

РОССИЙСКАЯ АКАДЕМИЯ НАУК
ИНСТИТУТ ОБЩЕЙ ГЕНЕТИКИ им. Н.И. ВАВИЛОВА

На правах рукописи

КЛИМОВ

Евгений Александрович

**ХАРАКТЕРИСТИКА ГЕНОМНЫХ
ПОСЛЕДОВАТЕЛЬНОСТЕЙ ЧЕЛОВЕКА, ПРИЛЕГАЮЩИХ
К МЕСТАМ ВСТРАИВАНИЯ ВИРУСА ПАПИЛЛОМЫ**

Специальность 03.00.15 – генетика

ДИССЕРТАЦИЯ

на соискание ученой степени
кандидата биологических наук

Научный руководитель -
доктор биологических наук,
профессор Сулимова Г.Е.

Москва – 2002

ОГЛАВЛЕНИЕ

| | |
|---|----|
| Список сокращений | 4 |
| Введение | 5 |
| Обзор литературы | 8 |
| I. Роль вирусов папиллом человека в патогенезе шейки матки | 8 |
| 1. Заболеваемость и факторы риска рака шейки матки (РШМ) | 8 |
| 2. Ассоциация РШМ с инфекцией вирусами папиллом | 8 |
| 3. Структура вирусного генома | 11 |
| 4. Механизмы трансформирующего действия вирусных онкогенов | 14 |
| 5. Персистенция вирусной ДНК | 15 |
| 6. Интеграция вирусной ДНК в клеточный геном | 17 |
| II. Подходы к картированию и анотированию геномов высших эукариот | 19 |
| 1. Генетическое картирование | 19 |
| 2. Физическое картирование | 21 |
| 2.1. Цитогенетическое картирование | 23 |
| 2.2. Картирование на основе гибридов соматических клеток | 24 |
| 2.3. Рестрикционное картирование генома человека | 26 |
| 2.4. Контиг-карты хромосом человека на основе перекрывающихся клонов геномной ДНК человека | 27 |
| 2.5. Картирование генома на основе радиационных гибридов соматических клеток (RH-картирование) | 30 |
| 3. Предсказание генов человека с помощью компьютерных методов анализа последовательностей биополимеров | 35 |
| 3.1. Молекулярно-биологические базы данных | 36 |
| 3.1. Строение генов и подходы к их предсказанию в последовательностях генома человека | 37 |
| Материалы и методы | 40 |
| Результаты и их обсуждение | 53 |
| I. Отработка метода радиационного картирования на примере локализации мозгоспецифичной последовательности Gfhb и маркера GC-богатой последовательности – cmgc30 | 53 |

| | |
|---|-----|
| 1. Подбор праймеров и оптимизация условий полимеразной цепной реакции | 55 |
| 2. Локализация исследуемых маркеров | 56 |
| II. Локализация последовательностей генома человека, прилегающих к местам встраивания вируса папилломы | 60 |
| 1. Физическая локализация последовательностей, прилегающих к местам встраивания HPV16 методом RH-картирования на хромосомах человека | 63 |
| 1.1. Оптимизация условий ПЦР, скрининг панели радиационных гибридов и определение локализации маркеров INT на RH-карте | 63 |
| 1.2. Совмещение RH-карты с другими типами карт и поиск генов, прилегающих к местам встраивания HPV16 | 65 |
| 2. Локализация <i>in silico</i> последовательностей, прилегающих к местам встраивания HPV16 в геноме человека | 68 |
| 3. Поиск гомологий последовательностей, прилегающих к местам встраивания HPV16, с экспрессирующимися последовательностями человека | 73 |
| 4. Анализ нуклеотидных последовательностей генома человека, прилегающих к местам встраивания HPV16 | 78 |
| 4.1. Анализ нуклеотидных последовательностей генома человека, прилегающих к местам встраивания HPV16, с использованием сервиса Genomatrix Suite | 78 |
| 4.2. Анализ геномных последовательностей человека, прилегающих с 3'-конца к интегрированной вирусной ДНК | 81 |
| 5. Анализ гипотетических генов | 83 |
| 5.1. Установление экзон-интронной структуры и поиск необходимых регуляторных элементов | 83 |
| 5.2. Определение предположительной функции белков, кодируемых описанными генами | 85 |
| 5.3. Анализ экспрессии изучаемых генов в различных органах и тканях | 86 |
| Заключение | 90 |
| Выводы | 92 |
| Список литературы | 93 |
| Приложения | 107 |

СПИСОК СОКРАЩЕНИЙ

п.н. – пар нуклеотидов

т.п.н. – тысяч пар нуклеотидов

ОТ-ПЦР – обратной транскрипции полимеразная цепная реакция

ПЦР – полимеразная цепная реакция

ТК – тимидинкиназа

АРОТ – amplification of papillomavirus oncogenes transcripts

ВАС – bacterial artificial chromosome

DDBJ – DNA Data Bank of Japan

EBI – The European Bioinformatics Institute

EMBL – European Molecular Biology Laboratory

EST – expressed sequence tag

FISH – fluorescence *in situ* hybridization

GDB – Genome Data Base

IFGT – irradiation and fusion gene transfer

HPRT – фосфорибозилтрансфераза

HPV – human papilloma virus

HR-HPV – high risk HPV

MAR – matrix attachment region

NCBI – National Center for Biotechnology Information

ORF – open reading frame

RH – radiation hybrids

S/MAR – scaffold/matrix attachment region

STS – sequenced tagged site

URR – upstream regulated region

UTR – untranslated region

YAC – yeast artificial chromosomes

ВВЕДЕНИЕ

Одной из важных задач современной генетики является изучение наследственных заболеваний человека и молекулярных механизмов онкогенеза. В этой связи активно ведутся исследования по поиску и изучению вирусных генов, связанных с развитием различных опухолей, а также последовательностей генома человека, в которые происходит встраивание вирусных генов.

Вирус папилломы человека (HPV) относится к ДНК содержащим вирусам. Известно около 100 представителей HPV человека, из которых типы 16, 18 и им родственные ассоциированы с развитием рака шейки матки [De Villiers, 1994]. В связи с его высокой онкогенностью и распространенностью в популяции человека изучение HPV представляет немалый интерес, как для прикладной, так и для теоретической науки.

HPV может существовать в клетке как в эписомальной, так и в интегрированной форме. Экспрессия вирусного генома происходит с использованием механизма сплайсинга, при этом экспрессируются два трансформирующих гена вируса *E6* и *E7*, а в случае интегрированной формы и слитые с вирусными клеточные последовательности, расположенные с 3'-конца встроенной ДНК вируса [Klaes et al., 1999; Киселев и др., 2001]. Часть интегрированных вирусных последовательностей остаются «молчащими» и не экспрессируются.

В случае интеграции вирус теряет собственный сайт полиаденилирования. До последнего времени оставалось неизвестным, где происходит терминация транскрипции ДНК вируса при ее интеграции в геном человека.

Как свидетельствуют литературные данные по локализации сайтов, в которые происходит встраивание вируса, интеграция HPV16 и других типов HPV может происходить в разные участки клеточного генома [el Awady et al., 1987; Cannizzaro et al., 1988; Sastre-Garau et al., 1990; Gilles et al., 1996; Wilke et al., 1996; Koopman et al., 1999; Thorland et al., 2000; Wentzensen, et al., 2002]. Многие авторы перечисленных работ указывают на близкое расположение к местам интеграции участков клеточного генома, часто подверженных

хромосомным перестройкам и делециям, а также генов, прямо или косвенно участвующих в развитии онкозаболеваний. Локализацию в указанных работах проводили с использованием *in situ* гибридизации или панели гибридов соматических клеток, т.е. методов, не позволяющих определить точную физическую локализацию интегрированных последовательностей HPV.

Несмотря на впечатляющие успехи в расшифровке первичной нуклеотидной последовательности, аннотирование генома человека далеко от завершения. Остаются не определенными функции и структура более 2/3 генов человека. Структура и предположительные функции большинства генов на настоящий момент определены *in silico* [Venter et al., 2001]. И хотя полученные компьютерными методами данные нельзя принимать как абсолютно достоверные, этап работы по аннотированию генома человека *in silico* является основой для дальнейших экспериментальных исследований и значительно облегчает планирование и постановку эксперимента.

Физическое картирование на хромосомах человека мест интеграции ДНК вируса и их *in silico* анализ имеют большое значение для понимания того, как интегрированный вирусный геном может нарушать генетическую программу клетки.

Целью данной работы были точная физическая локализация мест встраивания HPV16 в хромосомы человека и характеристика *in silico* геномных последовательностей, прилегающих к местам встраивания вирусной ДНК, транскрибируемой в клетках плоскоклеточного рака шейки матки.

Для достижения указанной цели были поставлены следующие задачи:

1) отработка метода картирования с использованием панели радиационных гибридов соматических клеток (RH-картирование);

2) определение физической локализации геномных последовательностей, прилегающих к местам встраивания транскрибируемой ДНК вируса папилломы тип 16 (маркеры INT) в хромосомы человека в клетках плоскоклеточного рака шейки матки методом RH-картирования и картирования *in silico*;

3) анализ *in silico* последовательностей генома человека, прилегающих к местам встраивания HPV16, и структурно-функциональный анализ гипотетических генов, в последовательности которых происходит интеграция ДНК HPV16;

4) поиск *in silico* в геномных последовательностях человека возможных сайтов терминации транскрипции интегрированной ДНК вируса;

5) экспериментальное подтверждение методом ОТ-ПЦР транскрипционной активности проанализированных гипотетических генов человека.

ОБЗОР ЛИТЕРАТУРЫ.

I. Роль вирусов папиллом человека в патогенезе шейки матки.

1. Заболеваемость и факторы риска рака шейки матки (РШМ).

РШМ является одним из ведущих в структуре онкологической заболеваемости и смертности среди женщин и остается одной из важных медицинских и социальных проблем в экономически развитых странах. Опухоли шейки матки представляют собой опухолевую патологию у женщин со средней частотой встречаемости около 440000 случаев в год, что составляет 5,8% от общей частоты заболеваемости раком [Киселев, 1997].

Первые исследования, посвященные эпидемиологии РШМ, появились еще в XIX веке. Rigoni-Stern в 1842 г. высказал гипотезу об инфекционном происхождении РШМ. Достаточно долгое время господствовало мнение, основанное на утверждении R.Meuer (1910), о травме как этиологическом моменте метапластических изменений эпителия шейки матки. G.Ward (1935) считал, что не разрывы шейки матки сами по себе, а последующая инфекция, ведущая к хроническому воспалению, является этиологическим моментом в развитии карцином. Установлено, что использование барьерных методов контрацепции приводит к достоверному снижению заболеваемости РШМ [Киселев, 1997; Zur Hausen, 1991; Zur Hausen, 1994], применение внутриматочных контрацептивов не оказывает влияния на заболеваемость РШМ. Данные о влиянии длительной стероидной контрацепции на частоту дисплазии и рака шейки матки разноречивы [цит. по: Сахарова, Нечушкин, 1999].

2. Ассоциация РШМ с инфекцией вирусами папиллом.

Концепция вирусной природы РШМ имеет длинную историю и неразрывно связана с исследованиями кондилом различных локализаций. Генитальные кондиломы описаны еще в трудах по античной медицине. Первые предположения об инфекционной природе рака шейки матки высказал в 1842 г. Rigoni-Stern. В 1903 г. Borrel и Bosc предложили вирусную теорию развития

опухолей; Rous, Kidd (1938), Rous, Friedewald (1941) получили доказательства усиления активности опухолеродных вирусов в присутствии канцерогенных веществ. Достижения в экспериментальной вирусологии позволили Л.А. Зильберу (1945), позже Horsfall (1963) и Southam (1964) сформулировать положения об опухолеродных вирусах как агентах, превращающих нормальную клетку в опухолевую посредством воздействия на ее генетический материал, действие этих агентов отличается от инфекционного, и в дальнейшем размножении уже возникших опухолевых клеток опухолеродные вирусы роли не играют.

К этому времени относится демонстрация Melnick (1952) и А.Д. Тимофеевским (1961) с помощью электронно-микроскопической техники постоянного присутствия вирусоподобных частиц в папилломах. Dunn и Ogilvie (1968) выделили вирусные частицы из генитальных бородавок. Позже с помощью электронно-микроскопических, иммуногистохимических и ДНК-гибридизационных исследований была подтверждена папилломавирусная этиология кандилом [цит. по: Сахарова, Нечушкин, 1999].

Впервые серьезные предположения о возможной роли вирусов папиллом в развитии РШМ были высказаны в середине 70-х годов. Следует подчеркнуть, что несмотря на то, что из зараженных клеток не удается выделить инфекционный агент, именно вирус выступает в качестве этиологического фактора развития опухоли. Это утверждение основано на сочетании следующих критериев: 1) регулярное обнаружение в опухолевых клетках вирусной ДНК (как в интегрированной, так и в эписомальной форме); 2) клонированные вирусные гены в клеточных системах *in vitro* способны индуцировать злокачественную трансформацию клеток, включая клетки человека; 3) регулярная экспрессия вирусных генов в опухолевых клетках; 4) существование в природе сходных вирусов, способных индуцировать образование опухолей у экспериментальных животных; 5) эпидемиологические данные, подтверждающие связь между опухолевым процессом и наличием вирусного генетического материала в опухолевых клетках [Киселев, 1997]. Методика молекулярного клонирования позволила создать физическую карту вирусов с помощью эндонуклеаз, определить местоположение и последовательность генов, кодирующих вирусные белки [De Villiers, 1994]. Было показано, что в клетках РШМ геном вируса активно транскрибируется, а решающее значение

для развития РШМ имеет персистенция вирусного генома [Zur Hausen, 1994]. Клеточные линии, полученные из опухолевой ткани шейки матки, позволяют проводить сравнительный анализ линий и опухолей на персистенцию вирусной ДНК, функцию и экспрессию генов ДНК вируса папилломы человека (HPV – human papilloma virus). В тканях эпителиальных дисплазий и карцином шейки матки в процессе опухолевой трансформации нарушается регуляция транскрипции вирусных онкогенов, что проявляется количественными различиями в экспрессии ранних генов E6 и E7, играющих основную роль в индукции злокачественного роста клеток [Zur Hausen, 1991].

Современные молекулярно-диагностические тесты быстро продвигаются в направлении повышения их чувствительности и специфичности. В настоящее время для идентификации вирусной ДНК используются различные методики гибридизации ДНК и молекулярного клонирования. Наиболее чувствительным методом в настоящее время признана полимеразная цепная реакция с типоспецифическими и видоспецифическими праймерами, позволяющая выявлять вирусные последовательности в геноме клеток опухолей шейки матки в 95-100% случаев [Zur Hausen, 1994; Киселев, 1997]. В настоящее время идентифицировано более 100 типов HPV, подробно описаны более 70 типов, твердо установлен факт, что определенные типы HPV могут инфицировать строго определенный вид эпителия и вызывать характерные изменения. Выявление многовариантности генотипов HPV и идентификация специфических генотипов, накопление данных о злокачественной трансформации генитальных кондилом позволили рассматривать папилломавирус как возможный этиологический фактор развития РШМ [Zur Hausen, Gissman, 1980; De Villiers, 1994; Zur Hausen, 1994].

Из всех идентифицированных типов вирусов папиллом 34 ассоциированы с поражением аногенитальной области. Вирусы папиллом инфицируют базальные слои эпителия (наиболее "уязвимым" участком является зона перехода многослойного плоского эпителия в цилиндрический эпителий). Онкогенный потенциал папилломавирусов существенно варьирует; по способности инициировать диспластические (предраковые) изменения и рак папилломавирусы условно разделены на группы "высокого" и "низкого" риска возникновения опухолевой трансформации инфицированного эпителия. Типы HPV 6,11, 42, 43, 44 были классифицированы как типы низкого риска развития

рака, типы 16, 18, 48, 56 и другие – высокого риска. Так, HPV6 и HPV11 являются причиной остроконечных кондилом, часто идентифицируются при дисплазиях легкой и средней степеней тяжести. HPV16 и HPV18 преобладают над другими типами папилломавирусов при РШМ: HPV16 выявляется в 50-70% случаев, HPV18 – в 10-20% случаев, а остальные типы HPV высокого риска выявляются значительно реже [Zur Hausen, 1991; De Villiers, 1994]. Вирус папилломы типа 18 ассоциирован с развитием аденокарцином, имеет более высокий онкогенный потенциал, с ним связывают быстрый темп опухолевой прогрессии, низкий уровень дифференцировки инфицированного эпителия, неблагоприятный прогноз по сравнению с другими онкогенными типами. [цит. по: Сахарова, Нечушкин, 1999]

В первых обзорах представлены данные о широком распространении HPV в популяции человека и его выраженном онкогенном потенциале [Zur Hausen, Gissman, 1980]. Авторы обращают внимание, что одного только инфицирования вирусом папилломы недостаточно для индукции опухолевого роста, и указывают на роль ко-факторов в HPV-зависимом канцерогенезе. Таким образом, инфекция вирусом папилломы является необходимым, но не достаточным фактором развития злокачественного процесса, в процессы иммортализации и трансформации вовлечены клеточные факторы, участвующие в регуляции клеточного цикла и дифференцировки клеток [Zur Hausen, Gissman, 1980; Zur Hausen, 1994; Киселев, 1997].

3. Структура вирусного генома.

Вирусы папиллом относятся к семейству паповавирусов (Papovaviridae) и представляют собой группу вирусов, поражающих крупный рогатый скот, птиц и человека; они способны инфицировать базальные клетки кожи и плоского эпителия. Папилломавирусы – одна из наиболее гетерогенных групп вирусов, критерием дифференцировки которой является степень генетического родства вирусов по данным молекулярной гибридизации: она может колебаться от 10 до 85%. Диаметр вирусных частиц 55 нм. Вирус не имеет внешней оболочки. Капсид вируса состоит из 72 капсомеров. Детальный анализ молекулы ДНК HPV стал возможен после разработки методики расщепления ДНК с использованием рестрицирующих эндонуклеаз и анализа этих

фрагментов с помощью гель-электрофореза. Данный метод позволил определить характерные карты рестрикции ДНК и создать физическую карту расположения сайтов рестрикции в геноме различных папилломавирусов [Zur Hausen, Gissman, 1980].

Известные типы папилломавирусов человека сходны по своей генетической структуре. Генетический материал вируса представлен кольцевой двухцепочечной молекулой ДНК длиной около 8000 пар оснований. Одна из нитей ДНК содержит 9 открытых рамок считывания (ORF - open reading frames), которые потенциально кодируют до 10 протеинов, и регуляторный участок генома (URR - upstream regulatory region). Другая нить ДНК не кодирующая [цит. по: Сахарова, Нечушкин, 1999]. Для собственной репликации вирус использует клеточные белки.

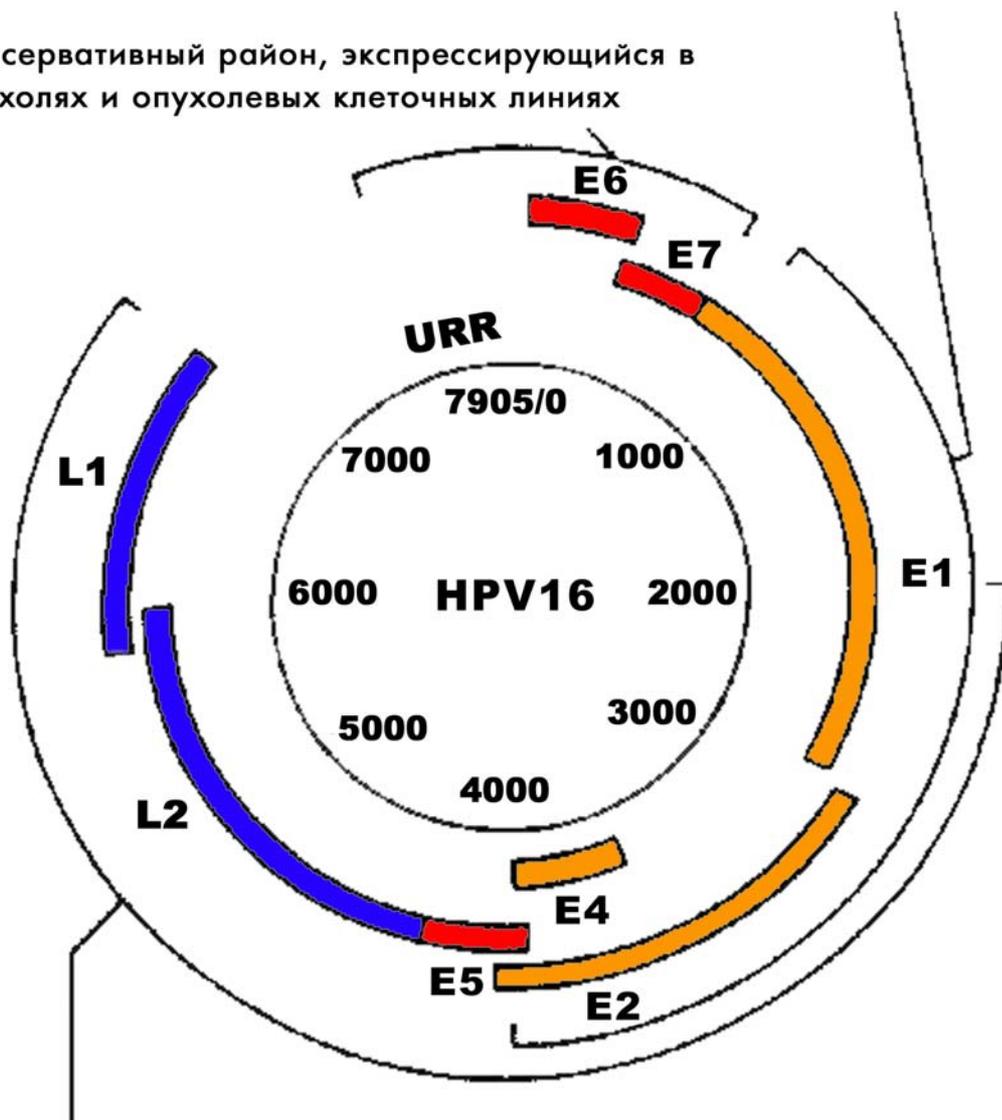
Регуляторный участок генома содержит область вирусного промотора, locus репликации, энхансерные элементы, активирующие последовательности, необходимые для репликации вирусной ДНК, сайты связывания для вирусного белка E2 и клеточных факторов транскрипции. В URR также содержатся сайты связывания для рецепторов прогестерона и глюкокортикоидных гормонов, увеличивающих транскрипцию в 2-3 раза [Khare et al., 1997; Chen et al., 1996]. В регуляторной области генома располагается промотор p97 (HPV16), p105 (HPV18), регулирующий экспрессию вирусных генов; активность самого промотора регулируется вирусными и клеточными белками.

URR вирусного генома располагается между концом области поздних генов и началом области ранних генов. Открытые рамки считывания генома вируса разделены на ранний (early, E) и поздний (late, L) участки (рис. 1). Ранний фрагмент включает гены E1-E7, кодирующие синтез белков, ответственных за различные функции в процессе репликации вируса и трансформации клеток. Гены E1 и E2 ответственны за репликацию вируса, а также участвуют в регуляции транскрипции вирусных генов. Продукт гена E1 также отвечает за поддержание персистенции вирусного генома в эписомальной форме. Ген E2 кодирует продукты, которые могут как транс-активировать, так и подавлять экспрессию ранних генов и энхансерных участков URR [Antson et al., 2000]. Клеточный белок p53 способен модулировать активность белка E2 [Mantovani, Banks, 1999]. Ген E4 участвует в процессе созревания вирусных частиц, гены E5-E7 обладают трансформирующим потенциалом (транскрипция

контролируется URR). Маленький вирусный белок E5 связывается с одной из субъединиц вакуолярной АТФазы [Briggs et al., 2001]. Также, действуя в цис-положении, он повышает в 4-10 раз эффективность клеточной immortalization белками E6 и E7 [Stoppler et al., 1996]. Гены E6 и E7 всегда определяются и экспрессируются в опухолях шейки матки и полученных из опухолей шейки матки клеточных линиях. Поздний фрагмент генома состоит из генов L1 и L2, кодирующих структурные белки вириона [Киселев, 1997].

Район, в котором при интеграции происходят разрывы

Консервативный район, экспрессирующийся в опухолях и опухолевых клеточных линиях



Район теряющийся или не экспрессирующийся при интеграции

Рис. 1. Схематическое изображение генома вируса папилломы человека тип 16. Объяснения в тексте. [Лаасри Мажид, 1998].

4. Механизмы трансформирующего действия вирусных онкогенов.

Онкобелок Е6 играет роль коактиватора, взаимодействующего с факторами транскрипции и элементами основного транскрипционного комплекса [Etscheid et al., 1994]. В С-концевой части он содержит также домен «цинковые пальцы», способный связываться с ДНК [Ristriani et al., 2000]. Направление действия белка Е6 (супрессорное или трансактивирующее) определяется характером, положением и числом сайтов связывания транскрипционных факторов в промоторах. Промотор p97 (HPV типа 16) может играть роль в поддержании латентного состояния вирусной инфекции. С онкобелком Е6 HPV типов 16 и 18 взаимодействуют, как минимум, 7 клеточных белков, которые еще плохо охарактеризованы. В настоящее время из них идентифицировано 3 клеточных белка: p53, Е6-АР, Е6-ВР. Белок p53 ассоциирован как с контролем клеточного роста, так и с неопластической трансформацией; делеции или мутации в некоторых случаях превращают p53 в активный онкоген. Не все злокачественные опухоли имеют мутации в p53, однако доказан альтернативный механизм инактивации p53 путем связывания с онкобелком Е6 HPV типов 16 и 18, что приводит к потере контроля за пролиферацией, повреждением ДНК и хромосомной нестабильности. Клеточный белок Е6-АР (Е6 associated protein) в комплексе с онкобелком Е6 участвует в деградации p53, а также, возможно, укорачивает период «полужизни» и уменьшает уровень p53 в HPV-иммортиализованных клетках. Белок Е6-ВР (Е6-binding protein) взаимодействует с геном Е6 HPV типов 16 и 18, что в результате приводит к ингибированию процесса дифференцировки клеток и, возможно, создает условия для репликации вирусной ДНК [Mantovani, Banks, 1999; Vogelstein, Kinzler, 1992; Hartwell, 1992]. Белок Е6 способен также нарушать функции фактора некроза опухолей (TNF) [Duerksen-Hughes et al., 1999]. Показана высокая корреляция уровня экспрессии вирусного белка Е6 с трансформирующим потенциалом вируса [Liu et al., 1995].

Таким образом, ген *Е6* кодирует мультифункциональный белок, трансактивирующая активность которого предполагает участие в регуляции транскрипции, взаимодействие с белком p53 и его деградацией, в нарушении механизма контроля клеточного роста и процесса дифференцировки клеток [Galloway, McDougll, 1996; Zur Hausen, 1999a, 1999b]. Показано, что Е6-

зависимая деградация p53 необходима, но не достаточна для трансформации клеток HPV [Nakagawa et al., 1995].

Вирусный белок E7 взаимодействует с клеточным белком Rb105 (супрессор клеточного роста), Rb-родственными белками и H1-киназой. Эти взаимодействия являются независимыми процессами, необходимыми для эффективной трансформации клеток белком. Высокая экспрессия гена E7 ведет к быстрому и эффективному изменению фенотипа клеток [Liu et al., 1995]. Белок HPV E7 может ослаблять активность некоторых циклин-зависимых ингибиторов киназ, ассоциироваться с циклин/циклин зависимыми киназными комплексами и дестабилизировать онкосупрессорный ген ретинобластомы [Jones, Munger, 1996]. E7 также связывается с актином, что приводит к его более активной полимеризации [Rey et al., 2000a; 2000b], и ингибирует функциональную активность альфа-интерферона, отвечающего за ответ клетки на вирусную инфекцию [Barnard, McMillan, 1999; Barnard et al., 2000].

Таким образом, мультифункциональный вирусный белок E7 способен нарушать регуляцию контроля клеточного цикла, образуя стабильный неактивный комплекс с серией регулирующих клеточный цикл белков; стимулирует развитие событий в S-фазе клеточного цикла и нарушения регуляции пролиферации трансформированных клеток; способен отменять остановку клеток в G1-фазе клеточного цикла, оказывать митогенное влияние и стимулировать неконтролируемый синтез ДНК [Chen et al., 1993; De Villiers, 1994; Galloway, McDougll, 1996; Zur Hausen, 1999].

5. Персистенция вирусной ДНК.

Геном вируса существует в клетке хозяина в двух формах: эписомальной и интегрированной. Данные о формах персистенции вирусной ДНК на разных стадиях опухолевого процесса противоречивы. По мнению некоторых авторов, интеграция необходима для развития опухоли, т.к. по их данным более 80% опухолей содержат вирусную ДНК в интегрированной форме [Klaes et al., 1999; Wentzensen et al., 2002]. Другие авторы предполагают, что обе формы персистенции ДНК HPV являются значимыми, а интеграция не является определяющим звеном в механизме индукции онкогенеза [Zur Hausen, 1999; Szarka et al., 2000; Киселев и др., 2001].

Транскрипция эписомальной вирусной ДНК инициируется в URR, продолжается последовательно через гены *E6*, *E7*, *E1*, *E2*, *E4* и *E5* (при этом может формироваться несколько сплайсированных вариантов РНК) и terminates в конце всего раннего участка генома в полиА-сайте на конце гена *E5* (Рис. 2).

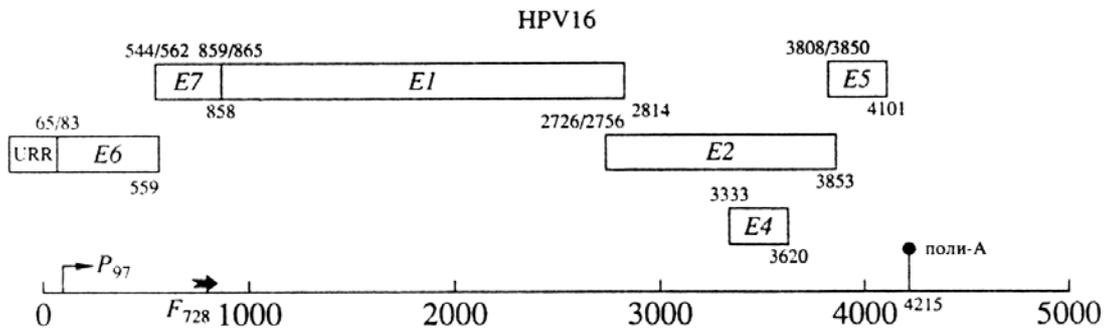


Рис. 2. Схема транскрипции эписомального вирусного генома. Объяснения в тексте. [Киселев и др., 2001]

В случае интеграции вирусного генома его транскрипция также инициируется в URR [Schwarz et al., 1985], продолжается через гены *E6* и *E7*, а также через часть участка *E1* + *E2*, затем продолжается в клеточных последовательностях и terminates в неизвестном клеточном поли(А)-сайте. Очевидно, что размер транскрибируемого продукта может существенно варьировать, если интеграция происходит в различные участки хозяйского генома [Киселев и др., 2001].

Разрыв кольцевой молекулы вирусной ДНК может происходить в любом месте, однако, наиболее часто – в области *E1/E2*. Область URR-*E6-E7* остается интактна в опухолевых клетках [Schneider-Mannoury et al., 1987; Wagatsuma et al., 1990]. При разрыве в области *E6-E7* опухоль не возникает, поэтому эти случаи редко регистрируются. Исчезновение супрессорной функции вирусного белка E2 обуславливает суперэкспрессию вирусных онкогенов *E6* – *E7*, трансформирующее действие которых способствует прогрессии неоплазий.

Предполагается, что ген *E1* участвует в подавлении вирусного промотора, следствием мутации гена *E1* являются активация транскрипции вирусного генома и увеличение трансформирующей активности вирусных онкогенов в случае кольцевой формы [Matlashewski, 1989; Zur Hausen, 1994].

6. Интеграция вирусной ДНК в клеточный геном.

Биологическое значение интеграции вирусной ДНК в геном до сих пор остается неясным. Процесс интеграции сопровождается частичной потерей генетического материала вируса, при этом вирусный геном независимо от делеций теряет способность к полной репликации, и в клетках, содержащих интегрированный клеточный геном, отсутствует продукция вирусных частиц. Предполагалось, что интеграция представляет собой активационный механизм прогрессии неоплазии от дисплазии тяжелой степени к раку.

Как свидетельствуют литературные данные по локализации участков генома, в которые происходит встраивание вируса, интеграция HPV16 и других типов HPV может происходить в разные участки клеточного генома: 1q22, 3p14.2, 3p21, 3q26, 8q21-q22.3, 9p13, 10q24, 12q14-q15, 13q33, 13 [Koopman et al., 1999; Sastre-Garau et al., 1990; Gilles et al., 1996; Wilke et al., 1996; Cannizzaro et al., 1988; el Awady et al., 1987]. Многие авторы перечисленных работ указывают на близкое расположение к местам интеграции ДНК вируса участков клеточного генома, часто подверженных хромосомным перестройкам (сайты ломкости хромосом – fragile sites) и делециям, а также генов, прямо или косвенно участвующих в развитии онкозаболеваний. Локализацию в указанных работах проводили с использованием *in situ* гибридизации или панели гибридов соматических клеток, т.е. методов, не позволяющих определить точную физическую локализацию интегрированных последовательностей HPV. Авторы перечисленных работ предполагают, что интеграция вирусной ДНК индуцирует нестабильность клеточного генома и, как следствие, нарушения в различных хромосомах, тестируемых по потере гетерозиготности [Mitra, 1999]. Также показано, что различные онкогенные вирусы интегрируют в так называемые матрикс прикрепляющие области генома человека (MAR) [Shera et al., 2001] или вблизи повторяющихся последовательностей, например, Alu повторов [Carmody et al., 1996].

Более детальный анализ трех последовательностей генома человека, прилегающих к местам встраивания HPV16, был проведен Thorland с соавторами [Thorland et al., 2000]. Авторы использовали для локализации секвенированных ими участков интеграции HPV16 современные подходы: поиск гомологичных последовательностей в базах данных и метод RH-

картирования, подтвержденный гибридизацией *in situ*. Все три последовательности были локализованы вблизи сайтов ломкости хромосом.

Авторы другой работы [Wentzensen, et al., 2002] проанализировали 68 экспрессирующихся вирус-клеточных транскриптов, полученных с использованием метода APOТ (подробнее о методе см. раздел «Результаты и обсуждение»). В этой работе локализацию клеточных последовательностей проводили *in silico* с использованием программы BLASTN. Авторы показали, что места интеграции HPV16 случайно распределены в геноме человека, однако, приблизительно в 70% случаев находились вблизи известных сайтов ломкости хромосом, а в 50% интеграция происходит в транскрибируемые гены. (анализ EST).

В заключение следует отметить, что РШМ является уникальной моделью канцерогенеза, ассоциированного с вирусной инфекцией. Ее уникальность определяется постоянным присутствием в опухолевых клетках экзогенной генетической информации, которая обладает трансформирующим потенциалом, а также наличием экспериментального подтверждения роли трансформирующих вирусных генов в индукции трансформированного фенотипа и накоплением большого количества эпидемиологических данных, подтверждающих связь между опухолевым процессом и наличием вирусного генетического материала.

II. Подходы к картированию геномов высших эукариот.

1. Генетическое картирование.

Изучение генома человека до недавнего времени было возможно только путем генетического анализа – построения генетических карт или карт сцепления (linkage map) [Лобашев, 1967]. Основой для их построения является изучение частоты рекомбинации генных локусов в мейозе. Карты сцепления показывают порядок линейного расположения генов и маркеров на хромосоме и генетическое расстояние между ними, выраженное в процентах рекомбинации – сантиморганах (сМ). Размер генома человека оценивается примерно в 2,91 млрд.п.н. [Venter et al., 2001]. Карта генетического сцепления составляет около 2809 сМ для мужчин и 4782 сМ у женщин [Morton, 1991]. Следовательно, на 1 сМ генетической карты, что соответствует 1% рекомбинации, приходится в среднем 1 млн.п.н. ДНК на физической карте генома.

Поскольку частота рекомбинации в разных точках генома различна (“горячие точки” рекомбинации, районы генома, где рекомбинация подавлена – центромерные и теломерные участки хромосом, блоки конститутивного гетерохроматина и др.), размерность сМ может существенно варьировать и, в результате, карты сцеплений не отражают реальных физических расстояний между маркерами и генами на хромосомах.

Построение генетических карт человека до начала 70-х годов XX века продвигалось очень медленными темпами. Небольшой размер семей, длительный период одного поколения, ограниченное число информационных родословных и отсутствие методов эффективного цитогенетического анализа всех пар хромосом затрудняло целенаправленное картирование генов человека. Так, первый ген человека был картирован на X-хромосоме в 1911 г., а первый аутосомный ген – только в 1968 г. К середине 70-х годов на хромосомах человека было картировано менее 100 генов, значительная часть которых оказалась локализованными на X-хромосоме [Горбунова, 1999].

Дальнейший прогресс в области генетического картирования в значительной мере связан с деятельностью крупных научно-исследовательских центров по созданию банков клеточных культур, представляющих наиболее

интересные и обширные родословные. Так, в Центре по Изучению Полиморфизма Человека – СЕРН (от фр. Centre d'Etudes du Polymorphisme Humain) – была создана уникальная коллекция перевиваемых клеточных линий, полученных от членов семей, многоступенчатые родословные которых насчитывают десятки и даже сотни индивидуумов. Перевиваемые линии клеток получали путем трансформации их вирусом Эпштейн-Барра. Такие лимфобластные линии клеток способны к неограниченному росту в условиях культивирования [Cohen et al., 1993; Chumakov et al., 1993].

СЕРН-коллекции родословных представляют собой идеальные системы для генетического анализа наследственных признаков. Эти коллекции были использованы исследователями во всем мире для локализации генов человека и различных типов маркеров. В результате исследования этих клеточных линий определены генотипы членов СЕРН-семей одновременно по тысячам полиморфных локусов и построены соответствующие генетические карты. Материал таких линий в виде клеточных клонов или образцов ДНК используется, в частности, для анализа сцепления сегрегирующих генов или с вновь описанными полиморфными локусами.

Огромный вклад в систематизацию и обобщение информации о генетических картах хромосом человека, о локализации и функциях отдельных генов и о структуре генома в целом вносят исследования, проводимые на протяжении последних 30 лет под руководством профессора Виктора Мак-Кьюсика в Университете Джона Хопкинса в Балтиморе. Результатом этих исследований является систематическое (с двухгодичным интервалом между последними шестью публикациями) издание энциклопедий под названием: “Менделевское наследование у человека: каталог генов человека и генетических болезней” (“Mendelian inheritance in men. Catalog of autosomal dominant, autosomal recessive, and X-linked phenotypes”). Эти издания содержат сводные данные обо всех картированных генах человека и связанных с ними наследственных болезнях. В настоящее время все эти данные доступны в Internet (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>). Появление и развитие компьютерных баз данных, возможность совмещения различных типов карт позволило перейти на качественно новый уровень анализа картированных последовательностей.

В настоящее время построены генетические карты хромосом человека, мыши, сельскохозяйственных животных и многих других организмов [Захаров, 1979; Захаров и др., 1993; Weissenbach et al., 1992; NIH/CEPH Collaborative Mapping Group, 1992; Shuler et al., 1996; компьютерные базы данных: Genethon (carbon.wi.mit.edu:8000/cgi-bin/contig/phys_map), CEPH-Genethon (www.cephb.fr/bio/ceph-genethon-map.html) GDB (Genome Data Base) (<http://www.gdb.org/>), GenBank (<http://www.ncbi.nlm.nih.gov/>) и др.].

2. Физическое картирование.

В отличие от генетических карт, построенных на основе групп сцепления и дающих статистические расстояния между ДНК маркерами и генами, физическое картирование дает физическое расстояние между маркерами в каждой хромосоме.

К методам физического картирования относят цитогенетическое и рестрикционное картирование, картирование с использованием гибридов соматических клеток, РН-картирование, клонирование в YAC (искусственные хромосомы дрожжей), космидах и плазидах и контиг-картирование на их основе, а также ДНК секвенирование. Использование искусственных хромосом создает основу для проведения генетического картирования, как на хромосомном, так и на субхромосомном уровне.

Основой физического картирования генома является построение физических карт, т.е. определение последовательности физических маркеров вдоль молекулы ДНК. В качестве физических маркеров могут выступать сами гены, анонимные фрагменты ДНК (D-сегменты), STS-маркеры (от англ. «sequenced tagged site») и EST-маркеры (от англ. «expression sequenced tag»), а также точки расщепления ДНК рестриктазами.

Однако при развитии работ по физическому картированию геномов млекопитающих исследователи столкнулись с трудностями при совмещении данных по картированию. Для преодоления этой проблемы в 1989 г. было предложено стандартизовать все обозначения меченых последовательностей ДНК в геноме и включить в себя все типы картированных последовательностей будь то просто картированный сегмент ДНК с неизвестной функцией (D-

сегменты), последовательность с необычными сайтами рестрикции, проба, выявляющая полиморфизм или последовательность, гибридизующаяся с определенным "бэндом" при гибридизации *in situ* [Olson et al., 1989].

Основной особенностью STS-маркеров, а также и основным требованием к ним, является уникальность их нуклеотидной последовательности в геноме. Эти маркеры облегчают перевод различной информации по картированию на единый «язык» STS для анализа и хранения генетической и молекулярной информации, кроме того, оптимизируется процесс насыщения физической карты генома человека маркерами. Созданные на сегодняшний день электронные молекулярно-генетические базы данных значительно облегчают поиск информации по картированию и секвенированию последовательностей любого изучаемого вида, позволяют оценивать степень гомологии и эволюционной связи между геномами различных видов. В настоящее время одно из основных направлений в данной области это перевод всех STS-маркеров на основу ПЦР для более удобного использования. К исходному секвенированному участку ДНК подбирается пара праймеров таким образом, чтобы расстояние между ними не превышало 1 т. п. н. Если участок и праймеры к нему не имеют аналогов в базе данных, содержащих секвенированные на данный момент последовательности ДНК, то праймеры синтезируют и с ними проводят ПЦР, где в качестве матрицы используют ДНК всего генома. Если полученный амплификат представляет собой единственный фрагмент на геном, то эта последовательность может считаться уникальной и может быть использована как STS. Вся информация, касающаяся каждого маркера, хранится в базах данных (GenBank и др.). Она включает в себя сведения о нуклеотидной последовательности праймеров, условия реакции ПЦР, длину продукта амплификации и его нуклеотидную последовательность. Одним из результатов международного проекта «Геном человека» явилось создание такого количества STS-маркеров, которое позволило покрыть участки через каждые 50 т. п. н. вдоль каждой хромосомы человека (на сегодняшний день в базах данных зарегистрировано более 60000 STS-маркеров). Однако пока геном человека насыщен STS-маркерами неравномерно.

Объединение карт возможно благодаря локализации многих клонированных генетических маркеров на физических картах, то есть привязке к определенным хромосомным локусам. Эти локусы служат для взаимосвязи

генетических карт с физическими и дают возможность выяснения соотношения между сантиморганами и парами нуклеотидов.

2.1. Цитогенетическое картирование.

Цитогенетические карты показывают локализацию маркера с точностью до определенной хромосомы, плеча или хромосомного сегмента. Этот тип карт показывает линейный порядок маркеров в хромосоме, расстояние между которыми пропорционально числу пар нуклеотидов, их разделяющих, т.е. относятся к физическим картам генома. По своей разрешающей способности они занимают промежуточное положение между генетическими картами и собственно физическими картами. Цитогенетические карты в большей степени основываются на расположении генов без определения их варибельности, тогда как генетические карты зависят от аллельного полиморфизма локусов. Построение цитогенетической карты облегчает развитие других типов физических карт, а именно, дает «скелет», на котором помещаются маркеры или контиги перекрывающихся клонов.

Для построения цитогенетических карт млекопитающих в настоящее время используется ряд подходов и методов. Определение хромосомной, а в некоторых случаях – субхромосомной, локализации маркеров проводят с использованием гибридов соматических клеток между различными видами млекопитающих. Широкое применение в формировании этого вида карт находит непосредственная гибридизация *in situ* уникальных молекулярных зондов с митотическими хромосомами. К основным методам формирования цитогенетических карт относятся также – хромосомный сортинг (проточная цитометрия), микродиссекции и микроклонирование определенных геномных фрагментов и сравнительное генетическое картирование (сравнительная цитогенетика).

За последние годы метод гибридизации *in situ* получил развитие в модификациях: FISH (гибридизация *in situ* с использованием флюоресцентной метки) [Fan et al., 1990; Hozier et al., 1992] и PRINS (метод, сочетающий гибридизацию *in situ* на метафазных хромосомах специфических праймеров с последующей ПЦР, включающий меченый биотином нуклеотид в продукт

амплификации) [Koch et al., 1989; Troyer et al., 1994]. Этот подход стал основным методом для построения цитогенетических карт.

Разрешение этого метода составляет около 3 млн. п. н., поэтому гибридизация *in situ* является методом картирования с низким уровнем разрешения.

2.2. Картирование на основе гибридов соматических клеток.

В последнее время все большее распространение получает метод картирования геномов млекопитающих с помощью гибридов соматических клеток. Соматические гибриды были получены спонтанно в 1960 году, с тех пор развитие работ по гибридам соматических клеток шло по следующим направлениям: а) поиск безопасных, удобных и эффективных агентов для слияния клеток; б) создание селективных систем, позволяющих выделять гибриды соматических клеток; в) изучение феномена сегрегации хромосом, направления и степени, а также возможности направленной сегрегации в соматических гибридах; г) использование гибридов соматических клеток для создания хромосомных карт млекопитающих [Эфрусси, 1976; Методы генетики..., 1985; Захаров и др., 1993].

Известно, что сформировавшийся при слиянии клеток грызуна и человека межвидовой гибрид спонтанно теряет хромосомы человека. Как правило, утрата хромосом происходит случайным образом, и это позволяет конструировать гибридные линии клеток, в которых содержатся разные хромосомы человека. Корреляция между присутствием конкретной хромосомы человека и экспрессией генетического маркера является основой для отнесения соответствующего гена к определенной группе сцепления. Из 1300 генов человека, картированных к 1996 году, примерно треть локализована на конкретных хромосомах с помощью методов генетики соматических клеток [Allikmets et al., 1996].

Было также показано, что единичные хромосомы донорного вида могут фрагментироваться в клетках хозяина, таким образом, получались гибридные клетки, несущие субхромосомные фрагменты, встроенные в геном хозяина. Были отобраны клоны, несущие небольшие фрагменты (50 - 100 т. п. н.) ДНК человека на фоне генома грызуна, что позволило физически картировать

небольшие сегменты генома с высоким уровнем разрешения, а также определять порядок и относительное расстояние между генами, входящими в одну группу сцепления.

Для тонкого картирования разработаны два метода: перенос генов, опосредованный хромосомой (CMGT) и перенос генов в процессе слияния облученной клетки-донора с необлученным реципиентом (IFGT, от английского – irradiation and fusion gene transfer). Первый способ предполагает инкубацию очищенных митотических хромосом с клетками реципиента в присутствии фосфата кальция. При этом происходит встраивание фрагментов донорных хромосом в хромосомы клетки-реципиента. Для идентификации гибридов, содержащих нужные фрагменты ДНК донора, применяют соответствующие методы селекции. К сожалению, встроенные фрагменты хромосом зачастую претерпевают реорганизацию, кроме того, есть достоверные данные о предпочтительном проникновении в клетку последовательностей из центромерных областей. Эти недостатки не позволяют использовать CMGT для картирования, хотя данный метод весьма эффективен для обогащения специфическими фрагментами хромосом – составной части стратегии клонирования, называемой «обратной генетикой». Фрагменты хромосом человека, появляющиеся в его гамма-облученной клетке, можно сохранить при слиянии этой клетки с клеткой грызуна. Углубленный анализ метода IFGT показал, что полученные фрагменты реорганизованы и для них характерна та же тенденция представлять центромерные области, что и при CMGT [Cox et al., 1990; Walter et al., 1994].

Наиболее широко распространенный пример селекции соматических гибридов – НАТ-селекция (от англ. – hipoxanthine, aminopterin, thymidine) [Littlefield, 1964]. В присутствии аминоптерина (или схожего с ним метотрексата) ингибируется синтез новых предшественников ДНК. Клетки, лишенные фермента тимидинкиназы (ТК), не могут утилизировать экзогенный тимидин и гибнут в присутствии аминоптерина. Аналогично, клетки, лишенные гипоксантин-фосфорибозилтрансферазы (НРРТ), не могут усваивать гипоксантин и, следовательно, нежизнеспособны в присутствии аминоптерина. Соматические гибриды, полученные при слиянии ТК-, НРРТ+-клеток с клетками ТК+, НРРТ-, могут оказаться одновременно ТК+, НРРТ+ [Эфрусси, 1976]. Гибридные клетки с таким генотипом способны расти в присутствии

аминоптерина, если гипоксантин и тимидин добавлены в среду (среда HAT). Распространен вариант HAT-метода (полуселекция), когда один партнер по гибридизации не может размножаться *in vitro*, а другой - является дефицитным по ТК или HPRT.

2.3. Рестрикционное картирование генома человека.

Рестрикционная карта – вид физической карты, на которой указаны расстояния между соседними сайтами расщепления ДНК определенной рестриктазой. Для относительно небольших геномов (бактерии, вирусы) или участков генов маркерами такой карты являются рестрикционные фрагменты определенного размера (обычно до 20 т.п.н.). Рестрикционной картой является также карта, на которой указан порядок следования сайтов рестрикции для одной или нескольких рестриктаз. При анализе более крупных геномов или отдельных хромосом строят макрорестрикционные карты, на которых указан порядок следования сайтов рестрикции крупнощеплящих рестриктаз (узнающих участок в 8 п.н.). При построении макрорестрикционных карт хромосом человека чаще всего используют рестриктазу *NotI* (сайт узнавания – GCGGCCGC) [Zabarovsky et al., 1990; 1991; Забаровский и др., 1994; Allikmets et al., 1996]. Эта рестриктаза удобна для физического картирования, поскольку *NotI*-сайты относительно равномерно распределены в хромосомах человека.

Кроме того, рестриктаза *NotI* имеет для рестрикционного картирования особое значение, так как более 90% STS, полученных секвенированием прилегающих к сайтам рестрикции регионам (400 - 500 п. н.), имеют полную или частичную гомологию с известными генами человека или других организмов, что свидетельствует об очень высоком уровне ассоциации сайта *NotI* с 5'-областями генов [Allikmets et al., 1996; Fedorova et al., 1997; Kashuba et al., 1999; Zabarovsky et al., 2000; Rakhmanaliev et al., 2001]. В настоящее время построены рестрикционные карты хромосомы 3, 21 и других хромосом человека [Zabarovsky et al., 1990; 1991, 2000; Забаровский и др., 1994; Fedorova et al., 1997; Wang et al., 1994; Ichikawa et al., 1993; Kashuba et al., 1999].

2.4. Контиг-карты хромосом человека на основе перекрывающихся клонов геномной ДНК человека.

Методом высокого разрешения является построение контиговых карт перекрывающихся клонов. Контиговая карта или контиг (от contiguous - протяженный) – включает в себя набор перекрывающихся клонированных фрагментов геномной ДНК в комплексе с информацией о порядке их перекрывания. Применяя методы гибридизации с пробами ДНК, локализация которых известна, и фингерпринтинга (fingerprinting), можно локализовать клонированный фрагмент. В настоящее время создание контигов является общепринятым подходом к построению физической карты генома человека. Результатом картирования по этому методу является не только информация об относительном расположении маркеров и расстоянии между ними (выраженном в парах нуклеотидов), но и непосредственная доступность каждого картированного фрагмента ДНК для последующих структурных исследований, вплоть до расшифровки нуклеотидных последовательностей фрагментов ДНК и выявления их функциональной роли.

Для изучения геномов млекопитающих используют набор клонов (бактерий, бактериофагов или дрожжей), которые содержат рекомбинантную ДНК, представляющую в совокупности весь геном. Такой набор клонов называется клонотекой или библиотекой клонов.

Современные методы клонирования позволяют большие хромосомы млекопитающих, фрагментированные на небольшие отрезки ДНК, встраивать в бактериальные, фаговые или дрожжевые векторы, которые вводятся в клетки "хозяина", которые, в свою очередь, можно выделить и использовать как индивидуальные клоны для дальнейших исследований. Переход к молекулярному уровню генетического анализа – анализу библиотек клонов позволяет упростить и унифицировать процесс картирования. Применение клонотек делает возможным тонкое картирование генома и поиск в нем различных, в том числе экспрессирующихся, последовательностей. Немаловажной особенностью клонотек является также их универсальность, т.е. возможность их использования различными научными коллективами, а следовательно, появляется возможность координирования результатов экспериментов, проводимых разными лабораториями.

Первоначально библиотеки делались на основе плазмидных векторов для клонирования в *E. coli*. Максимальный размер вставки достигал 10 kb, что явно недостаточно, например, для клонирования гена человека, так как длина участка хромосомы, соответствующего одному гену составляет в среднем 40 – 50 kb. Затем были предложены векторы на основе фага лямбда: вектор замещения лямбда Charon (около 20 Kb вставки), космидные векторы – плазмиды, несущие cos-последовательности, распознаваемые компонентами системы упаковки фага лямбда. Размер вставки в космидных векторах в среднем около 40 Kb. Эти векторы представляют собой удобный инструмент для анализа больших фрагментов геномов, картирования хромосом и клонирования генов [цит. по: Аитова, 1997].

В 1987 г. был описан метод получения так называемых искусственных дрожжевых хромосом (YAC-клоны), которые способны включать в себя крупные фрагменты чужеродной ДНК, например ДНК человека, и размножать эти фрагменты одновременно с размножением дрожжевых клеток. Фрагменты, которые можно было бы таким образом клонировать, составляли 0,2 - 0,3 млн. п. н., и при размере генома человека примерно в 3 млрд. п. н. около 10000 YAC-клонов могли "покрыть" весь геном человека, что резко упростило бы задачу картирования этих фрагментов [Burke et al., 1987].

Позже другой группе исследователей удалось усовершенствовать процедуру получения YAC-клонов, способных нести в себе фрагменты чужеродной ДНК размером до 1,2 млн. п. н. [Green, Olson, 1990]. Это позволило сократить число клонов, покрывающих геном человека, до нескольких тысяч.

Их упорядоченное расположение вдоль хромосомы представляет хотя и трудоемкую, но реальную задачу, так как чем крупнее клонированные фрагменты ДНК человека, тем меньше требуется усилий и времени, чтобы установить их взаимное расположение. В настоящий момент пользуются единичными эталонными клонотеками: CEPH [Chumakov et al., 1992], ICRF [Larin et al., 1991].

Тем не менее, необходимо учитывать тот факт, что техника клонирования в YAC-клонах, несмотря на многочисленные преимущества, имеет ряд слабых сторон. Это в первую очередь касается подверженности YAC-клонов физической нестабильности (перестройкам, делециям) и химеризму. Известно, что до 30 - 50 процентов клонов большинства YAC-библиотек являются

химерными, т.е. они содержат ДНК из двух и более несмежных областей генома [Barilott et al., 1991; Cohen et al., 1993]. Кроме того, контиговая карта, состоящая только из YAC-клонов, не обеспечивает доступности клонированных фрагментов ДНК, пригодных для тонкого структурного анализа и секвенирования. Вследствие этого, для более детального картирования специфических объектов, таких как CpG островки, гены, области хромосомных перестроек, горячие точки рекомбинации, а также для секвенирования генома необходимы контиги, построенные из относительно небольших перекрывающихся клонов, например, космид. Эти требования могут быть выполнены при помощи объединения двух контиговых карт - YAC и космидной – путем гибридизации YAC-зондов с космидами представительных хромосомспецифических библиотек, что позволяет создать физическую карту в виде групп космид, расположенных в известном порядке [Holland et al., 1993; Nizetic et al., 1994; Chumakov et al., 1992].

Основным достоинством использования контигов YAC-клонов как основы для построения космидных контигов является необходимость картирования лишь ограниченного числа космид в пределах каждого YAC-клона, уже привязанного к определенному расположению на карте. Порядок и ориентация отдельных космидных контигов (как друг относительно друга, так и всех вместе вдоль хромосомы) определяется последовательной гибридизацией с несколькими YAC-клонами, по-разному их перекрывающимися. Потенциальная разрешающая способность такой карты определяется количеством интервалов между концами YAC-клонов, приходящимися на исследуемый участок хромосомы. Поэтому в последнее время в построение первичного YAC-контига вовлекаются библиотеки YAC-клонов с относительно малыми средними размерами вставки - миди-YAC-клоны (размер вставки 0,2 – 0,4 млн. п. н.); таким путем увеличивается количество интервалов перекрывания. Для того, чтобы проверить соответствие построенного YAC-контига истинному расположению участков генома применяется "Bottom Up"-подход, когда используется *in situ* гибридизации, а чаще полимеразная цепная реакция для сверки с последовательностью известных физических маркеров на хромосоме.

2.5. Картирование генома на основе радиационных гибридов соматических клеток (RH-картирование).

Накопленные современной радиобиологией сведения о влиянии гамма-лучей на живые клетки находят применение в самых разных областях биологии. В частности ионизирующее облучение клеток было использовано для создания нового метода картирования геномов человека и животных с высоким уровнем разрешения - метода радиационного картирования (RH-картирования).

Идея использования совместной сегрегации генов в облученных клеточных гибридах для картирования генов принадлежит Госсу и Харрису [Goss, Harris, 1975]. Метод заключается в следующем: донорные клетки подвергают воздействию высоких доз гамма-облучения, способного вызывать разрывы хромосом и приводить к различным хромосомным перестройкам. Вероятность разрыва между двумя сцепленными генами тем больше, чем больше расстояние между ними. В качестве облучаемого партнера используют как диплоидные клетки исследуемого вида, так и монохромосомные клеточные линии. Затем летально облученные клетки гибридизуют с нормальной необлученной клеточной линией грызуна, в образовавшихся гибридах фрагменты хромосом донора оказываются встроенными в хромосомы реципиента (грызуна), что обеспечивает стабильную сегрегацию гибридных хромосом в митозе и сохранение встроенных фрагментов. Количество встроенных фрагментов может составлять до 30% генома клетки реципиента. Для формирования панели радиационных гибридов соматических клеток (RH-панели) отбирается минимальное количество клонов, в которых представлены все последовательности генома донора (например, человека) с достаточной степенью перекрытия (с 6-7-кратным перекрытием генома). Затем методом полимеразной цепной реакции (ПЦР) проводится скрининг RH-панели с использованием маркеров с уже известной локализацией и на их основе формируется так называемая рабочая рамка маркеров. Локализация новых детектируемых маркеров устанавливается относительно маркеров рабочей рамки. ДНК, изолированная из панели независимых клонов радиационных гибридов, обеспечивает картирование STS маркеров и анализ расстояния между ними в геноме человека. Частота вызванных радиацией разрывов между двумя маркерами используется как мера расстояния, и положение маркера

определяется тем же способом, что и при построении карт сцепления.

Вероятность возникновения разрывов и других перестроек хромосом в живых клетках под воздействием ионизирующей радиации описывается распределением Пуассона, являющегося частным случаем биномиального распределения. При Пуассоновском распределении вероятность возникновения каждого конкретного события очень мала, в то время как общее число наблюдаемых событий очень велико и стремится к бесконечности. Многие расчеты в современной радиобиологии основываются на анализе пуассоновского распределения, в частности при применении так называемой теории мишени. Этот же подход использован и для теоретического обоснования RH-картирования. Метод RH-картирования основывается на следующих перечисленных ниже предпосылках:

1) согласно распределению Пуассона разрывы хромосом происходят случайным образом и независимо друг от друга;

2) встраивание фрагментов хромосом реципиента в хромосомы клеток донора также происходит случайным образом и независимо друг от друга, что обеспечивает присутствие в панели гибридных клеток всех последовательностей генома реципиента;

3) вероятность образования разрывов между двумя маркерами зависит только от расстояния между ними и дозы облучения, что позволяет оценивать физические расстояния между маркерами [Cox et al., 1990; Boehnke et al., 1991; Lange, Boehnke, 1992; Lange et al., 1995; Walter, Goodfellow, 1993; Walter et al., 1994; Newell et al., 1998].

Существенным является также тот факт, что типирование панели радиационных гибридов соматических клеток проводится только по принципу наличия или отсутствия конкретных маркеров, независимо от числа их копий в гибридных клетках.

Для анализа радиационных гибридов создан специальный статистический аппарат, основанный на разработке вероятностных моделей, описывающих процессы разрыва хромосом и сохранения фрагментов в гибридных клетках [Cox et al., 1990; Boehnke et al., 1991; Lange, Boehnke, 1992; Lange et al., 1995]. Основой этого статистического аппарата является метод, который был разработан для картирования генома человека с использованием популяционно-генеалогических данных, то есть он является традиционным гибринологическим

методом генетики. Суть его заключается в статистическом анализе результатов скрещивания особей и использовании этих результатов для получения оценок истинных вероятностей рекомбинации между исследуемыми локусами. В силу сложившейся под влиянием англоязычной литературы традиции, этот метод принято называть методом lod score, хотя, может быть, правильнее было бы говорить о методе изучения сцепления с помощью оценок (score), представляющих собой логарифм отношения шансов (lod - аббревиатура от *logarithm of the odds ratio*, цит. по Захаров и др., 1993). В настоящее время этот метод чаще называют методом максимального правдоподобия [Boehnke et al., 1991; Lange et al., 1995; Walter, Goodfellow, 1993; Walter et al., 1994; Newell et al., 1998].

Этот статистический метод позволяет использовать уже имеющиеся неполные карты: специфические радиационные гибриды отбираются по точкам разрывов, которые обозначены уже известными маркерами. Идентификация хромосомных фрагментов картируемого вида в исследуемых гибридах позволяет ориентировать синтенную группу относительно хромосомы и определять точное положение маркеров в хромосоме. Создание таких мультимегабазных хромосомных фрагментов (от 5×10^2 до 2×10^3 т. п. н.), по крайней мере, в 10 раз больших, чем фрагменты, клонированные в искусственных хромосомах дрожжей (мега-YACs), дает новые возможности для физического картирования хромосом с высоким уровнем разрешения и получения сравнительных карт геномов млекопитающих. Ценность RH-карт заключается в том, что они являются связующими между цитогенетическими картами и физическими, между картами контигов и STS-картами [Gyapay et al., 1996; Schuler et al., 1996; Stewart et al., 1997; Deloukas et al., 1998].

Преимуществом картирования с использованием радиационных гибридов соматических клеток, является то, что, используя различные дозы гамма-облучения можно конструировать панели радиационных гибридов разного разрешения, а также создавать панели, содержащие фрагменты индивидуальных хромосом [Cox et al., 1990; Stewart et al., 1997; Beasley et al., 1997; Walter et al., 1994]. В отличие от других стратегий картирования, когда определяют положение основных, или каркасных, маркеров и распределяют все остальные маркеры относительно этих локусов, используя метод RH-картирования сначала определяют наиболее вероятный порядок всех маркеров на хромосоме, а затем

рассматривают 20-35 групп сцепления маркеров, одновременно определяя наиболее вероятный порядок маркеров в пределах каждой группы сцепления. Полученный порядок сравнивают с порядком уже охарактеризованных маркеров на карте сцепления. Сравнение карт сцепления и карт, построенных с помощью YAC-клонов, с картами радиационных гибридов показывает, что ошибка при их совмещении не более 5%. А анализ 322 случайных кДНК показал, что RH-карта покрывает бóльшую часть генома человека.

Для картирования генома человека было сконструировано три коммерческие RH-панели с разным уровнем разрешения: GeneBridge 4 [Walter, Goodfellow, 1993; Walter et al., 1994], Stanford G3 panel [Stewart et al., 1997] и TNG4 [Beasley et al., 1997] (табл. 1). Все три панели созданы методом облучения диплоидных клеток человека с последующим их слиянием с нормальными клетками китайского хомячка.

Таблица 1. Характеристика стандартных панелей радиационных гибридов соматических клеток: GeneBridge 4 (G4), Stanford G3 panel (G3) и TNG4 (цит. по [Newell et al., 1998]).

| RH-панель | Число клонов | Доза облучения, Rad | Частота разрывов на 1 млн.п.н. | Максимальный размер фрагмента, млн.п.н. | Уровень разрешения, т.п.н. | Плотность маркеров на 1 млн.п.н. |
|-----------|--------------|---------------------|--------------------------------|---|----------------------------|----------------------------------|
| G4 | 93 | 3000 | 0,04 | 25 | 1000 | 0,4 |
| G3 | 83 | 10000 | 0,19 | 5,2 | 500 | 0,17 |
| TNG4 | 90 | 50000 | 3,6 | 0,28 | 50 | - |

Минимальный уровень разрешения RH-панели определяется как минимальная дистанция между маркерами (>0), которая может быть установлена с помощью данной панели в единицах среднего числа разрывов (R , ray). (Дистанция, равная 0, наблюдается, когда разрешаемые маркеры идентичны.) Единица расстояния 1 cR (centiRay) соответствует 1% частоты разрывов между двумя маркерами после облучения гамма-лучами и является функцией дозы облучения. Для каждой хромосомы соотношение между величинами cR и рекомбинационной единицей расстояния cM варьирует. Для хромосомы 3 человека, например, это соотношение в среднем составляет 4 cR на 1 cM. Соотношение величины cR с физическими расстояниями, выраженными в т.п.н., определяют на основании карты маркеров рабочей рамки и физических параметров генома человека в соответствии с оценками Мортонна [Morton, 1991].

Реальный уровень разрешения конкретной RH-панели зависит не только от способа ее получения, но и от степени ее насыщенности маркерами (плотности маркеров). На сегодняшний день достаточно высокая плотность маркеров получена только для панели GeneBridge 4. Для RH-панели Stanford G3 плотность маркеров составляет в среднем 1 маркер на 5,7 млн.п.н., а максимальное расстояние, которое может быть измерено между маркерами равно 1R, что соответствует 3,6 млн.п.н. (табл. 1). Это может создавать определенные трудности при локализации маркеров, лежащих в областях хромосом с низкой плотностью маркеров и значительно удаленных от маркеров рабочей рамки. К таким областям чаще всего относятся прицентромерные и прителомерные области хромосом. Панель TNG4 еще недостаточно насыщена маркерами и пока не может быть использована для тотального картирования.

В последнее время RH-картирование получает все большее распространение и применяется исследователями в различных целях: для тотального картирования [Guaray et al., 1996; Schuler et al., 1996; Stewart et al., 1997; Deloukas et al., 1998], для составления карт индивидуальных хромосом человека на основе хромосом-специфичных RH-панелей [Naylor et al., 1994], для установления мест локализации конкретных генов или STS [Jacquemin et al., 1999; Сулимова и др., 2000]. Это связано с такими преимуществами RH-картирования как отсутствие делетирования или перестроек клонируемого материала, поскольку размножение фрагментов генома человека происходит в клетках млекопитающего (клетки китайского хомячка), присутствие в RH-панели всех последовательностей генома, в том числе и тех, которые не могли быть клонированы в мега-YAC-клонах. Кроме того, RH-картирование – менее трудоемкий метод, поскольку весь геном человека представлен в панели ДНК из 93 клеточных линий радиационных гибридов, являющихся коммерческим препаратом, что облегчает проведение скрининга, а результат о локализации картируемого маркера на RH-карте можно получить через Интернет в течение 1 - 2 часов. К сожалению, в нашей стране этот метод не получил пока заслуженного признания и исследования с использованием RH-картирования сводятся к единичным [Свищева, 1999; Сулимова и др., 2000; Sulimova et al., 2001; Rakhmanaliev et al., 2001].

Использование унифицированных коммерческих RH-панелей (GeneBridge 4 и Stanford G3) в исследованиях лабораторий всего мира позволило за короткий

срок получить информативные RH-карты хромосом человека и совместить их с другими типами карт (генетическими, цитогенетическими, YAC-картами и др.), т.е. получить интегрированные карты хромосом человека. В настоящее время данный подход эффективно используется для построения физических карт геномов других млекопитающих: мыши, крысы, коровы, свиньи, лошади, бабуина, собаки, а также рыбки *Danio regia* и курицы (<http://compgen.rutgers.edu/rhmap/>) [см. обзор: Жданова, 2002].

Радиационное картирование используется также для верификации данных по секвенированию генома человека, выявлению неправильной состыковки контигов и несеквенированных районов генома [Olivier et al., 2001; Venter et al., 2001].

Таким образом, метод RH-картирования является на сегодняшний день самым совершенным и широко применяемым в мире методом физического картирования, и может быть использован для локализации генов, STS-маркеров и функциональных последовательностей. В данной работе метод RH-картирования использовался для локализации мест интеграции HPV16 в геном клеток опухолей шейки матки. Локализация мест интеграции ДНК вируса на хромосомах человека и анализ прилегающих геномных последовательностей имеет большое значение для понимания того, как интегративный вирусный геном может нарушать генетическую программу клетки.

3. Предсказание генов человека с помощью компьютерных методов анализа последовательностей биополимеров.

Лавинообразное накопление информации о нуклеотидных и аминокислотных последовательностях различных организмов привело к возникновению такой важной дисциплины как биоинформатика (или вычислительная биология).

Первые работы по компьютерному анализу последовательностей биополимеров появились еще в 60-70-х годах, однако, формирование вычислительной биологии как самостоятельной области началось в 80-х годах после развития методов массового секвенирования ДНК. Можно выделить пять направлений вычислительной биологии: непосредственная поддержка

эксперимента (физическое картирование, создание контиг и т.п.), организация и поддержание банков данных, анализ структуры и функции ДНК и белков, эволюционные и филогенетические исследования, а также собственно статистический анализ нуклеотидных последовательностей. Разумеется, границы между этими направлениями в значительной мере условны: результаты распознавания белок-кодирующих областей используются в экспериментах по идентификации генов, одним из основных методов предсказания функций белков является поиск сходных белков в базах данных, а для осуществления детального предсказания клеточной роли белка необходимо привлекать филогенетические соображения [цит. по: Гельфанд, Миронов, 1999].

3.1. Молекулярно-биологические базы данных.

В настоящее время экспериментальные данные различных авторов хранятся в электронных базах данных. Это позволяет использовать компьютеры для их сравнения, анализа, выявления новых закономерностей. Среди многих сотен баз данных молекулярно-биологической направленности можно выделить содержащие первичную биологическую информацию. К ним относятся базы данных:

- а) геномов (человека и других видов), в них содержится информация о нуклеотидных последовательностях (EMBL, GeneBank, DDBJ), полиморфизме, мутациях, наследственных болезнях (SNP, OMIM) и генетических картах;
- б) белковые: аминокислотные последовательности всех нативных белков (SwissProt, PIR, TrEMBL), пространственные структуры белков, ДНК, РНК и многомолекулярных комплексов (PDB, NDB);
- в) базы данных о строении РНК: первичная структура (EMBL), вторичная структура и третичная структура РНК (PDB, NDB)
- г) информация о экспрессии генов (dbEST – входит в EMBL, SAGE) олигонуклеотидные биочипы и 2D электрофорез;
- д) Medline – банк статей по биологии, медицине и др. наукам о живом.
- е) и многие другие.

Можно выделить следующие основные характеристики баз данных: основные структурные единицы базы данных; структура базы данных (таблицы, записи, поля); источники новых поступлений, правила внесения новой записи в

базу данных; объем и статистика данных в базе, регулярность обновления; правила корректировки информации в записях и форматов базы данных; доступ, интерфейс, язык запросов, связи с другими базами данных; оценка достоверности информации в базе данных.

Молекулярно-биологические базы данных являются неотъемлемым элементом современных биологических исследований. Они позволяют обмениваться данными, сопоставлять полученные результаты и публиковать их.

3.1. Строение генов и подходы к их предсказанию в последовательностях генома человека.

Ген можно определить как единицу наследственной информации, занимающую определенное положение в геноме или хромосоме и контролирующую определенную функцию в организме [Жимумев, 2000].

Гены эукариот имеют довольно сложную структурную организацию. Белок-кодирующие участки генов эукариот (экзоны) разделены некодирующими интронами. Тип экзон – интронной структурной организации обнаружен для большинства эукариотических генов (в том числе и для многих генов органелл), а также для некоторых генов РНК- и ДНК-содержащих вирусов. Интроны отделяются от экзонов канонической парой нуклеотидов: GT на 5'-конце и AG на 3'-конце. Число интронов и их внутренняя локализация характерны для каждого гена. Довольно часто в пределах интрона одного гена может находиться последовательность другого гена. В табл. 2 приведены усредненные параметры генов человека.

Инициация транскрипции генов происходит в промоторном участке, находящемся в 5'-области гена. Поскольку РНК-полимераза эукариот не может самостоятельно инициировать транскрипцию, для ее активации необходимо большое число белков, называемых факторами транскрипции (ФТ). В промоторном участке гена содержатся последовательности, связывающиеся с ФТ, что способствует формированию комплекса РНК-полимеразы и ФТ, необходимого для начала транскрипции. Следует отметить, что любой промотор может находиться под влиянием регуляторных последовательностей, расположенных за тысячи пар нуклеотидов от промотора (энхансеры).

Таблица 2. Усредненные характеристики генов человека [International Human Genome Sequencing Consortium, 2001].

| | |
|--------------------------------------|-----------------|
| Внутренний экзон, длина | 122 п.н. |
| Число экзонов | 7 |
| Интрон, длина | 1023 п.н. |
| 3' UTR, длина | 400 п.н. |
| 5' UTR, длина | 240 п.н. |
| Кодирующая последовательность, длина | 1100 п.н. |
| Белковый продукт, длина | 367 аминокислот |
| Протяженность в геноме | 14 т.п.н. |

Идентификация новых генов в геномной ДНК человека очень сложна. Это связано с несколькими особенностями генома человека: во-первых, процент кодирующих последовательностей невелик (~1,5% генома); во-вторых, внутренние экзоны могут быть очень малы (до 4 п.н. [Zang, 1998]); и, в-третьих, некоторые гены могут иметь много продуктов за счет альтернативного сплайсинга.

Современные методы предсказания генов основаны на трех подходах: а) наличие мРНК или EST; б) наличие сходства изучаемой последовательности с последовательностью известного гена; в) предсказание генов *de novo*, с помощью программ, основанных на HMM (hidden Markov model), примерами таких программ могут служить Genscan, Genie и FGENES.

Первый подход, основанный на экспериментальных данных, может приводить к ошибкам, основанным на артефактах экспериментальной работы. В основном загрязнение это баз данных EST несплайсированной мРНК или геномной ДНК, сходной по первичной структуре с мРНК. Второй подход позволяет более корректно предсказывать новые гены на основе сходства их последовательности с уже известными генами. Однако есть вероятность, что предсказанный этим методом ген окажется псевдогеном. Чувствительность и специфичность третьего метода – предсказание генов с использованием компьютерных программ – сильно зависит от отношения «сигнал – шум». Для генома человека корректное предсказание индивидуальных экзонов возможно в 70% случаев, а предсказание всех кодирующих экзонов – в 20% случаев [International Human Genome Sequencing Consortium, 2001]. Этими трудностями

обуславливается тот факт, что несмотря на впечатляющие успехи в расшифровке первичной нуклеотидной последовательности, аннотирование генома человека далеко от завершения. Остаются не определенными функции и структура более 2/3 генов человека. Структура и предположительные функции большинства генов на настоящий момент определены *in silico* [Venter et al., 2001]. И хотя полученные компьютерными методами данные нельзя принимать как абсолютно достоверные, этап работы по аннотированию генома человека *in silico* является основой для дальнейших экспериментальных исследований и значительно облегчает планирование и постановку эксперимента.

МАТЕРИАЛЫ И МЕТОДЫ

Реактивы и оборудование, использованные в работе.

В работе были использованы следующие реагенты: коммерческий препарат панели радиационных гибридов GeneBridge 4 RH-Panel (фирма “Research Genetics, Inc.”); набор реагентов для выделения ДНК *DIAtom*TM DNA Prep 200 (фирма “Biokom”, Россия); набор реагентов для выделения РНК Trizol RNA Prep 100 (фирма “IsoGen”, Россия); набор реагентов для амплификации ДНК *Maxi-Taq*TM (фирма “Biokom”, Россия); набор реагентов для постановки ОТ-ПЦР *GenePak*TM RT-PCR Universal (фирма “IsoGen”, Россия); агароза “Chemapol”; плаزمида pBR322, рестриктазы и маркер молекулярных весов 100bp DNA Ladder Plus фирмы “MBI Fermentas”; маркер молекулярных весов M50 (фирма “IsoGen”, Россия).

Панель радиационных гибридов соматических клеток (человек/хомяк) GeneBridge 4 RH-Panel (“Research Genetics Inc.”, США) содержит 93 образца ДНК индивидуальных гибридных клонов (25 нг/мкл) и два контрольных образца (ДНК линии фибробластов человека – HFL и ДНК китайского хомячка – A23). Данная панель полностью перекрывает геном человека (6 – 7 кратное перекрытие). Радиационные гибриды были получены путем слияния летально облученной (3000 Rad) линии клеток фибробластов человека (HFL) в качестве донора с тимидин-зависимой линией клеток китайского хомячка (A23) в качестве реципиента [Walter et al., 1994].

В работе были использованы пять пар олигонуклеотидных праймеров, ранее подобранные к 5 последовательностям, прилегающим к местам встраивания вируса папилломы человека тип 16 в геном человека, и синтезированные ЗАО “Синтол” (праймеры серии INT). Праймеры к мозгоспечифичной последовательности Ghfb (G2 и G4), синтезированные А.С. Володиным (Институт молекулярной генетики РАН). Праймеры к фрагментам кДНК гипотетических генов были подобраны «вручную» и с использованием программы “OligoCalculator” (синтезированы в НПФ “Литех”). Последовательности использованных в работе праймеров и температуры отжига, определенные с помощью соответствующих программ, представлены в табл. 3.

Таблица 3. Характеристика олигонуклеотидных праймеров, использованных в работе.

| Название | Последовательность праймеров (5'-3') | Температура и время отжига, °С/сек | Размер продукта (п.н.) |
|-----------|---|------------------------------------|-------------------------------------|
| INT254 | F: CTTTCAGCAAAGGATGTCATTG R: GCAAGGCTTATTAAGCTTCAGC | 58 / 50 | 350 |
| INT259 | F: ATCTACAGACAGCATGGTACGG R: TCAGAACACAACAAAAGGATGC | 63 / 45 | 213 |
| INT431 | F: TCCAAAAAATGAGGCAATCTC R: GGACTGAGGGAATAAACGACAC | 62 / 50 | 193 |
| INT407 | F: AGCAGGAGATGACAAGATGTTG R: CACAAGTAGCTGAGACCACAGG | 57 / 35 | 150 |
| INT290 | F: AAGTTCCACCATTCAAGACCAG R: TCATGATATCCTTGGTGTCTG | 56 / 45 | 272 |
| Ghfb | G2: CACTTGGAGCCATCGGGTAGC G4: GAATTGGCTTCTCACCAAAGGC | 64 / 35 | 186 |
| cmgc30 | F: GGCCTGGGAGGAGATGGCTG R: GGTGAGGAGCCGCTGGCGTT | 65 / 50 | 275 |
| LOC161154 | F1: GCACTGTGCTGGGCCAGGAG F2: GGTGCTTGCTGCCATCTGTTTC R: TGAAATCCAAACAAGCCAGAGC | 63.9 65.7 64.6 | для РНК F1/R – 712 F2/R – 521 |
| KIAA1808 | F1: ACTCCGACCTCACGGTC F2: CAGCTTTCACCTACAATACA R: GCTTCTTCTTAAGGTCATT | 49.8 50 54.9 | для РНК F1/R – 578 F2/R – 204 |
| KIAA0887 | F: AACGATGCCAAAAGGGAG R1: ATCCCTCAGGTTTGTGTTGTAG R2: GATTGGGCCACTCCTCTG | 56.8 54.9 56.5 | для РНК F/R1 – 166 F/R2 – 778 |
| LOC151128 | F1: TACTTAGAAAGCAAAATGGAG F2: AGTAATAGTGTCAGAGAAATC R: ATGAAGGAAGAGAAATAGAAG | 50 50 50 | для РНК F1/R – 329 F2/R – 228 |

Реакцию ПЦР проводили на термоциклерах PTC-100™ “MJ Research, Inc.” (USA) и Amply-4L (фирма “Биокон”, Россия). Также в работе были использованы камеры для вертикального и горизонтального электрофореза, источники питания постоянного тока Б5–50, настольные микроцентрифуги (фирма “Eppendorf”), термостаты, вортекс, наборы автоматических пипеток (фирмы “Biohit” и “High Tech Lab.”), трансиллюминатор с длиной волны 254 нм и фотокамерой и прибор “Gel Doc 1000” (фирма “Bio-Rad Laboratories, Inc.”) для сканирования и фотографирования гелей в ультрафиолетовом свете.

Нуклеотидные последовательности маркеров INT и праймеры к ним, а также праймеры к последовательности cmgc30 любезно предоставлены лабораторией молекулярной биологии вирусов Онкологического научного

центра им. Н. Н. Блохина. Праймеры к мозгоспецифичной последовательности Ghfb любезно предоставлены отделом молекулярных основ генетики человека Института молекулярной генетики РАН.

Выделение ДНК из цельной крови.

ДНК выделяли из цельной крови человека и китайского хомячка с использованием коммерческого набора реагентов *DIAtomTM DNA Prep200* согласно прописи, представленной изготовителем (фирма “IsoGen”, Россия). Выделение ДНК проводили под ламинаром во избежание загрязнения помещения и последующей контаминации инкубационной смеси чужеродной ДНК при постановке полимеразной цепной реакции.

Набор реагентов *DIAtomTM DNA Prep200* основан на использовании *Лизирующего реагента* с гуанидинтиоционатом, который предназначен для лизиса клеток, солюбилизации клеточного дербиса, а также денатурации клеточных нуклеаз. В присутствии *Лизирующего реагента* ДНК активно сорбируется на сорбенте *NucleoSTM*, затем отмывается от белков и солей спиртовым раствором. ДНК элюировали из сорбента *ЭкстраГеномTM* (10% смесь ионообменников (типа Chelex), 0,02% красителя Orange G и 0,01% тритона X-100) или дистиллированной водой. Хранили при -10°C .

Выделение РНК из цельной крови.

РНК выделяли из цельной крови человека с использованием коммерческого набора реагентов *Trizol RNA Prep 100*. Выделение РНК проводили под ламинаром согласно прописи, представленной изготовителем (фирма “IsoGen”, Россия).

Набор реагентов *Trizol RNA Prep 100* основан на использовании *Trizol* реагента с гуанидинтиоционатом и фенолом. Набор позволяет выделять из свежего биологического материала высокомолекулярную и нативную РНК. Выделенную РНК хранили при -70°C .

Проведение полимеразной цепной реакции с использованием термостабильной ДНК-полимеразы Taq-pol [Saiki et al., 1988, с модификациями].

Составление инкубационной смеси проводили в стерильных условиях под ламинаром.

Аmplификацию проводили согласно прописи, представленной фирмой-изготовителем набора для ПЦР Max-i-Taq™ (фирма “IsoGen”, Россия). Реакцию проводили в объеме 10 мкл. Инкубационная смесь содержала 1 мкл 10-кратного буфера для полимеразы (10-кратный буфер: 670 mM Трис-НСl, 160 mM (NH₄)₂SO₄, 0,1% Tween 20), прямой и обратный праймеры (по 50 нг каждого), по 0,2 mM каждого дезоксинуклеозидтрифосфата (dNTP), 2,0 – 2,5 mM MgCl₂, 20-25 нг геномной ДНК или 8 – 10 нг ДНК клона радиационных гибридов и 1 ед. активности Taq-полимеразы. В качестве положительного контроля использовали ДНК, выделенную из лимфоцитов крови человека, отрицательным контролем служила вода. Для всех пар праймеров, использованных для скрининга панели радиационных гибридов GeneBridge4, проводили ПЦР с геномной ДНК хомячка для определения наличия отсутствия в геноме китайского хомячка последовательностей, гомологичных исследуемым последовательностям человека.

Для повышения чувствительности и специфичности амплификации при проведении полимеразной цепной реакции, применяли ПЦР с “горячим стартом” (“Hot Start” PCR). С этой целью для механического разделения компонентов реакционной смеси до начала процесса ПЦР вместо обычно используемого жидкого минерального масла использовали твердое парафиновое масло “DinOil” в качестве барьера, первоначально разделяющего реакционную смесь на две неактивные части (в нижней – ДНК, dNTP и MgCl₂; в верхней – буфер, праймеры, Taq-полимераза и деионизованная вода), а далее играющего роль барьера против испарения смеси. При высокой температуре, 75 – 80⁰С, парафиновый слой плавится и всплывает, а две искусственно разделенные фазы одной реакционной смеси перемешиваются и реакция стартует при 80⁰С. Амплификацию начинали с денатурации (95⁰С, 2 мин.) и заканчивали на стадии синтеза (74⁰С, 8 мин.). Всего проводили 30 – 35 циклов (денатурация – 94⁰С, 0,6-1 мин.; отжиг – 0,3-0,8 мин., температура указана в таблице 2; синтез – 73⁰С, 0,8-1,5 мин.).

Проведение обратной транскрипции с использованием обратной транскриптазы и последующей полимеразной цепной реакции с использованием термостабильной ДНК-полимеразы.

Составление инкубационной смеси проводили в стерильных условиях под ламинаром.

Реакцию обратной транскрипции сопряженную с ПЦР проводили согласно прописи, представленной фирмой-изготовителем коммерческого набора реагентов GenePak™ RT-PCR Universal (фирма “IsoGen”, Россия).

Набор реагентов GenePak™ RT-PCR Universal представляет собой лиофилизированные сухие смеси, готовые для проведения реакции обратной транскрипции и ПЦР. Каждая пробирка для проведения реакции обратной транскрипции содержит готовую смесь необходимых реагентов: 100 ед. обратной транскриптазы M-MLV (Promega), 20 ед. ингибитора РНК-аз RNasin (Promega), смесь гексануклеотидных праймеров со случайной последовательностью, дезоксинуклеозидтрифосфаты и оптимизированную буферную систему. В каждую пробирку необходимо добавить 5 мкл исследуемой РНК и 5 мкл *OT Растворителя*. Для проведения ПЦР отдельная пробирка с готовой смесью одержит ингибированную OligoTaq™ ДНК полимеразу, переходящую в активную форму при “горячем старте”, и оптимизированную буферную систему для проведения одной стандартной ПЦР. В эту пробирку добавляли 13 мкл *OT Растворителя* и 2 мкл смеси праймеров. Амплификация проводится в той же пробирке, что и реакция обратной транскрипции без дополнительных процедур переноса ревертазной реакционной смеси из одной пробирки в другую. Конечные концентрации компонентов в ПЦР составляют: OligoTaq™ – 1 ед., dNTP – 200 мкМ, Mg²⁺ - 2,5 мМ. Объем реакционной смеси составлял около 25 мкл

Реакцию обратной транскрипции проводили при 37⁰С в течении 40-50 минут. ПЦР начинали с предварительной денатурации (95⁰С, 1 мин.), затем следовали 35 циклов, состоящие из стадии денатурации (94⁰С, 0,5 мин.), отжига (49⁰С, 0,5 мин.) и синтеза (73⁰С, 0,8 мин.). Заканчивали амплификацию стадией синтеза (73⁰С, 5 мин.).

Приготовление 6%-ного (9%-ного) полиакриламидного геля

[Маниатис и др., 1984].

Готовили 30%-ный раствор акриламида (29 г акриламида и 1 г N,N'-метиленабисакриламида растворяли в 90 мл дистиллированной воды и доводили объем до 100 мл). Для приготовления 6%-ного (9%-ного) полиакриламидного геля отбирали 6 мл (9 мл) 30%-ного акриламида, добавляли 3 мл 10-кратного трис-боратного буфера (ТВЕ), 200 мкл свежеприготовленного 10%-го персульфата аммония (100 мг персульфата аммония растворить в 1 мл дистиллированной воды), доводили до объема 30 мл дистиллированной водой, добавляли 50 мкл TEMED, перемешивали и заливали в камеру для вертикального электрофореза. Полимеризация геля проходила за 25 – 35 мин. при комнатной температуре.

Приготовление 2%-ного агарозного геля [Маниатис и др., 1984].

Для приготовления 2%-ного агарозного геля к 1,2 г агарозы добавляли в 54 мл дистиллированной воды и прогревали на водяной бане в течение 30 – 40 минут до полного растворения агарозы, затем добавляли 6 мл 10-кратного трис-боратного буфера (ТВЕ), 1,5 мкл бромистого этидия (10 мг/мл), перемешивали и заливали в камеру для горизонтального электрофореза. Раствор агарозы застывал за 15 – 20 мин. при комнатной температуре.

Приготовление маркера молекулярных масс [Маниатис и др., 1984].

Отбирали 10 мкг плазмиды pBR322 или pUC19, добавляли 10 мкл 10-кратного низкосолевого буфера для рестрикции (100 мМ трис-НСl, рН 7,5, 100 мМ MgCl₂, 10 мМ дитиотейтола), объем доводили бидистиллированной водой до 90 мкл и перемешивали. Добавляли 10 мкл (50 ед. акт.) рестриктазы *MspI*, перемешивали и инкубировали в термостате при 37⁰С в течение двух часов. В качестве маркера молекулярных весов использовали рестриктные фрагменты указанных плазмид.

Детекция продуктов амплификации в 6%-ном полиакриламидном геле

[Маниатис и др., 1984].

Для электрофоретического анализа отбирали по 10 мкл инкубационной смеси из каждой пробы после проведения амплификации. В качестве маркера

молекулярных весов использовали продукты рестрикции эндонуклеазой *MspI* плазмиды pBR322 или маркеры молекулярных весов M50 (“IsoGen”, Россия) и 100bp DNA Ladder Plus (“MBI Fermentas”, Россия). Электрофорез проводили при 10 В/см до тех пор, пока краситель (бромфеноловый синий) не продвигался на 6 – 8 см. После электрофореза гель прокрашивали бромистым этидием и фотографировали или сканировали гель в коротковолновом УФ-свете (длина волны 254 нм).

Детекция продуктов амплификации в 2%-ном агарозном геле [Маниатис и др., 1984].

Для электрофоретического анализа отбирали по 10 мкл инкубационной смеси из каждой пробы после проведения амплификации. В качестве маркера молекулярных весов использовали продукты рестрикции эндонуклеазой *MspI* pBR322 или маркеры молекулярных весов M50 (“IsoGen”, Россия) и 100bp DNA Ladder Plus (“MBI Fermentas”, Россия). Электрофорез проводили при 10 – 12 В/см до тех пор пока краситель (бромфеноловый синий) не продвигался на 8 – 10 см. После электрофореза гель фотографировали в коротковолновом УФ-свете (длина волны 254 нм) или сканировали на приборе для сканирования гелей в УФ – свете “Gel Doc 1000” (Bio-Rad Laboratories, Inc) с выводом данных на экран монитора компьютера.

Обработка данных и использование Internet для локализации ДНК маркеров на картах радиационных гибридов.

Для получения локализации картируемого маркера результаты скрининга панели радиационных гибридов отправляли через Internet на сервер Whitehead Institute/MIT Center for Genome Research (<http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper>).

На этом сервере имеется специальная форма для отправки данных, где можно задать название картируемого маркера и степень точности при определении локализации маркера (LOD score) на карте радиационных гибридов. Величину значения LOD score можно выбрать в пределах от 9 до 21. При локализации *NotI*-STS-маркеров, использовали значение 15, рекомендуемое разработчиками программного обеспечения данного сервиса, как оптимальное. Необходимой частью формы является заполнение графы с результатами

скрининга: положительные ответы отмечаются знаками “1”, отрицательные – “0”, если ответ является сомнительным, то он обозначается как “2”.

Необходимо так же указать адрес электронной почты (e-mail), на который должен прийти ответ о локализации маркера на карте радиационных гибридов. Ответ о локализации картируемого маркера приходит из Whitehead Institute в течение одного часа.

Использованные в работе компьютерные программы, методы биоинформатики и компьютерные базы данных.

Программу “**Multi-Analyst**” (version 1.1) являющуюся программным обеспечением прибора “Gel Doc 1000”, использовали для сканирования полиакриламидных и агарозных гелей в УФ свете (длина волны 260 нм) и редактирования полученных изображений.

Программы для подбора праймеров.

1) “**Oligo**” – позволяет подбирать праймеры к имеющимся нуклеотидным последовательностям, а также на основе нуклеотидной последовательности праймеров и их концентрации в о.е./мл проводить перерасчет концентрации на нг/мл и рассчитывать температуру их отжига;

2) “**Primer**” – аналогична программе “Oligo”, но позволяет точнее рассчитывать температуру отжига праймеров в результате введения в программу дополнительных поправок;

3) “**OligoCalculator**” – определяет температуру отжига для введенных нуклеотидных последовательностей (до 33000 нуклеотидов) и позволяет конвертировать нуклеотидные последовательности, т.е. генерировать последовательности, комплиментарные введенным последовательностям в направлениях 5’-3’ и 3’-5’, что удобно при подборе обратных праймеров.

Программы для анализа нуклеотидных последовательностей.

1) Программу анализа нуклеотидных последовательностей **BLASTN** (<http://www.ncbi.nlm.nih.gov/BLAST/>) использовали для поиска гомологий исследуемых последовательностей с последовательностями, размещенными в международных базах данных.

2) Для получения более детальных сведений о местах локализации картированных маркеров и при построении интегрированных карт хромосом человека использовали сервис **MapView** (http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/hum_srch), позволяющий совмещать различные типы карт: цитогенетические, генетические, карты радиационных гибридов, карты генов, YAC, а также совмещать все типы карт с контигами геномных последовательностей.

3) Для картирования *in silico* исследуемых нуклеотидных последовательностей генома человека использовали программу **BLASTN**, размещенную на странице **The Human Genome Resource** сервера NCBI (<http://www.ncbi.nlm.nih.gov/genome/guide/human/>). Данная программа отличается от описанной ранее тем, что она совмещена с сервисом **MapView** и позволяет на основе гомологии локализовать исследуемую последовательность на любых картах генома человека.

4) Для распознавания генов в исследуемых нуклеотидных последовательностях и построения экзон-интронной структуры генов использовали программу **GENSCAN** (<http://genes.mit.edu/GENSCAN.html>) [Burge, Karlin, 1997].

5) Для предсказания промоторных областей в исследуемых протяженных геномных последовательностях, использовали программы **PromoterInspector** (<http://genomatix.gsf.de/cgi-bin/promoterinspector/promoterinspector.pl>) [Scherf et al., 2000] и **PromoterScan** (http://www.fruitfly.org/seq_tools/promoter.html). Предсказание основано на контекстно-специфичных особенностях, которые определяются по выборке обучающих последовательностей (всех млекопитающих).

6) Поиск сайтов связывания факторов транскрипции на основе заданной нуклеотидной последовательности осуществляли с помощью программы **MatInspector V2.2** (<http://transfac.gbf.de/cgi-bin/matSearch/matsearch.pl>), размещенной в базе данных факторов транскрипции TRANSFAC 4.0.

7) Предсказание потенциальных сайтов полиаденилирования осуществляли с помощью программы **POLYA SCAN** (<http://www.gene-regulation.com/cgi-bin/pub/programs/polyascan/polyascan.cgi>), размещенной на сайте **Gene Regulation**.

8) Для детального анализа геномных последовательностей человека, прилегающих к местам встраивания HPV16, использовали сервис **Genomatrix Suite** (<http://genomatix.gsf.de>). Этот сервис включает в себя программы, осуществляющие поиск гомологий с геномными и экспрессирующимися последовательностями человека (BLASTN) и позволяющие проводить их анализ на предмет наличия промоторов (PromoterInspector) и функциональных сайтов (белок-ДНК связывающих сайтов), а также повторяющихся элементов генома. Для проведения анализа можно ввести номер исследуемой последовательности, если она зарегистрирована в базах данных EMBL или GDB, или ввести саму последовательность. Результаты выдаются в виде табличного материала и в интерактивно-графическом виде, что более наглядно [Heinemeyer et al., 1999; Wingender et al., 2000; Wingender et al., 2001; Frisch et al., 2002; Liebich et al., 2002].

9) Аминокислотные последовательности белков, кодируемые изучаемыми генами, получены трансформацией нуклеотидных последовательностей генов (*in silico* трансляцией) с помощью сервиса **DNA=>Protein** (<http://cn.expasy.org/tools/dna.html>).

10) Поиск сходных аминокислотных последовательностей и построение карт локального сходства для гипотетических белков, а также построение филогенетических деревьев проводили с использованием программ сервера **GeneBee** (<http://genebee.msu.ru/>).

Компьютерные базы данных, использованные в работе.

В работе использовали следующие компьютерные базы данных и Internet-ресурсы, размещенные на сервере **NCBI (National Center for Biotechnology Information)** – Национальный центр информации по биотехнологии (USA, <http://www.ncbi.nlm.nih.gov/>):

1) **GenBank** – база данных генетических последовательностей, поддерживаемая NIH (National Institutes of Health, USA, <http://www.nih.gov/>). Содержит аннотированную коллекцию всех общедоступных последовательностей ДНК, РНК и белков, снабженных литературными ссылками и различной биологической информацией. Обновляется каждые два месяца. Доступна через сервер NCBI: <http://www.ncbi.nlm.nih.gov/>. Является частью International Nucleotide Sequence Database Collaboration, которая

объединяет три крупнейшие коллекции нуклеотидных последовательностей: DDBJ (DNA Data Bank of Japan) (Japan, <http://www.ddbj.nig.ac.jp/>), EMBL (EBI) и GenBank (NCBI) [Benson et al., 1999].

2) **UniGene** – база данных, которая содержит кластеры похожих последовательностей и позволяющая совместить сразу несколько карт и, следовательно, более точно локализовывать маркеры на генетических картах и определять их положение на хромосоме (USA, <http://www.ncbi.nlm.nih.gov/UniGene/>).

3) **The Human Genome Resource** (<http://www.ncbi.nlm.nih.gov/genome/guide/human/>) – сервер, на котором размещены базы данных нуклеотидных последовательностей генома человека, SNP (от англ. Single Nucleotide Polymorphism), генов и EST человека, интерактивные карты хромосом человека и другие Интернет-ресурсы, связанные с характеристикой генома человека.

4) **LocusLink** (<http://www.ncbi.nlm.nih.gov/LocusLink/index.html>) – поисковая система, позволяющая получать краткое описание любого зарегистрированного генетического локуса и набор ссылок на все имеющиеся документы, содержащие информацию об этом локусе.

5) **OMIM** (<http://www.ncbi.nlm.nih.gov/OMIM/>) – база данных, содержащая информацию обо всех известных генах и кодируемых ими белках (функция, участие в развитии заболеваний).

6) Крупнейшая интегрированная поисковая система **ENTREZ** (<http://www.ncbi.nlm.nih.gov/Entrez/>) для нуклеотидных и аминокислотных последовательностей, библиографии (PubMed), полных геномов (Genomes), а также трехмерных структур белков (MMDB) создана и поддерживается NCBI. При этом поиск последовательностей ДНК и белков не ограничивается только ресурсами GenBank, используются также и другие доступные в Internet базы данных.

Другие базы данных:

GDB (Genome Data Base) – содержит полную информацию обо всех картированных генах, ДНК-маркерах и ДНК-зондах человека. (USA, <http://www.gdb.org/>).

EMBL (The European Bioinformatics Institute) – база данных нуклеотидных последовательностей Европейской Молекулярно-Биологической Лаборатории (UK, <http://www.ebi.ac.uk/embl/>) пополняется большей частью непосредственно авторами, определившими первичную структуру фрагмента ДНК или РНК и, кроме последовательности нуклеотидов, содержит разнообразную информацию о каждом фрагменте, включая литературные ссылки, перекрестные ссылки на документы других баз данных, таблицы особенностей и др. Существует с 1982 года. База данных – продукт сотрудничества EMBL (Великобритания), GenBank (США) и DDJP (Япония), каждая из этих трех групп собирает свою порцию информации из всех возможных мировых источников, ежедневно обмениваясь новыми и обновленными документами друг с другом.

Genethon – содержит информацию о маркерах, полученных с помощью YAC-клонов, генетические карты хромосом человека (**CEPH-Genethon** (France, <http://www.cephb.fr/bio/ceph-genethon-map.html>)), мыши, сельскохозяйственных животных, и многих других организмов, а также позволяет совмещать карту радиационных гибридов с картой YAC-контигов (http://carbon.wi.mit.edu:8000/cgi-bin/contig/phys_map).

Whitehead Institute (UK, <http://www-genome.wi.mit.edu/>) и **Sanger Centre** (<http://www.sanger.ac.uk/>) – содержат информацию о маркерах, локализованных в геноме человека с помощью радиационных гибридов соматических клеток человека и китайского хомячка (или мыши).

RHDB (RHdb) Radiation Hybrid DataBase – общедоступная база данных по радиационным гибридам [Rodriguez-Tome, Lijnzaad, 1999]. Данные в этой базе структурируются по панелям RH-клонов. В базе содержатся данные о геномах человека (с 1995), мыши (с 1997) и крысы (с 1998 года). Поддерживается EMBL (EBI) (UK, <http://corba.ebi.ac.uk/RHdb/>)

TIGR (The Institute for Genomic Research) – база данных нуклеотидных последовательностей полученных Институтом Исследования Геномов и компанией Celera Genomics (USA, <http://www.tigr.org/>). В настоящее время большая часть данных является коммерческими.

TRANSFAC – база данных содержащая все известные факторы транскрипции с их кратким описанием и последовательностями, служащими сайтами их посадки на молекуле ДНК, а также информацию о других ДНК-

связывающих белках и их свойствах (<http://transfac.gbf.de/TRANSFAC/>) [Heinemeyer et al., 1999; Wingender et al., 2000; Wingender et al., 2001; Liebich et al., 2002].

Поскольку панель радиационных гибридов GeneBridge4 является коммерческим препаратом, обработка данных и получение информации о результатах скрининга RH-панели проводится через Internet с сервера – Whitehead Institute/MIT Center for Genome Research (<http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper>) с использованием программы для обработки данных “RHMAPPER”.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

I. Отработка метода радиационного картирования на примере локализации мозгоспецифичной последовательности Gfhb и маркера GC-богатой последовательности smgc30.

Метод радиационного картирования (РН-картирования) позволяет локализовать исследуемую последовательность ДНК на хромосомах человека и установить порядок сцепленных с ней известных ДНК-маркеров (маркеров рабочей рамки), а также оценить физическое расстояние между маркерами. РН-карты совмещены со всеми типами карт генома человека, что облегчает задачу поиска клонов, несущих картируемую последовательность, а, следовательно, ее выделение для дальнейших исследований.

Метод РН-картирования с середины 90-х годов стал основным методом тонкого физического картирования геномов человека и других позвоночных животных [Жданова, 2002]. Однако в нашей стране этот метод используется крайне ограниченно: картирование отдельных хромосом норки и свиньи, данные лаборатории сравнительной генетики животных ИОГен РАН по картированию хромосомы 3 человека [Сулимова и др., 2000; Rakhmanaliev et al., 2001]. К началу наших исследований отечественных работ с использованием РН-картирования не проводилось. Это обстоятельство потребовало проведения дополнительных экспериментов для отработки метода РН-картирования на модельных системах.

Отработку метода РН-картирования проводили на примере локализации в геноме человека двух ДНК-маркеров: к мозгоспецифичной последовательности Hfb1 (маркер Gfhb) и к GC-богатой последовательности, гиперметирированной в опухолях шейки матки (маркер smgc30).

Выбор мозгоспецифичной последовательности Hfb1 для постановки модельного опыта по ее локализации в геноме человека методом РН-картирования обусловлен следующими причинами. Ранее в отделе молекулярных основ генетики человека Института молекулярной генетики РАН была исследована структура геномного клона, полученного из космидной библиотеки хромосомы 5 человека, гибридизующегося с мозгоспецифической

последовательностью Hfb1. Анализ нуклеотидной последовательности клона позволил сделать предположение, что данная последовательность представляет собой участок не описанного ранее экзона гена синаптического белка комплексина 2 человека, кодирующего протяженную 3'-нетранслируемую область соответствующей мРНК.

Комплексины 1 и 2 – эволюционно консервативные высоко гомологичные цитоплазматические белки, локализующиеся в пресинаптических окончаниях нейронов млекопитающих, были получены из гомогената мозга крыс в комплексе с другими белками, участвующими в экзоцитозе синаптических пузырьков [McMahon et al., 1995; Ishizuka et al., 1995; Takahashi et al., 1995]. В настоящее время показано, что комплексины участвуют в формировании SNARE-комплекса, рецептора для белка α -SNAP, состоящего из белков синтаксина 1, SNAP-25 и синаптобревина, обеспечивающего освобождение нейромедиаторов из синаптических везикул в синаптическую щель [Pabst et al., 2000]. Однако роль комплексинов в механизме синаптической передачи окончательно не выяснена и активно исследуется [Ishizuka et al., 1997; Ono et al., 1998; Takahashi et al., 1999; Itakura et al., 1999].

Для комплексина 2 была известна физическая локализация в геноме человека. Совпадение места локализации маркера Gfhb, полученного нами методом РН-картирования, с локализацией гена комплексина 2, приведенной в базе данных, позволило бы подтвердить предположение о том, что последовательность Hfb1 – является частью неопisanного ранее экзона гена комплексина 2, кодирующего один из вариантов 3'-нетранслируемого конца соответствующей мРНК. Кроме того, совпадение собственных и литературных данных о локализации гена позволило бы судить о точности полученной нами локализации.

Другой последовательностью, выбранной нами для отработки метода РН-картирования, была CG-богатая последовательность, гиперметилированная в опухолях шейки матки (маркер – cmgc30).

Данная последовательность (AccN AJ431613) была получена с использованием модифицированного метода метилчувствительной ПЦР (МЧ-ПЦР) со статистическими праймерами к GC богатым последовательностям ДНК из опухолей шейки матки в лаборатории молекулярной биологии вирусов НИИ

канцерогенеза Российского онкологического научного центра им. Н. Н. Блохина, РАМН.

Основной целью нашей работы было определение мест интеграции в геноме человека вируса папилломы тип 16. Нельзя было исключать при этом, что вирус будет встраиваться в промоторные области клеточных генов, нарушая тем самым нормальное функционирование клетки. Как известно, промоторные области генов содержат CpG-островки, т. е. являются GC-богатыми последовательностями. Однако, проведение ПЦР на GC-богатых последовательностях имеет определенные сложности и требует введения дополнительных модификаций (например, использование глицерина, бетаина и других компонентов для снижения температуры денатурации ДНК и повышения специфичности реакции). Поэтому мы сочли необходимым отработать условия ПЦР скрининга RH-панели и для GC-богатых ДНК-маркеров на примере маркера cmgс30.

1. Подбор праймеров и оптимизация условий полимеразной цепной реакции.

Температуру отжига для каждой пары праймеров использованной в работе рассчитывали, используя компьютерные программы “Primer” и “Oligo”. Однако она не всегда оказывалась оптимальной для проведения ПЦР. Поэтому для всех праймеров температуру отжига подбирали опытным путем, принимая за исходную – температуру, рассчитанную с использованием программы “Primer”.

Оптимизацию условий ПЦР проводили, используя геномную ДНК человека и геномную ДНК китайского хомячка в качестве матриц. Одним из главных критериев, предъявляемых к ДНК-маркерам при работе с RH-панелью, является отсутствие в геноме китайского хомячка последовательностей, гомологичных исследуемым последовательностям человека, поэтому для всех пар праймеров проводилась ПЦР с геномной ДНК хомячка. Оптимальные условия проведения ПЦР для каждой пары праймеров приведены в таблице 2 (см. раздел “Материалы и методы”).

2. Локализация исследуемых маркеров.

Скрининг RH-панели проводили с праймерами к соответствующим ДНК-маркерам. Результатом ПЦР с пробами RH-панели является либо наличие, либо отсутствие амплифицированных фрагментов, соответствующих по длине фрагментам ожидаемой длины для данной пары праймеров (рис.3).

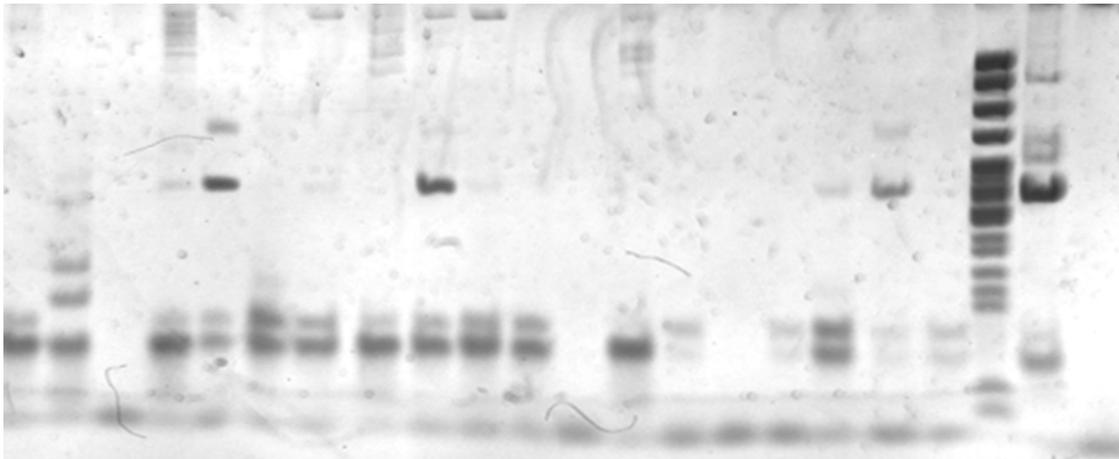


Рисунок 3. Электрофоретическое разделение продуктов амплификации в 6%-м ПААГ при скрининге RH-панели с парой праймеров Ghfb. Дорожки: 1 - 19 - ДНК клонов RH-панели. отрицательный контроль (вода вместо матричной ДНК); 20 - маркер молекулярных масс (pBR322/MspI); 21 - положительный контроль (тотальная ДНК человека); 22 - отрицательный контроль (геномная ДНК китайского хомячка в качестве матрицы).

На рис. 3 в качестве примера представлен результат электрофоретического анализа скрининга RH-панели с праймерами к ДНК-маркеру Ghfb после окрашивания бромистым этидием. Видны положительные ответы (дорожки 5, 9, 18) и отрицательные ответы (дорожки 1 – 4, 6 – 8, 10 – 17, 19). Для маркера smgc30 был проведен повторный скрининг RH-панели, который позволил уточнить первичные данные.

Результаты скрининга маркеров Ghfb и cmgc30 представлены в табл. 4 в виде матрицы, которая вводится в форму на сервере Whitehead Institute для обработки программой RHMAPPER, позволяющей определить уровень сцепления исследуемых маркеров с маркерами рабочей рамки и локализовать их на RH-карте.

Таблица 4. Результаты ПЦР-скрининга RH-панели GeneBridge4 (положительный ответ – 1; отрицательный – 0; сомнительный – 2).

| |
|---|
| Ghfb |
| 100001110011001000000200112010000100012002020110201000000001000200002000000000000202000210101 |
| cmgc30 |
| 001000000000010000000000000100000020000200001000010010001101111000000001201000001000000000000 |

Для получения информации о локализации картируемого маркера на RH-карте, результаты скрининга отправляли через Интернет в соответствии с формами, приведенными на сервере Whitehead Institute. При отправлении данных необходимо указать величину значения LOD score (от 13 до 17), представляющего собой логарифм отношения шанса (правдоподобия) сцепления двух локусов к шансу (правдоподобию) отсутствия сцепления (см. обзор литературы). Значение LOD score при ответе не должно превышать 3, так как эта величина составляет вероятность 1:1000 того, что между двумя исследуемыми локусами не произошел разрыв. Величина LOD score < 3 считается достаточной для доказательства сцепления маркеров. Величина LOD score > 3 означает, что вероятность разрыва между исследуемым маркером и маркером рабочей рамки больше 1:1000. Такой результат считается не достоверным. Обработка результатов ведется с учетом всех маркеров, ранее локализованных на RH-картах, и каждый из них проверяется на возможность сцепления с исследуемым маркером.

На сервер Whitehead Institute необходимо отправлять результаты по всем 93 клонам (положительный – 1; отрицательный – 0; сомнительный – 2) (см. табл. 3). Ответ о локализации из Whitehead Institute приходит в течение одного часа на адрес электронной почты (e-mail) отправителя, который необходимо указать при отправлении формы. В случае единичной локализации также присылается карта радиационных гибридов с включенными в нее локализованными исследуемыми маркерами.

В ответе также указывается хромосома, на которой программа обработки результатов скрининга RHMAPPER (разработана Whitehead Institute) локализовала маркер, маркеры рабочей рамки (framework) с которыми сцеплен исследуемый маркер, а также расстояние до них и до ближайших маркеров, выраженное в **centiRay** (cR). **cR** – это мера расстояния, означающая частоту разрывов между двумя маркерами, вызванных радиацией. Для каждой хромосомы соотношение между cR и cM варьирует. Кроме того, в электронном письме указывается значение LOD score, т.е. вероятность сцепления картированного маркера с маркерами рабочей рамки.

Использование радиационного картирования не позволяет использовать картируемую последовательность в качестве исходного материала для выделения и последующего экспериментального анализа, поскольку каждый индивидуальный клон RH-панели содержит множество фрагментов генома человека (до 30% от генома клеток-реципиентов), встроенных в хромосомы реципиента случайным образом. Уровень разрешения RH-картирования с использованием панели GeneBridge4 сопоставим с размером вставки в мега-УАС-клонах (1 млн.п.н.) или ВАС-клонах, что позволяет совместить RH-карты с УАС- и ВАС-картами, а также картами геномных контигов. От этих карт можно перейти к транскрипционным и генным картам. Наличие же индивидуальных УАС- или ВАС-клонов для исследуемых маркеров позволяет получить физический материал для дальнейшего выделения и изучения картированных последовательностей.

Результаты скрининга с праймерами к маркерам Gfhb и smgc30 позволили определить их локализацию на карте радиационных гибридов Whitehead Institute (WI-RH-Map). Чтобы иметь возможность сравнить полученные данные по локализации с мировыми базами данных, было определено также положение картированных маркеров на универсальной карте GeneMap99-GB4. Такой перерасчет возможен, так как значительная часть маркеров рабочей рамки WI-RH-Map локализована на RH-карте GeneMap99-GB4. Были определены физические интервалы и сублокусы хромосом, в которых находятся картированные нами последовательности. Установлены также УАС-контиги и геномные контиги, к которым принадлежат картированные маркеры, и определена их физическая локализация на хромосомах.

Программой RHMAPPER было определено физическое положение ДНК-маркера Gfhh: 3.78 cR от маркера AFMA109XA5 и 3.77 cR от маркера WI-6737 (626.67 и 633.18 cR₃₀₀₀, соответственно). Таким образом, картируемый маркер лежит в интервале 631.3 – 637.9 cR₃₀₀₀. Этот интервал находится между маркерами D5S504 – D5S677 (175.3-182.4 cM), цитогенетическое положение которых – 5q35. В том же интервале находится ген complexin II (636.91 cR₃₀₀₀), что подтверждает принадлежность изучаемой ДНК-последовательности к этому гену. Участок хромосомы 5, находящийся в интервале D5S504 – D5S677, клонирован в YAC-контиг WC5.13, что позволяет более детально изучить картируемую ДНК-последовательность.

Для маркера smgc30 также было определено физическое положение: 2.53 cR от CHLC.GATA65D11, 3.25 cR от CHLC.GATA64G07 и 10.87 cR от WI-9913. Таким образом, исследуемый маркер находится в районе около 347.4 cR₃₀₀₀. Этот интервал находится между маркерами D9S172 – D9S160 (112.0-117.7 cM), цитогенетическое положение которых – 9q31. В этом интервале нет описанных генов, что подтверждается поиском для маркера smgc30 в базах данных и последующего использования программы MapViewer. Также был определен геномный контиг, к которому принадлежит изучаемая последовательность, – NT_017568. Участок хромосомы 9, находящийся в интервале D9S172 – D9S160, клонирован в YAC-контиг WC9.14. Появившиеся позднее в базах данных нуклеотидные последовательности участка хромосомы 9 позволили подтвердить полученные нами результаты RH-картирования.

Таким образом, нами были оптимизированы условия ПЦР-скрининга RH-панели, позволяющие локализовать ДНК-маркеры на хромосомах человека с высокой степенью точности (lod score <3). Разработанные нами условия скрининга были использованы для локализации мест интеграции ДНК HPV16 в хромосомах клеток плоскоклеточного рака шейки матки.

II. Локализация последовательностей генома человека, прилегающих к местам встраивания вируса папилломы.

В представленной работе были изучены последовательности генома человека, прилегающие к местам встраивания вируса папилломы человека тип 16 (HPV16) с 3'-конца. Эти последовательности были ранее выделены с помощью метода АРОТ (amplification of papillomavirus oncogene transcripts) и секвенированы в лаборатории молекулярной биологии вирусов НИИ канцерогенеза Российского онкологического научного центра им. Н. Н. Блохина, РАМН.

Метод АРОТ основан на структурных различиях между 3'-концами онкогенных вирусных транскриптов. Интеграция HR-HPV геномов в клетки опухолей рака шейки матки обычно происходит с нарушением открытых рамок считывания для генов *E1* или *E2* [von Knebel Doeberitz, 1992] и с нарушением или делецией сигнала полиаденилирования ранних генов вируса (рис. 4). Это приводит к тому, что транскрипты, считываемые со встроенных *E6* и *E7* онкогенов, обычно содержат вирусные последовательности и примыкающие к их 3'-концам клеточные последовательности. Эти транскрипты обычно используют донорский сайт сплайсинга, расположенный правее *E1* старт-кодона (рис. 4, транскрипты типа А [Schwarz et al., 1985; Schneider-Gadicke et al., 1986]). В нескольких клеточных линиях карцином шейки матки интегрированные HPV геномы нарушаются в пределах гена *E4*, в результате чего образуются слитые вирус-клеточные транскрипты, которые охватывают *E6-E7-E1* последовательности в их 5'-концах, следующую за ними последовательность *E4* и клеточные последовательности в их 3'-концах (рис. 4, транскрипты типа В [Baker et al., 1987]). В случаях эписомальной формы HPV, наиболее распространенный HPV *E6-E7* содержащий транскрипт сплайсируется с образованием *E1-E4* формы (т.е. происходит вырезание гена *E2*) и завершается вирусным сайтом полиаденилирования (рис. 4, транскрипт эписомальной формы [Sherman et al., 1992]).

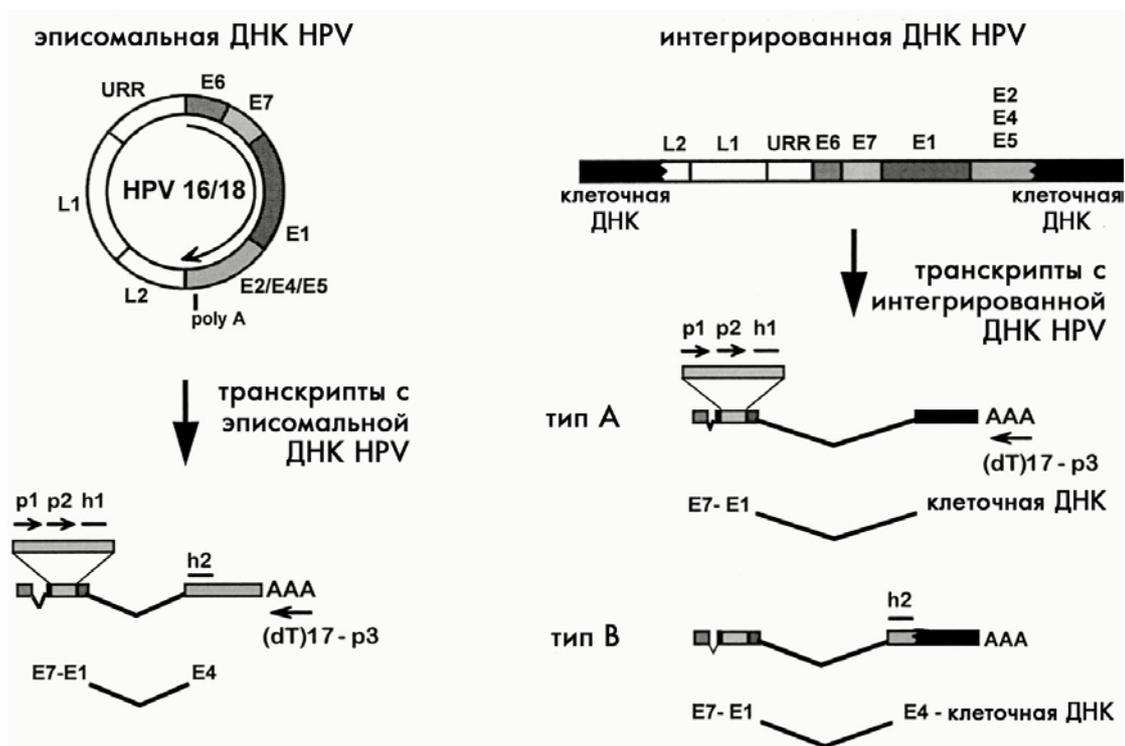


Рисунок 4. Структуры HPV16/18 геномов, эписомальный (слева) и интегрированный (справа), и образующихся транскриптов. (dT)17-p3 – праймер для ОТ-ПЦР; p1, p2 и p3 – ПЦР АРОТ праймеры; праймеры для гибридизации – h1 (специфичный к HPV E7) и h2 (специфичный к HPV E4); транскрипты интегрированной форм отличаются по отсутствию (type A) или наличию (type B) последовательности вирусного гена E4. [цит. по: Klaes et al., 1999].

Эти структурные особенности позволяют разделять транскрипты вирусных онкогенов интегрированной и эписомальной форм с помощью метода АРОТ. Используя oligo(dT)₁₇-праймер связанный с линкерной последовательностью ((dT)₁₇-p3), отбирается пул всех поли (А)-содержащих мРНК. Далее транскрипты онкогенов HPV с эписомальной и интегрированной форм амплифицируются с помощью ПЦР с использованием E7-специфичных праймеров в комбинации с p3 и (dT)₁₇-p3 праймерами, соответственно (рис. 4). Транскрипты интегрированной формы могут быть отделены от транскриптов эписомальной формы по размеру, поскольку размер транскриптов с эписомальной ДНК всегда составляет 1050 п.н., а размер транскриптов с интегрированной ДНК, часто вообще не содержащих E4, существенно варьируют. Кроме того, амплифицированные фрагменты были проверены с

помощью Southern blot гибридизации с использованием HPV E7- и E4-специфических олигонуклеотидов (рис. 4, h1 и h2).

Использование метода АРОТ позволило выделить 12 различных последовательностей хозяйской ДНК, прилегающих к местам встраивания HPV16 в геном клеток плоскоклеточного рака шейки матки и являющихся индивидуальными. Все 12 последовательностей получены из образцов опухолей различных пациентов. Размер таких клеточных последовательностей от точки состыковки вирусных и клеточных последовательностей до поли(А)-сайта варьирует от 140 до 450 нуклеотидов. Клеточные последовательности, прилегающие к 3'-концу интегрированной вирусной ДНК, были обозначены как маркеры INT. Все двенадцать последовательностей использовались для картирования *in silico*. Однако к началу нашей работы (1999 год) последовательности генома человека были слабо представлены в базах данных: для 4-х маркеров не было найдено достоверного сходства с геномными последовательностями (INT254, INT259, INT431 и INT407). Для маркера INT290 была показана двойная локализация (1p36 и Xp11.3). Это послужило основанием для локализации этих маркеров в геноме человека методом RH-картирования, который ранее не использовался для локализации интегрированной ДНК вируса папилломы. К этим пяти маркерам INT были подобраны соответствующие праймеры, которые использовались для физической локализации методом RH-картирования (табл. 3, раздел «Материалы и методы»).

1. Физическая локализация последовательностей, прилегающих к местам встраивания HPV16 методом RH-картирования на хромосомах человека.

1.1. Оптимизация условий ПЦР, скрининг панели радиационных гибридов и определение локализации маркеров INT на RH-карте.

Для локализации мест встраивания HPV16 в геном человека использовали метод RH-картирования. Температуру отжига для каждой пары праймеров рассчитывали, как описано в Главе I. Оптимизацию условий ПЦР проводили, используя геномную ДНК человека и геномную ДНК китайского хомячка в качестве матриц. Условия ПЦР для праймеров, которые были использованы для скрининга RH-панели, и длины амплифицируемых фрагментов приведены в табл. 2 (см. “Материалы и Методы”).

Ни одна пара праймеров, использованная в работе, не давала фрагментов такой же длины с геномом китайского хомячка, что и исследуемые маркеры. Праймеры, дающие фрагменты с ДНК китайского хомячка, отличающиеся по длине от продуктов амплификации с ДНК человека, также использовались для скрининга.

Скрининг RH-панели с праймерами к маркерам INT254, INT259, INT431, INT407, INT290 проводили методом, описанным в главе I. Для маркеров INT259 и INT431 был проведен повторный скрининг RH-панели, который позволил уточнить первичные данные. Конечные результаты скрининга представлены в табл. 5 в виде матрицы.

Таблица 5. Результаты ПЦР-скрининга RH-панели GeneBridge4
(положительный ответ – 1; отрицательный – 0; сомнительный – 2).

| | |
|---------------|---|
| INT254 | 001020010000002220100000120000020000000000101000000001000001010011010000110000001021221002000 |
| INT259 | 02000001010000000100001101001001100001000000111020110000000011100111000000000000002000001110 |
| INT431 | 000002200011020221100000010001100000000011001010000000100100011010001000001000100000000010121 |
| INT407 | 100001110000001000000010110010000100010001000110001000000001000100001000000000000002000010001 |
| INT290 | 00100001000000000110001000001000001000001020010010010000000001100000001112000110001001001011 |

Результаты локализации пяти INT маркеров с использованием программы RHMAPPER (Whitehead Institute) суммированы в табл. 6.

Таблица 6. Данные о локализации пяти картированных маркеров, предоставленные Whitehead Institute Centre, с использованием программы RHMAPPER.

| | |
|---|--|
| 1 | INT254 Chromosome 14 Places 7.80 cR from WI-5202 Places 5.23 cR from D14S277 (lod 1.46 relative to most likely) |
| 2 | INT259 Chromosome 4 Places 10.87 cR from CHLC.GATA42E03 Places 2.02 cR from WI-4048 (lod 0.57 relative to most likely) Places 5.23 cR from D4S394 (lod 0.70 relative to most likely) |
| 3 | INT431 Chromosome 13 Places 20.33 cR from WI-5867 |
| 4 | INT407 Chromosome 5 Places 9.98 cR from WI-6737 |
| 5 | INT290 Chromosome X Places 12.44 cR from DXS6745 Places 16.25 cR from WI-5587 (lod 1.90 relative to most likely) |

Для всех картированных маркеров значение LOD score находилось в пределах от 0,00 до 2,00, т.е. не превысило значение 3, что свидетельствует о достоверности сцепления исследуемых маркеров с маркерами рабочей рамки, и, следовательно, о достоверности их локализации на RH-карте.

1.2. Совмещение RH-карты с другими типами карт и поиск генов, прилегающих к местам встраивания HPV16.

По результатам проведенного скрининга была определена локализация пяти маркеров INT на карте радиационных гибридов Whitehead Institute (WI-RH-Map) (табл. 7). Чтобы иметь возможность сравнить полученные данные по локализации с мировыми базами данных, было определено также положение картированных маркеров группы INT на универсальной карте GeneMap99-GB4 (табл. 7). А использование программы MapViewer (сервер NCBI), содержащую интерактивные карты генома человека, позволило получить более полную информацию из баз данных и совместить все типы карт. Были определены физические интервалы и сублокусы хромосом, в которых находятся картированные нами последовательности. Установлены также YAC-контиги и геномные контиги, к которым принадлежат картированные маркеры (табл. 7).

Таблица 7. Локализация маркеров INT на хромосомах человека.

| INT | Локализация на карте WI-RH, cR | Локализация на карте GM99-GB4, cR | Физический интервал | Сублокус хромосомы | Позиция на хромосоме, т.п.н. / контиг | YAC-контиги |
|-----|--------------------------------|-----------------------------------|---------------------------------|--------------------|---------------------------------------|-----------------|
| 254 | 244.7 – 260.4 | 188.2 - 200.3 | D14S1028-D14S71 (69.8-73.7 cM) | 14q23 - q24 | 517,1 / NT_010028 | WC14.1 |
| 259 | 41.2 – 49.4 | 46.9 - 56.2 | D4S394-D4S1599 (14.5-22.0 cM) | 4p16 | 8.5 / NT_030667 | WC4.1 |
| 431 | 175.5 | 161.7 | D13S153-D13S163 (47.5-49.2 cM) | 13q21 | 55.5 / NT_024524 | WC13.2 |
| 407 | 541.9 | 645.1 | D5S498-D5S408 (184.7- 195.8 cM) | 5q35 | 171.6 / NT_006725 | WC5.13 |
| 290 | 48.5 – 61.4 | 140 | DXS8023-DXS991 (84.9-86.9 cM) | Xp11.21 - p11.4 | 441.6 / NT_011609 | WCX.7 WCX.10 |

Процедура совмещения карт позволила определить гены, лежащие в непосредственной близости от мест встраивания HPV16. На рис. 5 представлены карты хромосом с указанием мест локализации маркеров INT и генов, лежащих в непосредственной близости от картированных маркеров.

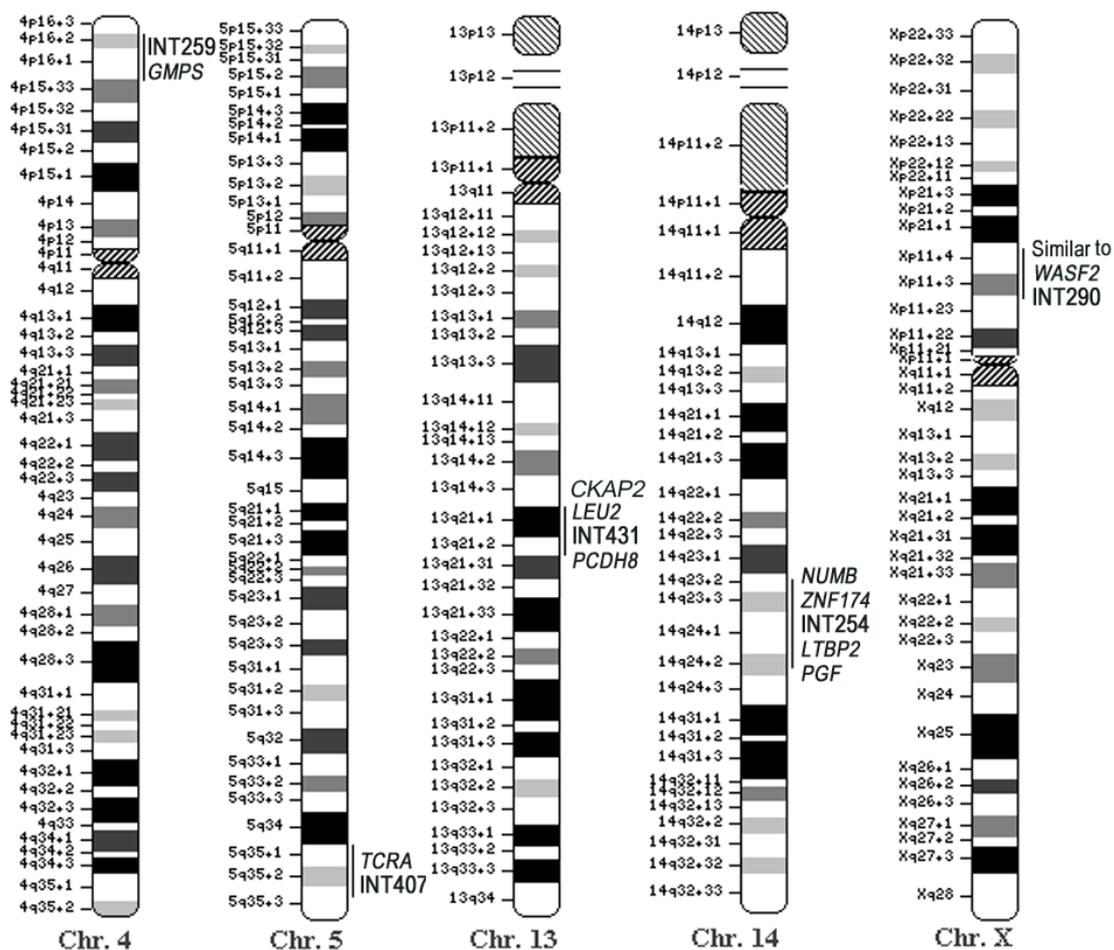


Рисунок 5. Локализация маркеров INT на хромосомах человека методом RH-картирования. Жирным выделены маркеры INT. Курсивом обозначены гены, лежащие в непосредственной близости от маркеров INT: WAS – ген синдрома Висконца-Олдрича; ZNF174 – «цинковые пальцы» 174; NUMB – гомолог гена NUMB дрозофилы (ростовой фактор нейронов); LTBP2 – латентный трансформирующий ростовой фактор бета-связывающегося белка 2; PGF – плацентарный ростовой фактор; EDG2 – ген эндотелиальной дифференцировки; PCPH8 – ген протокадгерина 8; CTNNAL1 – ген, кодирующий белок гомологичный альфа-катенину; TCRA – ген альфа-цепи T-клеточного рецептора; SKAP2, LEU2 – супрессоры онкогенеза.

Для всех картированных маркеров INT установлена единичная локализация в геноме человека. Сайты интеграции HPV16 распределены на разных хромосомах человека. Выраженной закономерности в их распределении в геноме не наблюдается. Однако, при сравнении наших данных с картами насыщения хромосом генами, оказалось, что места интеграции вируса локализованы в районах с высоким количеством генов (рис. 6).

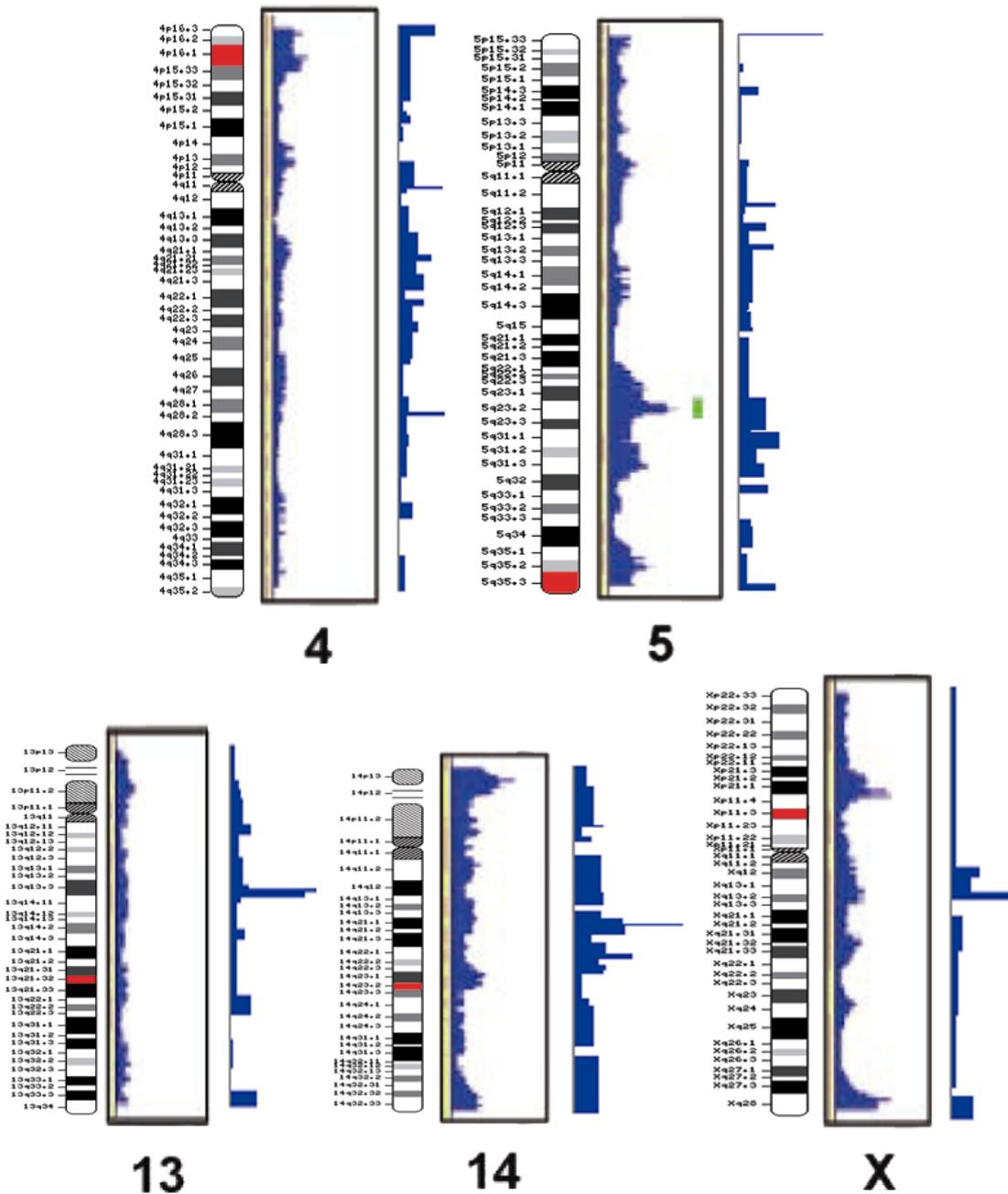


Рисунок 6. Физическая локализация маркеров INT по данным RH-картирования (красные бенды на идиограмме) в районах хромосом, насыщенных генами. Слева на право: идиограмма (G-banding), транскрипционная карта, опубликованная в журнале «Science» [Caron et al., 2002], карта насыщения генами, размещенная на сервере GeneMap'99.

Обращает на себя внимание тот факт, что один из сайтов интеграции (INT431) локализован в области 13q14-q21, насыщенной онкосупрессорами (*SCAP2*, *LEU2*). Многие гены, лежащие в местах локализации маркеров INT, могут также претендовать на роль онкосупрессоров. Среди них в первую очередь можно

отметить гены, участвующие в процессах развития и дифференцировки тканей (*NUMB*, *LTBP2*, *PGF*, *EDG2*) и гены, кодирующие сигнальные белки (*ZNF174*, similar to *WASF2*) и белки цитоскелета (*PCPH8*, *CTNNA1*).

2. Локализация *in silico* последовательностей, прилегающих к местам встраивания HPV16 в геноме человека.

С появлением и развитием общедоступных баз данных, содержащих геномные последовательности нуклеотидов, а также программ, позволяющих проводить сравнение нуклеотидных последовательностей, появилась возможность проводить *in silico* локализацию изучаемых последовательностей нуклеотидов. Картирование *in silico* можно считать самым точным, т.к. оно позволяет установить положение изучаемого ДНК-маркера с точностью до одного нуклеотида. Однако эффективно применять данный тип картирования можно только для организмов, для которых секвенирована большая часть генома. Несмотря на незаконченность секвенирования и аннотирования генома человека, для многих ДНК-маркеров метод локализации *in silico* дает достоверные результаты.

Нуклеотидные последовательности всех 12 маркеров INT были зарегистрированы в базе данных EMBL с присвоением номера (AccN, см. табл. 8) и доступны через Интернет. Это облегчает работу с нуклеотидными последовательностями через Интернет, т.к. многие программы для анализа нуклеотидных последовательностей (BLASTN, Genomatrix Suite и др.) позволяют вводить не всю последовательность, а только номер, под которым она зарегистрирована в базах данных.

Для локализации маркеров INT был проведен поиск сходных геномных последовательностей с использованием программы BLASTN, размещенной на странице **The Human Genome Resource** сервера NCBI и совмещенной с сервисом MapViewer. Данная процедура позволила не только установить геномные контиги, нуклеотидным последовательностям которых принадлежат изучаемые маркеры, но и локализовать маркеры INT на различных типах карт хромосом человека.

Ответы, дающие достоверный уровень сходства, а, следовательно, и верную локализацию, рассматривали как положительные. Достоверным считали сходство более 90% при длине идентичной последовательности не менее 100 нуклеотидов.

На основе полученных данных определена локализация семи ранее не картированных маркеров INT и подтверждена локализация пяти маркеров, картированных нами методом ПЦР-скрининга RH-панели (табл. 7). Из 12 маркеров INT два имеют множественную локализацию на хромосомах человека. Результат поиска подобных EST человека для маркера INT467 (см. главу 3) позволил определить, что данная последовательность сходна с мРНК для многокопийного гена 40S рибосомального белка S27 (*RPS1*), копии которого представлены на разных хромосомах (табл. 8). По результатам RH-картирования маркер INT290 локализован на Xp11.3, однако, он имеет высокий уровень сходства с двумя последовательностями генома, локализованными на хромосомах 1 (область p36) и X (область p11.3) (см. табл. 8).

Отдельно следует рассмотреть маркер INT466. Нуклеотидная последовательность этого маркера имеет достоверный уровень сходства с двумя участками нуклеотидной последовательности геномного контига NT_005567. На рис. 7 изображена примерная схема расположения маркера INT466 и контига NT_005567 друг относительно друга (также см. табл. 8).

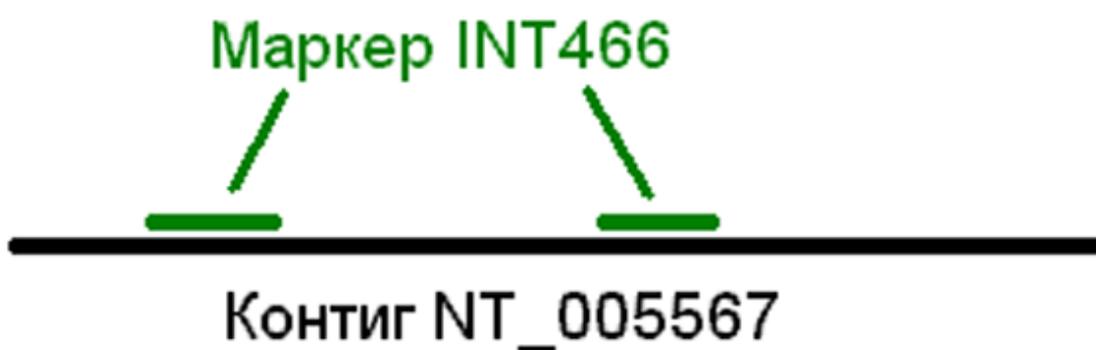


Рисунок 7. Схема расположения маркера INT466 относительно контига NT_005567.

Этот феномен может объясняться либо результатом неправильной состыковки клонов, составляющих контиг, либо результатом интеграции HPV16 в экзон транскрибируемого гена. Транскрипт, полученный в последнем случае, будет гомологичен экзонным областям гена, разделенным интроном в геномной последовательности. Как будет видно из главы 3, вирусная ДНК, в описываемом нами случае, интегрировалась в 5-й экзон гена *LOC131626* и ее транскрипция закончилась после 6-го среднего экзона, что служит дополнительным подтверждением возможности сплайсинга «слитых» вирус-клеточных транскриптов.

Таблица 8. Характеристика маркеров INT, гомологии маркеров INT с геномными последовательностями человека.

| INT маркеры | | | Сходства маркеров Int с геномными последовательностями человека (HTGS) | | | |
|-------------|-------------|----------|--|----------|-------------------------|----------------------------------|
| № | Длина, п.н. | AccN | Длина, п.н. | % | Геномные контиги (AccN) | Положение на хромосоме |
| 254 | 349 | AJ431608 | 352 | 95 | NT_010028 (AL589168) | 14q23.2 |
| 259 | 213 | AJ431609 | 210 | 98 | NT_030667 (AC067804) | 4p16.1 |
| 431 | 175 | AJ431610 | 175 | 100 | NT_024524 (AL138964) | 13q21.23 |
| 407 | 150 | AJ431611 | 148 | 98 | NT_006725 (AC010316) | 5q35.3 |
| 290 | 272 | AJ431612 | 268 | 98 | NT_011548 (AL022578) | Xp11.3 |
| | | | 272 | 100 | NT_028053 (AL096774) | 1p36 |
| 505 | 249 | AJ431614 | 248 | 93 | NT_008803 (AL137223) | 10q23.32 |
| 466 | 385 | AJ431617 | 125 149 | 99 92 | NT_005567 (AC023402) | 3q22.3 |
| 477 | 384 | AJ431615 | 372 | 94 | NT_023558 (AC004848) | 7q21.11 |
| 467 | 296 | AJ431616 | 280 | 96 | - | 1,2,3,4,5,6,7, 11,12,15,18,19 |
| 421 | 322 | AJ431618 | 326 | 98 | NT_005102 (AC079248) | 2q22.1 |
| 475 | 320 | AJ431619 | 306 | 99 | NT_029311 (AL390074) | 6q21 |
| 423 | 337 | AJ431620 | 338 | 98 | NT_005413 (AC067945) | 2q32.3 |

Три из двенадцати INT маркеров локализованы поблизости от сайтов ломкости хромосом (fragile sites): INT254 – FRA14B (14q23), INT431 – FRA13C (13q21.2) и INT505 – FRA10A (10q23.3 или 10q24.2). Другие авторы также указывают на наличие сайтов ломкости хромосом вблизи мест интеграции HPV [Koopman et al., 1999; Wilke et al., 1996; Cannizzaro et al., 1988; Thorland et al., 2000; Wentzensen et al., 2002]. В большинстве случаев сайты ломкости хромосом локализовывались цитогенетически с точностью до определенного «бэнда» (а в некоторых случаях и с меньшей точностью). При таком уровне разрешения совместная локализация мест интеграции вируса и «fragile»-сайта может быть совпадением. Однако, в одной из последних работ, посвященных этому вопросу [Wentzensen et al., 2002], на основе сравнительного анализа нуклеотидных последовательностей, прилегающих к сайтам интеграции HPV с базой данных черновых последовательностей генома человека, было показано, что интеграция вируса вблизи сайтов ломкости хромосом происходит почти в 70% случаев. По нашим данным этот процент существенно ниже (25%).

Данные по сходству маркеров INT и геномных последовательностей человека подтверждают сделанное по результатам RH-картирования предположение, что интеграция вируса хотя и не носит хромосом-специфичного характера, но происходит в области генома, насыщенные генами (рис. 8).

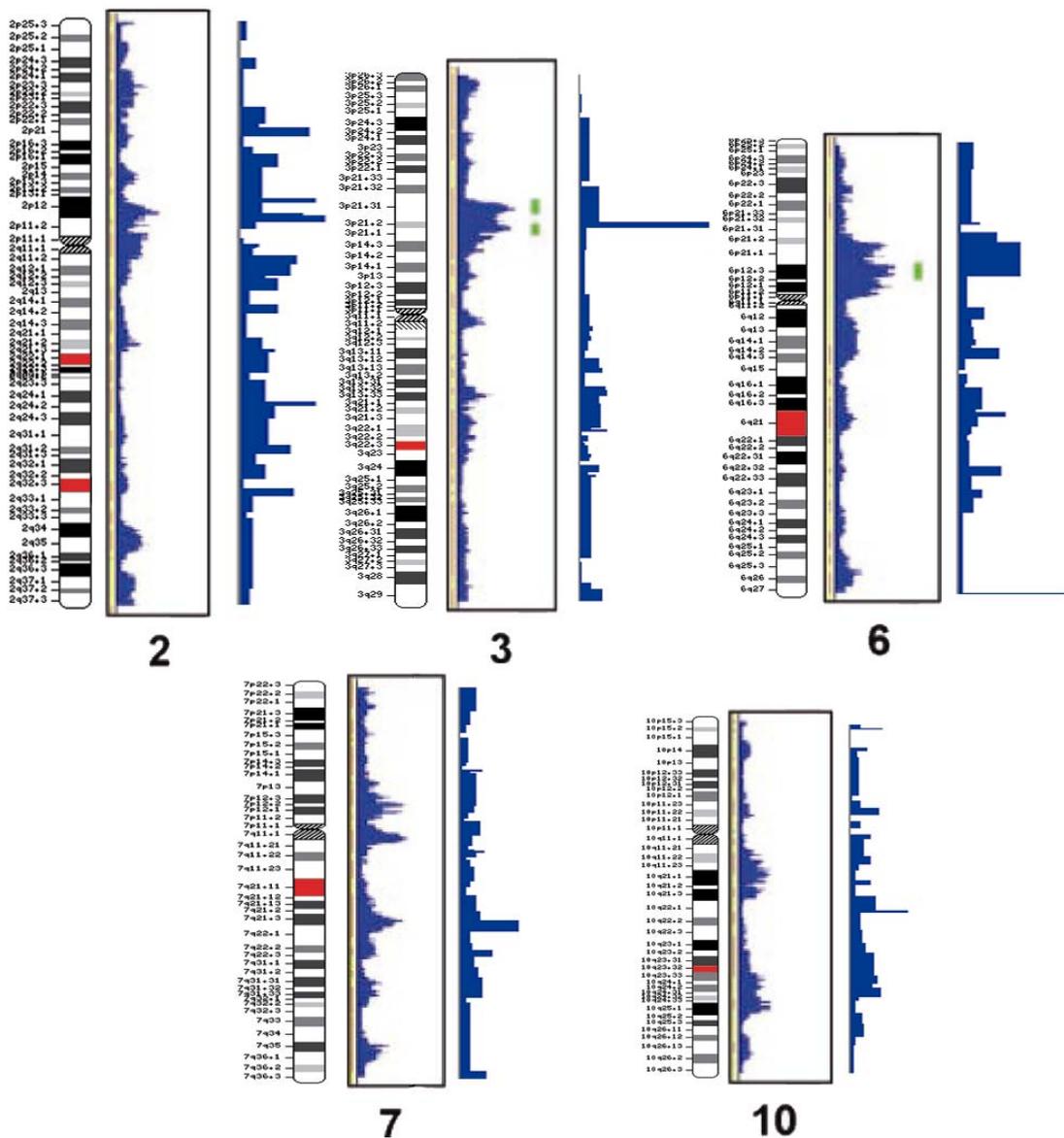


Рисунок 8. Локализация маркеров INT по данным картирования *in silico* (красные бенды на идиограмме) в районах хромосом, насыщенных генами. Слева на право: идиограмма (G-banding), транскрипционная карта, опубликованная в журнале «Science» [Caron et al., 2002], карта насыщения генами, размещенная на сервере GeneMap'99.

Учитывая насыщенность районов интеграции вирусной ДНК генами, был проведен поиск гомологичных маркерам INT экспрессирующихся последовательностей.

3. Поиск гомологий последовательностей, прилегающих к местам встраивания HPV16, с экспрессирующимися последовательностями человека.

Поиск экспрессирующихся нуклеотидных последовательностей человека, сходных с маркерами INT, проводили с помощью программы BLASTN по базам данных EST. Достоверным считали уровень сходства более 90% при длине идентичной последовательности не менее 100 нуклеотидов.

Только для пяти из 12 проанализированных маркеров INT были обнаружены клоны кДНК, имеющие достоверное сходство, что может быть в какой-то мере обусловлено небольшим размером маркеров (150-350 п.н.). При этом практически для всех маркеров были выявлены сходные геномные последовательности (см. главу 2, табл. 8). Это позволило существенно расширить возможности поиска гомологичных EST, за счет использования в анализе более протяженным (4 т.п.н.) клеточных последовательностей, фланкирующих интегрированную вирусную ДНК с 3'- и 5'- концов. Поиск проводили следующим образом:

а) проводили поиск геномных последовательностей (глава 2), включающих последовательности, идентичных маркерам INT;

б) отбирали геномные фрагменты, фланкирующие маркер INT с 5'- и 3'- конца, протяженностью 2000 п.н. с каждого конца;

в) осуществляли поиск гомологий с генами и кДНК отобранных таким образом фрагментов геномной ДНК человека с использованием общедоступных баз данных (dbEST, GeneBank, UniGen).

В тех случаях, когда сходство с одним и тем же геном или EST выявляется как непосредственно с маркерами INT, так и с прилегающими к ним последовательностями, можно считать достоверным встраивание HPV16 в соответствующие гены человека в опухолевых клетках плоскоклеточного рака шейки матки (маркеры INT290, INT505, INT466, INT423). Наиболее яркие примеры – это встраивание вируса в экзоны генов *LOC152028* (INT466), *LOC158537* (INT290) и *GLS* (INT423). Особо следует отметить сайт интеграции вируса, маркированный INT466. В этом случае ДНК вируса интегрировала в пятый экзон гена, подобного гену альфа/бета рецептору интерферона. Одна

часть маркера идентична участку пятого экзона этого гена, другая – последовательности шестого экзона. Это говорит о том, что слитый вирус-клеточный транскрипт подвергся сплайсингу. Для маркера INT467 показан высокий уровень сходства с кДНК человека, подобной одному из мультикопийных генов рибосомальных белков (*RPS27*).

Для четырех маркеров (INT259, INT477, INT421 и INT475) сходство с последовательностями генов были выявлены только для протяженных последовательностей, фланкирующих сайт интеграции вируса. В этих случаях вирус встраивается в интроны генов. В трех случаях из 12 не было выявлено EST, гомологичных маркерам INT или прилегающим к ним геномным последовательностям (INT254, INT431 и INT407).

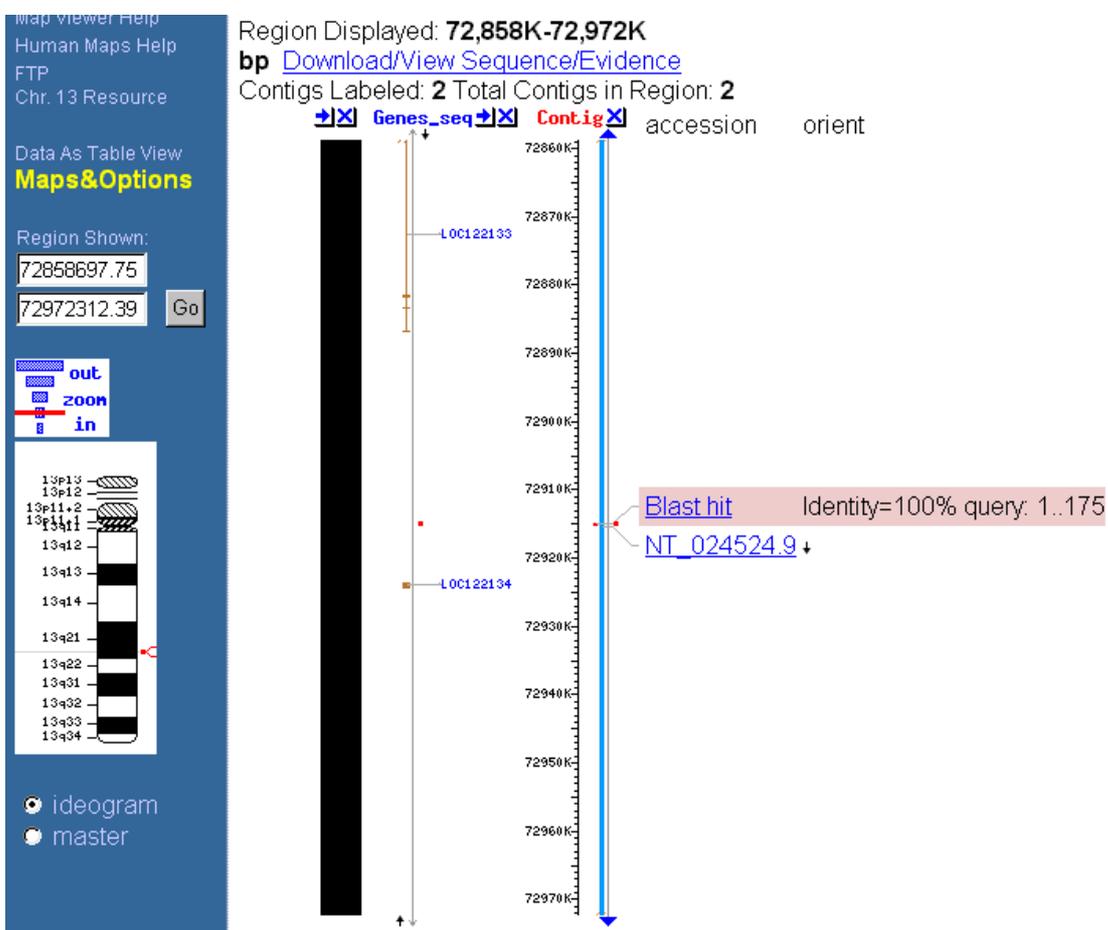


Рисунок 9. Пример построенной сервисом MapViewer для маркера INT431 интегрированной карты хромосомы 13 человека (фрагмент). Локализация маркера указана красным. Слева на право: идиограмма (G-бендинг), генная карта, карта геномных контигов.

Для всех маркеров INT с помощью сервиса MapViewer были построены интегрированные карты, а именно карта геномных контигов и генная карта. Это позволило определить номера экзонов и интронов, в которые интегрировала вирусная ДНК (табл. 9). Для маркеров INT254 и INT407 было установлено, что интеграция вирусной ДНК произошла в интроны гипотетических генов, а в случае с маркером INT431 – в межгенный участок (между гипотетическими генами *LOC122134* и *LOC122133*) (рис. 9).

Как видно из табл. 9, места интеграции равномерно распределены между интронами и экзонами. Только в одном случае вирусная ДНК интегрировала в межгенный участок генома, однако, вблизи генов (рис. 9).

Таблица 9. Гомологии маркеров INT и прилегающих последовательностей с генами, количество экзонов гомологичных генов, номера экзонов или интронов, в которые интегрировала ДНК HPV16.

| INT | Гомологичные гены человека | Кол-во экзонов | Экзон или интрон |
|-----|--|----------------|--------------------|
| 254 | <i>LOC161154</i> | 6 | Интрон 2 |
| 259 | <i>KIAA1808</i> | 17 | Интрон 12 |
| 431 | Между <i>LOC122134</i> и <i>LOC122133</i> | - | - |
| 407 | <i>KIAA0887</i> | 11 | Интрон 1 |
| 290 | <i>LOC158537</i> , подобный <i>WASF2</i> человека | 3 | Терминальный экзон |
| 505 | Ген миоферлина человека, <i>MYOF</i> | 53 | Терминальный экзон |
| 466 | <i>LOC152028</i> , подобный гену альфа/бэта рецептора 1 человека | 7 | Экзоны 5 и 6 |
| 477 | <i>LOC168295</i> | 17 | Интрон 5 |
| 467 | мРНК для гена рибосомального белка S27 (<i>RPS27</i>) | - | Экзон |
| 421 | <i>LOC151128</i> | 4 | Интрон 3 |
| 475 | NOGO-связывающий митохондриальный белок (<i>NIMP</i>) | 8 | Интрон 6 |
| 423 | Ген глутаминазы (<i>GLS</i>) | 18 | Терминальный экзон |

Все кДНК, гомологичные маркерам INT и прилегающим геномным последовательностям, представлены в базах данных в большом количестве (5-20 клонов), что указывает на высокий уровень их экспрессии, и получены из различных органов и тканей человека, как в норме, так и при патологиях (включая и опухолевые образования).

Хотя при анализе генов, в последовательности которых интегрирует ДНК HPV16, на первый взгляд, не обнаружено видимой закономерности, тем не

менее, обращает на себя внимание факт, что вирусная ДНК встраивается в участки генома, содержащие активно транскрибируемые последовательности. Это согласуется с выдвинутой ранее гипотезой о встраивании вируса в районы, насыщенные активно работающими генами.

Для проверки этой гипотезы информация о местах локализации интегрированной ДНК HPV различных типов, опубликованная в мировой литературе, была объединена с информацией, полученной в данной работе (рис. 10). Совмещение карты распределения интегрированной вирусной ДНК и транскрипционной карты ранее никем не проводилось.

Как видно из рис. 10, большинство мест интеграции приходится на пики активной транскрипционной активности последовательностей генома человека. К сожалению, методы, использованные авторами цитируемых работ [Koorman et al., 1999; Sastre-Garau et al., 1990; Gilles et al., 1996; Wilke et al., 1996; Cannizzaro et al., 1988; Thorland et al., 2000; Wentzensen et al., 2002], не позволяют указать локализацию последовательностей интегрированной ДНК вируса точнее определенного «бэнда» хромосомы. В работе Wentzensen с соавторами экспрессирующиеся вирус-клеточные последовательности локализовывали *in silico*. Однако авторы цитируемой работы указывают локализацию проанализированных ими последовательностей до хромосомного «бэнда», что не позволяет сопоставить пики транскрипционной активности на построенных нами интегрированных картах хромосом человека с полученными авторами данными о местах интеграции вируса. Отмечая наличие сходства 50% «слитых» вирус-клеточных транскриптов с EST человека, авторы не оценивают уровень транскрипции генов человека, расположенных в местах интеграции активно транскрибируемого в клетке интегрированного вирусного. Следует отметить, что за рамками анализа остались также нетранскрибируемые интегрированные вирусные последовательности, доля которых в клетке может варьировать [Van Tine et al., 2001]. Но поскольку решающим фактором для иммортализации клеток является активность вирусных генов *E6* и *E7*, предполагается, что нетранскрибируемая интегрированная ДНК HPV сама по себе не участвует в развитии опухоли [Zur Hausen, 1999; Szarka et al., 2000; Киселев и др., 2001].

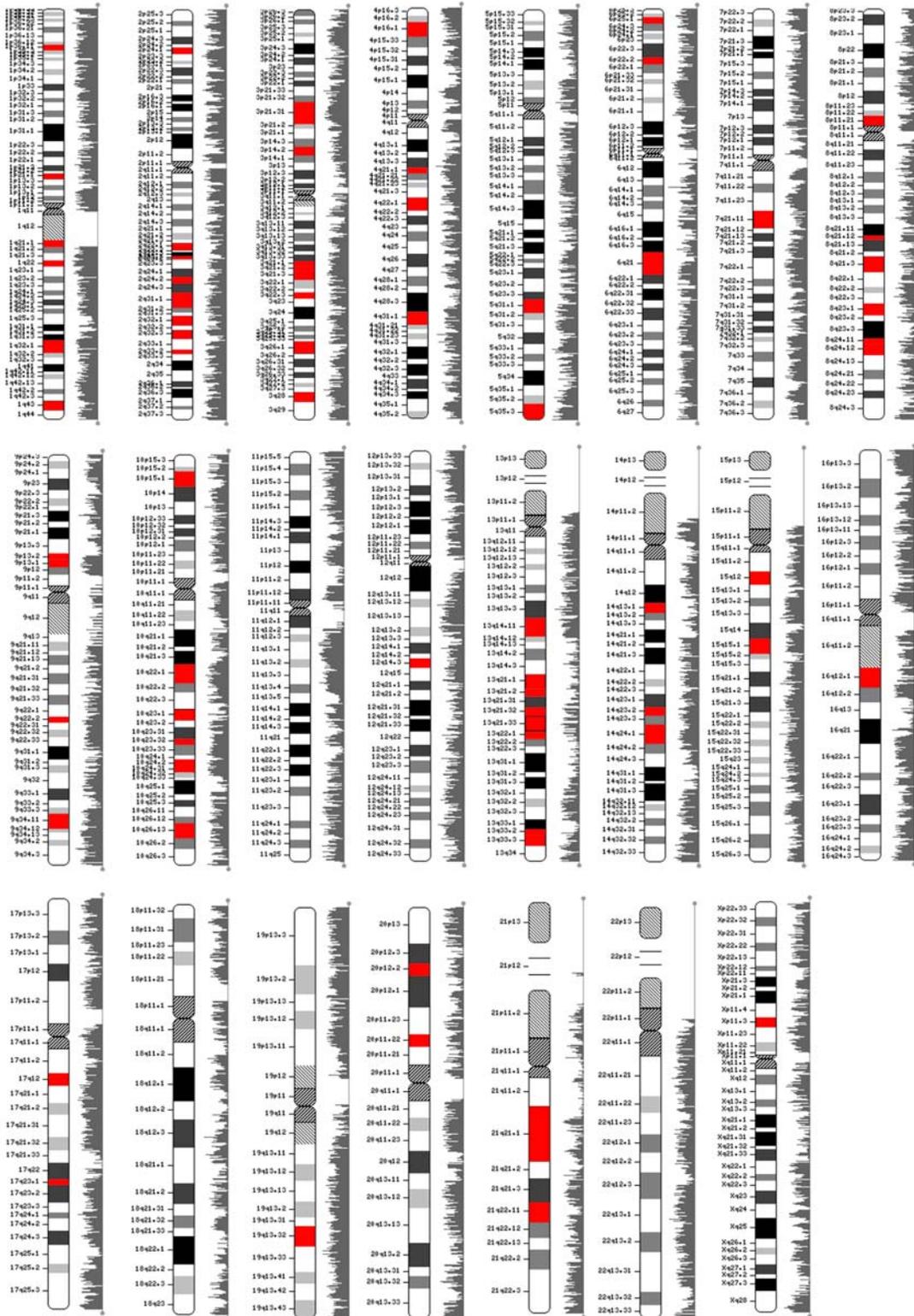


Рисунок 10. Локализация различных типов HPV (в основном HPV16): литературные и собственные данные. Карты (слева – идиограмма, справа – транскрипционная карта) совмещены сервисом MapViewer. Красным выделены бэнды, в которые интегрирует вирусная ДНК.

Учитывая выше изложенное, можно сделать вывод, что интеграция HPV16 может происходить в различные области генома человека. Встраивание ДНК вируса в активно транскрибируемые районы можно объяснить тем, что клеточная ДНК в таких районах находится в деконденсированном состоянии, что облегчает интеграцию чужеродного генетического материала и последующее связывание молекул транскрипционных факторов и РНК-полимеразы с интегрированной ДНК, что обеспечивает ее активную транскрипцию в таких районах. Остается неясным, происходит ли встраивание вирусной ДНК только с разрывом нуклеотидной цепи или с замещением части клеточной ДНК. Получение ответа на этот вопрос требует проведения дополнительных экспериментов.

4. Анализ нуклеотидных последовательностей генома человека, прилегающих к местам встраивания HPV16.

Как следует из предыдущих глав, все изученные нами места интеграции вирусной ДНК расположены в насыщенных генами районах. Однако остался открытым вопрос о возможной специфичности или особенностях нуклеотидных последовательностей в местах встраивания HPV16.

4.1. Анализ нуклеотидных последовательностей генома человека, прилегающих к местам встраивания HPV16, с использованием сервиса Genomatrix Suite.

Данный сервис включает в себя программы поиска и последующего анализа нуклеотидных последовательностей человека, подобных исследуемым. Результаты выводятся в виде таблиц или в графически-интерактивном виде, т.е. пользователь может менять вид изображения не внося изменения в исходные данные. Графическое изображение представляет собой совмещение карт: геномных контигов и наложенную на нее транскрипционную карту; на карте отмечены сайты связывания факторов транскрипции, промоторы (определяются программой PromotorInspector [Scherf et al., 2000]), повторяющиеся элементы генома и матрикс-ассоциированные последовательности (S/MAR) [Frisch et al.,

2002; Liebich et al., 2002]. От карты можно перейти на описание представленных на ней элементов. Вся процедура получения результатов полностью автоматизирована, от исследователя требуется только ввести исследуемую нуклеотидную последовательность или ее регистрационный номер, если она зарегистрирована в международных базах данных. На рис. 11 представлен пример графического отображения результатов работы сервиса Genomatrix Suite.

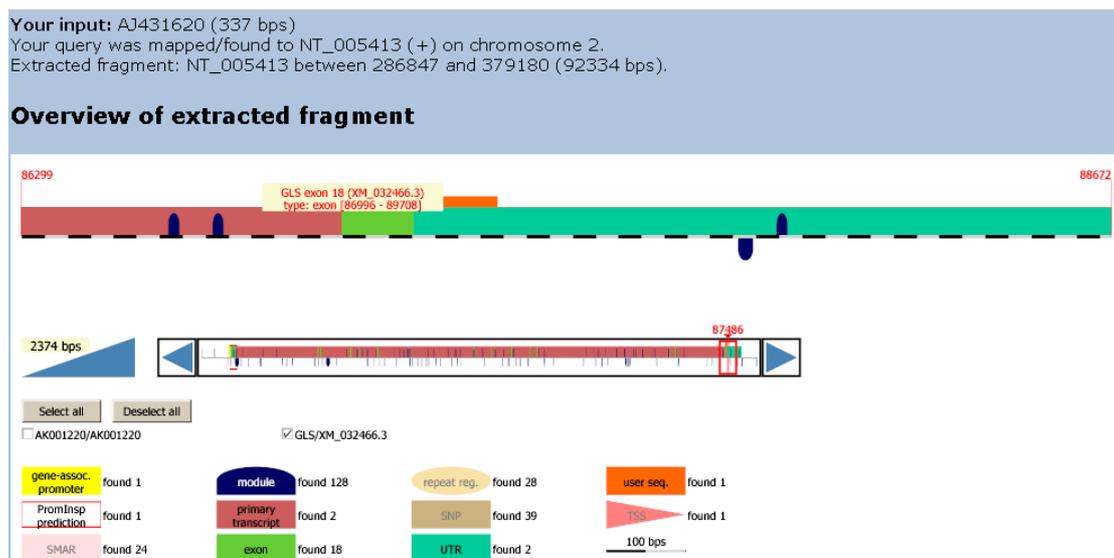


Рисунок 11. Результат работы сервиса Genomatrix Suite для маркера INT423.

Результаты поиска подобных маркерам геномных и экспрессирующихся нуклеотидных последовательностей подтвердились при использовании сервиса Genomatrix Suite (табл. 10). Следует отметить, что детальный анализ последовательностей генома проводится только при уровне сходства исследуемой последовательности с геномной не менее 95%. Из табл. 10 видно, что в тех случаях, когда уровень гомологии ниже 95%, анализ не проводился (в колонке «Ген...» стоит «no data»).

Таблица 10. Результаты поиска гомологичных геномных и экспрессирующихся нуклеотидных последовательностей с использованием сервиса Genomatrix Suite.

| INT | Гомология с последовательностями генома человека | | | | Ген, в последовательность которого произошла интеграция ДНК HPV16 | |
|-----|--|-----------------|------------------|---------|---|--------------------------------|
| | % | Хромосома | Геномные контиги | Цепь | Название | Экзон или интрон №/общее число |
| 254 | 95 | 14 | NT_010028 | + | <i>LOC161154</i> | Интрон 2/5 |
| 259 | 98 | 4 | NT_030667 | + | <i>LOC152752 (stKIAA1808)*</i> | Экзон 11/11 |
| 431 | 100 | 13 | NT_024524 | + | - | - |
| 407 | 97 | 5 | NT_006725 | + | <i>KIAA0887</i> | Интрон 1/10 |
| 290 | 100 | 1 | NT_021945 | + | <i>WASF2</i> | Экзон 3-4/8 |
| | 97 | X | NT_011584 | - | <i>LOC158537</i> | Экзон 1/1 |
| 505 | <95 | 10 | NT_030084 | + | no data | no data |
| 466 | <95 | 3 | NT_005567 | - | no data | no data |
| 477 | <95 | 7 | NT_023558 | - | no data | no data |
| 467 | 96 | 7 | NT_017168 | + | <i>LOC154648</i> | Экзон 1/1 |
| | 95 | 1 | NT_021933 | + | <i>RPS27 (MPSI)</i> | Экзон 2-4/4 |
| | <95 | 3,4,11,12,15,19 | no data | no data | no data | no data |
| 421 | 98 | 2 | NT_005102 | + | <i>LOC151128</i> | Интрон 3/4 |
| 475 | 99 | 6 | NT_029311 | + | <i>NIMP</i> | Интрон 6/8 |
| 423 | 98 | 2 | NT_005413 | + | <i>GLS</i> | Экзон 18/18 |

*st – similar to, т.е. подобный

Поскольку анализ с помощью сервиса Genomatrix Suite проводился позже, можно заметить некоторые отличия в названиях генов, которые, тем не менее, не являются существенными. Так для маркера INT259 подобной оказалась последовательность гена *LOC152752*, подобного гену *KIAA1808*. А для маркера INT467 – последовательность гена *RPS27*, альтернативное название – *MPSI*. Таким образом, эти изменения являются лишь следствием изменения номенклатуры генов в банках данных.

Ни для одного из маркеров INT не было обнаружено гомологий с повторяющимися элементами генома человека (наиболее близко встречались Aи-повторы и Lane-1 элементы), а также с матрикс-ассоциированными последовательностями генома человека (MAR). Данные участки генома были обнаружены на расстоянии не менее 500 нуклеотидов от маркеров INT. Другие авторы показывают наличие повторяющихся элементов генома вблизи мест интеграции HPV [el Awady et al., 1987; Carmody et al., 1996; Shera et al., 2001].

Геном человека содержит большое количество повторяющихся элементов, например, Alu-повторы представлены в количестве $\sim 1 \cdot 10^6$ копий на геном и должны встречаться через каждые 3 – 6 т.п.н. (что, не всегда соответствует действительности, поскольку наблюдается неравномерность их распределения в геноме) [Lane et al., 1992; International Human Genom Sequencing Consortium, 2001]. Line-элементы, представляющие класс ретротранспозонов, представлены $\sim 5 \cdot 10^5$ копиями на геном [Lengauer et al., 1990; International Human Genom Sequencing Consortium, 2001]. При подобной насыщенности генома повторяющимися элементами, вероятность интеграции в них вирусной ДНК достаточно велика, однако, по-видимому, эти вирусные ДНК не транскрибируются. На это может указывать тот факт, что нами не было отмечено ни одного случая интеграции вирусной ДНК вблизи повторов. Несмотря на небольшой размер выборки, проанализированной нами, мы должны были бы отметить эти случаи, учитывая высокую представленность различных типов повторов в геноме человека.

Тем не менее, в проанализированных нами случаях интеграция вируса происходит только в активно транскрибируемые последовательности генома. Это может служить дополнительным основанием для утверждения о некоторой избирательности интеграции ДНК HPV16 в геном человека.

4.2. Анализ геномных последовательностей человека, прилегающих с 3'-конца к интегрированной вирусной ДНК.

Из пункта 4.1 следует, что предпочтительное встраивание вируса в экзоны или интроны генов выявлено не было. Для тех случаев, когда вирус встраивается в терминальные экзоны генов, терминация транскрипции интегрированной ДНК вируса идет до сайта полиаденилирования, принадлежащего соответствующему гену. Так, по-видимому, происходит в случае маркеров INT290, INT505 и INT423, гомологичных терминальным экзонам генов *LOC158537*, *MYOF* и *GLS*, соответственно. Метод АРОТ, использованный для выделения «слитых» вирус-клеточных транскриптов, не позволяет выделять последовательности большой длины. Поэтому в случаях встраивания вируса во внутренние экзоны, мы можем дискриминировать только те транскрипты, терминация которых должна заканчиваться после внутреннего

экзона, в который произошла интеграция. Случай с маркером INT466, гомологичным последовательностям двух экзонов одного гена, доказывает это предположение: имеет место сплайсинг первичной вирус-клеточной РНК, но терминация транскрипции происходит не после последнего экзона, а после 6-го экзона, являющегося внутренним.

Однако остается неизвестным, где происходит терминация транскрипции в случае встраивания вирусной ДНК в интроны и средние экзоны генов, а также в межгенные пространства. Чтобы ответить на этот вопрос, был предпринят поиск сайтов терминации транскрипции на расстоянии до 2,5 т.п.н. от 3'-конца встроенной ДНК HPV16. Анализ первоначально проводился «вручную», т.е. без использования специальных программ, а затем полученные результаты были проверены с помощью программы POLYA SCAN, достаточно точно предсказывающей сайты полиаденилирования в последовательностях геномной ДНК высших эукариот.

Таблица 11. Наличие последовательностей, способных терминировать транскрипцию ДНК HPV16 при ее встраивании в интроны, внутренние экзоны и межгенные пространства.

| INT | Наличие необходимых последовательностей | | Терминальный экзон (GenScan) | Расстояние до PolyA-сайта, н.п. |
|-----|---|------------|------------------------------|---------------------------------|
| | анализ «вручную» | POLYA SCAN | | |
| 254 | + | + | + | 638 |
| 259 | + | + | - | 907 |
| 431 | + | + | - | 166 |
| 407 | + | + | - | 232 |
| 290 | + | + | - | 1612 |
| 505 | + | + | - | 552 |
| 466 | + | + | - | 1270 |
| 477 | + | + | - | 234 |
| 467 | - | - | - | - |
| 421 | + | + | - | 266 |
| 475 | + | + | - | 168 |
| 423 | + | + | + | 332 |

Во всех случаях в исследованных «вручную» последовательностях были найдены сайты полиаденилирования и тимидин-богатые участки, способные терминировать транскрипцию. Использование программы POLYA SCAN полностью подтвердило полученные результаты (табл. 11). С помощью программы GenScan был проведен поиск возможных терминальных экзонов в

исследуемых последовательностях. В табл. 11 также указано расстояние от места состыковки вирусной и клеточной ДНК до ближайшего с 3'-конца сайта полиаденилирования.

Применение программы GenScan позволило найти терминальный экзон только для двух последовательностей (табл. 11). Это можно объяснить либо недостаточностью данных в исследуемых последовательностях (малая длина), либо малыми возможностями предсказания участков терминации транскрипции использованной программой. Последнее более вероятно, так как в настоящее время накоплено еще мало данных о последовательностях, терминирующих транскрипцию у эукариот.

Тем не менее, показано наличие сайта полиаденилирования далее от 3'-конца встроенной вирусной ДНК на расстоянии до 1,7 т.п.н. (табл. 11). По-видимому, вирус при интеграции, теряя собственный сайт полиаденилирования, использует альтернативные клеточные сигналы терминации транскрипции. Вероятность этого уже обсуждалась в литературе по интеграции вирусной ДНК [Киселев и др., 2001], однако анализ клеточных последовательностей примыкающих к 3'-концу встроенной вирусной ДНК ранее не проводился.

5. Анализ гипотетических генов.

Для выяснения природы представленных в базах данных транскриптов (гипотетических генов), нуклеотидные последовательности пяти гипотетических генов *LOC161154*, *LOC151128*, *LOC158537*, *KIAA1808* и *KIAA0887* были проанализированы и описаны с целью установления их точной структуры, а также возможных функций кодируемых ими белков.

5.1. Установление экзон-интронной структуры и поиск необходимых регуляторных элементов.

С помощью программ BLASTN и GenScan для последовательностей этих генов построена экзон-интронная структура. Так как в двух случаях из пяти выявились ошибки в состыковке клонов, составляющих геномные контиги, большая часть работы была проведена «вручную», т.е. без использования компьютерных программ. Это позволило избежать появления программных

ошибок и более точно выверить последовательности экзонов и расположение регуляторных элементов. Используя программы PromotorInspector и MatInspector, в 5' областях генов найдены промоторные участки, содержащие сайты связывания для основных факторов транскрипции. Для всех генов найдены стартовые кодоны (ATG) и стоп кодоны, а также сайты полиаденилирования. Все экзоны описанных генов фланкированы каноническими сайтами сплайсинга (AG...GT). Результаты определения экзон-интронной структуры и необходимых для работы гена регуляторных элементов представлены в табл. 12 (в колонке «Длина гена» – расстояние в н.п. от промотора до конца сайта полиаденилирования). Факторы транскрипции и последовательности в промоторе, служащие местами их посадки, вынесены в приложение.

Таблица 12. Основные регуляторные элементы генов, количество экзонов, гомологии с белками.

| Гены | Длина гена, н.п. | Стоп-кодоны | PolyA-сайты | Число экзонов | Гомологичные белки |
|------------------|------------------|-------------|-------------|---------------|----------------------------|
| <i>LOC161154</i> | 35590 | TGA | AATAAA | 6 | - |
| <i>KIAA1808</i> | ~211602* | TAA | AATAAA | 17 | ABLIM 57% |
| <i>LOC158537</i> | 9185 | TAA | AATAAT | 3 | WASF2 88% |
| <i>KIAA0887</i> | ~35857* | TGA | AATAAA | 11 | Faf-P1 (<i>D.m.</i>) 37% |
| <i>LOC151128</i> | ?** | TAG | AATAAA | 5 | - |

* – поскольку нуклеотидные последовательности данных генов принадлежат к разным геномным контигам, точную длину генов определить невозможно;

** – в процессе построения экзон-интронной структуры гена выявилась ошибка в состыковке контига, вследствие чего определить длину гена невозможно.

Используя сервис Genomatrix Suite, удалось получить информацию об исследуемых последовательностях генов. Программа проводит выравнивание имеющихся в базах данных кДНК с геномными последовательностями (аналогично программе BLASTN). Как видно из табл. 13, использование компьютерных программ не дает достоверной информации о наличии промоторных областей и количестве экзонов. Меньшее количество экзонов связано с использованием программой представленных в базах данных кДНК. Однако известно, что получить полноразмерную кДНК очень сложно, особенно для протяженных транскриптов.

Таблица 13. Анализ гипотетических генов с использованием сервиса *Genomatrix Suite*.

| Ген | Промотор | | TSS | Модули в промоторной области (только для GAP) | Число экзонов | Наличие белкового продукта |
|------------------|----------|-----|-----|--|---------------|----------------------------|
| | GAP | PIP | | | | |
| <i>LOC161154</i> | 1s | 1 | 0 | - | 6 | + |
| <i>KIAA1808</i> | 1b | 0 | 0 | NFKB_APIF_01 | 11 | + |
| <i>LOC158537</i> | 1b | 0 | 0 | - | 1 | + |
| <i>KIAA0887</i> | 1s | 2 | 0 | RORA_GATA_01 APIF_NFAT_01 CEBP_STAT_01 ETSF_ETSF_01 ETSF_HAML_01 ETSF_CREB_01 APIF_ETSF_02 | 11 | + |
| <i>LOC151128</i> | 1b | 0 | 0 | - | 5 | + |

Примечания.

GAP – Gene-associated Promotor (промотор, ассоциированный с геном; g, s, b – градации достоверности, по степени уменьшения)

PIP – PromoterInspector Prediction (промотор, предсказанный программой PromoterInspector)

TSS – Transcription Start Site (сайт начала транскрипции)

Все эти элементы определяются программно, следовательно, они могут быть вероятными, но не достоверными.

Таким образом, нами построены наиболее вероятные экзон-интронные структуры для пяти гипотетических генов.

5.2. Определение предположительной функции белков, кодируемых описанными генами.

Аминокислотные последовательности белков, кодируемые изучаемыми генами, получены трансформацией нуклеотидных последовательностей генов (*in silico* трансляцией) с помощью программы DNA=>Protein.

Методом поиска гомологий (см. табл. 12) и построением карт локального сходства определена предположительная функция для трех гипотетических белков (продукта гена *LOC158537*, белки *KIAA1808* и *KIAA0887*).

Продукт гена *LOC158537* имеет высокий уровень сходства, с белком человека *WASF2* (88%), позволяющий считать эти белки гомологами, принадлежащими к одному семейству. Этот белок входит в семейство ГТФаз –

переносчиков сигнала к актину цитоскелета (WAS family). Белок KIAA1808 подобен последовательности белка человека ABLIM (LIM actin-binding protein 1) (57%), содержащего LIM домен, играющий ключевую роль в регуляции развития. Белок KIAA0887 имеет достоверный уровень сходства с белком Faf-P1 (37%) дрозофилы, содержащим UBX домен.

Полученные данные позволяют предположить, что продукты исследованных нами генов выполняют в клетках те же функции, что и сходные с ними белки.

На основании гомологии белковых продуктов генов *WASF2* и *LOC158537*, ген *LOC158537* назван нами *WASF4* (т.к. описано уже 3 гена семейства WAS) и его нуклеотидная последовательность зарегистрирована в базе данных EMBL как последовательность нового гена, согласно построенной в данной работе экзон-интронной структуре (AccN AJ509090).

5.3. Анализ экспрессии изучаемых генов в различных органах и тканях.

Для доказательства экзон-интронной структуры описанных генов и наличия их экспрессии в клетках человека была проведена ОТ-ПЦР с праймерами к экзонам четырех генов *LOC161154*, *LOC151128*, *KIAA1808* и *KIAA0887*. В качестве матрицы для синтеза первой цепи ДНК была использована РНК, выделенная из крови (1), из опухоли яичников человека (2), прилегающей к ней нормальной ткани (3) и нормальной ткани шейки матки (4). В табл. 14 отмечены предполагаемые длины продуктов ПЦР и ОТ-ПЦР с парами праймеров к экзонам изучаемых генов (последовательности праймеров см. в разделе «Материалы и методы», табл. 3).

Таблица 14. Пары праймеров к экзонам гипотетических генов, длины ожидаемых продуктов и наличие продуктов ОТ-ПЦР для использованных в работе проб РНК.

| Ген | Пары праймеров | Номера экзонов | Ожидаемый продукт для геномной ДНК, п.н. | Ожидаемый продукт для РНК, п.н. | Наличие продукта для РНК (ткань) |
|------------------|----------------|----------------|--|---------------------------------|----------------------------------|
| <i>LOC161154</i> | F1+R | 1 – 5 | 29790 | 712 | - |
| | F2+R | 3 – 5 | 7763 | 521 | + (1) |
| <i>KIAA1808</i> | F1+R | 11 – 18 | 184020 (?) | 578 | + (3) |
| | F2+R | 15/16 – 18 | - | 204 | + (3) |
| <i>KIAA0887</i> | F+R1 | 6 – 7 | 265 | 166 | + (3,4) |
| | F+R2 | 6 – 11 | 12875 | 778 | - |
| <i>LOC151128</i> | F1+R | 1 – 4 | 30601 (?) | 329 | - |
| | F2+R | 2 – 4 | 27576 (?) | 228 | - |

? – невозможность точно установить длину фрагмента геномной ДНК в связи с неправильной состыковкой геномных контигов.

Для гена *LOC151128* не было получено продуктов амплификации на пробах РНК, использованных в работе. В данных тканях этот ген не экспрессируется, однако, это не означает отсутствие его экспрессии в других, не изученных нами, тканях организма.

При амплификации с праймерами к экзонам трех других генов (*LOC161154*, *KIAA1808* и *KIAA0887*) получены фрагменты кДНК ожидаемой длины. Экспрессия гена *LOC161154* была обнаружена только в крови, для гена *KIAA1808* экспрессия показана только в нормальной ткани шейки матки, а для гена *KIAA0887* – в нормальных тканях яичника и шейки матки (рис. 12 и 13).

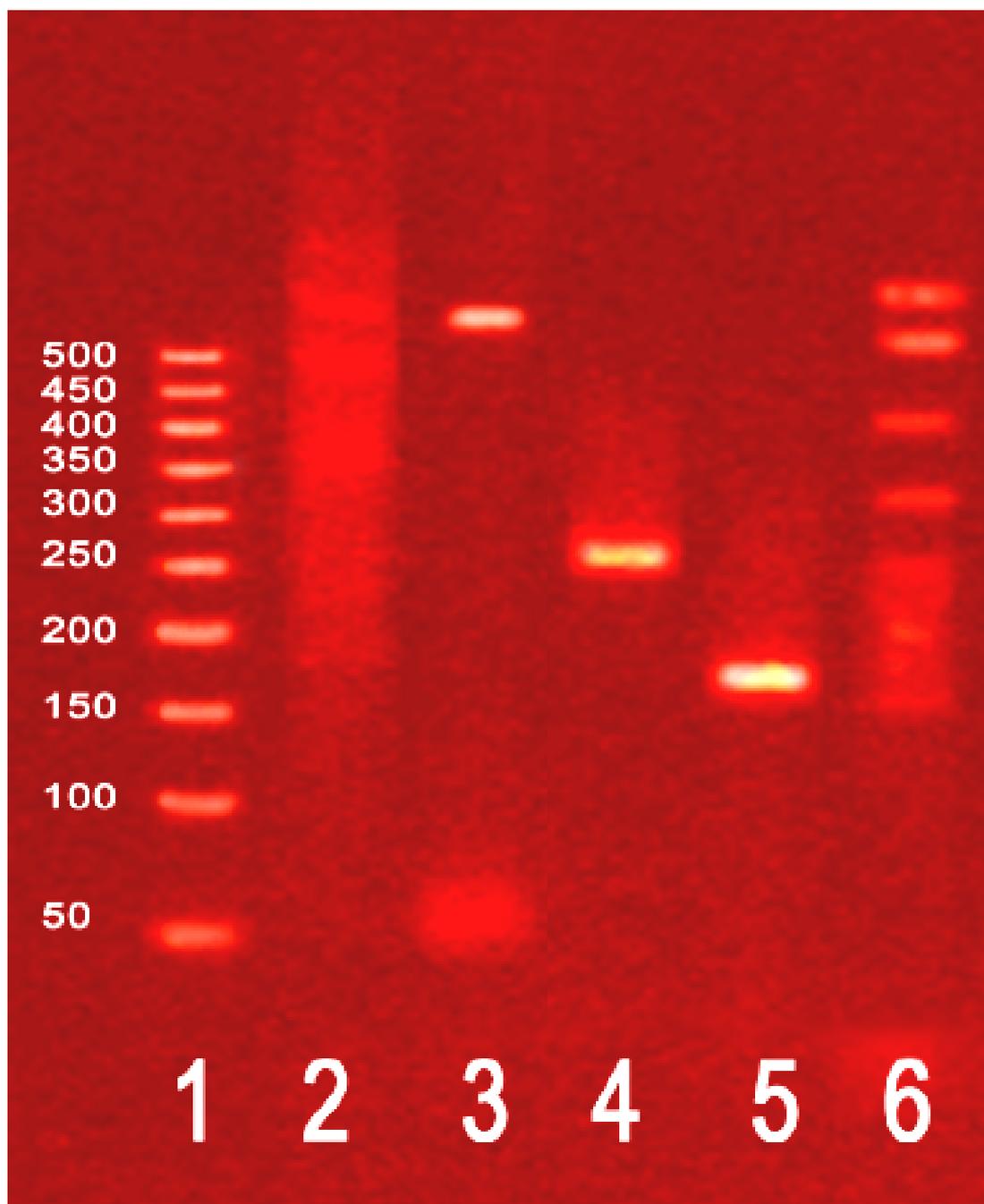


Рисунок 12. Электрофорез в 2%-м агарозном геле продуктов ОТ-ПЦР. Дорожка 1 – маркер молекулярных масс M50; праймеры к мРНК гена LOC161154 (экзоны 3 и 5); дорожка 2 – проба геномной ДНК человека, дорожка 3 – РНК из клеток крови; праймеры к мРНК гена KIAA0887 (экзоны 6 и 7); дорожка 4 – проба геномной ДНК человека, дорожка 5 – РНК из клеток опухоли молочной железы. Дорожка 6 – маркер молекулярных масс (pBR322 / MspI).

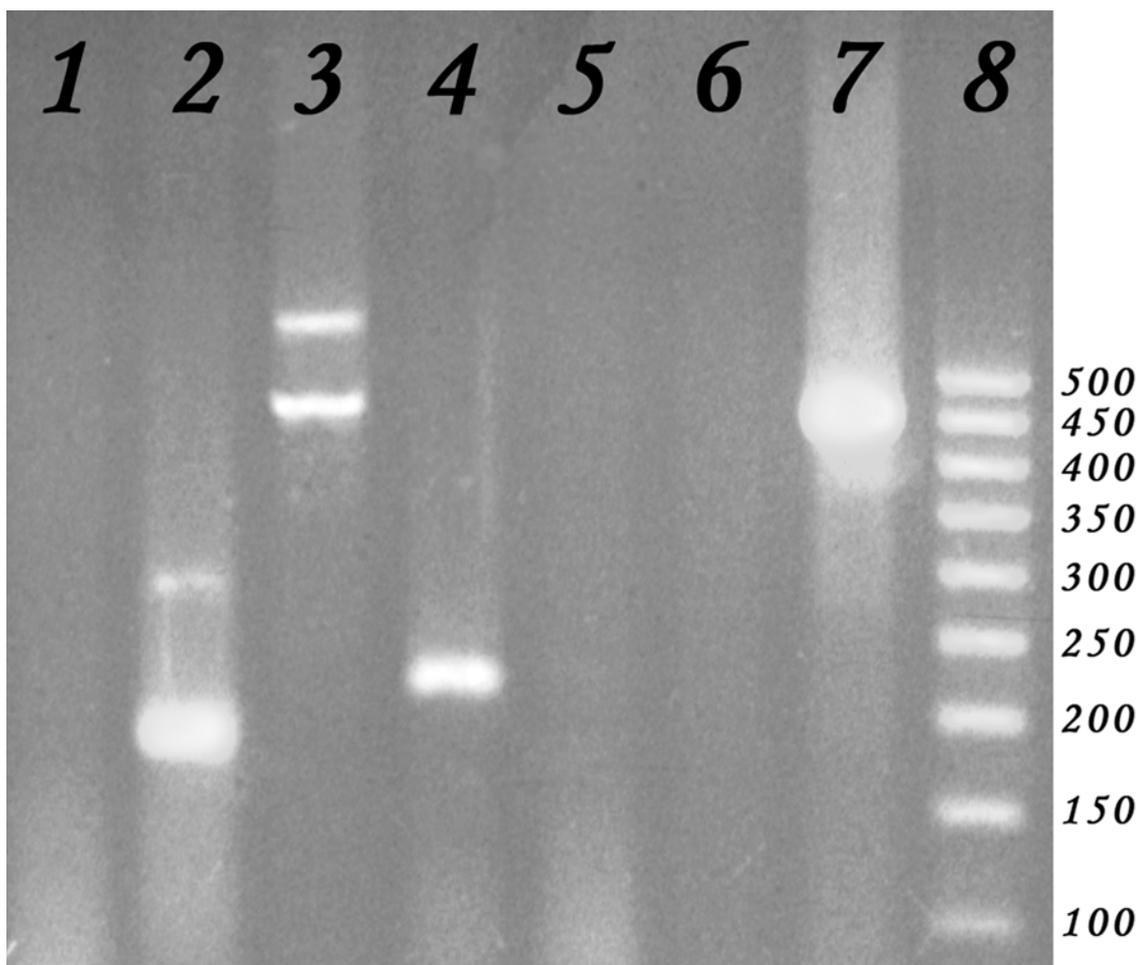


Рисунок 13. Электрофорез в 2%-м агарозном геле продуктов ОТ-ПЦР. В качестве матрицы – РНК и клеток шейки матки. Дорожка 1 – праймеры к мРНК гена LOC161154 (экзоны 3 и 5); дорожка 2 – праймеры к мРНК гена KIAA0887 (экзоны 6 и 7); дорожки 3 и 4 – праймеры к мРНК гена KIAA1808 (экзоны 11 и 18, 15/16 и 18, соответственно), дорожки 5 и 6 – праймеры к мРНК гена LOC151128 (экзоны 1 и 4, 2 и 4, соответственно). Дорожка 7 – праймеры к мРНК гена GAPDH (экспрессируется во всех тканях организма). Дорожка 8 – маркер молекулярных масс M50.

Таким образом, данные ОТ-ПЦР доказывают реальное существование и экспрессию гипотетических генов *LOC161154*, *KIAA1808* и *KIAA0887*. Совпадение размеров продуктов с ожидаемыми подтверждает правильность построенной нами экзон-интронной структуры этих генов.

ЗАКЛЮЧЕНИЕ

В процессе отработки метода картирования с использованием панели радиационных гибридов соматических клеток (RH-картирование) определена физическая локализация мозгоспецифичной последовательности Hfb1, кодирующей протяженную 3'-нетранслируемую область мРНК для белка комплексина 2 человека, а также последовательности smgc30, гиперметилированной в клетках рака шейки матки. Результаты RH-картирования мозгоспецифичной последовательности Hfb1 совпали с известной ранее локализацией гена комплексина 2 (5q35). Последовательность smgc30 была локализована в области 9q31, позднее было показано наличие в этой области сходной последовательности нуклеотидов. Эти данные подтверждают правильность выбора метода RH-картирования, как одного из самых точных и эффективных методов тонкого физического картирования.

С помощью метода RH-картирования была определена физическая локализация пяти геномных последовательностей, прилегающих к местам встраивания транскрибируемой ДНК вируса папилломы тип 16 в хромосомы человека (маркеры INT) в клетках плоскоклеточного рака шейки матки. Повторная локализация *in silico* этих маркеров подтвердила результаты RH-картирования. Семь других маркеров INT были локализованы только *in silico*, что стало возможным после появления в банках данных более полного варианта черновых последовательностей генома человека.

Анализ мест локализации маркеров INT и их сходства с экспрессирующимися последовательностями генома человека позволил нам высказать предположение, о предпочтительном встраивании ДНК вируса в районы хромосом, характеризующиеся активной транскрипцией. Высказанное нами предположение согласуется с недавно опубликованными данными других авторов [Wentzensen et al., 2002]. Необходимо иметь в виду, что предположение о предпочтительном встраивании ДНК HPV16 в активно транскрибируемые области генома человека правомочно только для транскрибируемой ДНК вируса папилломы. Этот феномен можно объяснить тем, что активно транскрибируемые последовательности находятся в деконденсированном

состоянии, облегчающем интеграцию чужеродного генетического материала и последующую его транскрипцию.

Анализ *in silico* клеточных последовательностей, расположенных с 3'-конца далее от места интеграции, показал наличие в них клеточного сигнала терминации транскрипции (содержащего сайт полиаденилирования). Эти данные позволяют нам утверждать, что транскрипция интегрированной вирусной ДНК в случае встраивания ее в интроны клеточных генов и межгенные последовательности, происходит до ближайшего с 3'-конца клеточного сигнала терминации транскрипции.

Результаты проведенного в данной работе анализа мест интеграции ДНК вируса и прилегающих к ним геномных последовательностей человека позволяют лучше понять, как происходит взаимодействие вирусного и клеточного геномов, приводящее к возникновению рака шейки матки.

Проведен детальный анализ структуры и функции пяти гипотетических генов, являющихся местами встраивания ДНК HPV16. С помощью компьютерных программ определены их экзон-интронные структуры, найдены необходимые для функциональной активности генов элементы в 5'- и 3'-регуляторных областях и на границах экзонов. Методом поиска сходства удалось определить предположительную функцию для трех белковых продуктов изучаемых генов. Экспериментально (методом ОТ-ПЦР) показана экспрессия двух генов (*LOC161154* и *KIAA0887*) в тканях человека. На основании гомологии белковых продуктов генов *WASF2* и *LOC158537*, ген *LOC158537* назван нами *WASF4*, его нуклеотидная последовательность зарегистрирована в базе данных EMBL как последовательность нового гена (AccN AJ509090).

ВЫВОДЫ:

1. Установлена физическая локализация (методом RH-картирование) на хромосомах человека мозгоспецифичной последовательности Hfb1 (5q35) и гиперметилированной в клетках рака шейки матки последовательности cmgc30 (9q31).
2. Показано, что места интеграции транскрибируемой в клетках плоскоклеточного рака шейки матки ДНК вируса папилломы (тип 16) локализованы в активно транскрибируемых областях генома (хромосомы 1, 2, 3, 4, 5, 6, 7, 10, 13, 14, X).
3. Впервые, с помощью анализа *in silico*, показано наличие с 3'-конца от встроенной ДНК HPV16 альтернативных клеточных последовательностей, в норме не используемых клеткой, но способных терминировать транскрипцию интегрированной вирусной ДНК.
4. Определена структура пяти гипотетических генов человека (*LOC161154*, *KIAA1808*, *LOC158537*, *KIAA0887* и *LOC151164*), являющихся местами встраивания транскрибируемой ДНК HPV16. Нуклеотидная последовательность одного из проанализированных генов (*LOC158537*) зарегистрирована в базе данных EMBL, как последовательность нового гена *WASF4* (AccN AJ509090).
5. Определена на основе анализа *in silico* предположительная функция трех белков, кодируемых гипотетическими генами человека:
 - а) продукт гена *LOC158537* имеет высокий уровень сходства, с белком человека WASF2 (88%), позволяющий считать эти белки гомологами, принадлежащими к одному семейству ГТФаз – переносчиков сигнала к актину цитоскелета (WAS family);
 - б) продукт гена *KIAA1808* имеет сходство (57%) с белком человека ABLIM, содержащим LIM домен;
 - в) продукт гена *KIAA0887* имеет достоверный уровень сходства с белком Faf-P1 (37%) дрозофилы, содержащим UBX домен.
6. Экспериментально показана экспрессия трех гипотетических генов (*LOC161154*, *KIAA1808* и *KIAA0887*) в тканях человека; это позволяет утверждать, что они являются реальными, а не гипотетическими генами человека.

СПИСОК ЛИТЕРАТУРЫ

1. Аитова С.С. ДНК-маркеры хромосомы 13 человека: локализация, полиморфизм и использование для картирования // Диссертация на соискание ученой степени кандидата биологических наук. Москва. 1997.
2. Горбунова В.Н. Молекулярные основы медицинской генетики. // СПб. 1999. 212С.
3. Жданова Н.С. Радиационное картирование геномов. Настоящее и будущее // Генетика. 2002. Т.38. С.581-594.
4. Жимумев И.Ф. Современные представления о структуре гена у эукариот // Соросовский образовательный журнал. 2000. Т.6. С.17-24.
5. Забаровский Е.Р., Домнинский Д.А., Киселев Л.Л. Физическое картирование генома человека: на пути к выработке оптимальной стратегии // Молекулярная биология. 1994. Т.28. .
6. Захаров И.А. Генетические карты высших организмов // Л. Наука. 1979. 158С.
7. Захаров И.А., Никифоров В.С., Степанюк Е.В. Генетическое картирование сельскохозяйственных животных // Успехи современной генетики. 1993. Т.18. С.36-75.
8. Киселев Ф.Л. Вирусы папиллом человека как этиологический фактор рака шейки матки: значение для практики и здравоохранения // Вопросы вирусологии. 1997 Т.6. С.248-251.
9. Киселев Ф.Л., Киселева Н.П., Кобзева В.К., Грицко Т.М., Семенова Л.А., Павлова Л.С., Клэс Р., фон Кнебель-Деборитц М..Статус ДНК вируса папиллом человека в опухолях шейки матки. Молекулярная биология. 2001. Т.35. С.470-476.
10. Лаасри Мажид. Биологический потенциал мутантных вариантов генов Е6 и Е7 вируса папиллом человека тип 18 // Диссертация на соискание ученой степени кандидата биологических наук. Москва. 1998.

11. Лобашев М.Е. Генетика. Курс лекций // Изд. Ленинградского Университета. 1967. 489С.
12. Маниатис Т., Фрич Э., Сембрук Дж. Молекулярное клонирование // М. Мир. 1984. 480С.
13. Методы генетики соматических клеток // Под Ред. Дж. Шея. М. Мир. 1985. Т.1. 311С.; Т.2. 629С.
14. Сахарова О.В., Нечушкин М.И. Роль вирусов папиллом человека в патогенезе шейки матки (обзор литературы) // 1999. www.rosoncweb.ru
15. Свищева Г.Р. Статистические методы радиационного гибридного картирования: селективные модели // Генетика. 1999. Т.35. С.95-100.
16. Сулимова Г.Е., Компанийцев А.А., Кунижева С.С., Климов Е.А., Рахманалиев Э.Р., Удина И.Г. Картирование в геноме человека EST- и STS-маркеров с использованием панели радиационных гибридов // Генетика. 2000а. Т.36. С.900-907.
17. Сулимова Г.Е., Компанийцев А.А., Мойсяк Е.В., Рахманалиев Э.Р., Климов Е.А., Удина И.Г., Захаров И.А. Радиационное картирование как один из основных методов построения физических карт геномов человека и животных с высоким уровнем разрешения // Радиационная биология. Радиоэкология. 2000б. Т.40. С.520-528.
18. Эфрусси Б. Гибридизация соматических клеток // М. Мир. 1976. 195С.
19. Allikmets R., Kashuba V.I., Huebner K., LaForgia S., Kisselev L.L., Klein G., Dean M., Zabarovsky E.R. Mapping of 22 *NotI* linking clones on human chromosome 3 by polymerase chain reaction and somatic cell hybrid panels // Chromosome Research. 1996. V.4. P.33-37.
20. Antson A.A., Burns J.E., Moroz O.V., Scott D.J., Sanders C.M., Bronstein I.B., Dodson G.G., Wilson K.S., Maitland N.J. Structure of the intact transactivation domain of the human papillomavirus E2 protein // Nature. 2000. V.403. P.805-809.
21. Baker, C.C., Phelps, W.C., Lindgren, V., Braun, M.J., Gonda, M.A., Howley, P.M. Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines // J. Virol. 1987. V.61. P.962-971.

22. Barillot E., Dausset J., Cohen D. Theoretical analysis of a physical mapping strategy using random single-copy landmarks // *Proc. Natl. Acad. Sci. USA*. 1991. V.88. P.3917-3921.
23. Barnard P., McMillan N.A. The human papillomavirus E7 oncoprotein abrogates signaling mediated by interferon-alpha // *Virology*. 1999. V.259. P.305-313.
24. Barnard P., Payne E., McMillan N.A. The human papillomavirus E7 protein is able to inhibit the antiviral and anti-growth functions of interferon-alpha // *Virology*. 2000. V.277. P.411-419.
25. Benson D.A., Boguski M.S., Lipman D.J., Ostell J., Ouellette B.F., Rapp B.A., Wheeler D.L. GenBank // *Nucleic Acids Res.* 1999. V.27. P.12-17.
26. Boehnke M., Lange K., Cox D.R. Statistical methods for multipoint radiation hybrid mapping // *Am. J. Hum. Genet.* 1991. V.49. P.1174-1188.
27. Briggs M.W., Adam J.L., McCance D.J. The human papillomavirus type 16 E5 protein alters vacuolar H(+)-ATPase function and stability in *Saccharomyces cerevisiae* // *Virology*. 2001. V.280. P.169-175.
28. Burge C., Karlin S. Prediction of complete gene structures in human genomic DNA // *J. Mol. Biol.* 1997. V.268. P.78-94.
29. Burke D.T., Carle G.F., Olson M.V. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors // *Science*. 1987. V.236. P.806-812.
30. Cannizzaro L.A., Durst M., Mendez M.J., Hecht B.K., Hecht F. Regional chromosome localization of human papillomavirus integration sites near fragile sites, oncogenes, and cancer chromosome breakpoints // *Cancer Genet. Cytogenet.* 1988. V.33. P.93-98.
31. Carmody, M.W., Jones, M., Tarraza, H., and Vary, C.P. Use of the polymerase chain reaction to specifically amplify integrated HPV-16 DNA by virtue of its linkage to interspersed repetitive DNA // *Mol. Cell. Probes*. 1996. V.10. P.107-116.
32. Caron H., van Schaik B., van der Mee M., Baas F., Riggins G., van Sluis P., Hermus M.C., van Asperen R., Boon K., Voute P.A., Heisterkamp S., van

- Kampen A., Versteeg R. The human transcriptome map: clustering of highly expressed genes in chromosomal domains // *Science*. 2001. V.291. P.1289-1292.
33. Chen T. M., Chen C. A., Hsieh C. Y., Chang D. Y., Chen Y. H., Defendi V. The state of p53 in primary human cervical carcinomas and its effects in human papillomavirus-immortalized human cervical cells // *Oncogene*. 1993. V8. P.1511-1518.
34. Chen Y.H., Huang L.H., Chen T.M. Differential effects of progestins and estrogens on long control regions of human papillomavirus types 16 and 18 // *Biochem Biophys Res Commun*. 1996. V.224. P.651-659.
35. Chumakov I., Bellanne-Chantelot C., LeGall I., Cohen D. Integrated mapping across the whole human genome // *Cold Spring Harbor Symp. Quant. Biol*. 1993. V.58. P.377-382.
36. Chumakov I., LeGall I., Billaut A. Isolation of chromosome 21-specific yeast artificial chromosomes from a total human genome library // *Nature Genet*. 1992. V.1. P.222-225.
37. Cohen D., Chumakov I. Weissenbach J.A. First-generation physical map of the human genome // *Nature*. 1993. V.366. P.698-701.
38. Cox D.R., Burmeister M., Price E.R., Kim S., Myers R.M. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes // *Science*. 1990. V.250. P.245-250.
39. De Villiers E.M. Human pathogenic papillomavirus types: an upgrade // *Curr Top Microbiol Immunol*. 1994. V.186. P.1-12.
40. Deloukas P., Schuler G.D., Gyapay G. Beasley E.M., Soderlund C., Rodriguez-Tome P., Hui L., Matisse T.C., McKusick K.B., Beckmann J.S., Bentolila S., Bihoreau M., Birren B.B., Browne J., Butler A., Castle A.B., Chiannikulchai N., Clee C., Day P.J., Dehejia A., Dibling T., Drouot N., Duprat S., Fizames C., Bentley D.R., et al. A physical map of 30,000 human genes // *Science*. 1998. V.282. P.744-746.
41. Duerksen-Hughes P.J., Yang J., Schwartz S.B. HPV 16 E6 blocks TNF-mediated apoptosis in mouse fibroblast LM cells // *Virology*. 1999. V.264. P.55-65.

42. el Awady M.K., Kaplan J.B., O'Brien S.J., Burk R.D. Molecular analysis of integrated human papillomavirus 16 sequences in the cervical cancer cell lines SiHa // *Virology*. 1987. V.159. P.389-398.
43. Etscheid B.G., Foster S.A., Galloway D.A. The E6 protein of human papillomavirus type 16 functions as a transcriptional repressor in a mechanism independent of the tumor suppressor protein, p53 // *Virology*. 1994. V.205. P.583-585.
44. Fan Y., Davis L.M., Shows T.B. Mapping small DNA sequences by fluorescence in situ hybridization directly on banded metaphase chromosomes // *Proc. Natl. Acad. Sci.* 1990. V.87. P.6223-6227.
45. Fedorova L., Kost-Alimova M., Gizatullin R.Z., Alimov A., Zabarovska V.I., Szeles A., Protopopov A.I., Vorobieva N.V., Kashuba V.I., Klein G., Zelenin A.V., Sheer D., Zabarovsky E.R. Assignment and ordering of twenty-three unique *NotI*-linking clones containing expressed genes including the guanosine 5'-monophosphate synthetase gene to human chromosome 3 // *Eur. J. Hum. Genet.* 1997. V.5. P.110-116.
46. Frisch M, Frech K, Klingenhoff A, Cartharius K, Liebich I, Werner T. *In silico* prediction of scaffold/matrix attachment regions in large genomic sequences // *Genome Res.* 2002. V.12. P.349-354.
47. Galloway D.A., McDougall J.K. The disruption of cell cycle checkpoints by papillomavirus oncoproteins contributes to anogenital neoplasia // *Semin Cancer Biol.* 1996. V.7. P.309-315.
48. Gilles C., Piette J., Ploton D., Doco-Fenzy M., Foidart J.M..Viral integration sites in human papilloma virus-33-immortalized cervical keratinocyte cell lines // *Cancer Genet. Cytogenet.* 1996. V.90. P.63-69.
49. Goss S.J., Harris H. New method for mapping genes in human chromosomes // *Nature.* 1975. V.255. P.680-684.
50. Green E.D., Olson M.V. Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction // *Proc. Natl. Acad. Sci. USA.* 1990. V.87. P.1213-1217.

51. Gyapay G., Schmitt K., Fizames C. Jones H., Vega-Czarny N., Spillet D., Muselet D., Prud'Homme J.F., Dib C., Auffray C., Morissette J., Weissenbach J., Goodfellow P.N. A radiation hybrid map of the human genome // *Human Molecular Genetic*. 1996. V.5. P.339-346.
52. Hartwell L. Defects in a cell cycle checkpoint may be responsible for the genomic instability of cancer cells // *Cell*. 1992. V.71. P.543-546.
53. Heinemeyer T., Chen X., Karas H., Kel A.E., Kel O.V., Liebich I., Meinhardt T., Reuter I., Schacherer F., Wingender E. Expanding the TRANSFAC database towards an expert system of regulatory molecular mechanisms // *Nucleic Acids Res*. 1999. V.27. P.318-322.
54. Holland J., Coffey A.J., Giannelli F., Bentley D.R. Vertical integration of cosmid and YAC resources for interval mapping on the X chromosome. // *Genomics*. 1993. V.15. P.297-304.
55. Hozier J.C., Davis M. Cytogenetic approach to genome mapping // *Anal. Biochem*. 1992. V.200. P.205-217.
56. Ichikawa H., Hosoda F., Arai Y. Shimizu K., Ohira M., Ohki M. A *NotI* restriction map of the entire long arm of human chromosome 21 // *Nature Genet*. 1993. V.4. P.361-366.
57. International Human Genome Sequencing Consortium. Initial sequencing and analysis of human genom // *Nature*. 2001. V.409. P.860-922.
58. Ishizuka T., Saisu H., Odani S., Abe T. Synaphin: a protein associated with the docking/fusion complex in presynaptic terminals // *Biochem. Biophys. Res. Commun*. 1995. V.213. P.1107-1104.
59. Itakura M., Misawa H., Sekiguchi M., Takahashi S., Takahashi M. Transfection analysis of functional roles of complexin I and II in the exocytosis of two different types of secretory vesicles // *Biochem. Biophys. Res. Com*. 1999. V.265. P.691-696.
60. Jacquemin P., Depetris D., Mattei M.G., Martial J.A., Davidson I. Localization of human transcription factor TEF-4 and TEF-5 (*TEAD2*, *TEAD3*) genes to chromosomes 19q13.3 and 6p21.2 using fluorescence in situ hybridization and radiation hybrid analysis // *Genomics*. 1999. V.55. P.127-129.

61. Jones D.L., Munger K. Interactions of the human papillomavirus E7 protein with cell cycle regulators // *Semin Cancer Biol.* 1996. V.7. P.327-337.
62. Kashuba V.I., Gizatullin R.Z., Protopopov A.I., Li J., Vorobieva N.V., Fedorova L., Zabarovska V.I., Muravenko O.V., Kost-Alimova M., Domninsky D.A., Kiss C., Allikmets R., Zakharyev V.M., Braga E.A., Sumegi J., Lerman M., Wahlestedt C., Zelenin A.V., Sheer D., Winberg G., Grafodatsky A., Kisselev L.L., Klein G., Zabarovsky E.R. Analysis of *NotI* linking clones isolated from human chromosome 3 specific libraries // *Gene.* 1999. V.239. P.259-271.
63. Khare S., Pater M.M., Tang S.C., Pater A. Effect of glucocorticoid hormones on viral gene expression, growth, and dysplastic differentiation in HPV16-immortalized ectocervical cells // *Exp Cell Res.* 1997. V.232. P.353-360.
64. Klaes R., Woerner S.M., Ridder R., Wentzetzen N., Duerst M., Schneider A., Lotz B., Melscheimer P., von Knebel Doeberitz M. Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes // *Cancer Research.* 1999. V.59. P.6132-6136.
65. Koch J.E., Kolvraa S., Petersen K.B., Gregersen N., Bolund L. Oligonucleotide - priming methods for the chromosome - specific labelling of alpha satellite DNA in situ // *Cromosoma.* 1989. V.98. P.259-265.
66. Koopman L.A., Szuhai K., van Eendenburg J.D., Bezrookove V., Kenter G.G., Schuurin E., Tanke H., Fleuren G.J. Recurrent integration of human papillomavirus 16, 45 and 67 near translocation breakpoints in new cervical cancer cell lines // *Cancer Res.* 1999. V.59. P.5615-5624.
67. Lange K., Boehnke M. Bayesian methods and optimal experimental design for gene mapping by radiation hybrids // *Ann. Hum. Genet.* 1992. V.56. P.119-144.
68. Lange K., Boehnke M., Cox D.R., Lunetta K.L. Statistical methods for polyploid radiation hybrid mapping // *Genome Res.* 1995. V.5. P.136-149.
69. Larin Z., Monaco A.P., Lehrach H. Yeast artificial chromosome libraries containing large inserts from mouse and human DNA // *Proc. Natl. Acad. Sci.* 1991. V.88. P.4123-4127.

70. Lengauer C., Riethman H., Cremer T. Painting of human chromosomes with probes generated from hybrid cell lines by PCR with Alu and L1 primers // *Hum Genet.* 1990. V.86. P.1-6.
71. Liebich I, Bode J, Frisch M, Wingender E. S/MARt DB: a database on scaffold/matrix attached regions // *Nucleic Acids Res.* 2002. V.30. P.372-374.
72. Littlefield J.W. Selection of hybrids from matings of fibroblasts in vitro and their presumed recombinants // *Nature.* 1964. V.256. P.495-497.
73. Liu Z., Ghai J., Ostrow R.S., Faras A.J. The expression levels of the human papillomavirus type 16 E7 correlate with its transforming potential // *Virology.* 1995. V.207. P.260-270.
74. Mantovani F., Banks L. The interaction between p53 and papillomaviruses // *Semin Cancer Biol.* 1999. V.9. P.387-395.
75. Matlashewski G. The cell biology of human papillomavirus transformed cells // *Anticancer Res.* 1989. V.9. P.1447-1456.
76. McMahon H.T., Missler M., Li C., Sudhof T.C. Complexins: cytosolic proteins that regulate SNAP receptor function // *Cell.* 1995. V.83. P.111-119.
77. Mitra A.B. Genetic deletion and human papillomavirus infection in cervical cancer: loss of heterozygosity sites at 3p and 5p are important genetic events // *Int. J. Cancer.* 1999. V.82. P.322-324.
78. Morton M.E. Parameters of the human genome // *Proc. Nat. Acad. Sci.* 1991. V.88. P.7474-7476.
79. Nakagawa S., Watanabe S., Yoshikawa H., Taketani Y., Yoshiike K., Kanda T. Mutational analysis of human papillomavirus type 16 E6 protein: transforming function for human cells and degradation of p53 in vitro // *Virology.* 1995. V.212. P.535-542.
80. Naylor S.L., Buys C.H.C.M., Carritt B. Report of the fourth international workshop on human chromosome 3 mapping 1993 // *Cytogenet. Cell Genet.* 1995. V.65. P.1-50.
81. Newell W., Beck S., Lehrach H., Lyall A. Estimation of distances and map construction using radiation hybrids // *Genome Res.* 1998. V.8. P.493-508.

82. NIH/CEPH Collaborative Mapping Group. A comprehensive genetic linkage map of the human genome // *Science*. 1992. V.258. P.67-86.
83. Nizetic D., Gellen L., Hamvas R., Mott R., Grigoriev A., Vatcheva R., Zehetner G., Yaspo M.L, Dutriaux A., Lopes C., et al. An integrated YAC-overlap and "cosmid-pocket" map of the human chromosome 21 // *Hum. Mol. Genet*. 1994. V.3. P.759-770.
84. Olivier M., Aggarwal A., Allen J., Almendras A.A., Bajorek E.S., Beasley E.M., Brady S.D., Bushard J.M., Bustos V.I., Chu A., Chung T.R., De Witte A., Denys M.E., Dominguez R., Fang N.Y., Foster B.D., Freudenberg R.W., Hadley D., Hamilton L.R., et al. A high-resolution radiation hybrid map of the human genome draft sequence // *Science*. 2001. V.291. P.1298-1302.
85. Olson M., Hood L., Carton C., Botstein D. A common language for physical mapping of human genome // *Science*. 1989. V.245. P.1434-1435.
86. Ono S., Baux G., Sekiguchi M., Fossier P., Morel N.F., Nihonmatsu I., Hirata K., Awaji T., Takahashi S., Takahashi M. Regulatory roles of complexins in neurotransmitter release from mature presynaptic nerve terminals // *Eur. J. Neurosci*. 1998. V.10. P.2143-2152.
87. Pabst S., Hazzard J.W., Antonin W., Sudhov T.C., Jahn R., Rizo J., Fasshauer D. Selective interaction of complexin with the neuronal SNARE complex. Determination of the binding regions // *J. Biol.Chem*. 2000. V.275. P.19808-19818.
88. Rakhmanaliev E.R., Kutsenko A., Kompaniytsev A.A., Klimov E.A., Mojsjak E.V., Udina I.G., Sulimova G.E. *NotI*-STSs as Genes Markers // 13th International Genome Sequencing and Analysis Conference (GSAC XIII). Program & Abstract Journal. San Diego, CA. USA. October 25 - 28. 2001. P.91.
89. Rey O., Lee S., Baluda M.A., Swee J., Ackerson B., Chiu R., Park N.H. The E7 oncoprotein of human papillomavirus type 16 interacts with F-actin in vitro and in vivo // *Virology*. 2000. V.268. P.372-381.
90. Rey O., Lee S., Park N.H. Human papillomavirus type 16 E7 oncoprotein represses transcription of human fibronectin // *J. Virol*. 2000. V.74. P.4912-4918.

91. Ristriani T., Masson M., Nomine Y., Laurent C., Lefevre J.F., Weiss E., Trave G.
HPV oncoprotein E6 is a structure-dependent DNA-binding protein that recognizes four-way junctions // *J. Mol. Biol.* 2000. V.296. P.1189-1203.
92. Rodriguez-Tome P., Lijnzaad P. The radiation hybrid database // *Nucleic Acids Res.* 1999. V.27. P.115-118.
93. Saiki R.K., Gelfand D.H., Stoffel S., Scharf S.J., Higuchi R., Horn G.T., Mullis K.B., Erlich H.A. Primer-directed enzymatic amplification of DNA with a thermostable DNA Polymerase // *Science.* 1988. V.239. P.487-491.
94. Sastre-Garau X., Schneider-Maunoury S., Couturier J., Orth G. Human papillomavirus type 16 DNA is integrated into chromosome region 12q14-q15 in a cell line derived from a vulvar intraepithelial neoplasia // *Cancer Genet. Cytogenet.* 1990. V.44. P.243-251.
95. Scherf M., Klingenhoff A., Werner T. Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach // *J. Mol. Biol.* 2000. V.297. P.599-606.
96. Schneider-Gadicke, A., and Schwarz, E. Different human cervical carcinoma cell lines show similar transcription patterns of human papillomavirus type 18 early genes // *EMBO J.* 1986. V.5. P.2285-2292.
97. Schneider-Mannoury S., Croissant O., Orth G. Integration of human papillomavirus type 16 DNA sequences: Possible early event in the progression of genital tumors // *J. Virol.* 1987. V.61. P.3295-3298.
98. Schuler G.D., Boguski M.S., Stewart E.A., Stein L.D., Gyapay G., Rice K., White R.E., Rodriguez-Tome P., Aggarwal A., Bajorek E., Bentolila S., Birren B.B., Butler A., Castle A.B., Chiannilkulchai N., Chu A., Clee C., Cowles S., Day P.J., Dibling T., Drouot N., Dunham I., Duprat S., East C., Hudson T.J., et al. A gene map of the human genome // *Science.* 1996. V.274. P.540-546.
99. Schwarz, E., Freese, U. K., Gissmann, L., Mayer, W., Roggenbuck, B., Stremlau, A., zur Hausen, H. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells // *Nature (Lond.).* 1985. V.314. P.111-114.

100. Shera K.A., Shera C.A., McDougall J.K. Small tumor virus genomes are integrated near nuclear matrix attachment regions in transformed cells // *J. Virol.* 2001. V.75. P.12339-12346.
101. Sherman, L., Alloul, N., Golan, I., Durst, M., and Baram, A. Expression and splicing patterns of human papillomavirus type-16 mRNAs in pre-cancerous lesions and carcinomas of the cervix, in human keratinocytes immortalized by HPV 16, and in cell lines established from cervical cancers // *Int. J. Cancer.* 1992. V.50. P.356-364.
102. Stewart E.A., McKusick K.B., Aggarwal A., Bajorek E., Brady S., Chu A., Fang N., Hadley D., Harris M., Hussain S., Lee R., Maratukulam A., O'Connor K., Perkins S., Piercy M., Qin F., Reif T., Sanders C., She X., Sun W.L., Tabar P., Voyticky S., Cowles S., Fan J.B., Cox D.R., et al. An STS-based radiation hybrid map of the human genome // *Genome Res.* 1997. V.7. P.422-433.
103. Stoppler M.C., Straight S.W., Tsao G., Schlegel R., McCance D.J. The *E5* gene of HPV-16 enhances keratinocyte immortalization by full-length DNA // *Virology.* 1996. V.223. P.251-254.
104. Sulimova G.E., Udina I.G., Kutsenko A.S., Kashuba V.I., Rakhmanaliev E.R., Kompaniytsev A.A., Protopopov A.I., Moisjak E.V., Klimov E.A., Muravenko O., Zabarovsky E.R., and Kiselev L.L. Physical and Gene *NotI*-Map of Human Chromosome 3 // Abstracts of International Symposium "Molecular Mechanisms of Genetic Processes and Biotechnology". Moscow, Minsk 18 - 21 November 2001, Minsk 22 - 24 November 2001. P.162-163.
105. Szarka K., Veress G., Juhasz A., Konya J., Sapy T., Soos G., Hernadi Z., Gergely L. Integration status of virus DNA and p53 codon 72 polymorphism in human papillomavirus type 16 positive cervical cancers // *Anticancer Res.* 2000. V.20. P.2161-2167.
106. Takahashi S., Ujihara H., Huang G.Z., Yagyu K., Sanbo M., Kaba H., Yagi T. Reduced hippocampal LTP in mice lacking a presynaptic protein: complexin II // *Eur. J. Neurosci.* 1999. V.11. P.2359-2366.
107. Takahashi S., Yamamoto H., Matsuda Z., Ogawa M., Yagyu K. Identification of two highly homologous presynaptic proteins distinctly localized at the dendritic and somatic synapses // *FEBS lett.* 1995. V.368. P.455-460.

108. Thorland E., Myers S., Persing D., Sarkar G., McGovern R., Gostout B., Smith D. Human papillomavirus type 16 integrations in cervical tumors frequently occur in common fragile sites // *Cancer Research* 2000. V.60. P.5916-5921.
109. Troyer D.L., Goad D.W., Xie H., Rohrer G.A., Alexander L.J., Beattie C.W. Use of direct in situ single - copy (DISC) PCR to physically map fine porcine microsatellites // *Cytogenet. Cell Genet.* 1994. V.67. P.199-204.
110. Van Tine B., Knops J., Broker T., Chow L., Moen Jr. P. *In situ* analysis of the transcriptional activity of integrated viral DNA using tyramide-FISH // *Dev Biol (Basel)* 2001. V.106. P.381-385.
111. Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O., Yandell M., Evans C.A., Holt R.A., Gocayne J.D., Amanatides P., Ballew R.M., Huson D.H., Wortman J.R., Zhang Q., Kodira C.D., Zheng X.H., Chen L., Skupski M., et al. The sequence of the human genome // *Science.* 2001. V.291. P.1304-1351.
112. Vogelstein B., Kinzler K.W. p53 function and dysfunction // *Cell.* 1992. V.70. P.523-526.
113. von Knebel Doeberitz M. Papillomaviruses in human disease: Part II. Molecular biology and immunology of papillomavirus infections and carcinogenesis // *Eur. J. Med.* 1992. V.1. P.485-491.
114. Wagatsuma M., Hashimoto K., Matsukura T. Analysis of integrated human papillomavirus type 16 DNA in cervical cancers: amplification of viral sequences together with cellular flanking sequences // *J Virol.* 1990. V.64. P.813-821.
115. Walter M.A., Goodfellow P.N. Radiation hybrids: irradiation and fusion gene transfer // *Trends Genet.* 1993. V.9. P.352-356.
116. Walter M.A., Spillett D.J., Thomas P., Weissenbach J., Goodfellow P.N. A method for constructing radiation hybrid maps of whole genome // *Nature Genetics.* 1994. V.7. P.22-28.
117. Wang Ji-Yi, Zabarovsky E.R., Talmadge C., Berglund P., Chan K.W., Pokrovskaya E.S., Kashuba V.I., Zhen D.K., Boldog F., Zabarovskaya V.I., et al. Somatic cell hybrid panel and *NotI* linking clones for physical mapping of human chromosome 3 // *Genomic.* 1994. V.20. P.105-113.

118. Weissenbach J., Gyapay G., Dib C., Vignal A., Morissette J., Millasseau P., Vaysseix G., Lathrop M. A second-generation linkage map of the human genome // *Nature*. 1992. V.359. P.794-801.
119. Wentzensen N., Ridder R., Klaes R., Vinokurova S., Schaefer U., von Knebel Doeberitz M. Characterization of viral-cellular fusion transcripts in a large series of HPV16 and 18 positive anogenital lesions // *Oncogene* 2002. V.21. P.419-426.
120. Wilke C.M., Hall B.K., Hoge A., Paradee W., Smith D.I., Glover T.W. FRA3B extends over a broad region and contains a spontaneous HPV16 integration site: direct evidence for the coincidence of viral integration sites and fragile sites // *Hum. Mol. Genet.* 1996. V.5. P.187-195.
121. Wingender E., Chen X., Fricke E., Geffers R., Hehl R., Liebich I., Krull M., Matys V., Michael H., Ohnhaus R., Pruss M., Schacherer F., Thiele S., Urbach S. The TRANSFAC system on gene expression regulation // *Nucleic Acids Res.* 2001. V.29. P.281-283.
122. Wingender E., Chen X., Hehl R., Karas H., Liebich I., Matys V., Meinhardt T., Pruss M., Reuter I., Schacherer F. TRANSFAC: an integrated system for gene expression regulation // *Nucleic Acids Res.* 2000. V.28. P.316-319.
123. Zbarovsky E.R., Boldog F., Erlandsson R., Kashuba V., Allikmets R., Marcsek Z., Kisselev L., Stanbridge E., Klein G., Sumegi J., Winberg G. New strategy for mapping the human genome based on a novel procedure for construction of jumping libraries // *Genomics*. 1991. V.11. P.1030-1039.
124. Zbarovsky E.R., Boldog F., Thompson T., Scanlon D., Winberg G., Marcsek Z., Erlansson R., Stanbridge E., Klein G., Sumegi J. Construction of a human chromosome 3 specific *NotI* linking library using a novel cloning procedure // *Nucleic Acids Res.* 1990. V.18. P.6319-6324.
125. Zbarovsky E.R., Gizatullin R., Podowski R.M., Zbarovska V.V., Xie L., Muravenko O.V., Kozyrev S., Petrenko L., Skobeleva N., Li J, Protopopov A., Kashuba V., Ernberg I., Winberg G., Wahlestedt C. *NotI* clones in the analysis of the human genome // *Nucleic Acids Res.* 2000. V.28. P.1635-1639.

126. Zhdanova N.S., Thomsen P.D., Astakhova N.M., Kuznetsov S.B., Jorgensen C.B., Plyusnina E.V., Serov O.L. Production of pig-mink cell hybrids with single pig chromosomes 2, 5, 12, or t(1,13) // *Mamm Genome*. 1996. V.7. P.613-615.
127. Zur Hausen H. 1991. Human papillomaviruses in the pathogenesis of anogenital cancer. *Virology*. V.184. P.9-13.
128. Zur Hausen H. Human pathogenic papillomaviruses, topics in microbiology and immunology. Ed. H. zur Hausen // Berlin. 1994. V.186. P.131-157.
129. zur Hausen H. Immortalization of human cells and their malignant conversion by high risk human papillomavirus genotypes // *Semin Cancer Biol*. 1999. V.9. P.405-411.
130. zur Hausen H. Papillomaviruses in human cancers // *Proc Assoc Am Physicians*. 1999. V.111. P.581-587.
131. Zur Hausen H., Gissman L. *Viral Oncology*. Ed. G.Klein // New York, 1980. 433P.

Приложение 1. Транскрипционные факторы (ТФ), способные связываться с последовательностью промотора гена *LOC161154*, и нуклеотидная последовательность, с которой происходит связывание. Анализ нуклеотидной последовательности проведен с использованием программы MatInspector V2.2, основанной на базе данных ТФ – TRANSFAC 4.0.

| Название ТФ (V\$ – позвоночные, SP1 – название_изоформа) | Последовательность для связывания (верхний регистр – необходимая последовательность) |
|---|---|
| V\$ER_Q6 | ctcagcccaggTGACcсаа |
| V\$AP4_Q5 | ctCAGCccag |
| V\$E47_Q2 | cagccCAGGtgaccca |
| V\$E47_Q1 | agcCCAGgtgaccca |
| V\$MYOD_Q1 | gccCAGGtgacc |
| V\$LMO2COM_Q1 | gccCAGGtgacc |
| V\$AP1FJ_Q2 | ggTGACcсааа |
| V\$AP1_Q4 | ggTGACcсааа |
| V\$AP1_Q2 | ggTGACcсааа |
| V\$VMYB_Q1 | ccaAACGgca |
| V\$VMYB_Q2 | ccaAACGgc |
| V\$TCF11_Q1 | GTCaactggcgc |
| V\$CEBPB_Q1 | actggcGCAAccg |
| V\$NF1_Q6 | actTGGCgcaaccggagc |
| V\$VMYB_Q2 | cgcAACCGg |
| V\$CP2_Q1 | gccaccCCAG |
| V\$AP2_Q6 | caCCCcagccc |
| V\$NF1_Q6 | gctTGGCcccacaatgcg |
| V\$SOX5_Q1 | ccaCAATgc |
| V\$CMYB_Q1 | acaatgcccGTTGtgcc |
| V\$TCF11_Q1 | GTCaggccgggga |
| V\$MZF1_Q1 | gccGGGGa |
| V\$IK2_Q1 | gccgGGGAttg |
| V\$GKLF_Q1 | ggagcagggaTGGG |
| V\$IK2_Q1 | agcaGGGAtggg |
| V\$IK2_Q1 | tgggGGGAatat |
| V\$IK3_Q1 | tgggGGGAatatt |
| V\$IK1_Q1 | tgggGGGAatatt |
| V\$MZF1_Q1 | tggGGGGa |
| V\$OCT1_Q2 | gggaatATTCgtcag |
| V\$GATA1_Q2 | cgtcaGATAttaa |
| V\$GATA1_Q3 | cgtcaGATAttaa |
| V\$GATA1_Q4 | gtcaGATAttaa |
| V\$TCF11_Q1 | GTCAgatatttaa |
| V\$LMO2COM_Q2 | caGATAtt |
| V\$GATA_C | aGATATttaa |
| V\$TATA_C | taTTTAAAc |
| V\$SR_Y_Q2 | gagaACAAagca |
| V\$AP4_Q5 | tgCAGCagg |
| V\$SP1_Q6 | gcagGGCGgccg |
| V\$IK2_Q1 | gcccGGGAgctg |
| V\$IK2_Q1 | agctGGGAgccg |
| V\$AP2_Q6 | ggCCCCaagca |

Приложение 2. Транскрипционные факторы (ТФ), способные связываться с последовательностью промотора гена *KIAA1808*, и нуклеотидная последовательность, с которой происходит связывание. Анализ нуклеотидной последовательности проведен с использованием программы MatInspector V2.2, основанной на базе данных ТФ – TRANSFAC 4.0.

| Название ТФ (V\$ – позвоночные, SP1 – название_изоформа) | Последовательность для связывания (верхний регистр – необходимая последовательность) |
|---|---|
| V\$IK2_01 | agcaGGGAcca |
| V\$AP4_Q6 | agCAGCaggg |
| V\$AP4_Q5 | agCAGCaggg |
| V\$AP4_Q6 | ctCAGCagca |
| V\$AP4_Q5 | ctCAGCagca |
| V\$RFX1_01 | tgtgagctcaGTAAccc |
| V\$RFX1_02 | ctgtgagctcaGTAAccc |
| V\$IK2_01 | ctcaGGGAcaag |
| V\$E2_Q6 | ttaccctcagGGGTct |
| V\$E2_01 | ttaccctcagGGGTct |
| V\$NF1_Q6 | accTGGCacgacctacc |
| V\$DELTAEF1_01 | cacgACCTtac |
| V\$DELTAEF1_01 | cttaACCTggc |
| V\$CETS1P54_01 | ccTGGAAgat |
| V\$DELTAEF1_01 | ccccACCTgga |
| V\$MYOD_Q6 | ccCACCTgga |
| V\$USF_C | cCACCTgg |
| V\$DELTAEF1_01 | catcACCTccc |
| V\$CEBPB_01 | aactgcgGCAAtga |
| V\$VMYB_01 | cacAACTgcg |
| V\$MYOD_Q6 | caCAACTgcg |
| V\$VMYB_02 | cacAACTgc |
| V\$MYCMAX_02 | atcCACAtgcac |
| V\$TCF11_01 | CTCAtccacatgc |
| V\$USF_C | cCACATgc |
| V\$NMYC_01 | gggccCGTGctt |
| V\$USF_C | gCCCGTgc |
| V\$GC_01 | gcagGGAGggggccc |
| V\$SP1_Q6 | gcagGGAGggggccc |
| V\$IK2_01 | tgcaGGGAgggg |
| V\$GKLF_01 | catgcagggaGGGG |
| V\$TH1E47_01 | aaggtcatCTGGtcaa |
| V\$TCF11_01 | GTCAtctgggtcaa |
| V\$MYOD_Q6 | gtCATCtgggt |
| V\$RORA1_01 | gggagaaGGTCat |
| V\$GKLF_01 | ccagaggagAAGG |
| V\$IK2_01 | cagaGGGAgaag |
| V\$OCT1_02 | cagattATGCagggg |
| V\$TCF11_01 | GTCAgattatgca |
| V\$CETS1P54_01 | acAGGAgtca |
| V\$AP2_Q6 | gaCCCAcaggag |
| V\$IK2_01 | cctgGGGAcca |
| V\$MZF1_01 | cctGGGGa |

| | |
|----------------|-----------------------|
| V\$CMYB_01 | gggtcgggtctGTTGtctc |
| V\$E2_01 | gcacgggggtCGGTct |
| V\$E2_Q6 | gcacgggggtCGGTct |
| V\$USF_Q6 | cgCACGgggc |
| V\$MYC_MAX_02 | cagCACGagccg |
| V\$USF_Q6 | agCACGagcc |
| V\$CP2_01 | gcaccaCCAG |
| V\$AP4_Q6 | gcCAGCgggc |
| V\$AP4_Q5 | gcCAGCgggc |
| V\$VMAF_01 | cacagcTGACcctgccagc |
| V\$AP1_Q2 | gcTGACcctgc |
| V\$AP1FJ_Q2 | gcTGACcctgc |
| V\$ER_Q6 | ggacgcacagcTGACcctg |
| V\$LMO2COM_01 | gcaCAGCtgacc |
| V\$MYOD_Q6 | caCAGCtgac |
| V\$AP4_Q6 | caCAGCtgac |
| V\$AP4_Q5 | caCAGCtgac |
| V\$NFAT_Q6 | gttggGAAAgag |
| V\$IK1_01 | tggtGGGAaagag |
| V\$CEBPB_01 | ctgttggGAAAgag |
| V\$IK2_01 | tggtGGGAaaga |
| V\$LYF1_01 | gttGGGAaa |
| V\$CMYB_01 | ctgcgacactGTTGggaa |
| V\$IK2_01 | ggctGGGAgcag |
| V\$AP4_Q6 | tcCAGCgggc |
| V\$AP4_Q5 | tcCAGCgggc |
| V\$E47_02 | tccaCACGtggtgca |
| V\$ARNT_01 | tcccacaCGTGttgca |
| V\$USF_01 | cccaCACGtggtgc |
| V\$MYC_MAX_01 | cccaCACGtggtgc |
| V\$MAX_01 | cccaCACGtggtgc |
| V\$LMO2COM_01 | ccaCACGtggtg |
| V\$MYC_MAX_02 | ccaCACGtggtg |
| V\$MYOD_01 | ccaCACGtggtg |
| V\$NMYC_01 | ccacaCGTGttg |
| V\$MYOD_Q6 | caCACGtggt |
| V\$USF_Q6 | caCACGtggt |
| V\$USF_C | aCACGTgt |
| V\$DELTAEF1_01 | cagcACCTegc |
| V\$AP4_Q5 | cgCAGCacct |
| V\$MYC_MAX_02 | aagCACTgatg |
| V\$MYOD_Q6 | agCACTgat |
| V\$USF_C | gCACTTga |
| V\$DELTAEF1_01 | actcACCTtta |
| V\$IK2_01 | ccggGGGAact |
| V\$MZF1_01 | ccgGGGGa |
| V\$IK2_01 | gcagGGGAccgg |
| V\$MZF1_01 | gcaGGGGa |
| V\$CMYB_01 | tgggaggcccGTTTcagg |
| V\$IK2_01 | cgetGGGAggcc |
| V\$LYF1_01 | gctGGGAgg |
| V\$STAF_02 | gcttCCCAggaggctttgcgg |
| V\$CETS1P54_01 | ccAGGAggct |

Приложение 3. Транскрипционные факторы (ТФ), способные связываться с последовательностью промотора гена *KIAA0887*, и нуклеотидная последовательность, с которой происходит связывание. Анализ нуклеотидной последовательности проведен с использованием программы MatInspector V2.2, основанной на базе данных ТФ – TRANSFAC 4.0.

| Название ТФ (V\$ – позвоночные, SP1 – название_изоформа) | Последовательность для связывания (верхний регистр – необходимая последовательность) |
|---|---|
| V\$STAT3_01 | cggcaattTCcgggaagtgttc |
| V\$STAT1_01 | cggcaatttccGGAAgtgttc |
| V\$NFKAPPAB65_01 | ggcaatTTCC |
| V\$CREL_01 | ggcaatTTCC |
| V\$CEBPB_01 | aatttccGGAAgtg |
| V\$STAT_01 | TTTCcggaa |
| V\$NRF2_01 | tccGGAAgtg |
| V\$CETS1P54_01 | tcCGGAgtg |
| V\$NKX25_01 | ggAAGTg |
| V\$IK2_01 | gttcGGGAtagg |
| V\$GATA1_02 | ttcggGATAggccc |
| V\$GATA1_03 | ttcggGATAggccc |
| V\$GATA1_04 | tcggGATAggccc |
| V\$LMO2COM_02 | ggGATAggc |
| V\$GATA_C | gGATAGgccccg |
| V\$AHRARNT_01 | aggccccgagCGTGgcg |
| V\$AP2_Q6 | cgCCCTctggcg |
| V\$NF1_Q6 | ctcTGGCggcggccagag |
| V\$SP1_Q6 | aggtGGCGggggg |
| V\$NF1_Q6 | aggTGGCggggggcgggg |
| V\$GC_01 | ggggGGCGggggga |
| V\$SP1_Q6 | ggggGGCGggggg |
| V\$MZF1_01 | cggGGGGa |
| V\$IK2_01 | cgggGGGAggtt |
| V\$CMYB_01 | cgcagacgcaGTTGcgcg |
| V\$E47_01 | gacGCAGttgcgcg |
| V\$LMO2COM_01 | acgCAGTtgcg |
| V\$ELK1_01 | cagaccGGATgtggtt |
| V\$E2_Q6 | agaccgatgTGGTg |
| V\$E2_01 | agaccgatgTGGTg |
| V\$CETS1P54_01 | acCGGAtgtg |
| V\$RORA1_01 | tgctgaGGTCag |
| V\$T3R_01 | ggctgaGGTCagtcc |
| V\$TCF11_01 | GTCAGttccggcg |
| V\$TAXCREB_01 | ggcgggTGACggtgc |
| V\$CREB_02 | cgggTGACggtg |
| V\$ATF_01 | gggTGACggtgcgg |
| V\$APIFJ_Q2 | ggTGACggtgc |
| V\$CREB_Q4 | ggTGACggtgcg |
| V\$API_Q2 | ggTGACggtgc |
| V\$CREB_Q2 | ggTGACggtgcg |

| | |
|----------------|--------------|
| V\$AP1_Q4 | ggTGACggtgc |
| V\$IK2_01 | gagcGGGAtcta |
| V\$CETS1P54_01 | gcGGGAtcta |
| V\$GATA3_03 | cgGGATctaa |

Приложение 4. Транскрипционные факторы (ТФ), способные связываться с последовательностью промотора гена *WASF4* (*LOC158537*), и нуклеотидная последовательность, с которой происходит связывание. Анализ нуклеотидной последовательности проведен с использованием программы MatInspector V2.2, основанной на базе данных ТФ – TRANSFAC 4.0.

| Название ТФ (V\$ – позвоночные, SP1 – название_изоформа) | Последовательность для связывания (верхний регистр – необходимая последовательность) |
|---|---|
| V\$OCT1_Q6 | ttaaagtAAAAtat |
| V\$OCT1_Q7 | taaagttaAATT |
| V\$S8_01 | aaatgtaaATTAata |
| V\$E4BP4_01 | aattatATAAtt |
| V\$HLF_01 | ATTAtataat |
| V\$VBP_01 | aTTATataat |
| V\$TATA_01 | ttaTATAattgctta |
| V\$NKX25_02 | taTAATtg |
| V\$HNF3B_01 | gcttaTATTctttc |
| V\$HFH3_01 | ttaTATTctttc |
| V\$HFH2_01 | ttaTATTcttt |
| V\$TCF11_01 | TTCAttctatatac |
| V\$HFH1_01 | tataGTTAatt |
| V\$NKX25_02 | ttTAATtt |
| V\$CETS1P54_01 | gcAGGAcact |
| V\$PADS_C | gGTGGTctt |
| V\$APIFJ_Q2 | acTGACccggt |
| V\$AP1_Q2 | acTGACccggt |
| V\$AP1_Q4 | acTGACccggt |
| V\$USF_Q6 | ctCACTtgca |
| V\$USF_C | tCACTTgc |
| V\$GATA1_03 | ctcaaGATAactgt |
| V\$GATA1_02 | ctcaaGATAactgt |
| V\$GATA1_04 | tcaaGATAactgt |
| V\$EV11_06 | tcaAGATAa |
| V\$GATA3_02 | caaGATAact |
| V\$GATA2_02 | caaGATAact |
| V\$GATA1_06 | caaGATAact |
| V\$GATA1_05 | caaGATAact |
| V\$LMO2COM_02 | aaGATAact |
| V\$GATA_C | aGATAActgta |
| V\$VMYB_01 | gatAACTgta |
| V\$VMYB_02 | gatAACTgt |
| V\$IK3_01 | ttctgGGAaggca |
| V\$IK2_01 | ttctGGGAaggc |
| V\$STAT_01 | TTCTgggaa |
| V\$IK1_01 | ttctGGGAaggca |
| V\$RFX1_01 | ttctgggaagGCAAc |

| | |
|---------------|------------------------|
| V\$LYF1_01 | tctGGAag |
| V\$MEF2_02 | gcaacatcttgaAATAggcagg |
| V\$MEF2_03 | gcaacatcttgaAATAggcagg |
| V\$CEBPB_01 | acatcttGAAAtag |
| V\$GKLF_01 | tgaaatagcAGGG |
| V\$IK2_01 | ggcaGGGActgg |
| V\$NF1_Q6 | gacTGGCtgaacagctg |
| V\$AP4_01 | tggaaCAGCtgggctct |
| V\$LMO2COM_01 | gaaCAGCtgggg |
| V\$MYOD_01 | gaaCAGCtgggg |
| V\$MYOD_Q6 | aaCAGCtggg |
| V\$AP4_Q6 | aaCAGCtggg |
| V\$AP4_Q5 | aaCAGCtggg |
| V\$RFX1_01 | gtcatgctaGAAAcAA |
| V\$TCF11_01 | GTCAtgcttagaa |
| V\$SRY_02 | agaaACAAtag |
| V\$GRE_C | gaaacaagaTGTCctt |
| V\$TCF11_01 | GTCAtgtatcccc |
| V\$AP2_Q6 | atCCCCcagggt |
| V\$CP2_01 | acaaaacCCAG |

Приложение 5. Транскрипционные факторы (ТФ), способные связываться с последовательностью промотора гена *LOC151128*, и нуклеотидная последовательность, с которой происходит связывание. Анализ нуклеотидной последовательности проведен с использованием программы MatInspector V2.2, основанной на базе данных ТФ – TRANSFAC 4.0.

| Название ТФ (V\$ – позвоночные, SP1 – название_изоформа) | Последовательность для связывания (верхний регистр – необходимая последовательность) |
|---|---|
| V\$TCF11_01 | GTCAttgatcta |
| V\$CETS1P54_01 | tcTGGAGgac |
| V\$IK2_01 | caggGGGActtg |
| V\$MZF1_01 | cagGGGGa |
| V\$VMYB_02 | cctAACAgA |
| V\$VMYB_01 | cctAACAgag |
| V\$CEBPB_01 | ccctgcaGCAAact |
| V\$CETS1P54_01 | ccTGGAcate |
| V\$OCT1_05 | gacatttCCATaca |
| V\$OCT1_06 | gacatttccATACa |
| V\$GF11_01 | tcctctgaAATCtaggtggaggtt |
| V\$CREL_01 | ggaggtTTCC |
| V\$DELTAEF1_01 | cccaACCTga |
| V\$S8_01 | caaccttgATTAttga |
| V\$HNF3B_01 | ttgatTATTgacttc |
| V\$HFH2_01 | gatTATTgactt |
| V\$APIFJ_Q2 | atTGACTtctg |
| V\$AP1_Q2 | atTGACTtctg |
| V\$AP1_Q4 | atTGACTtctg |
| V\$CP2_01 | gcaccaCAAG |
| V\$USF_Q6 | ccCACAgcc |
| V\$MYC MAX_02 | cacCACAtggaa |

| | |
|----------------|--------------------------|
| V\$USF_Q6 | acCACAtgga |
| V\$USF_C | cCACATgg |
| V\$RFX1_02 | ttgaccctctGAAAcca |
| V\$LMO2COM_01 | ccaCAGCtgagc |
| V\$MYOD_Q6 | caCAGCtgag |
| V\$AP4_Q5 | caCAGCtgag |
| V\$AP4_Q6 | caCAGCtgag |
| V\$NF1_Q6 | gagTGGCtgggacttacg |
| V\$IK2_01 | ggctGGGActta |
| V\$VBP_01 | cTTACggcac |
| V\$NF1_Q6 | cctTGGCtgcacacagca |
| V\$GKLF_01 | acacagcagaGGGG |
| V\$AP4_Q6 | caCAGCagag |
| V\$AP4_Q5 | caCAGCagag |
| V\$RFX1_02 | gtccagccacGAAAcca |
| V\$TH1E47_01 | ctaggcatCTGGtct |
| V\$MYOD_Q6 | ggCATCtgg |
| V\$IK2_01 | tgatGGGAgggt |
| V\$SP1_Q6 | gatgGGAGggtct |
| V\$GC_01 | gatgGGAGggtctg |
| V\$BARBIE_01 | ccacAAAGgtgtctg |
| V\$TH1E47_01 | aaaggtgtCTGAcatg |
| V\$AP1FJ_Q2 | tcTGACatggc |
| V\$NF1_Q6 | acaTGGCctggagacatt |
| V\$S8_01 | ccttggtgATTAacat |
| V\$ST1_01 | ggtgATTAacattta |
| V\$OCT1_06 | ctcctgttATTTa |
| V\$TCF11_01 | GTCAGaaaacggg |
| V\$VMYB_01 | gaaAACGggt |
| V\$VMYB_02 | gaaAACGgg |
| V\$GATA1_02 | tgccaGATAcacta |
| V\$GATA1_03 | tgccaGATAcacta |
| V\$GATA1_04 | gccaGATAcacta |
| V\$LMO2COM_02 | caGATAcac |
| V\$GATA_C | aGATACactaa |
| V\$GF11_01 | cactaaatAATCtctctcaaggtc |
| V\$POLY_C | cACTAAAtaatctctctc |
| V\$RORA1_01 | ctctcaaGGTCaa |
| V\$TCF11_01 | GTCaaagtccac |
| V\$GATA3_03 | acAGATctct |
| V\$RFX1_02 | ctagggcagggGCAAcat |
| V\$USF_Q6 | gcCACCagtc |
| V\$TCF11_01 | TTCAtatagcaag |
| V\$CEBPB_01 | tcatataGCAAgag |
| V\$STAT_01 | TTCCtaaaa |
| V\$MYOD_Q6 | tcCATCtgag |
| V\$DELTAEF1_01 | gaccACCTcgg |
| V\$S8_01 | tggactttATTAttca |
| V\$OCT1_02 | tttattATTCatatt |
| V\$AP1_C | aTTATTCAat |
| V\$TCF11_01 | TTCAtattactat |
| V\$RORA2_01 | gcattttGGTCaa |
| V\$RORA1_01 | gcattttGGTCaa |

| | |
|------------------|-------------------------|
| V\$SR_Y_02 | ttcaACAAgtct |
| V\$CEBPB_01 | aagtctaGGAAgtt |
| V\$CEBPB_01 | gtctctaGGAAgtt |
| V\$NFKAPPAB65_01 | aggaagTTCC |
| V\$CREL_01 | aggaagTTCC |
| V\$GATA1_02 | ccaaaGATAttgcc |
| V\$GATA1_03 | ccaaaGATAttgcc |
| V\$GATA1_04 | caaaGATAttgcc |
| V\$GATA3_03 | aaAGATattg |
| V\$LMO2COM_02 | aaGATAttg |
| V\$GATA_C | aGATATtgcca |
| V\$SREBP1_02 | caGCACcccac |
| V\$DELTAEF1_01 | cactACCTtgg |
| V\$NFY_01 | tggttaCCAAtttagtc |
| V\$NFY_Q6 | gtaCCAAttta |
| V\$TCF11_01 | TTCAtgctgttaa |
| V\$FREAC7_01 | tgtaaTAAAgacata |
| V\$XFD2_01 | ttaaTAAAgacata |
| V\$TATA_01 | ttaTAAAggaaagag |
| V\$NFAT_Q6 | taaagGAAAgag |
| V\$ER_Q6 | agagtttaagTGACcccac |
| V\$NKX25_01 | ttAAGTg |
| V\$AP1_Q4 | agTGACccaca |
| V\$AP1FJ_Q2 | agTGACccaca |
| V\$AP1_Q2 | agTGACccaca |
| V\$USF_Q6 | tcCACAcggc |
| V\$MZF1_01 | gctGGGGa |
| V\$IK2_01 | gctgGGGAggcc |
| V\$GF11_01 | ggcctcacAATCatggtggaagc |
| V\$SR_Y_02 | cctcACAAtcat |
| V\$SOX5_01 | ctcaCAATca |
| V\$TCF11_01 | ATCATggtggaag |
| V\$CEBPB_01 | tcatggtGGAAgac |
| V\$BARBIE_01 | aggcAAAGgagaaga |
| V\$GATA1_02 | caagaGATAactta |
| V\$GATA1_03 | caagaGATAactta |
| V\$GATA1_04 | aagaGATAactta |
| V\$GATA3_02 | agaGATAact |
| V\$GATA2_02 | agaGATAact |
| V\$GATA1_06 | agaGATAact |
| V\$GATA1_05 | agaGATAact |
| V\$LMO2COM_02 | gaGATAact |
| V\$GATA_C | aGATAActtat |
| V\$SR_Y_02 | acttACAAaacc |
| V\$VBP_01 | cTTACaaaac |
| V\$TCF11_01 | GTCAgatcttgtg |
| V\$GATA3_03 | tcAGATcttg |
| V\$CEBPB_01 | gtgtgggGGAAact |
| V\$MZF1_01 | gtgGGGGa |
| V\$IK1_01 | gtggGGGAaactg |
| V\$IK2_01 | gtggGGGAaact |
| V\$NFAT_Q6 | tggggGAAActg |
| V\$S8_01 | acgactcaATTAtctc |

| | |
|----------------|-------------------|
| V\$DELTAEF1_01 | ctccACCTggc |
| V\$USF_Q6 | tcCACCTggc |
| V\$MYOD_Q6 | tcCACCTggc |
| V\$USF_C | cCACCTgg |
| V\$NF1_Q6 | accTGGCcctgccttga |
| V\$GATA1_02 | cccttGATAtttgg |
| V\$GATA1_03 | cccttGATAtttgg |
| V\$GATA1_04 | ccttGATAtttgg |
| V\$LMO2COM_02 | ttGATAtt |
| V\$GATA_C | tGATATttggg |
| V\$SR_Y_02 | tattACAAtta |
| V\$SOX5_01 | attaCAATt |
| V\$SP1_Q6 | tttGGTGgggac |
| V\$GC_01 | tttGGTGgggaca |
| V\$LYF1_01 | tttGGTGgg |
| V\$IK2_01 | ggtgGGGAcaca |
| V\$MZF1_01 | ggtGGGGa |
| V\$OCT1_06 | ctctttgaaATTCT |
| V\$CEBPB_01 | aaattctGTAActc |
| V\$E4BP4_01 | aattctGTAAct |
| V\$HNF3B_01 | gtaacTCTTactcc |
| V\$GATA1_02 | ccacaGATAgttat |
| V\$GATA1_03 | ccacaGATAgttat |
| V\$CDPCR3_01 | cacagatagttATAG |
| V\$GATA1_04 | cacaGATAgttat |
| V\$LMO2COM_02 | caGATAgtt |
| V\$GATA_C | aGATAGttata |
| V\$AP4_Q6 | ttCAGCagga |
| V\$AP4_Q5 | ttCAGCagga |
| V\$CETS1P54_01 | gcAGGAtgga |
| V\$CEBPB_01 | atgtgaaGAAAaaa |
| V\$VMYB_01 | aaaAACTgag |
| V\$TCF11_01 | GTCAacagagtga |
| V\$VMYB_02 | gtcAACAg |
| V\$CEBPB_01 | gagtgaaGTAAtag |
| V\$TCF11_01 | GTCAGagaaatca |