

На правах рукописи

ЖАБЕРЕВА Анастасия Сергеевна

**РАЗРАБОТКА КЛАССИФИКАЦИИ  
НЕСТ-ДОМЕННЫХ УБИКВИТИН-ПРОТЕИН ЛИГАЗ  
НА ОСНОВАНИИ ФИЛОГЕНЕТИЧЕСКОГО И СТРУКТУРНО-  
ФУНКЦИОНАЛЬНОГО АНАЛИЗА**

03.00.28 – биоинформатика

**АФТОРЕФЕРАТ**

диссертации на соискание ученой степени  
кандидата биологических наук

Москва - 2009

Работа выполнена в Государственном учреждении высшего профессионального образования «Нижегородская государственная медицинская академия федерального агентства по здравоохранению и социальному развитию»

**Научный руководитель:**

Кандидат медицинских наук

Гайнуллин Мурат Рушанович

**Официальные оппоненты:**

Доктор биологических наук

Лисица Андрей Валерьевич

Кандидат биологических наук

Афонников Дмитрий Аркадьевич

**Ведущая организация:**

Учреждение Российской академии наук Институт цитологии РАН

Защита состоится 18 июня 2009 года в 11 часов на заседании Диссертационного совета Д 001.010.01 при Учреждении Российской академии медицинских наук Научно-исследовательском институте биомедицинской химии им. В.Н.Ореховича РАМН (ИБМХ РАМН) по адресу: 119121, Москва, ул. Погодинская, 10.

С диссертацией можно ознакомиться в библиотеке ИБМХ РАМН по адресу: 119121, Москва, ул. Погодинская, 10.

Автореферат разослан «        »                      2009 года.

Ученый секретарь Диссертационного совета  
кандидат химических наук



Е.А.Карпова

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность проблемы.** Убиквитилирование - многоступенчатый процесс, который катализируется сложно организованной мультиферментной системой - убиквитин-протеин лигазным комплексом, состоящим из компонентов E1 (убиквитин-активирующий фермент), E2 (убиквитин-конъюгирующий фермент) и E3 – собственно убиквитин-протеин лигазы, которая катализирует финальное присоединение молекулы убиквитина к субстрату, а также наращивание мультиубиквитиновой цепи (Glickman and Ciechanover, 2002; Mann and Jensen, 2003). E3 лигазы являются центральными детерминантами процесса убиквитилирования, поскольку именно они распознают и убиквитилируют субстрат, что определяет молекулярные и/или клеточные эффекты убиквитилирования. Описанный выше трехступенчатый механизм инициирует все известные реакции убиквитилирования, независимо от того, предназначен ли субстрат для протеасомальной деградации или убиквитилирование играет регуляторную роль, не связанную напрямую с протеолизом (Wiesner et al., 2007). Во многих случаях субстрат, убиквитин и убиквитин-конъюгирующий фермент E2 должны присутствовать вместе в реакционном центре E3. Поэтому, в дополнение к функции передачи убиквитина на лизин белка-мишени, E3 должен также скоординировать взаимодействие большого количества белковых компонентов системы в своем реакционном центре (Pickart, 2001; Pickart and Eddins, 2004; Petroski and Deshaies, 2005).

Полифункциональность убиквитин-протеин лигаз опосредуется их структурными особенностями. В настоящее время известны 2 основные разновидности E3 лигаз: ферменты E3, содержащие каталитический RING домен (Really Interesting New Gene), и HECT-доменные E3 лигазы (Homologous to E6-associated protein C-Terminus, гомологичный C-концевой последовательности белка E6).

RING-доменные убиквитин-протеин лигазы координируют белок-мишень и E2-убиквитин-тиоэфирный комплекс таким образом, чтобы обеспечить прямую передачу убиквитина от E2 к белку-мишени. В данную группу, помимо мономерных RING-доменных белков, входят многокомпонентные белковые комплексы, например, куллин-содержащие убиквитин-протеин лигазные комплексы (SCF-комплексы), для которых каждая из молекулярных функций обеспечивается специфической субъединицей (Robinson and Ardley, 2004). Такая неоднородность состава группы существенно усложняет их структурно-функциональный анализ.

Напротив, все HECT-доменные E3 лигазы – мономерные белки, сами распознающие субстрат и сами же его убиквитирующие. В отличие от RING-доменных убиквитин-протеин лигаз, при взаимодействии между E2 и HECT-доменной E3 происходит перенос молекулы убиквитина сначала на остаток цистеина активного центра домена HECT. После этого HECT E3 взаимодействует с белком-мишенью, и передает на него убиквитин (Kee and Huibregtse, 2008). Не смотря на относительную простоту структурной организации HECT-доменных лигаз, для многих белков показана полифункциональность. Она, в свою очередь, определяется наличием в структуре E3 лигазы, помимо каталитического домена HECT, других доменов.

Убиквителирование – один из важнейших механизмов посттрансляционной регуляции свойств и функций белков. Регуляторный потенциал убиквителирования превосходит все известные посттрансляционные модификации, а по разнообразию контролируемых функций сопоставим с фосфорилированием. В частности, система убиквитина вовлечена в регуляцию клеточного цикла, передачу внутриклеточного сигнала, апоптоз, репарацию ДНК и др. (Haglund and Dikic, 2005; Belgareh-Touzer et al., 2008). В свою очередь, дефекты системы

убиквитина приводят к развитию различных патологических процессов, а сами компоненты системы убиквитина рассматриваются в качестве потенциальных мишеней новых лекарственных средств (Scheffner and Staub, 2007; Bernassola et al., 2009). Разнообразие биологических эффектов убиквитилирования в норме и вовлечение системы убиквитина в патогенез заболеваний подчеркивают актуальность исследований системы убиквитина, как в рамках фундаментальной науки, так и ее прикладных аспектов.

После завершения проектов по расшифровке геномов ряда эукариот, стало накапливаться большое количество данных о новых белках. При этом соотношение неохарактеризованных белков к экспериментально охарактеризованным белкам растет экспоненциально. Именно поэтому одним из актуальных вопросов является расшифровка функций неохарактеризованных белков.

По данным базы данных белков UniProt протеомного сервера ExPASy (<http://www.expasy.org/>), домен HECT присутствует в структуре 1246 белков из различных организмов. Однако 90 из них экспериментально охарактеризованы как белки, обладающие убиквитин-протеин лигазной активностью. Среди них структурно-функциональные взаимосвязи изучены лишь для белка RSP5 (и его гомологов) и белка HERC (и его гомологов). Остальные аминокислотные последовательности, содержащие домен HECT, являются функционально неохарактеризованными, и представлены в базе данных транслянтов с нуклеотидных последовательностей EMBL (TrEMBL). На сегодня отсутствует единая номенклатура HECT-доменных убиквити-протеин лигаз. Таким образом, актуальной является задача инвентаризации данного семейства белков с использованием различных инструментов компьютерного анализа.

**Цель работы:** Разработать классификацию НЕСТ-доменных убиквитин-протеин лигаз на основании комплексного анализа их структурных и функциональных особенностей.

**Задачи исследования:**

1. Провести филогенетический анализ каталитического домена НЕСТ НЕСТ-доменных убиквитин-протеин лигаз;
2. Провести анализ доменного состава полной аминокислотной последовательности НЕСТ-доменных убиквитин-протеин лигаз;
3. На основании совокупности филогенетических и структурно-функциональных данных создать классификацию НЕСТ-доменных убиквитин-протеин лигаз;
4. Установить наличие корреляции между профилями экспрессии НЕСТ-доменных убиквитин-протеин лигаз из организма *Homo sapiens* и предлагаемой классификацией.

**Научная новизна.** В работе впервые представлена комплексная классификация НЕСТ-доменных убиквитин протеин лигаз из семи организмов с расшифрованным геномом.

1. Для данной группы белков впервые проведен филогенетический анализ каталитического домена НЕСТ, который позволил разделить исследуемые белки на группы близких гомологов. Для белков выделенных гомологичных групп проведен анализ доменной организации.

2. На основании совокупности данных о сходстве первичных структур и общности доменной организации впервые создана структурно-функциональная классификация НЕСТ-доменных убиквитин-протеин лигаз.

3. В работе впервые проведен анализ профилей экспрессии НЕСТ-доменных убиквитин-протеин лигаз из организма *Homo sapiens* в норме и при патологии, а также сопоставление данных экспрессии исследуемых белков в норме и при патологии для ряда тканей.

**Практическая значимость.** Полученные данные могут быть использованы для прогнозирования функций неохарактеризованных белков, содержащих домен HEST. Полученные данные вошли в базы данных UbiProt и Ubiquitomix, посвященных системе убиквитина и разработанных в группе постгеномных технологий Научно-исследовательского института прикладной фундаментальной медицины при Государственном образовательном учреждении высшего профессионального образования «Нижегородская государственная медицинская академия федерального агентства по здравоохранению и социальному развитию». Проведенный комплексный анализ исследуемой группы белков представляет научный интерес с точки зрения решения фундаментальных вопросов механизмов и роли системы убиквитина как для клетки, так и для организма в целом.

**Апробация работы.** Основные положения диссертационной работы докладывались и обсуждались в ходе следующих конференций:

- Научно-практическая конференция, посвященная 90-летию чл.-корр. РАМН Н.Ю. Беленкова, Нижний Новгород, 2007;
- Proceedings of the 3-rd Moscow Conference on Computational Molecular Biology, Moscow, Russia, July 27-31, 2007;
- 4th International Conference “Genomics, Proteomics, Bioinformatics and Nanobiotechnologies for Medicine”, Moscow - Nizhny Novgorod – Moscow, 2008;
- Sixth International Conference on Bioinformatics of Genome Regulation and Structure, Novosibirsk, Russia, June 22-28, 2008;
- The BGRS'2008 International Summer School for Young Scientists "Evolution, Systems Biology and High Performance Computing Bioinformatics", Novosibirsk, Russia, June 29 - July 2, 2008.

**Публикации.** Материалы диссертационной работы отражены в 14 публикациях: 5 статей и 9 публикаций в сборниках докладов международных и российских научных конференций.

**Объем и структура диссертации.** Работа состоит из введения, обзора литературы, описания материалов и методов, собственных результатов, их обсуждения, выводов, списка литературы, включающего 180 источников, приложения. Диссертация изложена на 130 страницах машинописного текста, включая 14 таблиц и 24 рисунка.

## **МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ**

Разработка классификации НЕСТ-доменных Е3 лигаз включала три этапа:

1. предварительное распределение исследуемых белков на семейства на основании эволюционного родства их домена НЕСТ;
2. характеристика доменной организации исследуемых белков;
3. объединение данных филогенетического и структурного анализа.

Критериями разделения исследуемых белков на подсемейства служили:

- показатель идентичности, получаемый при парном и множественном выравнивании аминокислотных последовательностей белков;
- распределение белков на эволюционные кластеры на филогенетическом дереве домена НЕСТ и значения бутстреп-поддержки этих кластеров;
- наличие в структуре белков, входящих в состав кластера, других доменов, характерных только для данного кластера.

**Алгоритм поиска гомологов.** Исследуемые белки предварительно оценивались по критерию гомологии. Белки с тем или иным значением идентичности и сходства выявлялись с помощью алгоритма PSI-BLAST (Position-Specific Iterated BLAST) (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). В качестве последовательности запроса использовались аминокислотные последовательности дрожжевых представителей семейства НЕСТ-доменных белков. Параметры для поиска гомологии:

- матрица замен – BLOSUM62
- штраф за открытие вставки – 7, за продолжение вставки – 2
- ограничение на вероятность случайного совпадения – 10.0

При обработке результатов учитывались такие параметры как счет и E-value.

**Парное выравнивание.** Парное выравнивание аминокислотных последовательностей проводилось с помощью алгоритма BLAST. В качестве последовательности сравнения использовались аминокислотные последовательности дрожжевых представителей семейства НЕСТ-доменных белков, то есть все исследуемые белки последовательно сравнивались с каждой из дрожжевых последовательностей. Во всех случаях использовались следующие параметры:

- матрица замен – BLOSUM62
- штраф за открытие вставки – 7, за продолжение вставки – 2
- ограничение на вероятность случайного совпадения – 10.0

При обработке результатов учитывались такие параметры как счет, E-value, идентичность и схожесть.

**Множественное выравнивание.** Для проведения множественных выравниваний использовались интерактивные ресурсы MAFFT (<http://align.bmr.kyushu-u.ac.jp/mafft/online/server/>) и MUSCLE (<http://www.ebi.ac.uk/Tools/es/cgi-bin/muscle/>). Множественное выравнивание рассчитывалось со следующими параметрами:

- матрица замен – BLOSUM62
- штраф за открытие вставки – 1.53, за продолжение вставки – 0.00

Редактирование полученных множественных выравниваний осуществлялось в программе BioEdit (версия 7.0.5.2).

**Филогенетический анализ.** Филогенетические деревья были получены с помощью дистанционно-матричного метода (метод ближайших соседей, Neighbor-Joining, NJ) (Saitou and Nei, 1987) и метода максимальной экономии

(Maximum Parsimony, MP) (Eck and Dayhoff, 1966; Felsenstein, 1993) пакета программ PHYLIP (версия 3.66) с коррекцией многократных замен и 1000 псевдорепликами. Программа TreeView Win32 (версия 1.6.6) использовалась для визуализации филогенетических деревьев.

**Доменный анализ.** Данный этап проводился с использованием баз данных консервативных доменов NCBI CDD (<http://0-www.ncbi.nlm.nih.gov.library.vu.edu.au/Structure/cdd/cdd.shtml>), Pfam (<http://www.sanger.ac.uk/Software/Pfam/>) и SMART ([http://smart.embl-heidelberg.de/smart/set\\_mode.cgi?NORMAL=1](http://smart.embl-heidelberg.de/smart/set_mode.cgi?NORMAL=1)).

Аминокислотные последовательности каждого из исследуемых белков анализировались в базах данных на наличие структурно-функциональных доменов. Выявление доменов в исследуемых белках происходит посредством алгоритма BLAST. Последовательность запроса сравнивается с позиционно-специфической матрицей счета, полученной при выравнивании основных консервативных доменов. Результаты могут быть показаны как парные выравнивания последовательности запроса с имеющимися последовательностями доменов или как множественные выравнивания.

**Транскриптомный анализ.** Был проведен транскриптомный анализ НЕСТ-доменных убиквитин-протеин лигаз, экспрессирующихся в организме *Homo sapiens*. Поиск профилей экспрессии белков велся по транскриптомным библиотекам в базе данных UniGene молекулярно-биологического Интернет-сервиса NCBI. База данных UniGene (<http://www.ncbi.nlm.nih.gov/entrez/unigene>) представляет собой автоматизированную аналитическую систему, в которой собрана информация о геномах и транскриптомах различных организмов.

Для проведения транскриптомного анализа в UniGene информация представляется в виде таблицы или гистограммы, отражающей профиль экспрессии гена, кодирующего конкретный белок. Профилирование ведется по органам, стадиям развития организма и различным патологическим состояниям и представляет собой оценку уровня экспрессии гена. Уровень экспрессии гена

выражается через количество его транскриптов на миллион (TPM, transcripts per million). В UniGene при подобной форме представления данных для транскриптомного анализа учитываются все имеющиеся на данный момент транскриптомные библиотеки.

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

**Формирование выборки.** С использованием базы данных PROSITE была сформирована выборка, в которую вошли представители НЕСТ-доменных убиквитин-протеин лигаз из семи организмов с расшифрованным геномом: *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Arabidopsis thaliana*, *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*. По данным UniProt список содержал 207 белков. На основании аннотаций, представленных в записях по каждому белку, из полученного списка были исключены дубликаты одного гена и сплайсинг-варианты. На основании парного выравнивания, белки с высокой идентичностью (по крайней мере 95%) и короткие белковые фрагменты также были исключены из анализа. Конечный список содержал 122 белка, 71 из которых является экспериментально охарактеризованными белками из базы данных SWISS-PROT, 51 являются неохарактеризованными белками из базы данных TrEMBL.

Парные выравнивания исследуемых аминокислотных последовательностей показали, что белки семейства НЕСТ-доменных убиквитин-протеин лигаз имеют, по крайней мере, 30% идентичность. В большей степени такой уровень сходства объясняется тем, что все белки данного семейства содержат на С-концевом участке домен НЕСТ.

С целью определения уровня сходства между белками семейства НЕСТ-доменных E3 лигаз, наряду с вычислением процента идентичных аминокислотных остатков в их последовательностях, анализировали также порядок появления белков при множественных итерациях программы PSI-BLAST. В качестве исходной использовали последовательности НЕСТ-

доменных убиквитин-протеин лигаз из организма *Saccharomyces cerevisiae* (HUL4\_YEAST, HUL5\_YEAST, TOM1\_YEAST, UFD4\_YEAST, RSP5\_YEAST). Представители данного семейства были получены после первой-второй итерации из полной базы данных аминокислотных последовательностей. При последующих итерациях были получены белки из прокариотических организмов и не охарактеризованные как НЕСТ-доменные Е3 лигазы. Эти данные подтверждают то, что анализ включал всех представителей НЕСТ-доменных Е3 лигаз из семи анализируемых нами организмов.

**Филогенетический анализ НЕСТ-доменных убиквитин-протеин лигаз.** На данном этапе анализируемые белки были распределены на группы близких гомологов с помощью методов множественного выравнивания и филогенетического анализа.

Учитывая то, что в процессе эволюции большинство белков структурно усложнялись (что обуславливается появлением в их структуре новых функциональных доменов), множественное выравнивание и филогенетический анализ исследуемой группы белков был проведен по домену НЕСТ белков группы. Для этого на основании данных базы консервативных доменов NCBI CDD с помощью программы BioEdit из первичной структуры исследуемых белков «вырезался» домен НЕСТ. Коллекция доменов НЕСТ исследуемых белков была сформирована в отдельном файле программы BioEdit для дальнейшего анализа.

С целью минимизации ошибок при построении множественного выравнивания, было использовано два метода для его вычисления. Далее производился поиск лучшего выравнивания на основании данных о консервативных областях домена НЕСТ и его правка (удаление переменных участков) в программе BioEdit.

На основании полученного множественного выравнивания были построены филогенетические деревья, которые иллюстрируют картину распределения белков на группы и схемы родственных отношений между ними (белками и группами белков).

Филогенетические деревья, построенные методами ближайших соседей (Neighbor-Joining, NJ) и максимальной экономии (Maximum Parsimony, MP) имеют сходную топологию. Этот результат свидетельствует об объективности распределения исследуемых белков на подсемейства. Кроме того, для каждого метода были рассчитаны значения бутстреп-поддержки. Результаты представлены в таблице 1.

На филогенетическом дереве все исследуемые белки образуют 9 кластеров. Группы, содержащие дрожжевого представителя, были названы соответственно (UFD4, HUL4, HUL5, RSP5, TOM1), прочие группы были названы по представителю из организма *Homo Sapiens* (UBR5, LHERC (Large, большие), SHERC (Small, малые), K0317).

**Таблица 1.** Значения бутстреп-поддержки узлов кластеров на филогенетических деревьях, построенных методами ближайших соседей (NJ) и максимальной экономии (MP).

Название кластера	Бутстреп-поддержка NJ (%)	Бутстреп-поддержка MP (%)
<b>UBR5</b>	100	100
<b>UFD4</b>	100	99
<b>LHERC</b>	97	94
<b>SHERC</b>	99	98
<b>HUL4</b>	98	98
<b>HUL5</b>	75	40
<b>RSP5</b>	67	51
<b>TOM1</b>	78	57
<b>K0317</b>	100	100

На филогенетическом дереве все исследуемые белки образуют 9 кластеров. Группы, содержащие дрожжевого представителя, были названы соответственно (UFD4, HUL4, HUL5, RSP5, TOM1), прочие группы были

названы по представителю из организма *Homo Sapiens* (UBR5, LHERC (Large, большие), SHERC (Small, малые), K0317).

Кластер I, соответствующий подсемейству UBR5, имеет высокую бутстреп-поддержку - 100%. Кластер II (группа UFD4) также имеет достаточно устойчивое положение (бутстреп-поддержка 100%). Похожая картина наблюдается и для кластера III (группа LHERC) для которого бутстреп-поддержка составляет 97%. Кластер IV, соответствующий подсемейству SHERC, имеет высокую бутстреп-поддержку - 99%. Кластер V (группа HUL4) также имеет достаточно устойчивое положение (бутстреп-поддержка 98%). Бутстреп-поддержка общего узла SHERC и HUL4 составляет 98%. Кластер VI (группа HUL5) имеет значения бутстреп-поддержки 75%. Значение бутстреп-поддержки для кластера VII (группа RSP5) составляет 67%. Кластер VIII (группа TOM1) также устойчив (значения бутстреп-поддержки узла составляет 78%). Кластер IX (группа K0317) имеет высокую бутстреп-поддержку (100%).

Эволюция белков семейства НЕСТ-доменных E3 убиквитин-протеин лигаз была предположительно сопряжена с множественными дупликациями, потерей и слиянием генов, а также их горизонтальными переносами, что подтверждается, например, сложной многодоменной архитектурой ряда охарактеризованных представителей. В результате филогенетическое дерево НЕСТ-доменных E3 лигаз из семи полногеномных организмов имеет достаточно сложную многоуровневую топологию.

На основании филогенетического анализа НЕСТ доменов, исследуемые НЕСТ-доменные убиквитин-протеин лигазы из 7 отобранных организмов сгруппированы в 9 кластеров. Белки в этих кластерах характеризуются высоким уровнем сходства первичных последовательностей и имеют общее эволюционное происхождение. В 5 подсемействах прослеживается

эволюционное родство белков высших эукариот с ортологичными белками из дрожжей *Saccharomyces cerevisiae*.

**Анализ доменной организации НЕСТ-доменных убиквитин-протеин лигаз.** Анализ доменной организации показал, что, помимо С-концевого каталитического домена НЕСТ, N-концевой участок многих представителей исследуемой группы белков включает и другие домены.

Среди всех найденных доменов удалось выявить те, которые сохранились у большинства представителей каждого филогенетического кластера. Таким образом, был сформирован список доменов, характерных для НЕСТ-доменных белков.

Для представителей первого филогенетического кластера характерны домены ZnF\_UBR1 и PolyA, основной функцией которых являются белок-белковые взаимодействия. Так, домен ZnF\_UBR1, с одной стороны отвечает за связывание ДНК или РНК, белок-белковые взаимодействия, ассоциацию с мембраной, а с другой стороны участвует в узнавании N-конца субстратов для деградации по правилу N-конца. Домен PolyA вовлечен в гомодимеризацию, инициацию трансляции у эукариот и стабилизацию/деградацию мРНК.

Для представителей второго филогенетического кластера наиболее характерны домены ANK, Sad1, MIB\_HERC2, ARM, SRP1 и WWE. В процессе филогенетического анализа белки данного подсемейства организовали три подкластера, для которых характерна собственная доменная организация.

Так, представители 1-ого подкластера содержат домены ARM, SRP1 и WWE. Домены ARM и WWE участвуют в распознавании убиквитин-протеин лигазой белков-субстратов. Домен SRP1 участвует во внутриклеточном транспорте и секреции.

Представители 2-ого подкластера имеют практически идентичную доменную организацию, характеризующуюся доменами ANK, Sad1 и MIB\_HERC2. Анкириновые повторы (домен ANK) отвечают за белок-белковые взаимодействия и встречаются в большом количестве разнообразных белковых семейств. Анкириновые повторы являются одними из наиболее общих мотивов, ответственных за белок-белковые взаимодействия, и встречаются функционально разных белках, главным образом, эукариотических организмов. В основном это белки - инициаторы транскрипции, регуляторы клеточного цикла, цитоскелета, ионные транспортеры (Ju et al., 2007).

Домен Sad1 (или С-концевой домен, гомологичный белку UNC, UNC-like C-terminal) гомологичен белку UNC-48 из организма *Caenorhabditis elegans*, который участвует в процессах миграции ядра в процессе развития.

Домен MIB\_HERC2 взаимодействует с внутриклеточным доменом Delta, иницируя его убиквитилирование и интернализацию. Функция MIB важна в процессе передачи клеточного сигнала для эффективной активации белка Notch.

Повторы ARM белка UFD4 распознают убиквитиновые сигналы для деградации субстратов UFD4. Показано, что блокирование повторов ARM ведет к ингибированию убиквитилирующей активности UFD4.

Третий филогенетический кластер также состоит из двух подкластеров. Для первого из них характерно наличие таких доменов, как APC10, RCC1 и ATS1. Данные домены задействованы в процессе деления клетки.

В частности, APC10, являясь компонентом комплекса APC, отвечает за переход от метафазы к анафазе и выход клетки из митоза.

RCC1 участвует в регуляции конденсации хромосом. Данный домен связывает хроматин и взаимодействует с белком gap - ядерным ГТФ-связывающим белком, блокирующим ГТФазу и действующим как

стимулятор диссоциации гуанин-нуклеотида (GDS). Взаимодействие RCC1 с белком *gap* играет важную роль в регуляции экспрессии генов.

Домен ATS1 является родственным домену RCC1. Установлено, что данный домен участвует в регуляции процесса деления клетки и разделения хромосом.

Другие домены, такие как MIB\_HERC2, Cyt-b5 и ZZ-Herc2, выявленные у ортологичных белков HERC2, являются модулями белок-белковых взаимодействий в процессах передачи клеточного сигнала и регуляции клеточного цикла.

Для представителей четвертого кластера характерно наличие доменов ATS1 и RCC1, описанных выше для первого подкластера третьего филогенетического кластера. Белки данных групп входят в одно семейство и являются HERC белками, отличающимися по длине аминокислотой последовательности (Hochrainer et al., 2005).

Для белков пятого и шестого кластеров характерна монодоменная организация. В данные кластеры, в частности, входят такие НЕСТ-доменные лигазы, как HUL4 и HUL5 из организма *Saccharomyces cerevisiae*, и их гомологи из организма *Homo sapiens* UBE3A и UBE3C, UBE3B соответственно.

Белки, входящие в состав седьмого кластера являются мультидоменными и содержат C2 и WW домены, за исключение представителя Pub2 из организма *Schizosaccharomyces pombe*, который является уникальным представителем. Данная уникальность обусловлена тем, что белок Pub2 в своем составе не имеет домен C2. Вместе с тем показано, что данный белок является одним из прототипов данной группы белков (Tamaï and Shimoda, 2002).

Домен C2 является кальций-связывающим мотивом. Данный мотив связывает фосфолипиды, инозитолфосфаты и внутриклеточные белки. Этот

модуль найден во многих белках, вовлеченных в трансдукцию сигнала или мембранный транспорт.

Домен WW является модулем белок-белковых взаимодействий, и представляет собой два консервативных триптофановых домена, известных также как домены WWP и RSP5. Функционирует данный домен как модуль взаимодействия для разнообразного набора сигнальных белков. Он связывает специфические богатые пролином последовательности, но с низкой специфичностью по сравнению с другими белками, распознающими этот пептид, такими как антитела и рецепторы. Домен WW связывает белки со специфическими пролиновыми мотивами и/или мотивами, содержащими фосфосерин-фосфотреонин. Он также связывается с другими доменами белков, участвующих в процессе передачи сигнала. В рамках убиквитилирования домены WW отвечают за распознавание белков-субстратов.

Белки, входящие в состав восьмого подкластера, содержат в своем составе домены DUF908 и DUF913. Эти домены участвуют в транскрипции и процессинге мРНК, обладают убиквитин-протеин лигазной активностью. В данную группу входят такие представители, как белки PTR1\_SCHPO и TOM1\_YEAST, которые в свою очередь характеризуются высоким процентом идентичности и сходства аминокислотных последовательностей при парном выравнивании данных белков.

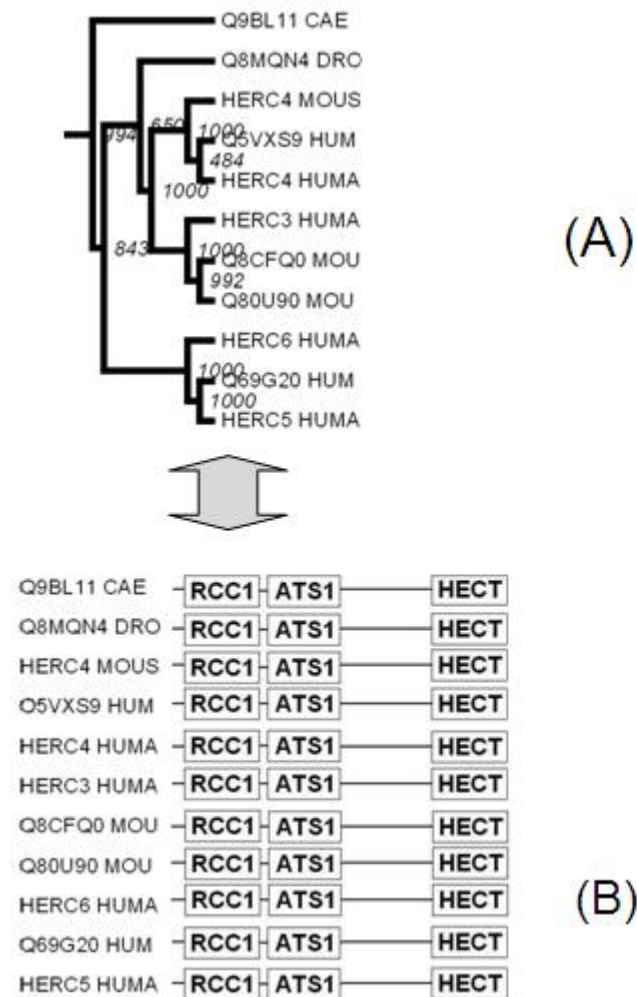
Доменный анализ представителей девятого кластера показал наличие домена Filamin в составе их первичной структуры. Домен Filamin (или АВР280 повтор) входит в состав домена rod, и состоит из 24 tandemных повторов, содержащих по 100 аминокислот и богатых глицином и пролином.

**Формирование классификации НЕСТ-доменных убиквитин-протеин лигаз.** При совместном анализе филогенетических данных по изолированному НЕСТ домену и данных о доменном составе полной

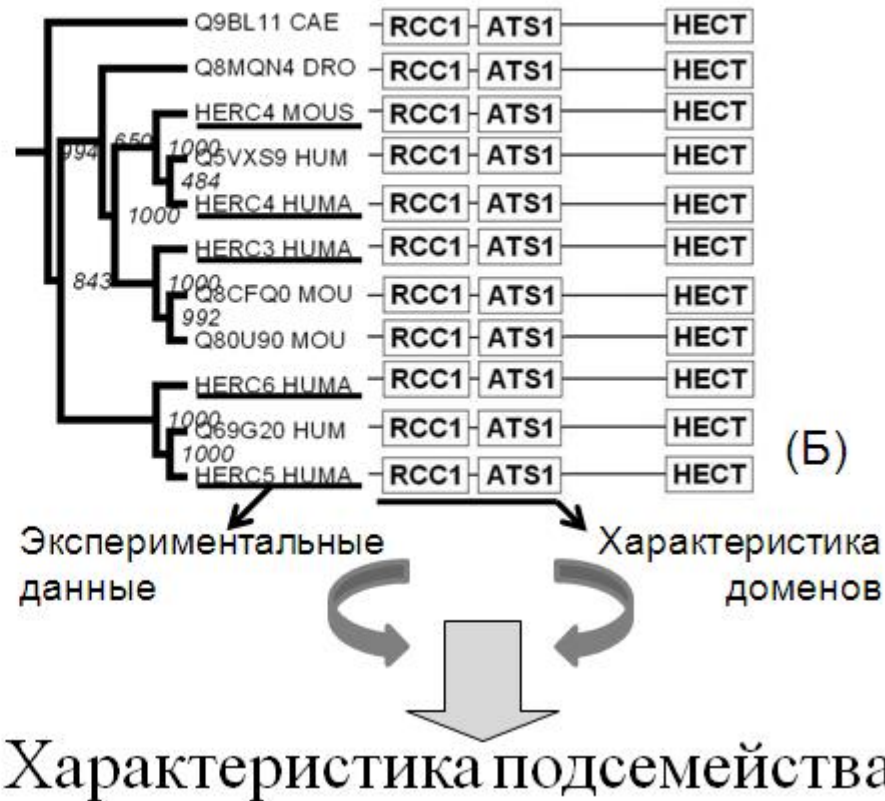
аминокислотной последовательности исследуемых белков было выделено девять групп гомологов, характеризующихся своим уникальным набором доменов (Рис. 1).

На основании этого филогенетические кластеры были выделены в качестве подсемейств НЕСТ-доменных убиквитин-протеин лигаз, была дана характеристика доменной организации каждого подсемейства, а также приведены функции некоторых белков подсемейств на основании анализа литературы. Таким образом, семейство НЕСТ-доменных убиквитин-протеин лигаз включает в себя 9 подсемейств: UFD4, HUL4, HUL5, RSP5, TOM1, UBR5, LHERC, SHERC, K0317 (Таблица 2).

**Анализ профилей экспрессии НЕСТ-доменных убиквитин-протеин лигаз.** Проведен анализ профилей экспрессии НЕСТ-доменных E3 лигаз из организма *Homo Sapiens*. Профиль экспрессии удалось установить для 37 белков. Затем был произведен количественный анализ уровня экспрессии в норме, профилирование велось по 47 органам. Также было произведено количественное сравнение уровней экспрессии белков в норме и при опухолевой патологии. При патологии профилирование для каждого белка велось по 26 органам. Сопоставить профили экспрессии белков в норме и при опухолевой трансформации удалось для 9 органов: надпочечники, пищевод, почки, печень, молочная железа, яичники, матка, ткани кожи, ткани шеи (рис.2). Установлено, что для НЕСТ-доменных убиквитин-протеин лигаз внутри подсемейств характерна сходная картина изменения уровня экспрессии при опухолевой патологии.



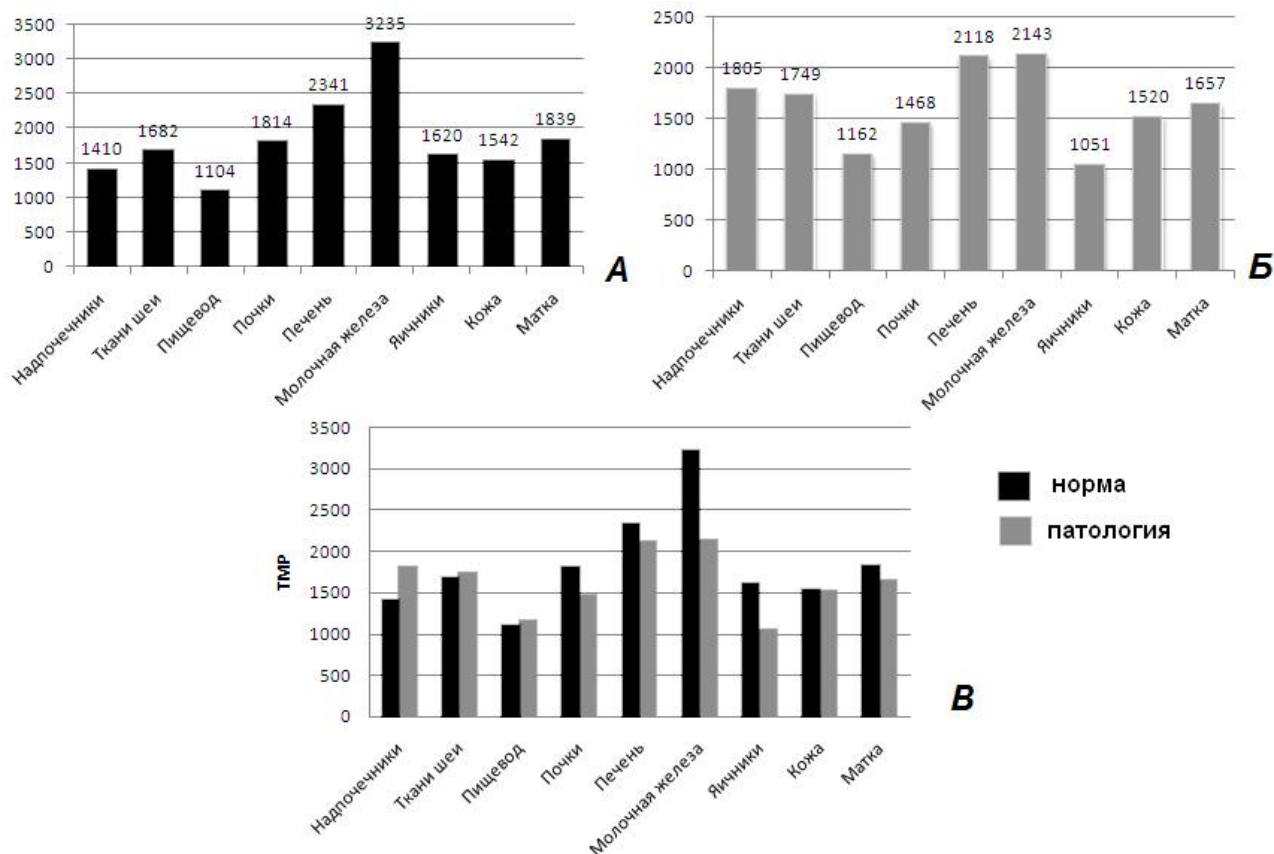
## Подсемейство S-HERC



**Рис. 1.** Создание классификации на примере подсемейства S-HERC. (A) Фрагмент филогенетического дерева, соответствующий кластеру S-HERC; (B) Результаты доменного анализа белков, входящих в кластер S-HERC; (B) Сопоставление данных филогенетического, доменного анализа и экспериментальных данных для создания классификации

**Таблица 2.** Классификация HECT-доменных убиквитин-протеин лигаз.

Суперсемейство
<b>Убиквитин-протеин лигазы</b>
Семейство
<b>HECT-доменные убиквитин-протеин лигазы</b>
Подсемейства
<b>UFD4</b>
<u>Характерны домены:</u> ANK, Sad1, MIB_HERC2, ARM, SRP1, WWE, HECT. <u>Функции:</u> убиквитилирование и деградация ключевых белков процессов секреции и транспорта, регуляции клеточного цикла.
<b>HUL4</b>
<u>Характерны домены:</u> HECT. <u>Функции:</u> регуляции транскрипции и экспрессии генов посредством убиквитилирования и деградации белков, непосредственное участие в развитии патологии.
<b>HUL5</b>
<u>Характерны домены:</u> HECT. <u>Функции:</u> регуляция активности 26S протеасомы, убиквитилирование и деградация белков, регуляция транскрипционной активности.
<b>RSP5</b>
<u>Характерны домены:</u> C2, WW, HECT. <u>Функции:</u> Регуляция работы рецепторов и клеточных каналов, регуляция клеточного цикла и работы факторов транскрипции. Протеасомальная деградация и эндцитоз. Связаны с мембраной.
<b>TOM1</b>
<u>Характерны домены:</u> DUF908, DUF913, HECT. <u>Функции:</u> регуляция транскрипции и процессинга мРНК, убиквитилирование ядерных белков.
<b>UBR5</b>
<u>Характерны домены:</u> ZnF-UBR1, PolyA, HECT. <u>Функции:</u> убиквитилирование и деградация ключевых белков процессов регуляции клеточного цикла, мембранного и секреторного транспорта и распознавания повреждения ДНК.
<b>LHERC</b>
<u>Характерны домены:</u> APC10, RCC1, ATS1, MIB_HERC2, Cyt-b5, ZZ_Herc2, HECT. <u>Функции:</u> Участие в процессах клеточного деления, передачи клеточного сигнала и внутриклеточного транспорта.
<b>SHERC</b>
<u>Характерны домены:</u> RCC1, ATS1, HECT. <u>Функция:</u> участие в процессе взаимодействия белков с убиквитин-подобным белком ISG15, регуляции клеточного цикла на уровне организации хромосом и экспрессии генов.
<b>K0317</b>
<u>Характерны домены:</u> Filamin, HECT. <u>Функции:</u> Локализованы в мембране. Регуляция мембранного транспорта?



**Рис. 2.** Значение экспрессии исследуемых НЕСТ-доменных убиквитин-протеин лигаз для ряда тканей в норме (А), при патологии (Б), и сопоставление профилей экспрессии в норме и при патологии (В).

## ВЫВОДЫ

1. Филогенетический анализ позволил объединить исследуемые белки в группы на основании гомологии и эволюционного родства первичных структур их HECT-домена: достоверно (97 – 100%) для групп UBR5, UFD4, LHERC, SHERC, HUL4, K0317, с меньшей долей достоверности (67 – 78%) для групп HUL5, RSP5, TOM1.

2. Доменный анализ позволил охарактеризовать доменную организацию исследуемых белков. Выявлено сходство доменной организации внутри каждой из филогенетических групп исследуемых белков семейства HECT-доменных убиквитин-протеин лигаз.

3. Объединение филогенетических и структурно-функциональных данных позволило сформировать классификацию HECT-доменных убиквитин-протеин лигаз. Семейство HECT-доменных убиквитин-протеин лигаз входит в суперсемейство убиквитин-протеин лигаз, и подразделяется на 9 подсемейств: UBR5, UFD4, LHERC, SHERC, HUL4, K0317, HUL5, RSP5 и TOM1, которые характеризуются собственным уникальным набором доменов.

4. Анализ профилей экспрессии показал разнообразие тканей, в которых представлены исследуемые белки. Для белков внутри подсемейств характерна сходная картина уровня экспрессии в норме, а также его изменения при опухолевой патологии.

## СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Гайнуллин М.Р. Шорина А.С. (Жаберева А.С.) Биоинформатический анализ особенностей пространственного строения белков – мишеней убиквитилирования // Нижегородский медицинский журнал. – 2003. - №1. - С. 66 - 73.
2. Chernorudskiy A.L., Garcia A., Eremin E.V., Shorina-Zhabereva A.S., Kondratieva E.V., Gainullin M.R. UbiProt: a database of ubiquitylated proteins // BMC Bioinformatics. – 2007. - №8. - P. 126 – 141.
3. Chernorudskiy A.L., Shorina A.S., Garcia A., Gainullin M.R. Prediction of direct effects of protein ubiquitylation using computational analysis // Biophysics. – 2006. - Vol. 51. - Suppl. 1. - P. 39.
4. Кондратьева Е.В., Шорина А.С., Чернорудский А.Л. Исследование динамики содержания убиквитилированных белков в плазме крови животных–опухоленосителей // Сибирский онкологический журнал. – 2008. - Приложение № 1. - С. 70 – 71.
5. Жаберева А.С., Чаплыгина Е.В., Гайнуллин М.Р. Филогенетический и структурно-функциональный подходы к разработке классификации НЕСТ-доменных убиквитин-протеин лигаз // Нижегородский медицинский журнал. – 2008. - № 3. - С. 123-124.
6. Гарсия А., Чернорудский А.Л., Шорина А.С. (Жаберева А.С.), Кондратьева Е.В., Еремин Е.В. База данных UbiProt: объектно-ориентированный Интернет-ресурс, посвященный убиквитилированным белкам // Международная конференция «Рецепция и внутриклеточная сигнализация». – Пущино. – 2007. - С. 12-15.
7. Chernorudskiy A., Garcia A., Eremin E., Shorina-Zhabereva A., Kondratieva E., Gainullin M.R. Ubiprot Database and analysis of protein ubiquitylation for application in pharmacological research // Fourth International Symposium on

- Computational Methods in Toxicology and Pharmacology Integrating Internet Resources. – Moscow. – 2007. - P. 88.
8. Шорина А.С. (Жаберева А.С.), Чернорудский А.Л., Гарсия А., Корягин А.С., Гайнуллин М.Р. Биоинформатический анализ влияния убиквитилирования на каталитическую активность ферментов // XVIII зимняя молодежная научная школа "Перспективные направления физико-химической биологии и биотехнологии". – Москва. – 2006. - С. 54.
  9. Chernorudskiy A.L., Shorina-Zhabereva A.S., Garcia A., Gainullin M.R. Direct influence of ubiquitylation on a target protein activity: “loss-of-function” mechanism revealed by computational analysis // Proceedings of the Fifth International Conference on Bioinformatics of Genome Regulation and Structure (Eds. N. Kolchanov, R. Hofestädt). – Novosibirsk. – 2006. - V. 1. - P. 252-255.
  10. Кондратьева Е.В., Шорина А.С. (Жаберева А.С.), Гайнуллин М.Р. Содержание убиквитина и убиквитин-белковых конъюгатов в плазме крови как диагностический критерий при малигнизации // Материалы конференции молодых ученых "Современные технологии диагностики и лечения в клинической медицине". - Санкт-Петербург. - 2006 , С. 24.
  11. Chernorudskiy A.L., Garcia A., Eremin E.V., Shorina-Zhabereva A.S., Kondratieva E.V., Gainullin M.R. UbiProt: a database of ubiquitylated proteins as a tool for computational analysis of ubiquitylation // A joint meeting of the Genetics Society and the British Societies for the Cell and Developmental Biology. – Edinburgh. – 2007. - P. 52.
  12. Chernorudskiy A., Shorina-Zhabereva A., Garcia, Gainullin M.R. A.Direct effects of protein ubiquitylation revealed by computational analysis // EMBO Conference “Ubiquitin and Ubiquitin-like modifiers in cellular regulation”. – Italy. – 2007. - P. 45.

13. Gainullin M.R., Zhabereva A.S., Chaplygina E.V., Okunev O.E. Classification and functional characterization of the HECT-domain ubiquitin-protein ligases // Sixth International Conference on Bioinformatics of Genome Regulation and Structure. – Novosibirsk. – 2008. - P. 266.
14. Chernorudskiy A.L., Eremin E.V., Astashev M.E., Zhabereva A.S., Garcia A., Gainullin M.R. A systems biology strategy for the studies of the ubiquitin system // 4th International Conference “Genomics, Proteomics, Bioinformatics and Nanobiotechnologies for Medicine”. – Moscow - Nizhny Novgorod – Moscow. – 2008. - P. 59.