

На правах рукописи

РЫБИНА
ОЛЬГА ЮРЬЕВНА

**Исследование транскрипции и природного полиморфизма
гена *Lim3*, участвующего в контроле продолжительности
жизни *Drosophila melanogaster***

03.02.07 – генетика

Автореферат
диссертации на соискание ученой степени
кандидата биологических наук

МОСКВА – 2010

Работа выполнена в лаборатории геномной изменчивости отдела молекулярной
генетики клетки Учреждения Российской академии наук
Института молекулярной генетики РАН

Научный руководитель: доктор биологических наук
Пасюкова Елена Генриховна

Официальные оппоненты: доктор биологических наук, профессор
Сулимова Галина Ефимовна
УРАН Институт общей генетики им. Н.И.
Вавилова РАН

доктор биологических наук, профессор
Тарантул Вячеслав Залманович
УРАН Институт молекулярной генетики РАН

Ведущая организация: Учреждение Российской академии наук Институт
биологии развития им. Н.К. Кольцова РАН

Защита состоится «__» _____ 2010 г. в __ часов на заседании диссертационного
совета Д 002.214.01 при Учреждении Российской академии наук Институте общей
генетики им. Н.И. Вавилова РАН по адресу: 119991, ГСП-1, г. Москва, ул. Губкина, д. 3.
Тел.: (499) 135-62-13
Факс: (495) 132-89-62
электронная почта: iogen@vigg.ru, aspirantura@vigg.ru
адрес в Интернете: www.vigg.ru.

С диссертацией можно ознакомиться в библиотеке ИОГен РАН

Автореферат разослан «__» _____ 20 г.

Ученый секретарь
диссертационного совета,
кандидат биологических наук

Синельщикова Татьяна Аркадьевна

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность проблемы

Старение как процесс возрастных изменений организма, ведущий к снижению его функциональной активности и увеличению вероятности смерти, свойственен большинству живых организмов, но темп и форма этого процесса могут резко отличаться у представителей, как различных таксонов, так и внутри одного вида. Такие различия в продолжительности жизни (ПЖ) обусловлены взаимодействием генетических факторов и факторов внешней среды. Температура, питание, загрязнение окружающей среды, являясь внешними факторами, запускают в организме каскады реакций, являющиеся компонентами генетически детерминированных сигнальных и метаболических путей, и в результате могут влиять на изменение ПЖ организма. Таким образом, в генетическом контроле длительности жизни может принимать участие множество генов. Действительно, наследуемость ПЖ не превышает 30%, что характерно для полигенного количественного признака.

В последнее время достигнут значительный прогресс в понимании того, какие гены определяют ПЖ у организмов различных систематических групп. Тем не менее, выявление новых генов, участвующих в контроле ПЖ, а главное, анализ особенностей их влияния на признак и характера их взаимодействия друг с другом являются актуальными задачами современной науки, решение которых необходимо для понимания причин и условий, обеспечивающих высокую ПЖ. Один из возможных подходов к решению этих задач заключается в изучении влияния структурного природного полиморфизма генов на изменение ПЖ организма.

Природный полиморфизм ранее был описан для сравнительно небольшого числа генов *Drosophila melanogaster* (Hasson et al., 1998; Balakirev, Ayala, 2003; Palsson et al., 2004). В ряде случаев была продемонстрирована взаимосвязь описанного природного полиморфизма с количественными признаками, включая ПЖ (De Luca et al., 2003; Carbone et al., 2006). Однако при этом не уделялось должного внимания изучению причинно-следственных связей между структурной вариабельностью гена и его экспрессией и, далее, изменением какого-либо признака организма. Очевидно, что именно такого рода информация имеет принципиально важное значение для понимания биологической роли генетической изменчивости. Хорошо известно, что ПЖ может существенно различаться у особей одного и того же вида. Понимание генетических причин такой внутривидовой и внутрипопуляционной изменчивости может быть достигнуто в результате исследования ассоциаций между структурной и функциональной вариабельностью генетического материала, приводящей к изменчивости ПЖ.

Ранее в Лаборатории геномной изменчивости Учреждения Российской академии наук Института молекулярной генетики РАН с помощью различных методов генетического

картирования было выявлено несколько групп генов, связанных с ранее неизвестными путями контроля ПЖ *Drosophila melanogaster*. В числе генов-кандидатов был ген *Lim3*.

Lim3, расположенный в левом плече второй хромосомы, в цитологической позиции 37B13-37C1, кодирует транскрипционный фактор РНК-полимеразы II. Белок *Lim3* является мишенью сложных регуляторных сетей и играет ключевую роль в регуляции развития нейронов дрозофилы. *Lim3* вместе с белками *Islet* и *Drifter* образуют «комбинаторный код», который определяет свойства мотонейронов, контролирует миграцию их аксонов и в итоге играет важную роль в обеспечении правильной мышечной иннервации (Thor et al., 1999; Certel, Thor, 2004; Landgraf, Thor, 2006). Таким образом, роль *Lim3* в развитии и функционировании нервной системы дрозофилы может являться существенным фактором для определения ПЖ дрозофилы.

Белок, кодируемый геном *Lim3*, обнаруживает высокий уровень гомологии с белками *LHX3/4* человека. Мутации в генах *LHX3/4* приводят к нарушениям секреции гормонов гипофиза, что является причиной различных нарушений, связанных с репродуктивной системой, ростом и метаболизмом. Таким образом, ген *Lim3* является геном-ортологом дрозофилы, перспективным для изучения не только ПЖ, но и механизмов нейроэндокринной регуляции человека.

Цель и задачи исследования

Целью нашей работы являлось исследование роли молекулярной изменчивости гена *Lim3* в определении уровня его транскрипции и контроле ПЖ *Drosophila melanogaster*.

В работе были поставлены следующие задачи:

1. Определить особенности организации 5' регуляторной области и инициации транскрипции транскрипта *Lim3A* гена *Lim3*.
2. Описать молекулярный полиморфизм 5' регуляторной области и начала структурной области транскрипта *Lim3A* гена *Lim3* в природной популяции *Drosophila melanogaster*.
3. Исследовать ассоциацию между молекулярным полиморфизмом 5' регуляторной области и начала структурной области транскрипта *Lim3A* гена *Lim3* и ПЖ.
4. Исследовать ассоциацию между молекулярным полиморфизмом 5' регуляторной области и начала структурной области транскрипта *Lim3A* гена *Lim3* и экспрессией *Lim3A*.

Научная новизна результатов исследования

В работе показано существование новой, ранее неаннотированной мРНК гена *Lim3*, *Lim3C*. Впервые определены особенности организации регуляторной области *Lim3A*, *Lim3C*: обнаружены регуляторные элементы в коровой и проксимальной промоторных областях, свидетельствующие о возможности инициации транскрипции этих мРНК со своего собственного промотора; показано наличие множественных стартов транскрипции с промоторов, не содержащих ТАТА-боксов.

В результате изучения молекулярной варибельности регуляторной области, экзона и части интрона *Lim3A* и *Lim3C* описано распределение полиморфных сайтов по функционально различным районам гена и продемонстрировано действие стабилизирующего отбора на область ДНК, соответствующую первому экзону *Lim3A* и *Lim3C*.

Показано, что варибельность регуляторной области гена, расположенной на расстоянии 680-380 п.н. от главного старта транскрипции, связана с изменением количества транскрипта на разных этапах развития и в разных частях тела взрослых мух. Этот результат свидетельствует о том, что в природной популяции полиморфизм выявляется в сайтах, существенных для регуляции общего, нетканеспецифического уровня транскрипции. Впервые показано, что природный полиморфизм регуляторной области гена может приводить к шестикратному изменению транскрипции гена.

Выявлен гаплотип, включающий два полиморфных маркера, расположенных в предсказанных на основе биоинформатического анализа сайтах связывания регуляторных белков, достоверно ассоциированный с количеством транскрипта *Lim3A* и ПЖ дрософилы. Показано, что составляющая гаплотип комбинация маркеров, ассоциированная с промежуточным уровнем транскрипции и высокой ПЖ, встречается в популяции с высокой частотой, в то время как комбинации, ассоциированные с низким или высоким уровнем транскрипции и низкой ПЖ, встречаются в популяции с низкой частотой, что указывает на функциональное и селективное преимущество среднего уровня экспрессии гена *Lim3*.

Таким образом, впервые показана функциональная связь между природным полиморфизмом гена *Lim3*, определяющего развитие нервной системы, и ПЖ дрософилы.

Научно-практическая значимость

Представленные в данной работе результаты расширяют представления о связи структурной изменчивости, в том числе природного полиморфизма регуляторной области гена с его функциональными возможностями: уровнем экспрессии и характером влияния на фенотип организма. Эти результаты должны быть учтены при создании генетических конструкций, используемых в различных биотехнологических проектах, с целью обеспечения требуемого уровня транскрипции используемых генов. Кроме того, полученные результаты могут быть использованы при чтении лекций и проведении практических занятий по общей генетике, генетике популяций, генетике количественных признаков, геронтологии в высших учебных заведениях биологического, медицинского и сельскохозяйственного профиля.

Апробация работы

Результаты, полученные в данной работе, были представлены на следующих научных семинарах и конференциях: на 41-ой конференции по популяционной генетике (Великобритания, 2007) на 40-ой американской ежегодной конференции по генетике дрософилы (Сан-Диего, США, 2008); на 4-ом съезде российского общества биохимиков и

молекулярных биологов (Новосибирск, Россия, 2008); 20-ом международном генетическом конгрессе (Берлин, Германия, 2008); на международной научной конференции по биоорганической химии, биотехнологии и бионанотехнологии, посвященной 75-летию со дня рождения академика Ю.А. Овчинникова (Москва, 2009), 5-ом съезде Вавиловского Общества Генетиков и Селекционеров (Москва, 2009), на 35-ом FEBS/YSF симпозиуме «Molecules of life» (Швеция, 2010).

Декларация личного участия автора

Вся экспериментальная часть выполнена автором самостоятельно. Анализ секвенированных образцов был проведен на автоматическом секвенаторе ABI Prism 310 на базе «Центра определения последовательности ДНК» в Учреждении Российской академии наук Института общей генетики им. Н.И. Вавилова РАН. Анализ данных выполнен совместно с Е.Г. Пасюковой.

Публикации

По материалам диссертации опубликовано 2 статьи в международных реферируемых журналах.

Структура и объём диссертации

Диссертационная работа изложена на 119 страницах машинописного текста и содержит главы: общая характеристика работы, обзор литературы, материалы и методы, результаты, обсуждение результатов, выводы, список литературы, приложение. Диссертация содержит 11 рисунков, 3 таблицы, 2 приложения и 206 литературных ссылок.

СОДЕРЖАНИЕ РАБОТЫ

Принципиальная особенность и новизна нашего подхода к изучению роли гена *Lim3* в контроле ПЖ заключалась в попытке связать выявленную в природной популяции структурную вариабельность гена не только с изменчивостью исследуемого признака (ПЖ), но и с особенностями экспрессии гена. В связи с этим для исследования полиморфизма необходимо было выбрать район генома, включающий не столько структурную часть гена, соответствующую его функционально важному транскрипту, сколько его 5' регуляторную область, строение которой определяет характер и уровень транскрипции. В связи с этим необходимо было, прежде всего, выбрать наиболее функционально важный транскрипт гена *Lim3* и определить, охарактеризовать его 5' регуляторную область.

Особенности организации проксимального промотора и инициации транскрипции *Lim3A* гена *Lim3*

Ген *Lim3* (28875 п.н.), образует две мРНК: *Lim3A* (2577п.н.) и *Lim3B* (2696п.н.) (Gen Bank accession nos. NM_165277 и NM_057258). *Lim3A* и *Lim3B* имеют одинаковые последовательности, за исключением первого экзона *Lim3A* и первого и второго экзонов *Lim3B* (Рис. 1).

Именно работы, сделанные на трансгенных конструкциях с *Lim3A*, показали участие гена *Lim3* в развитие мотонейронов (Thor et al., 1999). В связи с этим, данный транскрипт был выбран для дальнейшего исследования. Механизм образования *Lim3A* и строение его регуляторной области были неизвестны. Таким образом, необходимо было охарактеризовать особенности инициации транскрипции и строения регуляторной области *Lim3A*.

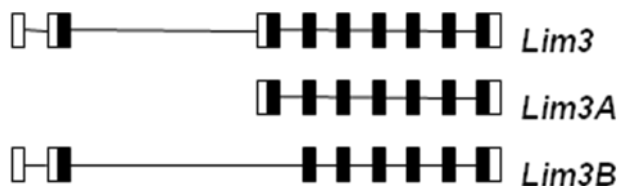


Рис. 1. Структура гена *Lim3 Drosophila melanogaster*.

Экзоны обозначены прямоугольниками; белые прямоугольники соответствуют нетранслируемой области, черные – транслируемой, интроны обозначены черными линиями.

Проведенный с помощью Neural Network Promoter Prediction, NNPP, version 2.2 биоинформатический анализ показал наличие двух типичных для нейрональных генов промоторных элементов с коэффициентом достоверности 0,98 и 1, расположенных от -540 до -491 п.н. и от -248 до -199 п.н. выше заявленного в базе данных старта транскрипции мРНК *Lim3A* (Рис. 2). Мы предположили, что найденные промоторы могут свидетельствовать о наличии старта транскрипции в непосредственной близости от них и, соответственно, могут участвовать в синтезе мРНК, отличных по длине от *Lim3A*. Для проверки этого предположения был использован Northern-блот анализ.

Действительно, Northern-блот анализ с зондом, комплиментарным части первого экзона *Lim3A* и 5' регуляторной области, непосредственно примыкающей к нему, показал наличие двух транскриптов вместо одного (Рис. 3А). Мажорный транскрипт по размеру (2,6 т.п.н.) соответствовал транскрипту *Lim3A* базы данных. Минорный транскрипт, имевший, вопреки ожиданиям, не больший, а меньший размер (2,4 т.п.н.) был новым, прежде не аннотированным, и был назван *Lim3C*.

Для того, чтобы выявить старт транскрипции *Lim3A* и *Lim3C*, а также подтвердить наличие второй неаннотированной мРНК мы провели 5'RACE анализ (Рис. 3Б).

Определение последовательности нуклеотидов 47 полученных клонов показало, что *Lim3A* и обнаруженный нами новый транскрипт *Lim3C* имеют множественные старты транскрипции (GenBank accession no. GU814523-GU814569).

Так, для *Lim3A* точка инициации транскрипции варьировала и была обнаружена в положениях -16 (3 клон), -2 (8 клонов) и +14 (4 клон) относительно старта, заявленного в базе данных, с преимущественным началом транскрипции в положении -6 (18 клонов) относительно ранее аннотированного (Рис. 2).

TATTTAAAAAAAACCTGTTATCGTTCAGTGTATAATGCCATTGTAATAAAAATCTGAGAAGCCTAGAGAAGCTGCAGAAATAAC
AAAGATACACATAT**TCCAACAGACTGTCAAGTCAAATTAC**AAAGAAGAGCCTTTATAAACTCTCCAAAGTCCGAATGCCAAATG
CGA<GTGTGT**T7A**AGGTTTCCTCGTTTGCATGCGTGTGTGTGTGGTAC**G46A,C**CACATTATTTGGCCTTTGGAAAA**A73G**CCGAAAC
C89AGAGAGCGCCGACCCTTGGCAGCTCATAGCCATAAGTCTCGTCTCGTCTCGCCTTACCAAC**G151G G156A G164A**
CAAGAGGG**G**GAGAGCGCCGACCCTTGGCAGCTCATAGCCATAAGTCTCGTCTCGTCTCGCCTTACCAAC**G**GACT**G**GCAAACT**G**
ACTGACGGACGGT**C**GAGAGCGCTCGAGAGTGTCTTAATGCAGTGAAGAAAATGTCCGGACCGACTTGGT**G**GACACCACATGTTGTTC**A**
GTTAGGCATTTATATTT**TCAATTAGCATGATCCAAG**AAAATTTATTAAATTTTATTATTGATGT**G**GATGGAGACAGCGTTTTT
G334A **A346T** **G358A** **G364A** **C375T T378C A382C T385C C386A**
GAAATTTGAATA**A**ACCTATAGCTT**A**AGTTC**G**TTAGAATTTA**G**GGTTT**C**CGT**A**TAAAAATCAAAGATCTCTAAGCATTACACT
A433T **Ubx/Exd/Hth** **T449G,C** **C474A** **491TAAACTTCTCA**
GAAAGAAAAGAGTAAT**A**GTCTGGGAATAGAA**T**TAATTGTGATATTTCCAATATGAC**C**ATTTCGATACTCCTGCA**A**ACAAATTTA

TTTTGAATTGTAGAATAATAGAATTTTGAATAATAACCATAAAATCCGGAGCATTGATTGAATTTTCTCGTGTGGGTTTTTTGGC
G586A **Kr** **G598T** **Kr** **G631T** **T643G**
TCAAAGGGTTAG**G**GGTTAAGGGCTAAGGGGGTTCGTTGGGGTTAG**G**GGTCAAGGGTA**T**GGGTAGGGGTTTTCTGGGTGGCAGG
TGCGC**AGACGTTGCTCTCGATCAGC**CGGTCCGGTCGGAGTCAGTCAGTCAGTATTTAACGGCTTTAAGGTCATTTTTCACAGTTA
CGAAACAATGACACAACAATAAGCAATGGCAGCCAAGTGCCTCAAACACATACACATTTGCACAAGCA**G**GCACACACACGAAAC
Hairy **CA845-** **G854A** **Kr** **G871C** **GRH**
GCACGGGCA**C**AGGATATGAC**G**GGGTAGAAGAAGAGT**C**ACCAAACTGCAGTGCAGAACAAATGCAAAACATATTTATGTGATTTT
A926G **GRH** **GAGA**
TAGTGCAG**A**CCAGTGTGTTTTCCAACCGAATTTGCCAACATGCGAAAGCGAGAGAGAGACCACGAAAGTGGAAATAGAGGTGGG
GAF/Psq **G1021A** **Sp1 -540** **C1046A** **A1050G**
CCGAGTGCAGAAAAGAGAG**G**CGGGTGTGGCAAAAAGCCGAG**G**TG**A**GCAGAGCTTTTCAAG**AGACCCTCCCCGAAAAC**TG
CAGGGCAGCTCGACGCACAAAAACAACCAATCCAGGCACAAACACAAAAAAGCCAAAAAGAAAACCCCCCTGTAAACACACA
C1177T -1180CA C1183T
CA/GT repeat binding protein
CACACACA**C**ACA**A**TA**C**GTACGCCAGACGCACACGCATGCCAGTGGAAAATGGAATGTGGCAAAGTGTACAATGCAATTTCCCC
-248 HSF A1320- HSF HB
GCTTTTTGTTTTGGCGACCAACCCAACTCAATTGTTGAGGCCAAAGGAAAAAATACAAAAAGAAA**A**CAGCGAAAAAAGGC

-199
CAGTCTGAAAACTGGCAGCCAGAGCAACAACAGCAGCACACATTTGGCACAGGGTGCAGCTGAATCGGCTTTCCAATTG
GAAGCGTAAGGGTTGGGTTCCCATTCAGCTGTGATT**ATGTCAGCCCTGTATGTG**CGCGCGTGTGTGCGTGTGCGGGGCT
HSF -27 -22 **+1** **+28 +33**
GCATTCAGTCCACCCCTTCTCCAATAAGCAACCGAATC**G**TGGCCCG**A**CT**G**TCAGTTGCTTTCCA**AGCCTCGAAGTTGC**
HB **T1658C HSF**
CGCTCGGTTTTTTGGCCGGTCACTTCGTGCAGCTAAAAGTCTCTGTGATAACGGTAGACAACAAGCCAAG**T**CGAAAAATACAAAT
BR-C Z4 BR-C Z3 HSF **HB** **-24 -19** **+1**
AAAAAGTAACTAAATCCGAAATTCGTTAGCGCAAAAAAAGTTTTTCCAAATCA**AG**TAAATGT**G**AT**TCAGC**CAACGGA

HSF
AACCCAAACAATGTGAAAC**GTGAAAGTGATGGTTGATTGCTCTGC**ATAGGAAAAGAA**TCCAATCAGAGT**CGCTCGGGATTTC
CACCGCCAAACAAGT**ATGGA**ACTTTTGAAGCTAATGATGTTCAAAGT**GACTTCCTGTC**CAATGGCAAAT**GTGACGATCGCGT**
ACCGCCGATTAACCTGAGTCAATTGCCAGATTTTGGCTATGTAAAGTAGCAAAAGGAGTTACCACCT**G**ATATGTTATACCTA**G**T
G2038A T2044G **C2058G** **G1991A** **G2006A**
GGGTATAGTGTGTTATGGATAATGATGATTAATAAAAAACATATCAATTAGAAGAGTAATCATATAATCTAT**GTATTTAGCG**
T>GTATAGTGTGTTATGGATAATGATGATTAATAAAAAACATATCAATTAGAAGAGTAATCATATAATCTAT**GTATTTAGCG**
TTATTTCTTTCTGCAATGTCAAAAAAAGAAGGACAAATCATCAATTCGCGGTATAAATATCGGTTTTTAATCACAATATGAG
TTATTTCTCAATTTAAATAATGAGTTTATAACAAATGTCTGTTTAAACAGATTCTTAATTTGCTT

Рис. 2. Структура 5' регуляторной области *Lim3A*, *Lim3C*.

Синими буквами обозначена некодирующая область первого экзона *Lim3A*, *Lim3C*; красными буквами - кодирующая область первого экзона *Lim3A*, *Lim3C*; фиолетовыми буквами обозначена последовательность праймеров для ПЦР анализа и секвенирования; зелеными буквами-праймеры для количественного ПЦР анализа в реальном времени. </> - границы анализируемой последовательности. Промоторные последовательности, предсказанные Neural Network Promoter Prediction, NNPP, version 2.2, показаны светло-серым цветом; числа в скобках обозначают положение нуклеотида промоторного элемента относительно раннее предсказанного старта транскрипции. Последовательности, выделенные зеленым прямоугольником обозначают инициаторы. Старты транскрипции отмечены звездочками и выделены жирным шрифтом. Подчеркнутая последовательность обозначает предсказанные элементы корового промотора. Буквы, выделенные квадратом – полиморфные сайты, встречающиеся в исследуемом ДНК фрагменте с частотой 0.06. Инсерции обозначены черным прямоугольником. Заглавные буквы, выделенные квадратом и жирным шрифтом, обозначают полиморфные сайты, ассоциированные с ПЖ. Заглавные буквы, выделенные розовым квадратом и жирным шрифтом, обозначают полиморфные маркеры, образующие гаплотип, достоверно ассоциированный с ПЖ. Заглавные буквы, выделенные голубым квадратом и жирным шрифтом, обозначают полиморфные сайты, ассоциированные с транскрипцией *Lim3A* и ПЖ. Полиморфные сайты, ассоциированные с транскрипцией *Lim3A*, отмечены буквами, выделенными голубым цветом. Обозначения полиморфных маркеров расположены сверху над ними. Сайты связывания транскрипционных факторов обозначены пунктирной линией.

Для мРНК *Lim3C* также наблюдался разброс значений. Минорные старты транскрипции были обнаружены в положениях +169 (3 клон), +179 (3 клон), относительно ранее заявленного старта, а мажорный старт – в положении +184 (8 клонов) (Рис. 2).

В определении сайта инициации транскрипции *Lim3A*, *Lim3C* учитывались лишь те клоны, старты которых встречались три и более раз, чтобы избежать ошибок, связанных с появлением в анализе усеченных, деградирующих при выделении фрагментов полноразмерных мРНК молекул.

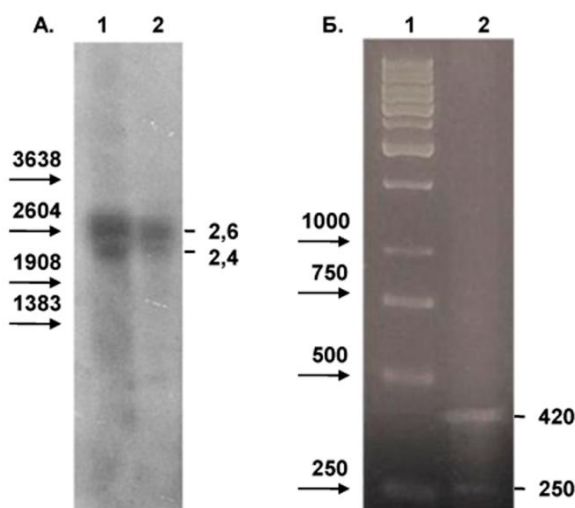


Рис. 3(А). Northern-блот анализ *Lim3A*. Размеры стандартных РНК маркеров (п.н., Promega) показаны стрелками слева, размеры найденных транскриптов (т.п.н.), отмечены справа. **(Б). 5'RACE анализ *Lim3A*.** Дорожка 1- стандартный 1т.п.н. ДНК маркер (Fermentas), дорожка 2- продукты 5'RACE анализа, размеры (п.н.) которых отмечены справа.

Таким образом, было обнаружено, что *Lim3C* короче *Lim3A* на 190 п.н. за счет некодирующей области первого экзона *Lim3C*, *Lim3A* и имеет структуру, идентичную *Lim3A* (все семь экзонов присутствовали в обоих транскриптах).

Далее с помощью различных биоинформатических ресурсов был проведен анализ коровой и проксимальной промоторных областей обоих транскриптов, который позволил нам выявить в них различные регуляторные элементы.

Было обнаружено, что инициатор транскрипта *Lim3A* (Т-Т-А₊₁Г-Т-С) соответствовал каноническому консенсусу старта транскрипции дрозофилы (Т-С-А₊₁-Г/Т-Т-Т/С) (Kutach, Kadonaga, 2000; Butler, Kadonaga, 2002). Такая же последовательность инициатора с отличием в один или несколько нуклеотидов была обнаружена у 13,2% генов, представленных в базе данных DCPD (*Drosophila* Core Promoter Database) (Табл. 1). Многие из этих генов, около 33% содержали только DPE (Downstream core promoter element). Их функция соответствовала функции гена *Lim3* и заключалась в кодировании транскрипционных факторов РНК-полимеразы II, участвующих в регуляции различных клеточных процессов.

Используя базы данных Eukaryotic Promoter Database, Current Release 100 (EPD) и DCPD, мы обнаружили последовательность DPE в положении +28 - (+33) относительно мажорного

старта транскрипции *Lim3A*. Эта последовательность (AGTTGC) полностью соответствовала консенсусу DPE (A/G/T₊₂₈-C/G-A/T-C/T-A/C/G-C/T) (Kutach, Kadonaga, 2000; Butler, Kadonaga, 2002) и встречалась в 0.8% из 1926 генов базы данных EPD.

Ген	Последовательность инициатора	Промоторный элемент
	* +1 *	
<i>Lim3A</i>	TCGTGGCCCGTT AGTC GTCAGTTGCTTTCCAAGC	
<i>CecA1</i>	TCAT TCAGTC	ТАТА-бокс и DPE
<i>CecA2</i>	TCAAC AGTC	ТАТА-бокс и DPE
<i>dhod</i>	TAGTC	DPE
<i>dpt</i>	TCAGTC	ТАТА-бокс
<i>E74B</i>	TCAGTC	Нет
<i>E74plb</i>	TCAGTC	DPE
<i>E74B p2</i>	TTAGTT	DPE
<i>G42</i>	GC AGTC	DPE
<i>G</i>	AC AGTC	DPE
<i>Gapdh1</i>	GC AGTC	Нет
<i>glass</i>	ACAAT TAGTC	DPE
<i>Gld</i>	GAAGCT GAGTC	ТАТА-бокс и DPE
<i>Gpdh</i>	GT AGTC	Нет
<i>hb (zyg)</i>	TCAG TCAGTC	ТАТА-бокс и DPE
<i>H2AvD</i>	GT AGTC	Нет
<i>hsp27</i>	AC AGTC	ТАТА-бокс
<i>hsp67bhs</i>	CA AGTC	ТАТА-бокс и DPE
<i>hsp83</i>	CG AGTC	ТАТА-бокс
<i>Labial</i>	GTAAT TCAGTC	DPE
<i>lsp1 alpha</i>	AC AGTC	ТАТА-бокс и DPE
<i>lsp1 beta</i>	AC AGTC	ТАТА-бокс и DPE
<i>Mst26 Ab</i>	TTAC AGTC	DPE
<i>primase</i>	TTAGTT	ТАТА-бокс
<i>sca</i>	TCAGTC	ТАТА-бокс и DPE
<i>Scr</i>	CGAGC AGTC	DPE
<i>sryAC</i>	TTAGTT	ТАТА-бокс
<i>yellow</i>	CC AGTC	ТАТА-бокс

Табл. 1. Структура и сходство компонентов 5' регуляторной области *Lim3A* и генов базы данных DCPD. Основной старт транскрипции обозначен (+1), минорные отмечены звездочками и выделены жирным шрифтом.

Регуляторных элементов, соответствующих ТАТА-боксу не было выявлено, но в положении (-27) - (-22) от старта *Lim3A* была обнаружена последовательность (СААТАА), часто встречающаяся в позиции (-36) - (-21) в регуляторных районах генов дрозофилы. Из 1926 генов базы EPD 0,6% содержали данную последовательность в своих промоторах. Например, СААТАА была обнаружена в регуляторных областях пяти *Enhancer of split [E(spl)]* генов (*HLHm3*, *HLHm5*, *HLHm8*, *HLHmβ*, *HLHmγ*) (Maeder et al., 2007), которые кодируют bHLH транскрипционные репрессоры. Экспрессия этих генов, как и гена *Lim3*, в основном приходится на эмбриональную стадию, а функция связана с развитием и функционированием нейронов.

В отличие от *Lim3A* транскрипт *Lim3C* имел менее вырожденный инициатор (TTG₊₁AGC). Для него не было обнаружено никаких известных промоторных элементов. Однако в положении (-24)-(-19) относительно мажорного старта транскрипции *Lim3C* была найдена последовательность (СААААТ), часто встречающаяся в регуляторных областях других генов. Такая последовательность была обнаружена в 0,6% из 1926 генов базы EPD в положении (-36)-(-18).

Кроме найденных двух мажорных и нескольких минорных стартов транскрипции *Lim3A* и *Lim3C* 5'RACE анализ показал наличие нескольких стартов, каждый из которых встречался не более чем в одном клоне, расположенных на расстояние около 250 п.н. выше мажорного старта *Lim3A*. В образовании этих одиночных транскриптов могли участвовать промоторы, предсказанные базой данных Neural Network Promoter Prediction или так называемые «скользящие промоторы», которые являются типичными и для ТАТА-содержащих генов, и для генов, не содержащих ТАТА-боксов, и при этом имеющих множественные старты транскрипции (Yasuhara et al., 2005).

Далее, с помощью программ TFSEARCH version 1.3, MOTIF и других биоинформатических ресурсов в исследуемой области генома были найдены потенциальные сайты связывания ряда транскрипционных факторов с процентом гомологии 90.2 - 100.0. К их числу относятся Heat Shock Factor (HSF), который может участвовать в контроле экспрессии не только генов, кодирующих белки теплового шока, но и, например, *even skipped*, гена, участвующего в развитии эмбрионов дрозофилы (Kuchar et al., 2007); Hunchback (HB), необходимый для ранней спецификации линий нейробластов (Isshiki et al., 2001), Broad-Complex Z3 и Broad-Complex Z4 (BR-C Z3/Z4), необходимые для метаморфной реорганизации ЦНС (центральная нервная система) дрозофилы (Spokony, Restifo, 2009) (Рис. 2). HB и BR-C Z3/Z4 являются специализированными транскрипционными факторами, участвующими в морфогенезе нервной системы дрозофилы и, возможно, играющими роль в регуляции транскрипции *Lim3*, который также необходим для развития и функционирования нейронов.

Таким образом, было показано, что структура коровой и проксимальной промоторных областей *Lim3A*, *Lim3C* типичны для транскрипционных факторов РНК полимеразы II, принимающих участие в развитии и функционировании нервной системы дрозофилы.

Исследование природного молекулярного полиморфизма в *Lim3* локусе

Для того чтобы определить, влияет ли природный молекулярный полиморфизм гена *Lim3* на его функцию, было проведено частичное определение последовательности нуклеотидов 50 аллелей гена, полученных из природной популяции Raleigh. Для этого было использовано 50 линий дрозофилы, каждая из которых содержала в гомозиготном состоянии вторую хромосому из популяции Raleigh в окружении хромосом изогенной лабораторной линии Samarkand дикого типа (De Luca et al., 2003). Для каждого из 50 аллелей *Lim3* была определена последовательность фрагмента ДНК длиной 2094 п.н., включающая 1557 п.н. регуляторной области *Lim3A*, 300 п.н. 5' некодирующей области (5' НТО), 109 п.н. кодирующей области первого экзона *Lim3A* и 128 п.н. следующего за ним первого интрона (GenBank accession nos 9GU814570, 33GU814571, 40GU814572, 44GU814573, 49GU814574, 58GU814575, 74GU814576, 76GU814577, 77GU814578, 87GU814579, 89GU814580, 98-21GU814581, 98-5GU814582, 100GU814583, 113GU814584, 115GU814585, 122GU814586, 161GU814587, 166GU814588,

180GU814589, 183GU814590, 200GU814591, 201GU814592, 207GU814593, 215GU814594, 226GU814595, 266GU814596, 273GU814597, 284GU814598, 285GU814599, 311GU814600, 316GU814601, 317GU814602, 325GU814603, 327GU814604, 336GU814605, 345GU814606, 351GU814607, 354GU814608, 361GU814609, 369GU814610, 376GU814611, 382GU814612, 407GU814613, 429GU814614, 434GU814615, 444GU814616, 461GU814617, 472GU814618, 473GU814619) (Табл. 2).

Всего было найдено 90 полиморфных маркеров, включающих 74 единичных нуклеотидных замены и 16 инсерций\делеций. Уровень полиморфизма оценивали с помощью следующих показателей: π – наблюдаемая средняя доля нуклеотидных различий между секвенированными аллелями (Nei, Tajima, 1981) и θ – среднее число сегрегирующих нуклеотидов на сайт (Watterson, 1975), используя программу DnaSP (Rozas, Rozas, 1999).

Район	Положение нуклеотида	Число инсерций\делеций + число единичных нуклеотидных замен	π (с.о.)	θ (с.о.)
Вся последовательность	1 – 2094	16+74	0.00709 (0.00046)	0.00864 (0.00098)
Регуляторная область	1 – 1557	14+58	0.00795 (0.00046)	0.00875 (0.00112)
	1 – 779	8+39	0.01168 (0.00069)	0.01118 (0.00179)
	780 – 1557	6+19	0.00421 (0.00043)	0.00631 (0.00135)
Экзон	1558 – 1966	1+7	0.00152 (0.00059)	0.00600 (0.00181)
5'НТО	1558 – 1857	1+5	0.00181 (0.00078)	0.00670 (0.00223)
Кодирующая область	1858 – 1966	0+2	0.00073 (0.00070)	0.00410 (0.00290)
Интрон	1967 – 2094	1+9	0.01443 (0.00131)	0.01570 (0.00523)

Табл. 2. Молекулярная изменчивость регуляторной области и части структурной области *Lim3A*. В скобках приведены значения стандартного отклонения.

Уровень изменчивости исследуемого района генома ($\pi = 0.00709 \pm 0.00046$ и $\theta = 0.00864 \pm 0.00098$) находился в пределах, характерных для соответствующих областей генома *Drosophila melanogaster* (Moriyama, Powell, 1996) (Табл. 2).

Наиболее вариабельным оказался интрон ($\pi = 0.01443 \pm 0.00131$ и $\theta = 0.01570 \pm 0.00523$). Тем не менее, если 5' регуляторную область разделить произвольно на две равные части, то дистальный район регуляторной области имел приблизительно тот же уровень вариабельности, что и интрон ($\pi = 0.01168 \pm 0.00069$ и $\theta = 0.01118 \pm 0.00179$) (Табл. 2).

Менее вариабельным были проксимальный район регуляторной области ($\pi = 0.00421 \pm 0.00043$ и $\theta = 0.00631 \pm 0.00135$) и 5' НТО первого экзона ($\pi = 0.00181 \pm 0.00078$ и $\theta = 0.00670 \pm 0.00223$). Самой консервативной оказалась кодирующая область первого экзона ($\pi = 0.00073 \pm 0.00070$ и $\theta = 0.00410 \pm 0.00290$), где было найдено только две несинонимичных замены, каждая с минимальной частотой 0.02 (Табл. 2).

При исследовании вариабельности данной области генома достоверно сцепленными между собой оказались немногие полиморфные маркеры, в основном расположенные на

небольшом расстоянии друг от друга (Рис. 4). Такой характер сцепления указывает на существование нормального уровня рекомбинации в исследуемом районе генома. Этот результат благоприятен для выявления причинно-следственных связей между молекулярным полиморфизмом и его фенотипическим проявлением.

Единичные нуклеотидные замены	T7A	A73G	C89A	C292T	G334A	G358A	G364A	C375T	T378C	A382C	T385C	C386A	A433T	C474A	G586A	G598T	C821A	G854A	T2014C	G2038A	T2044G	C2058G	
T7A																							
A73G																							
C89A																							
C292T																							
G334A																							
G358A																							
G364A																							
C375T																							
T378C																							
A382C																							
T385C																							
C386A																							
A433T																							
C474A																							
G586A																							
G598T																							
C821A																							
G854A																							
T2014C																							
G2038A																							
T2044G																							
C2058G																							

Рис. 4. Неравновесие по сцеплению между полиморфными сайтами в *Lim3* локусе.

Неравновесие по сцеплению представлено как матрица попарных сравнений, в которой расположение каждого полиморфного маркера представлено графически сверху вниз и слева направо. Полиморфные маркеры обозначены таким образом, что первым указан наиболее часто встречаемый нуклеотид, за ним следуют номер позиции в последовательности, где происходит замена, и далее наиболее редко встречаемый вариант нуклеотида. В таблицу включены только маркеры, имеющие, достоверное сцепление согласно двум подсчитанным с помощью программы DnaSP критериям, Фишера и χ^2 , с учетом коррекции Бонферрони (обозначено серым цветом).

Роль отбора в поддержании молекулярной вариабельности исследуемого района была оценена различными методами при помощи программы DnaSP. Было показано, что первый экзон *Lim3A* характеризуется достоверно отрицательными значениями D (Tajima, 1989), D* и F*(Fu, Li, 1993) и D* и F*, вычисленных с использованием внешней группы (Табл.3). Достоверные значения D, D* и F* были в большинстве случаев получены как для 5'НТО, так и для кодирующей области первого экзона *Lim3A*, проанализированных отдельно (Табл. 3). В связи с тем, что в кодирующей области было обнаружено только две несинонимичных замены, других тестов на селективную нейтральность, основанных на анализе соотношения

синонимичных и несинонимичных замен (McDonald, Kreitman, 1991; Stoletzki, Eyre-Walker, 2010), не применяли.

Район	D	D*	F*	D*, simulans	F*, simulans	D*, yakuba	F*, yakuba
Вся последовательность	-0.63216	-1.75232	-1.59703	-1.99481	-1.89078		
Регуляторная область	-0.31834	-1.03457	-0.92083	-1.35484	-1.28096		
Дистальный район	0.15386	-0.88538	-0.60895	-1.11283	-0.91096		
Проксимальный район	-1.07509	-1.02695	-1.23886	-1.29958	-1.51835	-1.55057	-1.67083
Экзон	-2.10512	<u>-3.73109</u>	<u>-3.79399</u>	<u>-3.40290</u>	<u>-3.56259</u>	<u>-2.78912</u>	<u>-3.07972</u>
5'НТО	-2.04396	<u>-3.26405</u>	<u>-3.37500</u>	<u>-2.78749</u>	<u>-3.02202</u>	-2.08900	-2.46833
Кодирующая область	-1.46443	-2.53305	-2.57464	-2.58770	-2.63091	-2.58770	-2.63091
Интрон	-0.22559	-1.28572	-1.10719	-0.44091	-0.35893	-1.99059	-1.81791

Табл. 3. Результаты тестов на селективную нейтральность для регуляторной области и части структурной области *Lim3A*. В качестве внешней группы использовали *D. simulans* и *D. yakuba* (соответствующие последовательности взяты из GenBank, accession nos. XM_002079849 и NT_167063) Достоверные значения D, D*, F*, выделены жирным шрифтом (P<0.05), выделены жирным шрифтом и подчеркнутые (P<0.02).

Таким образом, полученные результаты свидетельствуют о том, что первый экзон *Lim3A*, характеризующийся низким уровнем варибельности, подвергается действию стабилизирующего отбора. Поскольку уровень рекомбинации в исследуемом районе достаточно высок, можно предположить, что отбор действует именно на полиморфные сайты, расположенные в экзоне, а не какие-либо другие, сцепленные с ними.

Исследование ассоциации между природным молекулярным полиморфизмом в *Lim3* локусе и продолжительностью жизни

Для того чтобы оценить биологическую роль молекулярной варибельности, наблюдаемой в природной популяции Raleigh, мы проанализировали связь между обнаруженным нами полиморфизмом в *Lim3* локусе и изменчивостью ПЖ дрозофилы.

В данном анализе мы ограничились рассмотрением лишь тех полиморфных сайтов, для которых частота редкого варианта в исследованной нами выборке хромосом равнялась 0.06 и более (одинаковые замены нуклеотидов обнаружены в трех и более линиях). С одной стороны, такое ограничение давало уверенность в том, что мы не имеем дело со случайными ошибками, возникшими в результате реакций амплификации, использованных для получения фрагментов ДНК исследуемого гена, и их секвенирования. С другой стороны, это ограничение позволяло составить впечатление о том полиморфизме, который действительно распространен и сегрегирует в природной популяции.

Из 90 найденных нами полиморфных локусов принятому критерию соответствовало 44 (Рис. 2). Двухфакторный дисперсионный анализ позволил выявить достоверную ассоциацию между ПЖ и шестью маркерами, при этом не было выявлено достоверной ассоциации между ПЖ и полом особей, а также между ПЖ и взаимодействием двух главных факторов,

МАРКЕР*ПОЛ. На основании этого данные для самцов и самок были объединены, и для анализа ассоциации между полиморфными маркерами и ПЖ использован непараметрический анализ. В результате такого анализа достоверная ассоциация с ПЖ была выявлена для тех же шести полиморфных маркеров, четыре из них находились в регуляторной области (A433T; G871C; A1050G; C1177T), один в 5'НТО (T1658C), один в первом интроне (G1991A) (Рис. 2, табл. 4). Ни один из найденных полиморфных маркеров не был достоверно сцеплен с другими в популяции Raleigh.

В то же время, все особи внутри каждой из 50 исследованных линий имеют идентичную вторую хромосому и, соответственно, одинаковый набор маркеров, абсолютно сцепленных между собой и образующих определенные гаплотипы. Транскрипция гена и, как следствие, ПЖ могут зависеть от того, какая комбинация маркеров присутствует в регуляторной области транскрипта. Чтобы проверить это предположение и определить, существует ли достоверная ассоциация между ПЖ и различными гаплотипами, образованными достоверно ассоциированными с ПЖ полиморфными маркерами, расположенными в регуляторной области, был также использован непараметрический анализ. Гаплотипы, состоящие из четырех маркеров (A433T + G871C + A1050G + C1177T), из трех маркеров, расположенных в непосредственной близости от структурной части гена (G871C + A1050G + C1177T), и из двух полиморфных маркеров с наиболее значимой ассоциацией с ПЖ (G871C + C1177T) были достоверно ассоциированы с ПЖ (Табл. 4).

В целом в ходе анализа один и тот же пул данных был использован для проведения 47 тестов на ассоциацию с ПЖ. Суммируя полученные результаты, можно заключить, что после использования поправки Бонферрони, позволяющей скорректировать результаты с учетом множественности тестов, достоверная ассоциация была выявлена между ПЖ и гаплотипом (G871C + C1177T) ($P=0,0010$), а после использования менее консервативной поправки FDR (false discovery rate) – между ПЖ и двумя другими исследованными гаплотипами (A433T + G871C + A1050G + C1177T) ($P=0,0042$) и (G871C + A1050G + C1177T) ($P=0,0021$), и маркером G1991A ($P=0,0028$) (Табл. 4). Маркеры G871C и C1177T входят во все гаплотипы, причем ассоциация с ПЖ наиболее значима для гаплотипа, в котором присутствуют только эти два маркера. Можно предположить, что в рамках данного исследования именно комбинация двух маркеров, G871C и C1177T, в геноме особей из природной популяции Raleigh наиболее важна для ПЖ.

Один из аллелей каждого полиморфного маркера, образующего наиболее значимо ассоциированный с ПЖ гаплотип, имеет низкую частоту в популяции ($p_C = 0.08$ для C871G; $p_T = 0.06$ для C1177T). Из четырех возможных комбинаций аллелей только три (GC, CC, GT) были обнаружены в популяции.

Наблюдаемые частоты комбинаций аллелей хорошо согласуются с частотами, ожидаемыми на основе частот единичных аллелей в популяции ($\chi^2 = 0.0055$), что подтверждает отсутствие сцепления между маркерами в популяции.

Единичные нуклеотидные замены, инсерции/делеции, гаплотипы	Число линий с сегрегирующими аллелями, гаплотипами	Признак	Значение P ⁴	Среднее значение признака для сегрегирующих аллелей, гаплотипов/(с.о.)
A433T	38/12 ¹ ; 14/2 ²	ПЖ	0.0357	
		РНК, эмбрионы	0.0030/0.1384	0.7(0.06)/2.0(0.37)
		РНК, головы	0.0682/0.1543	
G586A, G598T	36/14 ¹ ; 12/4 ²	РНК, семенники	0.0325/0.3327	
		ПЖ	0.0992	
		РНК, эмбрионы	0.0131/0.0896	0.7(0.06)/1.4(0.29)
G871C	46/4 ¹ ; 13/3 ²	РНК, головы	0.1773/0.2230	
		РНК, семенники	0.0198/0.0896	
		ПЖ	0.0151	
A926G	47/3 ¹ ; 14/2 ²	РНК, эмбрионы	0.0021/0.0682	0.7(0.06)/2.0(0.37)
		РНК, головы	0.0052/0.0025	0.4(0.05)/0.8(0.07)
		РНК, семенники	0.0209/0.4588	
G1021A, C1046A	40/10 ¹ ; 10/6 ² , 42/8 ¹ ; 10/6 ²	ПЖ	0.7528, 0.6801	
		РНК, эмбрионы	0.0018/0.0195	1.1(0.20)/0.5(0.09)
		РНК, головы	0.0356/0.0091	
A1050G	26/24 ¹ ; 11/5 ²	РНК, семенники	0.0091/0.0014	
		ПЖ	0.0226	
		РНК, эмбрионы	0.5153/0.8548	
C1177T	47/3 ¹ ; 14/2 ²	РНК, головы	0.9029/0.2550	
		РНК, семенники	0.9190/0.1039	
		ПЖ	0.0084	
G1991A	46/4 ¹ ; 13/3 ²	РНК, эмбрионы	0.0033/0.0037	0.9(0.11)/0.3(0.04)
		РНК, головы	0.0021/0.0021	0.5(0.05)/0.1(0.02)
		РНК, семенники	0.0121/0.0044	
T1658C	45/5 ¹ ; 13/3 ²	ПЖ	0.0028	37(1)/29(2)
		РНК, эмбрионы	0.2992/0.2005	
		РНК, головы	0.2183/0.2274	
871 + 1177, CC/GC/GT	4/43/3 ¹ ; 3/11/2 ²	РНК, семенники	0.4115/0.2333	
		ПЖ	0.0195	
		РНК, эмбрионы	0.1842/0.2881	
		РНК, головы	0.1111/0.1475	
		РНК, семенники	0.1407/0.0535	
		ПЖ	0.0010	31(2)/38(1)/29(2)
		РНК, эмбрионы	0.0001/0.0014	1.8(0.27)/0.7(0.06)/0.3(0.04)
		РНК, головы	0.0011/0.0016	0.7(0.12)/0.4(0.04)/0.1(0.02)
		РНК, семенники	0.0053/0.0125	

Табл. 4. Генотип-фенотип ассоциация в *Lim3* локусе.

¹ Выборка из 50 линий.

² Выборка из 16 линий.

³ Этот маркер не сегрегирует в выборке из 16 линий.

⁴ Значения P приведены для непараметрического анализа средних значений ПЖ и непараметрического анализа количества транскрипта, охарактеризованного в нг/С(т). Достоверные значения показаны курсивом; значение P, оставшиеся достоверными после поправки FDR, обозначены курсивом и жирным шрифтом; значения P, оставшиеся достоверными после поправки Бонферрони, показаны курсивом, жирным шрифтом и подчеркнуты. Маркеры и гаплотипы, достоверно ассоциированные с ПЖ, выделены жирным курсивом. В скобках приведены значения стандартной ошибки.

Множественное сравнение средних позволило разделить линии с различными вариантами гаплотипа, состоящего из маркеров G871C и C1177T (GC, CC, GT), на две группы, достоверно различающиеся по ПЖ ($P < 0.05$). Первая группа, содержащая линии с наиболее часто встречающимся GC (86%) вариантом гаплотипа, имела среднюю ПЖ 38 (± 1) день. Вторая группа включала линии с редкими CC (8%) и GT (6%) вариантами гаплотипа, характеризующимися средней ПЖ 31 (± 2) день и 29 (± 2) дня.

Таким образом, комбинация двух маркеров (G871C и C1177T), расположенных в 5' регуляторной области *Lim3A*, наиболее важна для ПЖ; комбинация маркеров, ассоциированная с высокой ПЖ, широко представлена в популяции Raleigh, в то время как комбинация маркеров, ассоциированная с низкой ПЖ, является редкой.

Исследование ассоциации между молекулярным полиморфизмом в *Lim3* локусе и его экспрессией

Изменения в регуляторной области могут приводить к изменению уровня экспрессии гена, что, в свою очередь, может привести к изменению фенотипа особей, например, ПЖ мух. Для того, чтобы проверить это предположение, мы оценили связь между обнаруженным нами полиморфизмом в локусе *Lim3* и изменениями количества мРНК *Lim3A*.

Транскрипты *Lim3A*, *Lim3C* различаются только 5' НТО, которая в *Lim3C* на 190 п.н. короче, поэтому измерить можно или количество одного *Lim3A* или количество обоих транскриптов вместе. Было показано, что мажорный транскрипт *Lim3A* (Рис.3А) необходим для развития нейронов (Thor et al., 1999), поэтому именно *Lim3A* был использован для измерения экспрессии гена *Lim3* методом ПЦР в реальном времени в 16 линиях мух с различными вариантами найденного гаплотипа (G871C + C1177T). Так как основная экспрессия гена *Lim3* приходится на эмбриональную стадию, а в имаго она сосредоточена в основном в мозге и семенниках [FlyAtlas accession no. FBgn0002023], то для измерения транскрипции мы использовали эмбрионы, головы и семенники (Табл. 4).

В выборке из 16 линий корреляция между повторными независимыми измерениями количества *Lim3A* как в эмбрионах, так и в головах была высоко достоверна ($P < 0.0001$, $P = 0.0074$, соответственно), что подтверждало надежность результатов. Корреляции между независимыми измерениями в семенниках была недостоверна ($P = 0.1082$), возможно из-за низкого уровня транскрипции *Lim3A* в этой ткани. Корреляция наблюдалась также между количеством мРНК *Lim3A* в эмбрионах и головах ($P = 0.0064$), в эмбрионах и семенниках ($P = 0.0579$), в головах и семенниках ($P = 0.0204$).

Из 44 полиморфных сайтов 30 сегрегировали в выборке из 16 линий мух, 8 из них были сцеплены с другими маркерами. В результате проведенных 24 тестов с 22 маркерами и двумя гаплотипами (непараметрический анализ) была выявлена достоверная ассоциация между количеством мРНК *Lim3A* в эмбрионах и 14 полиморфными сайтами. Для четырех из них

(G871C, A926G, G1021A, C1046A) достоверность ассоциации сохранилась после применения поправки Бонферрони, а для четырех других маркеров (A433T, G586A, G598T, C1177T) – после применения поправки FDR (Табл. 4). Полиморфные сайты G1021A и C1046A, G586A и G598T были полностью сцеплены в выборке из 16 линий мух. Другой вариант непараметрического анализа (Yuan et al., 2006), основанный на использовании получаемых при измерении значений $C(t)$, которые пропорциональны логарифму количества субстрата, был применен для проверки полученных результатов. Достоверная (после применения поправки FDR) ассоциация была подтверждена для маркеров G871C и C1177T (Табл. 4). С помощью программы REST (Pfaffl et al., 2002), позволяющей при сравнении количества транскрипта в разных образцах учитывать различную эффективность реакции ПЦР в реальном времени для кДНК анализируемого и референсных генов, достоверная ассоциация с количеством мРНК *Lim3A* была подтверждена для маркеров G871C и C1177T ($P=0.0001$ для обоих).

Маркеры G871C и C1177T, образующие гаплотип, достоверно связанный с ПЖ мух, оказались также достоверно ассоциированными с количеством мРНК *Lim3A* в эмбрионах в результате анализа всеми перечисленными выше методами (Табл. 4), включая попарное сравнение, используя программу REST ($P=0.0001$ для каждого сравнения). Множественное сравнение средних позволило разделить линии с различными вариантами гаплотипа (G871C + C1177T), на три группы, достоверно ($P<0.05$) различающиеся по количеству транскрипта *Lim3A* в эмбрионах (CC: 1.8 ± 0.27 ; GC: 0.7 ± 0.06 ; GT: 0.3 ± 0.04). Различия между группами с максимальным и минимальным уровнем транскрипции достигают шести раз (Рис.5).

Стандартный непараметрический анализ позволил выявить достоверные ассоциации с количеством транскрипта *Lim3A* в головах взрослых мух для 5 полиморфных маркеров. Ассоциация с одним из них (C1177T) осталась достоверной после применения поправки Бонферрони, а для двух других (G871C, A926G) – после применения поправки FDR (Табл. 4). Непараметрический анализ, основанный непосредственно на величинах $C(t)$, выявил ассоциации, достоверные с учетом применения поправки FDR, для маркеров A926G и C1177T (Табл. 4). Однако анализ с помощью программы REST не подтвердил эти результаты. Достоверная ассоциация между гаплотипом (G871C + C1177T) и количеством мРНК *Lim3A* в головах была выявлена обоими методами непараметрического анализа ($P=0,0011$; $P=0,0016$; (Табл. 4). Программа REST обнаружила достоверное различие в количестве транскрипта *Lim3A* только между вариантами гаплотипа GT и CC ($P=0.023$). Множественное сравнение средних позволило разделить линии с различными вариантами гаплотипа (G871C + C1177T) на три группы, достоверно ($P<0.05$) различающиеся по уровню экспрессии *Lim3A* в головах взрослых мух (CC: 0.7 ± 0.12 ; GC: 0.4 ± 0.04 ; GT: 0.1 ± 0.02). Группы с максимальным уровнем транскрипции отличались от групп с минимальным уровнем транскрипции в шесть раз (Рис.5).

Стандартный непараметрический анализ позволил выявить достоверные ассоциации между количеством транскрипта *Lim3A* в семенниках и 16 полиморфными маркерами. Однако ни одна ассоциация не осталась достоверной после применения поправок Бонферрони или FDR (Табл. 4). Достоверная ассоциация между гаплотипом (G871C + C1177T) и количеством *Lim3A* мРНК в семенниках была выявлена обоими методами непараметрического анализа, однако ни одна из них не была подтверждена после использования поправок на множественное тестирование (Табл. 4).

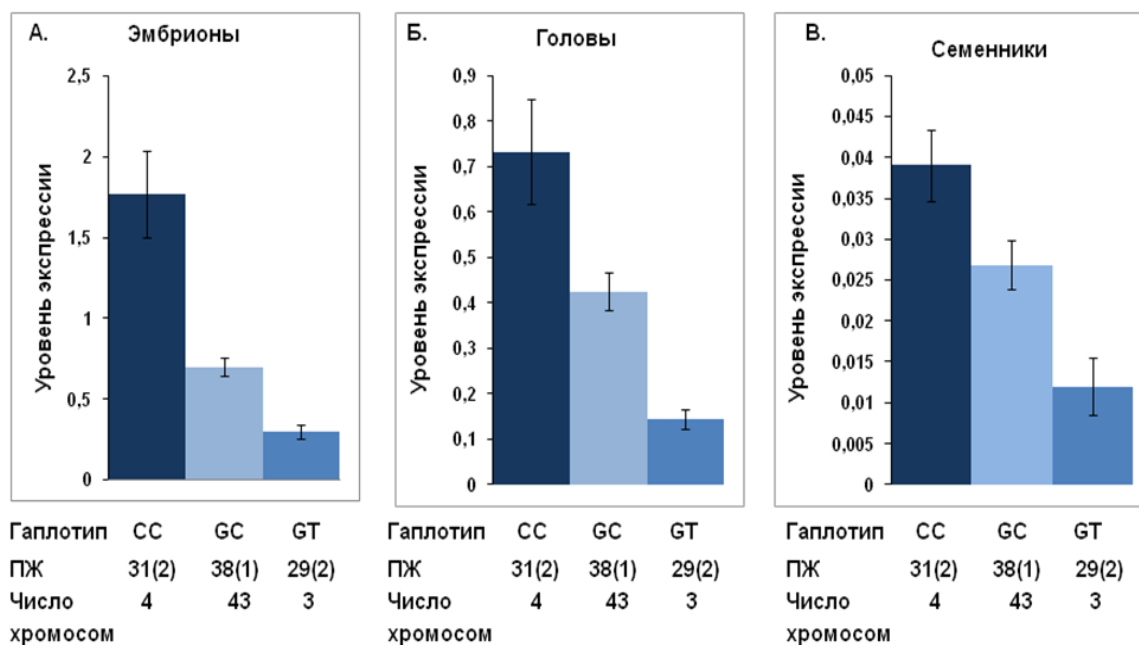


Рис. 5. Уровень мРНК *Lim3A* в линиях с различными вариантами гаплотипа. (А). Транскрипция *Lim3A* в эмбрионах. **(В).** Транскрипция *Lim3A* в головах взрослых мух. **(С).** Транскрипция *Lim3A* в семенниках. Для каждого варианта гаплотипа, состоящего из маркеров G871C, C1177T, обозначена средняя ПЖ в днях со стандартной ошибкой, указанной в скобках. Число хромосом с соответствующей комбинацией маркеров в популяции Raleigh обозначено внизу диаграммы.

Множественное сравнение средних показало, что транскрипция *Lim3A* в семенниках достоверно различается ($P < 0.05$) между линиями с CC (0.04 ± 0.004) и GT (0.01 ± 0.003) вариантами гаплотипа (Рис. 5).

Многие полиморфные сайты оказались достоверно ассоциированными с транскрипцией *Lim3A* в различных тканях. Разные методы анализа и поправки давали немного различные, но все же не противоречащие друг другу результаты (Табл. 4). Среди полиморфных маркеров самыми примечательными были G871C и C1177T, образующие гаплотип, достоверно ассоциированный с ПЖ ($P = 0.0010$), с транскрипцией *Lim3A* в эмбрионах ($P = 0.0001$), головах взрослых мух ($P = 0.0011$) и семенниках ($P = 0.0053$). Каждый из этих двух маркеров был достоверно ассоциирован с транскрипцией в эмбрионах (G871C: $P = 0.0002$; C1177T: $P = 0.0033$), головах ($P = 0.0105$, $P = 0.0021$), семенниках ($P = 0.0138$, $P = 0.0121$), а также с ПЖ ($P = 0.0151$, $P = 0.0084$). Полиморфные сайты A926G и G1021A+C1046A (сцеплены в выборке из 16 линий), расположенные между G871C и C1177T, были достоверно ассоциированы с транскрипцией

Lim3A в эмбрионах (A926G: P=0.0021; G1021A+C1046A: P=0.0018), головах (P=0.0052, P=0.0356) и семенниках (P=0.0209, P=0.0091), как и гаплотип (G871C + A926G + (G1021A + C1046A) + C1177T), состоящий из всех этих 5 маркеров: (P=0.0005, P=0.0058, P=0.0053, для эмбрионов, голов и семенников (Табл. 4).

Таким образом, полученные результаты позволяют предположить, что область ДНК, расположенная на расстоянии от 380 до 686 п.н. выше мажорного старта транскрипции *Lim3A* играет важную роль в регуляции экспрессии *Lim3A*, поскольку именно в ней расположено большинство маркеров, достоверно связанных с уровнем транскрипции, два из которых ассоциированы с изменением ПЖ мух.

Все пять полиморфных сайтов, описанные выше, находятся в предполагаемых сайтах связывания различных транскрипционных факторов: G871C и A926G в сайте связывания Grainy Head (Grh), G1021A в сайте связывания specificity protein-1 (Sp1)/Krüppel-like factor (KLF), C1046A в Zeste-подобном мотиве, C1177T в (CA/TG)₉-повторе (Рис. 2).

G871C и C1177T оказались наиболее значимыми полиморфными маркерами для экспрессии *Lim3A* и ПЖ дрозофилы. В линиях, в которых С, наиболее часто встречающийся аллель в Raleigh популяции, присутствовал в положении 1177, транскрипция *Lim3A* была средней, а ПЖ дрозофилы высокой. В линиях, в которых С был заменен на Т, более редкий аллель, транскрипция и ПЖ становились низкими (Табл. 5). Из этого следует, что в норме этот полиморфный сайт обуславливает активацию транскрипции *Lim3A* и, таким образом, должен находиться в мотиве связывания белка-активатора. C1177T оказался в предполагаемом сайте связывания белка, взаимодействующего с (CA/TG)₉-повтором, являющимся цис-регуляторным элементом (Sharma et al., 2007). Однако, ничего неизвестно про белок, связывающийся с повтором (CA)_n, кроме его молекулярной массы (65–72 кДа) (Papatsenko et al., 2002; Vashakidze et al., 1988).

Транскрипция *Lim3A* и ПЖ дрозофилы зависит от маркера G871C, при условии, что в положении 1177 находится С (Табл. 5). В линиях, в которых G, наиболее часто встречающийся в популяции Raleigh аллель, присутствовал в положении 871, экспрессия *Lim3A* была средняя, а ПЖ высокая. В линиях, в которых G был заменен на С, более редкий аллель, экспрессия становилась выше, а ПЖ ниже (Табл. 5). Из этого следует, что в норме этот полиморфный сайт обуславливает репрессию транскрипции *Lim3A* и, таким образом, должен находиться в мотиве связывания белка-репрессора. Grh, в предполагаемый мотив связывания которого попадает маркер G871C, принимает участие во многих регуляторных сетях, в том числе в комплексной регуляции спецификации нейробластов и апоптоза нейронов (Brody, Odenwald, 2000; Cenci, Gould 2005). Показано, что Grh, являясь репрессором, может взаимодействовать с Polycomb-group (PcG) белками (Dring, Pho) (Tuckfield et al., 2002, Blastya'k et al., 2006) и сайты связывания Grh часто находятся в составе Polycomb response elements (PREs) вместе с сайтами

связывания других белков: SP1/KLF, GAGA (GAF), Pho (Muller, Kassis 2006). Предполагаемые сайты связывания этих белков также были обнаружены в регуляторной области *Lim3A* (Рис. 2).

Еще один полиморфный маркер, A926G, находится в предполагаемом сайте связывания Grh. Однако наши данные показывают, что маркер A926G менее значим для транскрипции *Lim3A*, чем маркер G871C. Можно предположить, что оба сайта функционируют синергично, при этом G871C является ведущим в этом взаимодействии (Табл. 5).

Полиморфный сайт	G871C	A926G	G1021A	C1046A	C1177T			Средняя ПЖ в группах линий мух с различными вариантами гаплотипа (дни) (с.о.)
Предположительные сайты связывания транскрипционных факторов	Grainy head	Grainy head	Sp1	Zeste-подобный белок	Белок, связывающий CA/TG -повтор			
Эффект связывания транскрипционных факторов	СилР	СлР	СлР	СлА	СилА	Количество транскрипта <i>Lim3A</i> в линиях мух (нг) (с.о.)		
Консенсусный нуклеотид референсного генома дрозофилы	G	A	G	C	C			
						голова	эмбрионы	
284	C	A	G	C	C	0.910 (0.078)	2.578 (0.241)	31(2)
207	C	G	G	C	C	0.774 (0.195)	1.369 (0.116)	
285	C	G	G	C	C	0.514 (0.077)	1.361 (0.119)	
100	G	A	A	A	C	0.432 (0.061)	0.876 (0.097)	38(1)
40	G	A	G	C	C	0.402 (0.059)	0.847 (0.028)	
115	G	A	G	C	C	0.490 (0.034)	0.729 (0.085)	
336	G	A	G	C	C	0.420 (0.144)	0.481 (0.130)	
200	G	A	A	A	C	0.349 (0.009)	0.457 (0.138)	
461	G	A	G	C	C	0.394 (0.022)	0.452 (0.043)	
180	G	A	A	A	C	0.193 (0.022)	0.439 (0.134)	
74	G	A	G	C	C	0.301 (0.067)	1.045 (0.043)	
472	G	A	G	C	C	1.068 (0.364)	0.232 (0.027)	
201	G	A	A	A	C	0.757 (0.085)	0.676 (0.033)	
325	G	A	G	C	C	0.712 (0.057)	0.615 (0.148)	
76	G	A	A	A	T	0.127 (0.005)	0.358 (0.039)	29(2)
226	G	A	A	A	T	0.161 (0.049)	0.236 (0.049)	

Табл. 5. Анализ ассоциации полиморфизма регуляторной области *Lim3A* и его экспрессии. Колонки с маркерами, образующими гаплотип, выделены серым цветом. СлР-слабый репрессор, СилР-сильный репрессор, СлА-слабый активатор, СилА-сильный активатор. В скобках приведены значения стандартной ошибки.

В используемой нами выборке из 16 линий маркер C1046A полностью сцеплен с G1021A. C1046A вместе с A1050G попадает в Zeste-подобный мотив (Рис. 2). Наши данные

показывают, что полиморфизм маркера A1050G не связан с транскрипцией *Lim3A*. Таким образом, мы предполагаем, что этот Zeste-подобный мотив не функционален, по крайней мере, в эмбрионах, головах и семенниках. В этом случае ассоциация маркера 1046 с транскрипцией *Lim3A*, вероятно, объясняется его полным сцеплением с другим маркером – G1021A.

В линиях, в которых G, наиболее часто встречающийся в популяции Raleigh аллель, присутствовал в положении 1021, транскрипция *Lim3A* была средняя или высокая в зависимости от аллелей, располагающихся в позиции 871 и 1177. В линиях, в которых G был заменен на A, более редкий аллель, транскрипция становилась средней или низкой в зависимости от аллеля, располагающегося в позиции 871 и 1177 (Табл. 5). Мы предположили, что Sp1/KLF, который предположительно взаимодействует с мотивом, в котором локализован G1021A маркер, функционирует синергично с транскрипционными факторами, в предполагаемых сайтах связывания которых находятся маркеры G871C и C1177T. Sp1/KLF принимает участие в регуляции транскрипции гена человека *LHX3*, гомолога гена *Lim3* дрозофилы (Yaden et al., 2006). Sp1/KLF участвует в ремоделировании хроматина и как белок-активатор, являясь членом Trithorax group (TrxG) белков, и как белок-репрессор, являясь одним из PcG белков (Brown et al., 2005).

Наличие потенциальных сайтов связывания различных белков, встречающихся в составе PREs и TREs свидетельствует о возможной роли этих белков в поддержании активного и неактивного транскрипционного состояния *Lim3A*.

Таким образом, мы обнаружили, что полиморфные маркеры G871C и C1177T находятся в предполагаемых сайтах связывания белков репрессора и активатора, которые могут играть важную роль в регуляции транскрипции *Lim3* и определении ПЖ мух. При этом мы предположили, что белок-репрессор, Grh, и неизвестный белок-активатор, связывающийся с (CA/TG)₉-повтором, обеспечивают позитивно-негативную регуляцию транскрипции *Lim3A* и, вследствие этого, влияют на ПЖ дрозофилы. Разрушение равновесия между позитивной и негативной регуляцией приводит к отклонению уровня транскрипции *Lim3A* от нормы и уменьшению ПЖ мух. В целом, можно предположить, что средний уровень экспрессии, основанный на равновесии между активацией и репрессией гена и благоприятный для высокой ПЖ, может обеспечиваться взаимодействием PcG и trxG белковых комплексов, принимающих участие в активации и репрессии транскрипции различных генов. Было показано, что PcG и trxG белковые комплексы связываются с генами, кодирующими транскрипционные факторы, в том числе гомеодомен-содержащие белки, такие как Lim3, и, тем самым, участвуют в их регуляции (Schuettengruber et al., 2007).

ВЫВОДЫ

1. Ген *Lim3* образует новый, ранее неаннотированный транскрипт, названный *Lim3C*.
2. *Lim3A* и *Lim3C* имеют множественные старты транскрипции, которая обеспечивается с промоторов, не содержащих ТАТА-бокс. В коровой и проксимальной промоторных областях *Lim3A* и *Lim3C* присутствуют обнаруженные с помощью биоинформатического анализа регуляторные элементы, свидетельствующие о возможности инициации транскрипции этих мРНК со своего собственного промотора.
3. В природной популяции Raleigh первый экзон транскриптов *Lim3A* и *Lim3C* характеризуются наименьшим уровнем полиморфизма по сравнению с 5' регуляторной областью и интроном и подвергается действию стабилизирующего отбора.
4. Гаплотип, включающий два полиморфных маркера G871C и C1177T, расположенных в предсказанных на основе биоинформатического анализа сайтах связывания регуляторных белков, достоверно ассоциирован с количеством транскрипта *Lim3A* и ПЖ дрозофилы. Изменение гаплотипа может приводить к шестикратному изменению транскрипции гена и изменению ПЖ на 25%.
5. Вариабельность регуляторной области гена, расположенной на расстоянии 680-380 п.н. от главного старта транскрипции *Lim3A*, связана с изменением количества транскрипта на разных этапах развития и в разных частях тела взрослых мух, что свидетельствует о существовании в популяции полиморфизма в сайтах, важных для регуляции общего, нетканеспецифического уровня транскрипции.
6. Комбинация маркеров G871C и C1177T, ассоциированная с промежуточным уровнем транскрипции и высокой ПЖ, встречается в популяции с высокой частотой, в то время как комбинации, ассоциированные с низким или высоким уровнем транскрипции и низкой ПЖ, встречаются в популяции с низкой частотой, что указывает на функциональную и селективную ценность среднего уровня экспрессии гена *Lim3*.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи

1. Симоненко А. В., Рыбина О. Ю., Пасюкова Е. Г. Молекулярная изменчивость генов *shuttle craft* и *Lim3*, контролирующих развитие нервной системы, в природной популяции *Drosophila melanogaster* // Генетика, 2008, Т. 44, № 9. с.1172-1177.

2. Rybina O.Y., Pasyukova E.G. A Naturally occurring polymorphism at *Drosophila melanogaster Lim3* locus, a homolog of human *LHX3/4*, affects *Lim3* transcription and fly lifespan // PLoS ONE, 2010, V. 5(9): e12621.

Материалы всероссийских и международных конференций

1. Rybina O.Y., Pasyukova E.G., Molecular variation in the *Lim3* locus controlling neuron development is associated with *Drosophila melanogaster* lifespan. 41st Population Genetics Group meeting UK, Department of Plant Sciences, University of Oxford, United Kingdom. 2007. P. 44

2. Rybina O. Y., Pasyukova E.G. Molecular variation in the *Lim3* locus controlling interneuron and motorneuron development is associated with *Drosophila melanogaster* lifespan. 49 *Drosophila* Research Conference. San-Diego. USA. 2008.

3. Рыбина О. Ю., Пасюкова Е.Г. Молекулярный полиморфизм гена *Lim3* связан с изменчивостью продолжительности жизни *Drosophila melanogaster* в природной популяции. IV Съезд Российского общества биохимиков и молекулярных биологов. Новосибирск. Россия. 2008. С.115.

4. Pasyukova E.G., Rybina O. Y. Molecular variation in *Lim3* locus controlling neuron development is associated with *Drosophila melanogaster* lifespan. XX International Congress of Genetics. Berlin. Germany. 2008, С.193.

5. Рыбина О.Ю., Пасюкова Е.Г. Полиморфизм гена *Lim3* в природных популяциях *Drosophila melanogaster* сопряжен с уровнем экспрессии гена и продолжительностью жизни мух. Пятый съезд Вавиловского Общества Генетиков и Селекционеров. Москва. Россия. 2009. Ч. 2.С. 86.

6. Рыбина О.Ю., Пасюкова Е.Г. Природный полиморфизм 5' регуляторной области гена *Lim3 D. melanogaster*, гомолога генов человека *LHX3/4*, определяет различия в уровне его экспрессии и продолжительности жизни дрозофилы. Международная научная конференция по биоорганической химии, биотехнологии и бионанотехнологии, посвященная 75-летию со дня рождения академика Ю.А. Овчинникова. Москва. Россия. 2009. Т. 2. С. 206-207.

7. Rybina O. Y., Pasyukova E.G. Molecular variation in *Drosophila melanogaster Lim3* gene, a homolog of human *LHX3/4*, is associated with fly lifespan. 35th FEBS Congress "Molecules of life". Gothendurg. Sweden. 2010. P. 297.