛素化修饰的生物信息学研究具有重要意义。

1.2 hUbiquitome：实验验证的人类泛素化相关蛋白质数据库

hUbiquitome(http://202.38.126.151/hmdd/hubi/)[19]是北京大学发布的一个数据库，旨在收录高可信度的实验验证的人类泛素化相关蛋白质，该数据库共

收录了 1 个E1、12 个E2、138 个E3、279 个底物蛋白以及 17 个去泛素化酶。该数据库规模较小，但可信度较高，由于收录的全部为实验验证的数据，且为人工手动录入，提高了数据库所收录数据的可靠性，但同时降低了数据量，倘若加入质量控制体系，多渠道收集数据并对数据进行打分，将对提高数据量有所帮助。此外，该数据库的条目注释信息较少，每个条目包含的相关信息不多，但该数据库包含了数据提交功能，更为友好的是提供了原始数据下载的功能，这是相对于其它泛素化相关蛋白质数据库显著的优点。

1.3　E3Miner：使用文本挖掘方法建立的泛素化 E3 数据库

严格地讲，E3Miner(http://e3miner.biopathway.org/e3miner.html)[20]是一个提取泛素化修饰E3 信息的在线文本挖掘工具，由于其较好地整理并呈现了已挖掘到的泛素化修饰E3 相关信息，在此将其作为泛素化修饰相关蛋白质数据库之一进行介绍。

考虑到E3 在泛素化修饰通路中的重要性，关于E3 新的研究报道不断涌现，然而过去并没有针对泛素化修饰 E3 的特定的文本挖掘工具，针对这一问题，韩国科学技术院计算机技术系开发了 E3Miner 文本挖掘工具，旨在挖掘出现在海量文献中的 E3 信息，通过文本挖掘 MEDLINE 中泛素化修饰相关文献的摘要来获得泛素化修饰 E3 及其相应底物蛋白、E2、E1 以及 DUB 信息。该数据库的建设过程中首次将文本挖掘的方法应用于泛素化修饰相关蛋白数据的挖掘，并做成了在线工具。由于文本挖掘方法会补充大量的数据，将分散在各文献报道中的数据提取出来，该方法在各种数据库建设过程中一直扮演较为重要的角色，然而该方法同时也存在可信度较低的缺点。E3Miner 开创了泛素化修饰研究领域应用文本挖掘方法的先例，这对于后续相关数据库的建设是一个很好的基础。

如之前介绍，该数据库的建立是基于文本挖掘的方法，为了展示其挖掘算法具有较好的正确率，每个 E3 条目的注释信息中给出了对应句子的原文出处，这对使用者定位到具体出处的文献较为便利，然而经过测试发现该数据库提供的 E3 底物信息与 UniProt 数据库相应条目注释段提供的底物信息重合度较低，即该数据库采用的文本挖掘算法召回率较低，很多已被报道的 E3、底物蛋白数据并未被准确挖掘，此外该数据库同样存在更新维护不及时的缺点，2007 年 8 月 26 日后便再无更新，且不支持数据的批量下载，这使得其可用性降低，但是它首先将文本挖掘的方法引入泛素化相关蛋白数据的收集是具有启发意义的，是为后续相关工作奠定基础之举。

1.4　E3Net：升级的 E3Miner

针对E3Miner收录数据不完全，与UniProt注释信息重合度低的缺点，韩国科学技术院生物信息学实验室在E3Miner的基础上设计了E3Net(http://pnet.kaist.ac.kr/e3net/)[21]，该数据库在数据量方面较E3Miner有显著提升，且界面更为友好美观，从数据量及注释信息的丰富程度上来说，该数据库都是当前泛素化修饰相关蛋白数据库中最为出色的，共收录了 427 个物种中 2 201 个泛素化修饰E3 及 4 896 个底物蛋白信息，其中包含 493 个E3 与 1 277 个底物蛋白之间的 1 671 个特异选择关系，该数据库的数据来源主要是文本挖掘方法挖掘MEDLINE摘要得到的结果、UniProt相关条目注释信息、公共泛素化数据库收录数据以及高通量实验数据。

该数据库收录的数据侧重于描述泛素化修饰 E3 与底物蛋白之间的对应关系，同时给出了相应的生物学功能分析，然而该数据库缺少 E3 相对应的 E2 信息，从整个泛素-蛋白酶体系统的角度看数据不够全面，注释信息的完整性也有待于进一步完善，在数据的更新及下载方面同样存在需要改进之处，不过，该数据库是当前该领域最完全丰富的泛素化系统相关蛋白数据库，对于泛素化修饰领域的研究具有重要意义。

综合以上各数据库的特点可以发现，目前仍没有一个综合、全面的泛素-蛋白酶体系统相关蛋白质数据库，各个数据库均有所侧重，如 UbiProt 侧重收录泛素化底物信息，hUbiquitome 侧重人的泛素化数据，注重数据可靠性的同时降低了数据量，E3Miner 和 E3Net 侧重收录 E3 信息，数据库条目以 E3 为主键。以上各数据库都存在更新状况较差的问题，数据库发布后的维护工作不够，且注释信息都在不同程度上存在不够完善的问题，这其中，尽管 UbiProt 和 hUbiquitome 均在网站上设计了数据提交功能，

但鲜有应用，且除 hUbiquitome 外其他 3 个数据库均无数据批量下载功能，这反映了这些数据库不够开放的缺点，或与其发布的定位有关。倘若如 UniProt 等综合数据库一样，定位于专注做完整的丰富的数据库，将泛素-蛋白酶体系统相关蛋白数据予以全面收录，并完善更新维护工作，对泛素化领域的科研工作者来说将会是一个福音。

2 泛素化修饰网络的构建和分析

2.1 在酵母中尝试构建泛素化修饰网络

泛素化修饰网络可被描述为包括泛素、类泛素蛋白质(包括他们的结合、去结合状态)、所结合底物蛋白、关键酶以及蛋白酶体所构成的系统。较早时候部分研究致力于泛素化系统结构和功能的关系以及个别组分的进化信息，把泛素化系统作为一个体系进行整体研究的工作尚属空白。针对这一现象，美国国家卫生研究院的Venancio等[22]以模式生物酵母为研究对象构建了第一个全面的综合的泛素化修饰网络，并对网络进行了整体性质的分析，获得了诸多较有启发性的结果。Venancio开创了以生物学网络的方法研究泛素化修饰系统整体性质的工作，为后续相关研究奠定了良好的基础。

除了泛素分子之外，还存在很多类泛素分子，如SUMO、Nedd8、Urm1[23]，泛素化修饰与类泛素化修饰有多种形式且可对底物蛋白产生不同的影响。例如，通过第 48 位赖氨酸形成的多聚泛素链参与的泛素化修饰，以及neddylation修饰和urmylation修饰可使底物蛋白被蛋白酶体招募并水解，与之相对应，通过第 63 位赖氨酸形成的多聚泛素链以及单个泛素分子参与的泛素化修饰，以及sumoylation修饰，会改变底物蛋白的性质及相互作用关系，由此产生初级的调控作用[24]。Sumoylation已被证实参与若干生物学功能的调控，诸如核质运输、细胞周期进程、核孔复合体相关的相互作用、DNA修复和复制以及mRNA质量控制[25~27]，其他泛素分子如Apg12 形成的类泛素化修饰会调控特定的生物学进程，如自噬作用[28]。各种类泛素化修饰在细胞活动中同样扮演重要角色，Venancio研究的泛素化修饰网络是广义上的泛素化修饰网络，包括了对类泛素化修饰的研究。

Venancio将包含泛素化修饰核心组分(直接参与泛素结合、去除、循环利用以及相关复合体构成的组分)的系统称作泛素系统(Ub-system)，尽管之前有其他实验室研究报道泛素系统中个别组分的起源及进化信息[29~31]，但是鲜有研究试图从整体角度来描述泛素化系统。随着实验技术的发展，至少对于被深入研究的模式生物酵母，研究泛素化系统的整体性质已具备了可行性。Venancio研究了大量的泛素化系统相关数据来广泛地确定酵母的泛素系统组分，然后将它们已知的物理、遗传以及生化相互作用进行整合，并采用图表和网络的形式呈现这些数据，如此一来通过图算法的理论从整体上理解泛素化系统，这套思路在研究其他系统如转录调控网络和蛋白质相互作用网络已被证实可行[32~34]。研究据此呈现了泛素化系统的网络结构，通过分析泛素化网络，研究揭示了泛素化系统一些有趣的生物学特征，包括之前未知的泛素化系统组分之间的作用关系以及与其它调控机制之间的交互，最终，Venancio构建了第一个全面的综合的泛素化系统网络。

通过建立第一个综合的模式生物酵母的泛素化网络，Venancio 第一次对泛素化网络进行了整体的探索而非局限于个别组分的研究。结果，能够对泛素化网络里的子网络进行定量的了解，并从网络中发现了同源基因的多样化生化性质及具有类似生物学功能的组分。最终，研究还将泛素化修饰系统与其他调控系统诸如转录调控系统进行了对比，以及通过泛素化网络的结构性质理解特定组分的进化趋势。该研究意义重大，对其他真核生物的泛素化网络的构建具有指导意义。

2.2 人类泛素-蛋白酶体系统 E2 与 RING 类 E3 相互作用网络

针对泛素化修饰系统中关键酶之间的特异选择性，Timmers等[35]报道了E2 与RING类E3 的相互作用网络，这对于理解泛素化修饰系统中关键酶的级联具有指导意义。

泛素化修饰E2 与E3 之间选择性的相互作用是泛素化修饰通路的关键一步，目前有限的实验研究线索表明，E2 结构中的UBC折叠(fold)在E2 与E3 的相互作用中发挥重要作用[36~41]，另一方面，RING类E3 的一端通过RING-finger结构域结合搭载泛素分

子的E2，另一端结合特定的底物蛋白，起到类似脚手架的作用[42]，现有信息表明，E2、E3 之间的选择性结合取决于这两个结构域中的关键残基，这是 E2、E3 之间高度选择性的分子层次的原因。在 Timmers的研究中，特定的相互作用被限定在E2 的 UBC折叠和E3 的RING-finger结构域之间[43]，研究对象共包含 35 个E2 UBC折叠和 250 个E3 RING-finger结构域。基因组规模的蛋白质相互作用实验可帮助深入理解E2 与E3 的相互选择关系。Timmers旨在构建泛素化E2 与E3 之间的相互作用网络以便深入分析E2 与E3 之间的这种选择关系，研究使用了酵母双杂交实验手段进行筛选[39,44]。

研究结果显示超过 300 个 E2-E3 相互作用，其中大部分是之前所未知的，正如所预料，多数 E2、E3 可与多个 E3、E2 相互作用，其中一些 E2、E3 较为活跃，会与更多的 E3、E2 发生相互作用，进而联想到蛋白质相互作用网络中的节点蛋白。研究发现的 E2-E3 选择关系与之前体外泛素化实验的相关报道具有高度的一致性，其中一个与较多 E3 有相互作用关系的 E2，UBE2U，在实验中验证是存在最多相互作用 E3 的 E2。研究通过使用 GST pull down 实验方法验证了 E2, UBE2U 与 E3, MDM2 之间的相互作用，以及一些节点 E2 与 E3 之间的相互作用，证实了实验的高度可靠性。这个研究展示的相互作用提供了一个全局角度的 E2 与 E3 之间的选择性相互作用网络。

Timmers 的研究报道了高质量的聚焦于 E2 与 E3 选择性的相互作用网络框架，提供了详尽的 E2 与 E3 相互作用关系，这对于深入理解泛素化修饰系统中关键酶之间的选择关联很有意义，这项研究可以启发后续关于 E2 与 E3 选择性的研究，并可放大到整个泛素化修饰网络，用于鉴别泛素化修饰系统中相关组分之间的特异选择性，这是一项开创性的工作，对于认识泛素化修饰系统中关键酶之间的选择、E3 与底物之间的选择具有指导意义。

3　泛素化修饰位点的预测及泛素化修饰 motif 的研究

3.1　泛素化修饰位点的预测

由于泛素化修饰在调节细胞活动中发挥着重要作用，科研人员发展了大量的纯化底物蛋白的方法[45]。同时，大规模鉴定泛素化修饰底物蛋白和泛素相关蛋白质组学研究的研究热折射出鉴定泛素化修饰底物和修饰位点的重要性[46~53]。亲和纯化、蛋白水解消化、质谱分析被应用于大部分的研究中[54]。为了更为有效地探索更多未被发现的泛素化修饰位点，生物信息学预测的方法可为鉴定可能的泛素化位点带来帮助。这一方向已取得一些研究成果[55~59]，其中台湾交通大学生物信息学课题组Tung等[60]发展的泛素化修饰位点预测方法得到的结果较为理想。

正确使用有效分类特征和合适的分类器对设计有效的泛素化修饰位点预测体系是至关重要的。过去，大量的基于序列的特征被用来区分蛋白质的功能。例如，Auto-Motif程序使用了 6 类特征和支持向量机(SVM)来预测翻译后修饰[61]；POPI程序使用理化性质作为有效特征来预测肽段的免疫原性[62]。在 Tung的研究中，评估了 3 类可以从蛋白序列中提取的有效特征，分别为：传统的氨基酸序列信息[61,63]、进化信息[64,65]以及理化性质[62,66]。与此同时，评估了 3 种机器学习分类器：k-nearest neighbor、Naïve Bayes 以及 SVM。采用的泛素化数据集包含了 157 个泛素化修饰位点及 3 676 个推定非泛素化修饰位点，这些数据是从UbiProt数据库中 105 个蛋白质数据中提取出来的。结果显示最好的预测结果出自SVM分类器和全部理化性质的组合。

考虑到不相关的信息会影响分类器的分类效果，Tung 使用 IPMA 算法筛选出一个有益理化性质的子集来提高预测结果的准确度，同时这些经过筛选的理化性质作为分类指标有助于理解泛素化修饰的深层机制。通过使用 31 个筛选出的理化性质作为分类参数，研究建立了预测泛素化修饰位点的预测体系 UbiPred，其预测准确度达到 84.44%。相比之下，氨基酸序列与 SVM 结合的准确度为 65.57%，进化信息、全部理化性质对应的准确度分别为 66.33%、72.19%。除了预测准确度，受试者工作特征曲线(ROC 曲线)通常被用来评估分类器的区分能力。ROC 曲线下包含的面积越大，分类器的区分能力越高。UbiPred 的 ROC 值高达 0.85。

此外，Tung 应用了决策树方法 C5.0 来获得基于规则的知识并分析了 31 个有益的理化性质，5 个简洁易懂的规则提供了一个人类可以解读的方法来区

分泛素化位点和非泛素化位点。该预测方法是现有各种预测模型中结果较为理想的，这项工作对于相关的实验研究具有很好的指导，同时也有助于深入理解泛素化修饰的机制，该研究方向具有广阔的前景。

3.2 泛素化修饰 motif 的研究

随着泛素化修饰底物蛋白及修饰位点数据的不断产出，用生物信息学的方法分析这些数据、研究其内在规律已成为研究热点。在修饰位点的研究中，寻找泛素化修饰 motif 是一个有意义的方向。

翻译后修饰motif的研究在SUMO化修饰领域已取得较好结果。与泛素化修饰相似，SUMO化同样是一种重要的翻译后修饰，相比于泛素化，SUMO化已取得更为深入的研究进展。在SUMO化修饰中，已鉴定到修饰位点处一个一致的核心motif：ΨKXE[67]，其中Ψ代表一个氨基酸集合(异亮氨酸I、缬氨酸V、亮氨酸L、丙氨酸A、脯氨酸P或者蛋氨酸M)，K是SUMO化修饰位点，X代表该处可为任意氨基酸，E代表谷氨酸。值得注意的是，有些SUMO化修饰位点并不符合以上motif特征，符合上述motif特征的氨基酸残基也并不一定是SUMO化修饰位点。对于泛素化修饰，是否存在类似的motif是一个研究热点。

在Gygi等[68]的研究中，虽然证实泛素化修饰位点处并不存在motif特征，但修饰位点附近的氨基酸表达呈现出一定的趋势，相对于全部的赖氨酸位点，Gygi发现泛素化修饰位点处更倾向于呈现出局部净负电荷。在修饰位点左右 6 个氨基酸的距离内，酸性氨基酸如天冬氨酸(D)、谷氨酸(E)表达量增加，碱性氨基酸如赖氨酸(K)、精氨酸(R)、组氨酸(H)的表达量降低。尽管没有理想的具有预测性的修饰位点附近序列模体，但是Gygi发现修饰位点更倾向于集中在偏向N端一侧精氨酸表达量低、N端、C端赖氨酸和组氨酸表达量均较低的位置。

虽然 Gygi 对泛素化修饰 motif 的研究没有非常理想的结果，然而该方面的尝试不会就此终止，一方面，大量泛素化修饰位点还有待于发现，现阶段数据规模的限制是泛素化修饰 motif 不易被发现的原因之一；另一方面，现有寻找 motif 的方法也有待于优化提高，整体水平上泛素化修饰 motif 不存在并不表示泛素化修饰 motif 不会存在于特定的底物集团之中。新的寻找泛素化修饰 motif 的探索还在不断进行之中。

4 结语与展望

由于泛素化修饰对于调控细胞的生物学过程意义重大，对泛素化修饰机制的深入研究很有必要，尽管泛素化修饰已持续作为研究热点多年，但泛素化修饰的深层机制仍有待深入分析，如 E3 与底物的特异性选择问题、修饰位点的选择问题、泛素化修饰与其他翻译后修饰的交互应答(cross-talk)问题等，这些都有待于实验的研究，而在揭开泛素化修饰机制层层面纱的过程中，生物信息学方法会发挥重要的作用。随着实验产出的泛素化修饰相关蛋白数据越来越多，合理地组织、储存、呈现以及更新这些数据是首先要完成的工作，针对如此大量的数据，从生物学网络的角度分析这些数据间的内在联系和潜在规律也是研究的趋势所在，在以上工作的基础上，展开有效的底物蛋白及修饰位点的预测工作很有意义，这一研究结果将对实验研究起到很好的指导作用。用生物信息学的方法去探索数据中蕴含的规律已经在蛋白质相互作用网络等领域产生了意义非凡的影响，相信不断发展的实验技术与生物信息学方法的结合将会使我们对泛素化修饰的了解更为深入，实验与信息技术的结合将会给该领域带来深远的影响。

展望生物信息学在泛素化修饰领域可做出的贡献，我们认为有以下几个发展方向：

(1)综合的泛素化数据库的构建。

考虑到现有泛素化数据库的不完善之处，诸如：质控体系不够明确、收录对象过于局限(或针对底物蛋白或针对 E3)、条目注释信息不够丰富、更新维护不到位、数据散布于各个数据库之中等，综合的全面的泛素化修饰相关蛋白数据库有待于建立，这个数据库要具有数据来源可靠、数据量丰富、收录对象全面、条目注释信息完整、更新维护及时、访问下载方便等特点，这样的数据库将会给泛素化修饰研究领域的科研工作者带来极大的便利，有利于泛素化修饰相关蛋白质数据资源的发布和共享，对于推进泛素化修饰研究的发展有重要意义。

(2)泛素化修饰网络的整合分析。

考虑到泛素化相关蛋白数据的日益丰富，以生物学网络的方法组织、分析这些数据具备可行性与

必要性。这一研究方向具有重要意义，将会带来诸多重要发现：利用多种基因组尺度的数据集，重建人泛素化体系的蛋白质网络(通过公共数据库数据下载、文献检索和序列分析发现 UPS 的所有成员，建立一个最全的泛素化体系成员列表；利用遗传相互作用、蛋白质相互作用等数据建立泛素化体系成员之间的功能关联；整合利用蛋白组学核心技术平台产出的 E3-底物数据)，通过对网络拓扑结构、模块性质的分析发现泛素化体系中 E3 家族不同成员之间的功能关联，发现泛素化修饰和类泛素化修饰在功能关联上的差异，发现重要的泛素化系统成员，预测新的成员，同时在整个人蛋白网络中寻找泛素化体系成员与信号通路、调控通路的联系，为实验研究提供重要的研究线索。另外，泛素化修饰网络的构建将为研究泛素化体系的进化提供了条件，了解泛素化体系的跨物种保守性和物种特异性。

(3)泛素化修饰 E2 与 E3、E3 与底物之间特异性选择关系的揭示。

在泛素化系统中，E2 与 E3、E3 与底物之间具有特异选择性，特定的 E3 会对应特定的 1 个或多个底物，特定的底物又会对应特定的 1 个或多个 E3，同样的选择性存在于 E2 与 E3 之间，然而这种特异选择的机制还没有被揭示，基于怎样的分子结构形成的这些特异选择性还有待于深入研究，这其中生物信息学方法会给予一定的帮助。根据现有的 E2-E3、E3-底物之间的选择关系，通过生物信息学手段研究其内在规律，发现或者预测这种特异选择的机制，随着实验数据的不断产出及生物信息学技术的发展，这一设想终会成为可能，届时将会对深入理解泛素化修饰机制提供有意义的启示。

(4)泛素化修饰 motif 的研究及泛素化修饰位点的预测。

泛素化修饰 motif 是否存在尚未可知，但围绕其展开的探索从未停止。尽管 Gygi 在其研究结果中指出泛素化修饰 motif 并不存在，只是定性地表现为修饰位点附近酸性氨基酸表达量较高、碱性氨基酸表达量较低，但是对于泛素化修饰 motif 的探索不会就此结束，泛素化修饰 motif 的发现将对揭示底物选择机制具有重要意义，这是一个具有研究价值的领域，随着修饰位点数据规模的增加和寻找 motif 方法的完善，新的探索值得期待。

随着泛素化修饰相关蛋白数据的不断产出，泛素化修饰位点的预测工作可以更为深入地展开，之前相关工作由于数据量不够丰富、特征选择不够理想等原因，建立的预测模型用后续的实验数据检验发现预测结果并不理想，然而随着泛素化修饰数据的日益增多，建立更为精准的预测模型成为可能，而一个好的修饰位点预测模型又会为深入理解泛素化修饰的选择机制带来帮助，该方向的研究也是当前生物信息学方法分析泛素化修饰的热点所在。

(5)泛素化修饰与其他翻译后修饰之间交互应答(Cross-talk)的研究。

泛素化修饰作为一种重要的翻译后修饰，在蛋白质降解、转录调控、DNA 损伤修复等重要细胞生命活动过程中发挥重要作用，现有研究结果显示，各种翻译后修饰并不是孤立的，其彼此之间的交互应答是当前研究的热点。蛋白质翻译后的经历的一系列修饰过程对蛋白执行正常的生命活动起到至关重要的作用，而各种翻译后修饰之间的协调机制有待于进一步揭示。泛素化修饰作为当前一种被热门研究的翻译后修饰，在这样的交互应答背景下无疑会发挥至关重要的作用。采用生物信息学的方法分析各翻译后修饰之间的交互应答是备受关注的一个研究方向，随着泛素化修饰研究的深入，这一研究方向将具备更好的研究条件，泛素化修饰与其它翻译后修饰交互应答机制的揭示对于深入理解细胞整体水平的生命活动意义重大。

参考文献(References):

[1] Jennissen HP. Ubiquitin and the enigma of intracellular protein degradation. *Eur J Biochem*, 1995, 231(1): 1–30. DOI

[2] Pickart CM. Mechanisms underlying ubiquitination. *Annu Rev Biochem*, 2001, 70: 503–533. DOI

[3] Dye BT, Schulman BA. Structural mechanisms underlying posttranslational modification by ubiquitin-like proteins. *Annu Rev Biophys Biomol Struct*, 2007, 36: 131–150. DOI

[4] Ye YH, Rape M. Building ubiquitin chains: E2 enzymes at work. *Nat Rev Mol Cell Biol*, 2009, 10(11): 755–764. DOI

[5] Neutzner M, Neutzner A. Enzymes of ubiquitination and deubiquitination. *Essays Biochem*, 2012, 52(1): 37–50. DOI

[6] Muratani M, Tansey WP. How the ubiquitin-proteasome

system controls transcription. *Nat Rev Mol Cell Biol*, 2003, 4(3): 192–201. DOI

[7] Pornillos O, Garrus JE, Sundquist WI. Mechanisms of enveloped RNA virus budding. *Trends Cell Biol*, 2002, 12(12): 569–579. DOI

[8] Terrell J, Shih S, Dunn R, Hicke L. A function for monoubiquitination in the internalization of a G protein-coupled receptor. *Mol Cell*, 1998, 1(2): 193–202. DOI

[9] Rome S, Meugnier E, Vidal H. The ubiquitin-proteasome pathway is a new partner for the control of insulin signaling. *Curr Opin Clin Nutr Metab Care*, 2004, 7(3): 249–254. DOI

[10] Izzi L, Attisano L. Regulation of the TGFβ signalling pathway by ubiquitin-mediated degradation. *Oncogene*, 2004, 23(11): 2071–2078. DOI

[11] Hicke L. Protein regulation by monoubiquitin. *Nat Rev Mol Cell Biol*, 2001, 2(3): 195–201. DOI

[12] Pickart CM. Ubiquitin enters the new millennium. *Mol Cell*, 2001, 8(3): 499–504. DOI

[13] Herrmann J, Lerman LO, Lerman A. Ubiquitin and ubiquitin-like proteins in protein regulation. *Circ Res*, 2007, 100(9): 1276–1291. DOI

[14] Welchman RL, Gordon C, Mayer RJ. Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nat Rev Mol Cell Biol*, 2005, 6(8): 599–609. DOI

[15] Shaid S, Brandts CH, Serve H, Dikic I. Ubiquitination and selective autophagy. *Cell Death Differ*, 2012, doi: 10.1038/cdd.2012.72. DOI

[16] Zencheck WD, Xiao H, Weiss LM. Lysine post-translational modifications and the cytoskeleton. *Essays Biochem*, 2012, 52(1): 135–145. DOI

[17] Wagner SA, Beli P, Weinert BT, Nielsen ML, Cox J, Mann M, Choudhary C. A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. *Mol Cell Proteomics*, 2011, 10(10): M111. 013284. DOI

[18] Chernorudskiy AL, Garcia A, Eremin EV, Shorina AS, Kondratieva EV, Gainullin MR. UbiProt: a database of ubiquitylated proteins. *BMC Bioinformatics*, 2007, 8: 126. DOI

[19] Du YP, Xu NF, Lu M, Li TT. hUbiquitome: a database of experimentally verified ubiquitination cascades in humans. *Database (Oxford)*, 2011, 2011: bar055. DOI

[20] Lee H, Yi GS, Park JC. E3Miner: a text mining tool for ubiquitin-protein ligases. *Nucleic Acids Res*, 2008, 36(S2): W416–W422. DOI

[21] Han Y, Lee H, Park JC, Yi GS. E3Net: a system for exploring E3-mediated regulatory networks of cellular functions. *Mol Cell Proteomics*, 2012, 11(4): O111. 014076. DOI

[22] Venancio TM, Balaji S, Iyer LM, Aravind L. Reconstructing the ubiquitin network: cross-talk with other systems and identification of novel functions. *Genome Biol*, 2009, 10(3): R33. DOI

[23] Kerscher O, Felberbaum R, Hochstrasser M. Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu Rev Cell Dev Biol*, 2006, 22: 159–180. DOI

[24] Pickart CM, Eddins MJ. Ubiquitin: structures, functions, mechanisms. *Biochim Biophys Acta*, 2004, 1695(1-3): 55–72. DOI

[25] Palancade B, Doye V. Sumoylating and desumoylating enzymes at nuclear pores: underpinning their unexpected duties? *Trends Cell Biol*, 2008, 18(4): 174–183. DOI

[26] Seeler JS, Dejean A. Nuclear and unclear functions of SUMO. *Nat Rev Mol Cell Biol*, 2003, 4(9): 690–699. DOI

[27] Wilson VG, Heaton PR. Ubiquitin proteolytic system: focus on SUMO. *Expert Rev Proteomics*, 2008, 5(1): 121–135. DOI

[28] Ohsumi Y. Molecular mechanism of autophagy in yeast, Saccharomyces cerevisiae. *Philos Trans R Soc Lond B Biol Sci*, 1999, 354(1389): 1577–1580. DOI

[29] Hochstrasser M. Evolution and function of ubiquitin-like protein-conjugation systems. *Nat Cell Biol*, 2000, 2(8): E153–E157. DOI

[30] Iyer LM, Burroughs AM, Aravind L. The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like β-grasp domains. *Genome Biol*, 2006, 7(7): R60. DOI

[31] Iyer LM, Koonin EV, Aravind L. Novel predicted peptidases with a potential role in the ubiquitin signaling pathway. *Cell Cycle*, 2004, 3(11): 1440–1450. DOI

[32] Balaji S, Babu MM, Iyer LM, Luscombe NM, Aravind L. Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *J Mol Biol*, 2006, 360(1): 213–227. DOI

[33] Luscombe NM, Babu MM, Yu HY, Snyder M, Teichmann SA, Gerstein M. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*, 2004, 431(7006): 308–312. DOI

[34] Yu HY, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, Hao T, Rual JF, Dricot A, Vazquez A, Murray RR, Simon C, Tardivo L, Tam S, Svrzikapa N, Fan CY, de Smet AS, Motyl A, Hudson ME, Park J, Xin XF, Cusick

ME, Moore T, Boone C, Snyder M, Roth FP, Barabási AL, Tavernier J, Hill DE, Vidal M. High-quality binary protein interaction map of the yeast interactome network. *Science*, 2008, 322(5898): 104–110. DOI

[35] van Wijk SJ, de Vries SJ, Kemmeren P, Huang A, Boelens R, Bonvin AMJJ, Timmers HTM. A comprehensive framework of E2-RING E3 interactions of the human ubiquitin-proteasome system. *Mol Syst Biol*, 2009, 5: 295. DOI

[36] Huang L, Kinnucan E, Wang GD, Beaudenon S, Howley PM, Huibregtse JM, Pavletich NP. Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science*, 1999, 286(5443): 1321–1326. DOI

[37] Schulman BA, Carrano AC, Jeffrey PD, Bowen Z, Kinnucan ER, Finnin MS, Elledge SJ, Harper JW, Pagano M, Pavletich NP. Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. *Nature*, 2000, 408(6810): 381–386. DOI

[38] Zheng N, Wang P, Jeffrey PD, Pavletich NP. Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell*, 2000, 102(4): 533–539. DOI

[39] Christensen DE, Brzovic PS, Klevit RE. E2-BRCA1 RING interactions dictate synthesis of mono- or specific polyubiquitin chain linkages. *Nat Struct Mol Biol*, 2007, 14(10): 941–948. DOI

[40] Poyurovsky MV, Priest C, Kentsis A, Borden KL, Pan ZQ, Pavletich N, Prives C. The Mdm2 RING domain C-terminus is required for supramolecular assembly and ubiquitin ligase activity. *EMBO J*, 2007, 26(1): 90–101. DOI

[41] Xu Z, Kohli E, Devlin KI, Bold M, Nix JC, Misra S. Interactions between the quality control ubiquitin ligase CHIP and ubiquitin conjugating enzymes. *BMC Struct Biol*, 2008, 8: 26. DOI

[42] Ardley HC, Robinson PA. E3 ubiquitin ligases. *Essays Biochem*, 2005, 41: 15–30. DOI

[43] Lorick KL, Jensen JP, Fang SY, Ong AM, Hatakeyama S, Weissman AM. RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proc Natl Acad Sci USA*, 1999, 96(20): 11364–11369. DOI

[44] Winkler GS, Albert TK, Dominguez C, Legtenberg YIA, Boelens R, Timmers HTM. An altered-specificity ubiquitin-conjugating enzyme/ubiquitin-protein ligase pair. *J Mol Biol*, 2004, 337(1): 157–165. DOI

[45] Tomlinson E, Palaniyappan N, Tooth D, Layfield R. Methods for the purification of ubiquitinated proteins. *Proteomics*, 2007, 7(7): 1016–1022. DOI

[46] Denis NJ, Vasilescu J, Lambert JP, Smith JC, Figeys D. Tryptic digestion of ubiquitin standards reveals an improved strategy for identifying ubiquitinated proteins by mass spectrometry. *Proteomics*, 2007, 7(6): 868–874. DOI

[47] Hitchcock AL, Auld K, Gygi SP, Silver PA. A subset of membrane-associated proteins is ubiquitinated in response to mutations in the endoplasmic reticulum degradation machinery. *Proc Natl Acad Sci USA*, 2003, 100(22): 12735–12740. DOI

[48] Jeon HB, Choi ES, Yoon JH, Hwang JH, Chang JW, Lee EK, Choi HW, Park ZY, Yoo YJ. A proteomics approach to identify the ubiquitinated proteins in mouse heart. *Biochem Biophys Res Commun*, 2007, 357(3): 731–736. DOI

[49] Kirkpatrick DS, Weldon SF, Tsaprailis G, Liebler DC, Gandolfi AJ. Proteomic identification of ubiquitinated proteins from human cells expressing His-tagged ubiquitin. *Proteomics*, 2005, 5(8): 2104–2111. DOI

[50] Matsumoto M, Hatakeyama S, Oyamada K, Oda Y, Nishimura T, Nakayama KI. Large-scale analysis of the human ubiquitin-related proteome. *Proteomics*, 2005, 5(16): 4145–4151. DOI

[51] Peng JM, Schwartz D, Elias JE, Thoreen CC, Cheng DM, Marsischky G, Roelofs J, Finley D, Gygi SP. A proteomics approach to understanding protein ubiquitination. *Nat Biotechnol*, 2003, 21(8): 921–926. DOI

[52] Hagai T, Tóth-Petróczy Á, Azia A, Levy Y. The origins and evolution of ubiquitination sites. *Mol Biosyst*, 2012, 8(7): 1865–1877. DOI

[53] Udeshi ND, Mani DR, Eisenhaure T, Mertins P, Jaffe JD, Clauser KR, Hacohen N, Carr SA. Methods for quantification of *in vivo* changes in protein ubiquitination following proteasome and deubiquitinase inhibition. *Mol Cell Proteomics*, 2012, 11(5): 148–159. DOI

[54] Denison C, Kirkpatrick DS, Gygi SP. Proteomic insights into ubiquitin and ubiquitin-like proteins. *Curr Opin Chem Biol*, 2005, 9(1): 69–75. DOI

[55] Chen Z, Chen YZ, Wang XF, Wang C, Yan RX, Zhang ZD. Prediction of ubiquitination sites by using the composition of k-spaced amino acid pairs. *PLoS One*, 2011, 6(7): e22930. DOI

[56] Radivojac P, Vacic V, Haynes C, Cocklin RR, Mohan A, Heyen JW, Goebl MG, Iakoucheva LM. Identification, analysis, and prediction of protein ubiquitination sites. *Proteins*, 2010, 78(2): 365–380. DOI

[57] Cai YD, Huang T, Hu LL, Shi XH, Xie L, Li YX. Prediction of lysine ubiquitination with mRMR feature selection and analysis. *Amino Acids*, 2012, 42(4): 1387–1395. DOI

[58] Zhao XW, Li XT, Ma ZQ, Yin MH. Prediction of lysine

ubiquitylation with ensemble classifier and feature selection. *Int J Mol Sci*, 2011, 12(12): 8347–8361. DOI
[59] Feng KY, Huang T, Feng KR, Liu XJ. Using WPNNA classifier in ubiquitination site prediction based on hybrid features. *Protein Pept Lett*, 2012. DOI
[60] Tung CW, Ho SY. Computational identification of ubiquitylation sites from protein sequences. *BMC Bioinformatics*, 2008, 9: 310. DOI
[61] Plewczynski D, Tkacz A, Wyrwicz LS, Rychlewski L. AutoMotif server: prediction of single residue post-translational modifications in proteins. *Bioinformatics*, 2005, 21(10): 2525–2527. DOI
[62] Tung CW, Ho SY. POPI: predicting immunogenicity of MHC class I binding peptides by mining informative physicochemical properties. *Bioinformatics*, 2007, 23(8): 942–949. DOI
[63] Xue Y, Chen H, Jin CJ, Sun ZR, Yao XB. NBA-Palm: prediction of palmitoylation site implemented in Naïve Bayes algorithm. *BMC Bioinformatics*, 2006, 7: 458. DOI
[64] Jones DT. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics*, 2007, 23(5): 538–544. DOI
[65] Kaur H, Raghava GPS. A neural network method for prediction of β-turn types in proteins using evolutionary information. *Bioinformatics*, 2004, 20(16): 2751–2758. DOI
[66] Huang WL, Tung CW, Huang HL, Hwang SF, Ho SY. ProLoc: Prediction of protein subnuclear localization using SVM with automatic selection from physicochemical composition features. *Biosystems*, 2007, 90(2): 573–581. DOI
[67] Teng S, Luo H, Wang L. Predicting protein sumoylation sites from sequence features. *Amino Acids*, 2012, 43(1): 447–455. DOI
[68] Kim W, Bennett EJ, Huttlin EL, Guo AL, Li J, Possemato A, Sowa ME, Rad R, Rush J, Comb MJ, Harper JW, Gygi SP. Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol Cell*, 2011, 44(2): 325–340. DOI

•综合信息•

中国遗传学会荣获“全国优秀科技工作者”获奖者名单

中国人民解放军总医院袁慧军研究员荣获“全国十佳科技工作者提名奖”和“全国优秀科技工作者”称号，中国科学院遗传与发育生物学研究所杨崇林研究员和华中农业大学熊立仲教授荣获“全国优秀科技工作者”称号。

中国遗传学会
2012 年 12 月 27 日

1、袁慧军研究员简介

袁慧军，女，解放军总医院耳鼻咽喉研究所研究员，博士生导师，国家杰出青年基金获得者，《遗传》编委。从事聋病分子遗传学研究 17 年，在聋病新基因研究领域取得了多项有国际影响力的研究成果，将中国的遗传性耳聋研究推向国际前沿水平。在国际上首次报道了 3 个遗传性耳聋新基因(*PCDH15,PRPS1* 和 *SMAC*)，系统地揭示了致聋基因 *POU3F4, COCH, DFNA5, EYA4, WFS1, MYO7A, PAX3* 和 *MITF* 新突变及致聋机制，合作研制的药物性耳聋易感基因检测试剂盒，已获得专利并实现了临床应用转化。

2、杨崇林研究员简介

杨崇林博士 2005 年入选中国科学院“百人计划”。2010 年获“国家杰出青年基金”资助。多次获得“中国科学院优秀导师奖”等荣誉称号, 2010 获得“中国侨界贡献奖(创新人才”)。2011 年被评为“中国侨联特聘专家”。他在程序性细胞死亡的调控机制研究领域取得了重要突破，研究成果多次发表在 *Science* 等国际主流学术刊物上。主持或参与多项 973 或 863 等研究课题。杨崇林研究员在实验室建设和人才培养方面成绩斐然，他热心学术公益活动，积极组织和参加各种学术活动和学术交流。

3、熊立仲教授简介

熊立仲教授以“水稻抗旱性”为主题，系统地开展了水稻后期抗旱的分子基础和遗传改良研究，近年来以第一作者或通讯作者发表在 *Plant Cell*, *PNAS*, *Plant Journal*, *Plant Physiology* 等国际刊物上发表了 40 多篇学术论文，单篇文章最

高他引次数 5 年内超过 200 多次，申请或获得授权专利 20 多项。曾获得“国家杰出青年科学基金”、“中国青年科技奖”、“湖北青年五四奖章”、“教育部青年教师资助计划”、“教育部新世纪优秀人才计划”、“全国青年联合会第十届委员”等多个奖项及荣誉称号。他注重科教融合，培养了一批优秀人才。