

UniProt 数据库检索及数据条目注释信息

姓名 陈耿佳 学号 1301214752 组号 G01C

1. UniProt 蛋白质序列数据库

- 1) 参阅 Swiss-Prot 和 TrEMBL 统计报表 (Release Statistics), 列表说明这两个子库的总数据量, 以及不同蛋白质证据 (Protein Existence) 的数据条目数。

数据库	Swiss-Prot		TrEMBL	
总数据量	545388	100%	56010222	100%
蛋白水平证据	82702	15.2%	26770	0.05%
转录水平证据	62490	11.5%	854251	1.53%
同源预测	382436	70.1%	14322100	25.57%
其他预测	15871	2.9%	40807101	72.86%
不确定	1889	0.3%	0	0.00%

- 2) 列表说明 Swiss-Prot 和 TrEMBL 中数据条目数前 10 位的物种, 包括中文名、英文名和拉丁文学名。

Swiss-Prot:

排名	条目数	中文名	英文名	拉丁文学名
1	20264	人类	Human	Homo sapiens
2	16669	小鼠	Mouse	Mus musculus
3	12851	拟南芥	Mouse-ear cress	Arabidopsis thaliana
4	7897	大鼠	Rat	Rattus norvegicus
5	6621	酿酒酵母	Baker's yeast	Saccharomyces cerevisiae (strain ATCC 204508 / S288c)
6	5984	家牛	Bovine	Bos taurus
7	5103	裂殖酵母	Fission yeast	Schizosaccharomyces pombe (strain 972 / ATCC 24843)
8	4431	大肠杆菌	Colibacillus	Escherichia coli (strain K12)
9	4185	枯草杆菌	hay bacillus	Bacillus subtilis (strain 168)
10	4127	盘基网柄菌	Slime mold	Dictyostelium discoideum

TrEMBL:

排名	条目数	中文名	英文名	拉丁文学名
1	590031	人类免疫缺陷病毒	Human immunodeficiency virus 1	--
2	352018	海洋沉积物宏基因组	marine sediment metagenome	--
3	217903	非培养细菌	uncultured bacterium	--
4	115939	人类	Human	Homo sapiens
5	105994	小麦	Wheat	Triticum aestivum
6	96773	粳稻	Rice	Oryza sativa subsp. japonica
7	92711	丙型肝炎病毒	Hepatitis C virus	--
8	81523	乙型肝炎病毒	Hepatitis B virus	--
9	73928	大豆	Soybean	Glycine max
10	73055	矿山排水宏基因组	mine drainage metagenome	--

- 3) 列表说明以下已基本完成基因组测序的重要模式生物数据条目数总数 N、已审阅序列条目数 Nr、具有蛋白质证据的序列条目数 Np、在参考序列数据库 RefSeq 中具有 mRNA 序列的序列条目数 Nm、在蛋白质结构数据库 PDB 中具有结构的序列条目数 Nb。

物种	拉丁文学名	英文名	N	Nr	Np	Nm	Nb
人	Homo sapiens	Human	136,873	20,263	14,227	28,843	5,233
小鼠	Mus musculus	Mouse	73,499	16,664	8,416	30,703	1,474
大鼠	Rattus norvegicus	Rat	33,592	7,889	3,413	18,325	541
鸡	Gallus gallus	Chicken	23,745	2,257	706	8,634	182
非洲爪蟾	Xenopus laevis	African clawed frog	16,025	3,382	578	11,106	59
斑马鱼	Danio rerio	Zebrafish	55,758	2,919	396	21,330	65
黑腹果蝇	Drosophila melanogaster	Fruit fly	39,999	3,220	1,983	20,054	353
秀丽线虫	Caenorhabditis elegans	Worm	26,650	3,465	1,496	24,427	118
酿酒酵母	Saccharomyces cerevisiae	Baker's yeast	79,277	7,807	5,317	5,919	1,107
大肠杆菌 K12	Escherichia coli (strain K12)	Colibacillus	12,519	5,967	2,752	12,396	1,320
拟南芥	Arabidopsis thaliana	Mouse-ear cress	53,221	12,720	4,828	32,594	380
粳稻	Oryza sativa subsp. japonica	Rice	99,965	3,138	413	29,543	58
籼稻	Oryza sativa subsp. indica	Rice	40,542	650	85	160	2
玉米	Zea mays	Maize	61,132	735	345	22,792	35
棉花	Gossypium hirsutum	Upland cotton	2,030	150	35	154	1

2. 序列条目注释信息

- 1) 以人血红蛋白 α 亚基为例，说明一般注释信息（General Annotation）主要包括哪些内容。

UniProt 数据库中人血红蛋白 α 亚基条目 P69905 (HBA_HUMAN) 中，一般注释信息（General Annotation）包括以下各项：

Function 功能：介绍该蛋白质的主要功能。

人血红蛋白 α 亚基的主要功能为将氧气从肺部运输到外周组织。

Subunit structure 亚单元结构：介绍该蛋白质相关的亚单元构成。

成人血红蛋白 A(HbA)是由两条 α 链和两条 β 链组成的异四聚体；成人血红蛋白 A2(HbA2)由两条 α 链和两条 δ 链组成；早期胚胎血红蛋白(HbGower 2)由两条 α 链和两条 ϵ 链组成；胎儿血红蛋白 F 由两条 α 链和两条 γ 链组成。

Tissue specificity 组织特异性：介绍该蛋白质的组织特异性分布。

人血红蛋白 α 亚基主要存在于红细胞中。

Post-translational modification 翻译后修饰：介绍已知的（文献报道）该蛋白质相关的翻译后修饰。

人血红蛋白 α 亚基的 Thionville 变体 (V2E) 的起始甲硫氨酸没有被切除，且被乙酰化修饰。

Involvement in disease 相关疾病：介绍与该蛋白质（基因）相关的疾病，对于各相关疾病，给出其在 OMIM (Online Mendelian Inheritance in Man) 网站的链接，简要介绍其发病机制，并给出相应的参考文献。

人血红蛋白 α 亚基相关的疾病包括亨氏身体贫血症 (Heinz body anemias, HEIBAN)，地中海贫血症 (Alpha-thalassemia, A-TH) 和血红蛋白 H 病 (Hemoglobin H disease, HBH)。

Miscellaneous 其他信息：介绍该蛋白质在一般注释信息 (General Annotation) 中其他各项中未涉及但有价值的信息。

人血红蛋白 α 亚基使血液呈现红色。

Sequence similarities 序列相似性：介绍与该蛋白质序列相似性较高的蛋白质或蛋白家族。

人血红蛋白 α 亚基属于球蛋白家族 (Globin family)

Sequence caution 序列警示：介绍与该蛋白质来自相同基因但序列不同的其他来源的蛋白质链接及其原因。

EBI 数据库中 BAD97112.1 条目 Homo sapiens (human) partial alpha 2 globin variant 的序列与人血红蛋白 α 亚基不同，原因是起始错误。

- 2) 以人血红蛋白 α 亚基为例，说明序列特征注释信息 (Sequence Annotation) 主要包括哪些内容。

UniProt 数据库中人血红蛋白 α 亚基条目 P69905 (HBA_HUMAN) 中，序列特征注释信息 (Sequence Annotation) 包括以下各项：

Molecule processing 分子处理：介绍该蛋白质在合成到成熟过程中所包含的各种处理过程及各相关片段范围的标注，如起始甲硫氨酸的切除、信号肽、运输肽、前肽、成熟链、活性肽等，并给出相应的参考文献链接或其他来源信息（如预测）。

人血红蛋白 α 亚基序列第 1 位甲硫氨酸在被切除 (Initiator methionine), 余下 2-142 位组成成熟肽链 (Chain)。

Sites 位点: 介绍该蛋白质血序列中各种有意思的位点, 如活性位点、金属结合位点、其他分子结合位点等等。

人血红蛋白 α 亚基第 59 位和 88 位氨基酸为金属结合位点, 第 12、57、61、91、100 位为非糖化位点。

Amino acid modifications 氨基酸修饰: 介绍该蛋白质序列中发生氨基酸修饰的位点, 如糖基化、脂化、蛋白质交联、二硫键等等, 以及特殊氨基酸如硒代半胱氨酸和吡咯赖氨酸。

人血红蛋白 α 亚基第 8、17、41、62 位可发生糖基化修饰, 第 8、17、41 位赖氨酸可发生酰基化修饰。

Natural variations 自然变体: 介绍自然界存在的该蛋白质的突变位点, 给出其突变结果、来源、对蛋白质功能的影响、相关文献等。

人血红蛋白 α 亚基存在上百个自然变体, 如 Thionville 突变体 (V2E) 氧气结合能力下降。

Experimental info 实验信息: 介绍已有实验数据 (文献报道) 的该蛋白质序列相关的信息, 包括实验产生的变异位点、序列不确定区、序列冲突、不连续位点、非末端位点。

EBI 数据库中 BAD97112.1 条目 Homo sapiens (human) partial alpha 2 globin variant 的序列与人血红蛋白 α 亚基存在序列冲突, 差异位点为 N10H。

Secondary structure 二级结构: 介绍实验测定的该蛋白质的二级结构, 标注序列中各区域的二级结构, 包括螺旋、转角、 β 折叠。

人血红蛋白 α 亚基二级结构如下所示:



- 3) 以人血红蛋白 α 亚基为例, 说明数据库交叉链接 (Cross Reference) 主要包括哪些数据库。

Sequence databases 序列数据库: 收录各主要序列数据库中该蛋白质氨基酸序列、mRNA 序列、编码基因序列等等, 人血红蛋白 α 亚基在以下数据库中均有相应序列链接: EMBL、GenBank、DDBJ、PIR、RefSeq、UniGene。

3D structure databases 三维结构数据库: 收录各主要蛋白质三维结构及建模数据库中该蛋白质的三维结构信息和同源建模结果。人血红蛋白 α 亚基在 PDBe、RCSB-PDB、PDBj 数据库有约 200 个实验测定的结构数据; ProteinModelPortal、SMR (SWISS-MODEL Repository)、ModBase 主要为基于同源建模的三维结构预测平台, 给出该蛋白质在这些数据库中的建模结果; MobiDB 为较综合的蛋白质数据库, 主要提供该蛋白质相关的结构信息, 其提供的结构信息比 UniProt 更丰富。

Protein-protein interaction databases 蛋白-蛋白相互作用数据库: 收录以下数据库对该蛋白相关的实验测定或预测的相互作用蛋白数据: BioGrid、DIP、IntAct、MINT、STRING。

Chemistry 化学数据库: ChEMBL 数据库提供该蛋白质相关的基于药物发现的生

物活性数据，包括靶标生物活性、化验方法、配体效率、药物等。DrugBank 数据库提供以该蛋白质为靶标的药物信息，其收录的以人血红蛋白 α 亚基为靶标的药物包括：Amodiaquine、Chloroquine、Iron Dextran、Mefloquine、Primaquine、Quinine。

PTM databases 翻译后修饰数据库：收录该蛋白质的翻译后修饰数据。PhosphoSite 数据库收录该蛋白质的磷酸化修饰位点数据。

Polymorphism databases 多态性数据库：收录该蛋白质的单核苷酸多态性及相关疾病数据。DMDM (Domain mapping of disease mutations) 收录了人血红蛋白 α 亚基各自然突变位点 (来自 Swiss-Prot 数据库, 146 个条目) 和疾病相关突变位点 (来自 OMIM 数据库, 250 个条目)。

2D gel databases 双向凝胶电泳数据库：收录实验测定的该蛋白质相关的双向凝胶电泳图谱及相应的鉴定信息。以下数据库收录了人血红蛋白 α 亚基相关的双向凝胶电泳图谱：DOSAC-COBS-2DPAGE、REPRODUCTION-2DPAGE、SWISS-2DPAGE、UCD-2DPAGE。

Proteomic databases 蛋白质组学数据库：MaxQB、PaxDb、PeptideAtlas、PRIDE 等数据库收录了与人血红蛋白 α 亚基相关的蛋白质组学实验数据。

Protocols and materials databases 实验方案及材料数据库：DNASU Plasmid repository 收录与该蛋白质相关的可购买的重组质粒；Structural Biology Knowledgebase 收录了与该蛋白质相关的实验方法，如人血红蛋白铁含量测定方法等。

Genome annotation databases 基因注释数据库：Ensembl、GeneID、KEGG、UCSC 等数据库收录了与人血红蛋白 α 亚基相关的基因组注释信息，可了解该蛋白在基因组中的位置、基因结构、相邻基因等等基因组相关注释信息。

Organism-specific databases 组织特异性数据库：GeneCards、HPA (The Human Protein Atlas) 等数据库提供该蛋白的组织特异性分布数据，人血红蛋白 α 亚基主要分布在血液和组织间隙，在其他器官中也有分布。

Phylogenomic databases 系统发育树数据库：收录与该蛋白相关的系统发育树信息，在以下数据库均可检索到与人血红蛋白 α 亚基相关的系统发育树：eggNOG、HOVERGEN、InParanoid、KO、OMA、OrthoDB、PhylomeDB、TreeFam。

Enzyme and pathway databases 酶与信号通路数据库：Reactome 数据库收录了人血红蛋白 α 亚基参与的代谢信号通路和 Scavenger Receptor 结合并摄取配体的信号通路。

Gene expression databases 基因表达数据库：Bgee、CleanEx、Genevestigator 等数据库收录该蛋白质基因表达数据，包括不同组织表达情况、不同发育阶段表达情况等等。

Family and domain databases 蛋白家族与结构域数据库：Gene3D、InterPro、PANTHER、Pfam、PRINTS、SUPFAM、PROSITE、ProtoNet 等数据库收录该蛋白质所属蛋白家族、超家族及相关结构域信息，包括蛋白家族简介、成员、功能等。

Other 其它数据库：不属于上述各分类的其它数据库。ChiTaRS 数据库收录该蛋白

质相关的嵌合转录本及 RNA-Seq 数据；Evolutionary Trace 从进化角度分析该蛋白的功能位点等信息；GeneWiki 收录维基百科上该蛋白相关的条目；NextBio 收录该蛋白质的组织特异性分布、相关疾病、相关药物、相关转基因模型、相关研究等信息；PMAP-CutDB 收录以该蛋白质为底物的蛋白酶及其酶切位点等信息；PRO 数据库收录该蛋白质的 ontology 信息。

3. 数据库检索

- 1) 写出从 UniProt 数据库中检索已审阅的人珠蛋白 (globin) 家族 12 个亚基的步骤。列表说明这 12 个珠蛋白的登录号、蛋白质名称、和序列长度。其中与血红蛋白 alpha 亚基差异最大的序列是哪个？相同位点百分比？与血红蛋白 beta 亚基差异最小的序列是哪个？差异位点共多少个？

点击“Advanced Search”，Field 下拉菜单中选“Protein family”，Term 填入“globin”，点击“Add & Search”；点击“Advanced Search”，Field 下拉菜单中选“Organism [OS]”，Term 填入“human”，点击“Add & Search”；Results 中点击“Show only reviewed (12)”中的“reviewed (12)”链接，即得到已审阅的人珠蛋白 (globin) 家族 12 个亚基。如下表所示：

登录号 Entry	蛋白质名称 Protein names	序列长度 Length
P02008	Hemoglobin subunit zeta	142
P69905	Hemoglobin subunit alpha	142
P68871	Hemoglobin subunit beta	147
P02042	Hemoglobin subunit delta	147
P69891	Hemoglobin subunit gamma-1	147
P69892	Hemoglobin subunit gamma-2	147
P02144	Myoglobin	154
Q8WWM9	Cytoglobin	190
Q9NPG2	Neuroglobin	151
P09105	Hemoglobin subunit theta-1	142
P02100	Hemoglobin subunit epsilon	147
Q6B0K9	Hemoglobin subunit mu	141

其中与血红蛋白 α 亚基差异最大的序列是神经球蛋白 Neuroglobin，相同位点为 31/142，占 21.9%。与血红蛋白 β 亚基差异最小的序列是血红蛋白 δ (Hemoglobin subunit delta)，差异位点共 10 个。

- 2) 写出从 UniProt 数据库中检索以下序列条目的步骤和结果：拟南芥序列条目总数，其中已审阅条目数，已审阅序列中具有蛋白质证据全长序列条目数，上述序列中膜蛋白总数、注释为跨膜蛋白数且具有信号肽的序列数。
 - (1) 点击“Advanced Search”，Field 下拉菜单中选“Organism [OS]”，Term 填入“Arabidopsis thaliana”，点击“Add & Search”，得到拟南芥所有序列 53316 个条目；
 - (2) Results 中点击“Show only reviewed (12,851)”中的“reviewed (12,851)”链接，即得到已审阅的条目 12851 个；

- (3) 点击“Advanced Search”, Field 下拉菜单中选“Protein existence [PE]”, Choose 下拉菜单中选择“Evidence at protein level”, 点击“Add & Search”; 点击“Advanced Search”, Field 下拉菜单中选“Fragment (yes/no)”, Choose 下拉菜单中选择 no”, 点击“Add & Search”; 即得到具有蛋白质证据全长序列 4532 个条目;
 - (4) 点击 “Advanced Search”, Field 下拉菜单中选“Subcellular Location”, Term 填入“membrane”, 点击“Add & Search”, 上述序列中膜蛋白序列 1301 个条目;
 - (5) 点击 “Advanced Search”, Field 下拉菜单中选“General annotation [CC]”, Topic 下拉菜单中选择“Any”, Term 填入“transmembrane”, 点击“Add & Search”; 点击“Advanced Search”, Field 下拉菜单中选“All”, Term 填入“signal peptide”, 点击“Add & Search”; 即得到注释为跨膜蛋白数且具有信号肽的序列 3 个条目;
4. 利用高级检索功能, 从 UniProt 数据库中检索你课题相关或最感兴趣的蛋白质, 阅读其一般注释信息、序列特征注释信息, 相关文献, 并通过数据库交叉链接, 总结该蛋白质的研究进展。

UniProt Entry: Q6ZMU5 (TRI72_HUMAN) 已审阅

名称: Tripartite motif-containing protein 72 (TRIM72) 或 Mitsugumin-53 (MG53)

蛋白属性: 长度 477 AA, 蛋白水平证据。

General Annotation 一般注释信息

Function 功能: MG53 是一个肌肉特异性蛋白, 在细胞膜修复中起关键作用, 在损伤部位聚集启动修复装置。特异性结合磷脂酰丝氨酸。作为氧化的感应器: 发生膜损伤时, 细胞外氧化环境导致 MG53 形成二硫键并在损伤部位寡聚化。寡聚化的 MG53 作为核招募包含 MG53 的囊泡在损伤部位聚集, 启动膜修复。可能作为钙依赖膜释放过程的上游。是修复装置形成过程中运输 DYSF 到损伤部位所必需的。调节膜出芽和胞外分泌。可能参与包含 KCNB1 内吞囊泡的移动的调节。

Subunit structure 亚单元结构: 可形成同多聚体, 通过二硫键相连。可与 DYSF 及 CAV3 相互作用。

Subcellular location 亚细胞定位: 细胞膜 (肌纤维膜); 胞质囊泡膜。(通过与膜上的磷脂酰丝氨酸相互作用固定在膜上)。

Post-translational modification 翻译后修饰: 膜损伤发生时细胞外氧化环境进入导致 242 位半胱氨酸形成二硫键, 导致寡聚化。

Sequence similarities 序列相似性: 属于 TRIM/RBCC 家族。包含 1 个 B box 型锌指结构域; 包含 1 个 B30.2/SPRY 结构域; 包含 1 个 RING 型锌指结构域。

Ontologies 本体论

MG53 参与的生物学过程包括: 胞吐、肌肉器官发育、肌肉系统过程、质膜修复、蛋白质同寡聚化。可结合锌离子。可变剪切。

Alternative products 可变产物

MG53 可能存在 270-477 位氨基酸缺失的 isoform 2，尚无确切实验证据。

Sequence annotation 序列特征注释

Molecule processing 分子处理：所收录序列 1-477 位为 MG53 完整全长链。

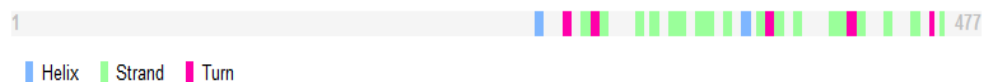
Regions 区域：

- 14-57 为 RING 型锌指结构域；
- 81-122 为 B box 型锌指结构域；
- 135-169 为 Coiled coil 结构域；
- 271-475 为 B30.2/SPRY 结构域。

Amino acid modifications 氨基酸修饰：242 位半胱氨酸可被氧化使 MG53 单体分子间形成二硫键。

Natural variations 自然变体：在 isoform 2 中 270-477 位氨基酸缺失。

Secondary structure 二级结构：仅收录了已经测定结构的 SPRY domain 的二级结构，如下图所示：

**References 参考文献**

仅列出 4 篇参考文献，前三篇为序列相关的，第四篇为 SPRY domain 的晶体结构。在 PubMed 中以“MG53 or TRIM72”为关键词进行搜索可搜到 40 篇文献，大部分为该蛋白质的功能研究，这些并未收录到 UniProt 该条目中。

Cross-references 数据库交叉链接

Sequence databases 序列数据库：EMBL、GenBank、DDBJ、RefSeq、UniGene 等数据库收录了 MG53 相关的各种来源的基因组 DNA、mRNA 及其翻译本、蛋白质氨基酸序列、表达序列标签等序列信息。其蛋白序列的 RefSeq 登录号为 NP_001008275，mRNA 登录号为 NM_001008274。

3D structure databases 三维结构数据库：PDBe、RCSB-PDB、PDBj 数据库收录了人 MG53 PRY-SPRY domain (278-470 位氨基酸) 晶体结构，PDB 登录号为 3KB5；ProteinModelPortal、SMR (SWISS-MODEL Repository)、ModBase、MobiDB 可以提供其他区域序列基于同源模建的三维结构预测结果。

Protein-protein interaction databases 蛋白-蛋白相互作用数据库：

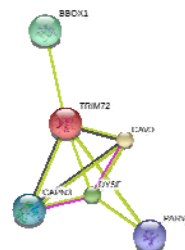
BioGrid 数据库收录了 MG53 的 6 个相互作用蛋白，有 western 等实验证据：

- UBE2H (ubiquitin-conjugating enzyme E2H)
- UBE2N (ubiquitin-conjugating enzyme E2N)
- PTK2 (protein phosphatase 1, regulatory subunit 71)
- MYH9 (non-muscle myosin heavy polypeptide 9)
- PTRF (RNA polymerase I and transcript release factor)
- UBE2I (ubiquitin-conjugating enzyme E2I)

IntAct 数据库收录的相互作用蛋白同样为上述 6 个分子。

STRING 数据库预测的 MG53 的 Functional Partners 为以下 5 个分子，关系网络图如下图所示（evidence view）：

CAV3 caveolin 3
 DYSF dysferlin
 BBOX1 butyrobetaine (gamma)
 CAPN3 calpain 3,
 PARVB parvin, beta



PTM databases 翻译后修饰数据库: PhosphoSite 数据库收录了 MG53 部分修饰位点，人 S307-ph, 小鼠 K137-ub, K178-ub, K398-ub, K462-ub。

Polymorphism databases 多态性数据库: DMDM (Domain mapping of disease mutations) 数据库收录了 MG53 如下 domain 信息，对每个 domain 可进一步分析其功能、属性、相关疾病、保守位点（序列图标）等。

Domain	CD Accession	E-Value	Start	End
RING	cd00162	3.1e-10	13	60
BBOX	cd00021	2.1e-10	84	122
zf-C3HC4	pfam00097	7.7e-06	14	56
zf-B_box	pfam00643	4.1e-11	81	122
SPRY	pfam00622	8.4e-11	342	472
RING	smart00184	2.4e-08	14	56
BBOX	smart00336	1.6e-08	81	122
PRY	smart00589	2.1e-16	288	340
SPRY	smart00449	1.9e-05	342	472

Proteomic databases 蛋白质组学数据库: MaxQB 数据库收录了 MG53 的蛋白表达、质谱肽段等信息。PaxDb 数据库收录了 MG53 在不同组织及不同物种中蛋白表达丰度。PRIDE 数据库收录多个实验室蛋白质组学鉴定中 MG53 相关信息。

Protocols and materials databases 实验方案及材料数据库: Structural Biology Knowledgebase 收录了与 MG53 相似的 PDB 蛋白结构、Protein Model Portal 预测结构模型、TargetTrack 靶标和实验方法、PSI Materials Repository 实验材料等。

Genome annotation databases 基因注释数据库: Ensembl 数据库收录了 MG53 基因注释及其转录本注释；MG53 转录本具有 7 个 exon，其中 6 个包含编码信息，标注了各个 domain 信息。GeneID 数据库收录了 MG53 的基因组信息、RNA-seq 数据、mRNA、相互作用、GO 等信息。KEGG 数据库收录了 MG53 的 motif 信息等。UCSC 数据库收录了 MG53 在多个数据库的各种序列链接、复杂疾病遗传关联研究、毒理基因组学数据、基因单体型、mRNA 二级结构、蛋白结构域与结构信息、直系同源基因、GO、基因模型等信息。

Organism-specific databases 组织特异性数据库: CTD (Comparative Toxicogenomics Database) 数据库收录了 MG53 在数十个物种中的同源基因的 NCBI 链接、相互作用化合物、相关疾病及相应药物、相互作用蛋白等数据。GeneCards 数据库收录了 MG53 的各种别称及数据库链接、RefSeq DNA 序列、调控原件、表观遗传修饰、基因组信息、各组织表达丰度、重组蛋白产物、抗体、Assay、结构域、功能、通路、相互作用蛋白、药物、转录本、多种来源的表达谱、直系旁系同

源基因、变体、疾病等等链接。H-InvDB 数据库收录了 MG53 基因结构、功能、表达、疾病等数据。HGNC 收录了 MG53 各种序列、临床等多种数据库链接。HPA 收录了 MG53 RNA 和蛋白的组织特异性表达谱。Nextprot 数据库收录了 MG5 功能、表达、互作、定位、序列、基因组、结构等信息。MIM (Online Mendelian Inheritance in Man) 以文献综述的形式阐述了 MG53 的功能、基因结构、动物模型等信息。PharmGKB 数据库收录了 MG53 基因组注释及多个数据库链接。GenAtlas 数据库收录了 MG53 基因组、RNA、表达、蛋白、疾病、动物模型等信息。

Phylogenomic databases 系统发育树数据库: eggNOG、HOVERGEN、InParanoid、KO、OMA、OrthoDB、PhylomeDB、TreeFam 等数据库收录了 MG53 直系、旁系同源基因等相关的系统发育树信息。

Gene expression databases 基因表达数据库: Bgee、CleanEx、Genevestigator 等数据库收录 MG53 基因表达数据, 包括不同组织表达情况、不同发育阶段表达情况等等。

Family and domain databases 蛋白家族与结构域数据库: Gene3D、InterPro、PANTHER、Pfam、PRINTS、SUPFAM、PROSITE、ProtoNet 等数据库收录了 MG53 所属蛋白家族、超家族及相关结构域信息, 包括蛋白家族简介、成员、功能等。

Other 其它数据库: ChiTaRS 数据库收录 MG53 相关的嵌合转录本及 RNA-Seq 数据; Evolutionary Trace 从进化角度分析 MG53 的功能位点等信息; NextBio 收录 MG53 的组织特异性分布、相关疾病、相关药物、相关转基因模型、相关研究等信息; GenomeRNAi 数据库收录了 MG53 相关的表型、RNAi 序列、基因信息等数据; PRO 数据库收录了 MG53 的 ontology 等信息。

总结:

- (1) UniProt 提供了蛋白的基本信息和丰富的数据库链接, 应有尽有, 可以获得较为全面的研究对象相关数据;
- (2) UniProt 条日本身更新略有滞缓, 部分数据库更新较为及时;
- (3) 数据库太多, 环环相扣, 信息量庞大, 看不过来, 需要慢慢消化;
- (4) 有些内容目前看不懂, 需要继续深入学习;
- (5) MG53 当前研究并不多, 有很多潜在的功能可以挖掘, 通过上述数据可以用以发现 MG53 的新功能。