

## СТРУКТУРНЫЕ ОСОБЕННОСТИ ДНК НА ГРАНИЦАХ АМПЛИКОНОВ ПРИ АМПЛИФИКАЦИИ ГЕНА *ERBB2* В КЛЕТКАХ ОПУХОЛИ РАКА МОЛОЧНОЙ ЖЕЛЕЗЫ

© 2013 г. Н. Ю. Матенко\*, С. П. Коваленко

Институт молекулярной биологии и биофизики Сибирского отделения Российской академии медицинских наук,  
Новосибирск, 630117

Поступила в редакцию 05.03.2013 г.

Принята в печать 11.04.2013 г.

В 25–30% случаев в клетках опухолей рака молочной железы (РМЖ) происходит амплификация фрагмента хромосомы вокруг гена *ERBB2*, что приводит к повышению экспрессии генов, располагающихся в данной области. Повышение экспрессии *ERBB2* и ряда других генов, прилежащих к *ERBB2*, могут оказывать влияние не только на рост и развитие опухоли, но и на ее чувствительность к различным видам терапии. До сих пор не установлены причины возникновения амплификации, точные границы *ERBB2*-содержащих ампликонов, а также специфические последовательности на границе амплифицируемых районов. Мы предположили, что границы амплифицируемой области определяются не столько последовательностью нуклеотидов сколько структурной организацией ДНК в этом районе и провели исследование по установлению границ *ERBB2*-содержащих ампликонов при РМЖ и структурных особенностей ДНК на границах этих ампликонов. С помощью ПЦР в режиме реального времени оценена доза 10 генов, прилежащих к *ERBB2*, а также самого гена *ERBB2* в 162 образцах опухолей РМЖ. Установлено несколько районов амплификации с границами внутри последовательности генов *IKZF3* и *RARA*. С помощью биоинформатического анализа нуклеотидной последовательности в районе гена *ERBB2* выявлены АТ-обогащенные участки с высокой степенью торсионной напряженности, способные формировать устойчивые вторичные структуры. Локализация идентифицированных сайтов совпадает с расположением границ *ERBB2*-содержащих ампликонов. Таким образом, показано, что границы *ERBB2*-содержащих ампликонов в опухолевых клетках РМЖ определяются структурными особенностями ДНК в этой области. Полученные результаты важны для понимания механизмов амплификации онкогенов.

**Ключевые слова:** ген *ERBB2*, рак молочной железы, амплификация, ПЦР в реальном времени, флексибельность ДНК.

**DNA STRUCTURAL FEATURES ON THE BORDERS OF *ERBB2* AMPLICONS IN BREAST CANCER**  
by N. U. Matsenko\*, S. P. Kovalenko (Institute of Molecular Biology and Biophysics, Siberian Division, Russian Academy of Medical Sciences, Novosibirsk, 630017 Russia; \*e-mail: N.matsenko@gmail.com). In 25–30% of breast cancer tumor cases amplification of chromosome fragment around the *ERBB2* underlies the increased expression of genes adjacent to *ERBB2*. Increased expression of genes within *ERBB2*-containing amplicons may have an impact not only on the growth and development of the tumor, but on the sensitivity of the tumor to different types of anti-cancer therapies. The initial cause of amplification and the exact borders of *ERBB2*-amplified chromosome fragment are still not completely characterized. No specific DNA sequences were found on the junction regions at intrachromosomal DNA amplification. We hypothesized that amplification borders can be specified by DNA structural peculiarities rather than the particular DNA sequence. This study focused on the mapping of *ERBB2* amplification borders in breast cancer and the search for unusual structural features of DNA at the borders of the identified amplicons. The copy number of 10 genes adjacent to *ERBB2* were evaluated by real time PCR in 162 breast cancer samples. Several *ERBB2*-containing amplicons of various lengths were revealed. In the majority of the analyzed samples, the borders of these amplicons were located within *ZNFN1A3* and *RARA* genes. A bioinformatics analysis of the nucleotide sequence peculiarities around

Принятые сокращения: aCGH (array-chromosomal genomic hybridization) — сравнительная геномная гибридизация на чипах; *ERBB2* - клеточный гомолог *v-erbB* онкогена вируса эритробластомы птиц типа 2; FISH (fluorescent *in situ* hybridization) — флуоресцентная *in situ* гибридизация; ИГХ — иммуногистохимический анализ; РМЖ — рак молочной железы.

\* Эл. почта: N.matsenko@gmail.com

*ERBB2* gene revealed the presence of AT-rich DNA regions with high degree of flexibility. These regions were able to form stable secondary structures. Positions of these sites strongly coincide with the positions of the *ERBB2*-containing amplicon borders found in real time PCR experiments. On the base of results obtained one can suppose that the structural features of DNA are involved in the formation of *ERBB2*-containing amplicon borders in breast cancer cells and the data are of importance for understanding the mechanisms of oncogene amplification.

**Keywords:** *ERBB2*, breast cancer, amplification, real time PCR, DNA flexibility.

DOI: 10.7868/S0026898413050145

## ВВЕДЕНИЕ

Увеличенная интенсивность экспрессии гена *ERBB2* показана в 25–30% случаев РМЖ и считается важнейшим прогностическим и диагностическим маркером этого заболевания [1, 2]. В подавляющем большинстве случаев в основе усиленной экспрессии лежит амплификация гена *ERBB2*. Причины и молекулярные механизмы амплификации гена *ERBB2* в опухолевых клетках РМЖ до сих пор не установлены. В литературе описано несколько механизмов внутривнутрихромосомной и внехромосомной амплификации онкогенов в опухолевых клетках [3–6]. Для реализации всех этих механизмов амплификации необходимо формирование двухцепочечного разрыва ДНК в качестве инициаторной точки, формирующей границу ампликона. Однако известно, что в эукариотической ДНК существуют сайты ломкости (фрагильные сайты, fragile sites) – специфические районы ДНК, которые обладают повышенной чувствительностью к появлению разрывов. Большинство сайтов ломкости картировано на цитологическом уровне в виде разрывов и брешей, появляющихся на метафазных хромосомах под действием условий репликативного стресса [7]. В ряде работ показано участие таких сайтов в процессах обмена между сестринскими хроматидами [8], в формировании транслокаций, делеций [9, 10], продемонстрирована их связь с рекомбинацией, интеграцией различных плазмидных и вирусных ДНК [11], амплификацией [12, 13]. Индукция сайта ломкости ведет к образованию двухцепочечного разрыва ДНК и может инициировать процесс амплификации. Более того, последовательная индукция двух фланкирующих сайтов ломкости может определять размер ампликона и расположение клона внутри амплифицированной области [14]. Следует отметить, что в литературе нет данных о наличии подобных сайтов ломкости в районе гена *ERBB2*.

Известно, что ДНК определенной группы сайтов ломкости обогащена повторяющимися АТ/ТА-динуклеотидными последовательностями, что обуславливает формирование фрагментов с высокой степенью торсионной напряженности (флексильности, гибкости). Такие фрагменты внутри

сайта ломкости названы флексильными сайтами. Флексильные сайты формируют устойчивые вторичные структуры, которые затрудняют работу репликативного комплекса и приводят к возникновению нестабильности и повышенной ломкости ДНК в этих районах [15–17].

Мы предположили, что флексильность не только определяет повышенную ломкость ДНК в описанных ранее сайтах ломкости, но сами по себе последовательности с высокой степенью флексильности могут быть причиной появления разрывов ДНК и последующих хромосомных перестроек в таких районах. В связи с этим возникновение *ERBB2*-содержащих ампликонов в геноме опухолевых клеток РМЖ может быть связано с присутствием последовательностей с повышенной чувствительностью к появлению ДНК-разрывов, в частности, участков высокой флексильности. В цели данного исследования входил поиск и идентификация сайтов высокой флексильности ДНК на границах ампликонов, содержащих ген *ERBB2* в опухолевых клетках РМЖ.

## ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

**Образцы ткани молочной железы.** В работе использованы 162 образца ткани опухоли молочной железы, из которых 34 получены в виде архивных срезов, а 128 – в ходе операций по мастэктомии в Новосибирской городской клинической больнице № 1 в период с 2003 по 2006 гг. (возраст пациентов 39–77, средний возраст 59). Забор операционного материала проводили с учетом согласия пациентов на использование их биологического материала для научных исследований. На проведение работ с клиническим материалом получено разрешение этического и научного комитета ФГБУ НИИ МББ СО РАМН.

Сразу после макродиссекции образец ткани опухоли РМЖ помещали в жидкий азот – до процедуры выделения ДНК. Образец считали пригодным для исследования, если содержание опухолевых клеток было выше 60%. Все образцы охарактеризованы по клинической классификации опухолей TNM и представляли собой панель различных гистологических форм РМЖ со 2-й и 3-й

степенью злокачественности: инфильтративные формы РМЖ (71.6%), неинфильтративные формы РМЖ (21.7%), воспалительный РМЖ (6.7%). В качестве контроля использованы 8 образцов нормальной ткани молочной железы и 5 образцов, взятых из прилегающей к опухоли нормальной ткани молочной железы.

**Выделение ДНК.** Выделение ДНК из архивных срезов проводили стандартным методом экстракции смесью фенол–хлороформ [18]. Выделение геномной ДНК из свежих образцов ткани проводили при помощи набора “Animal Tissue Genomic DNA Isolation kit” (“V-gene Biotechnology”, Китай) в соответствии с инструкцией производителя. Аликвоты выделенной геномной ДНК хранили при  $-20^{\circ}\text{C}$ . Концентрацию и качество полученных образцов ДНК оценивали по оптической плотности раствора (A) на спектрофотометре Hitachi-557 (“Hitachi”, Япония) при длинах волн 260, 230 и 280 нм.

**Праймеры и TaqMan зонды.** Последовательность праймеров и флуоресцентных TaqMan зондов подбирали с использованием Primer Express Software v2.0 (“Applied Biosystems”, США) и Primer3 (<http://frodo.wi.mit.edu>) (табл. 1). Для подтверждения специфичности ПЦР и отсутствия возможных полиморфизмов ДНК в районах гибридизации праймеров проведен анализ олигонуклеотидных последовательностей с помощью программы BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Для исключения возможности формирования праймер-димеров, шпилек и других вторичных структур последовательности олигонуклеотидов тестируют с помощью программы OligoAnalyzer v 3.0 (“Integrated DNA technologies”, Бельгия).

**ПЦР в реальном времени.** Определение количества копий генов проводили методом ПЦР в реальном времени с флуоресцентно-мечеными олигонуклеотидными пробами по типу TaqMan на приборе iQ 5 Real-Time PCR Detection System (“Bio-Rad”, США). Реакционная смесь с конечным объемом 25 мкл содержала 2 мкл раствора исследуемой геномной ДНК (5 нг/мкл), смесь для ПЦР (2×), 0.3 мкМ каждого прямого и обратного праймеров (ООО “Лаборатория Медиген”), 0.1–0.35 мкМ TaqMan пробы (ООО “Синтол”, Россия), 0.025 ед./мкл SmartTaq ДНК-полимеразы (ООО “Биолинк”, Россия). Условия проведения ПЦР: начальная денатурация при  $95^{\circ}\text{C}$  в течение 5 мин, затем 40 циклов при  $95^{\circ}\text{C}$  в течение 15 с и  $60^{\circ}\text{C}$  в течение 45 с. Значение флуоресценции рассчитывается автоматически в конце каждого цикла элонгации при температуре  $60^{\circ}\text{C}$  по каналам FAM (специфика) и ROX (контроль). Каждый образец опухолевой ткани анализировали как две пробы с разным количеством геномной ДНК (5 и 10 нг) с тремя повторами для каждой точки. Эффективность амплификации каждого ПЦР-про-

дукта определяли по калибровочным кривым, построенным по серии разведений раствора контрольной ДНК (2.5–40 нг). Строгая линейная зависимость между логарифмом начального числа копий и  $C_t$  наблюдалась для всех исследуемых генов ( $r^2 \geq 0.99$ ). Эффективность ПЦР составляла не менее 96%.

**Обработка результатов.** Дозу исследуемых генов в опухолевой ткани определяли относительно контрольного гена и нормировали по отношению к здоровой ткани молочной железы по методу, предложенному Пфафлом (Pfaffl) с соавт. [19]. В качестве контрольного гена использовали ген домашнего хозяйства UBC [20]. Для вычисления изменения дозы исследуемого гена в ДНК, выделенной из ткани опухолей РМЖ ( $N_{\text{иссл. ген}}$ ), относительно нормы использована следующая формула:

$$N_{\text{иссл. ген}} = \frac{(E_{\text{иссл. ген}})^{\Delta C_t(\text{норм. ДНК} - \text{опухолевая ДНК})}}{(E_{\text{контр. ген}})^{\Delta C_t(\text{норм. ДНК} - \text{опухолевая ДНК})}}$$

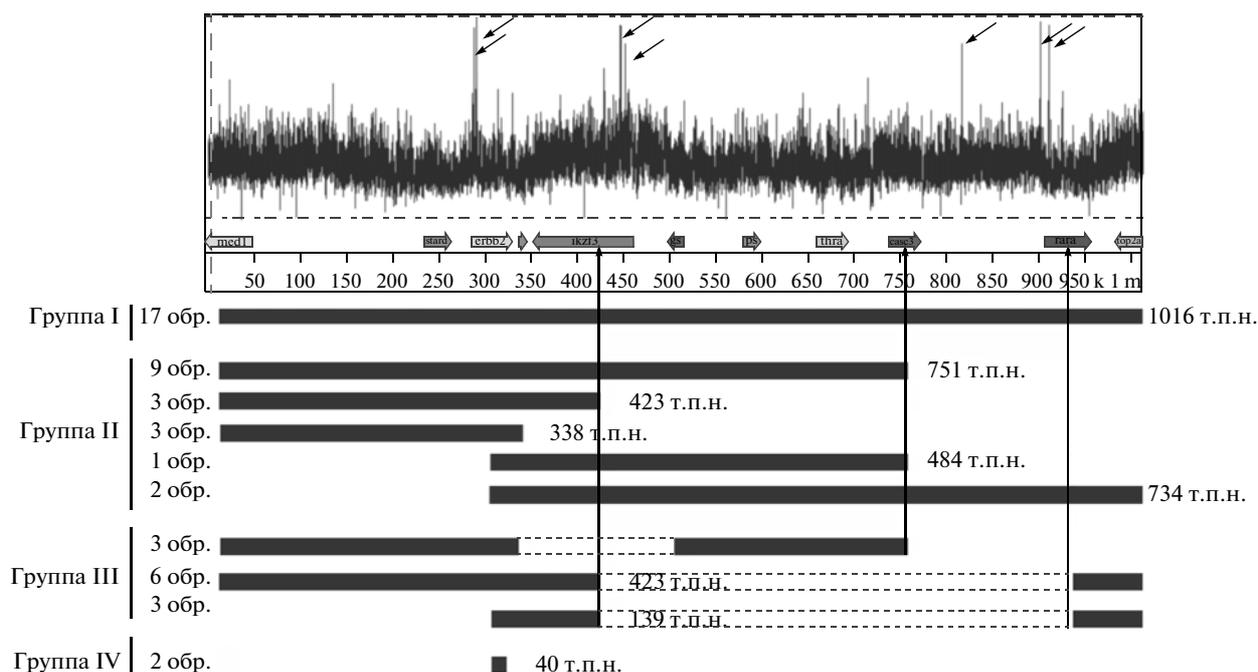
$C_t$  – точка в которой кривая флуоресценции пересекает заданное пороговое значение флуоресценции;  $\Delta C_t$  – разница между значениями  $C_t$  для гена в нормальном и опухолевом образцах ДНК. Эффективность ПЦР ( $E$ ) определяли по формуле  $E = 10^{1/m}$ , где  $m$  – это угол наклона калибровочной прямой для соответствующего гена. Дозу исследуемого гена в опухолевом образце считали увеличенной, если  $N_{\text{иссл. гена}} \geq 1.5$ . В пользу порогового значения 1.5, а не 2 (если принимать во внимание, что увеличение дозы гена происходит целыми числами) говорит то обстоятельство, что геномная ДНК, выделенная из образцов опухолевой ткани, содержала примеси ДНК нормальных клеток (не более 40%).

**Компьютерный анализ нуклеотидной последовательности.** Флексибельность ДНК в районе MED1–ERBB2–TOP2A анализировали при помощи универсальной биоинформатической платформы UGENE [21]. Программа рассчитывает флексибельность ДНК в углах скручивания между нуклеотидами в одной цепи ДНК. Расчет производится для окна размером 100 п.н. на протяжении заданного района со сдвигом рамки на 1 п.н. Внутри каждого окна значение флексибельности вычисляется для каждой соседней пары нуклеотидов и среднее значение всех пар в окне приписывается первой динуклеотидной паре. Соседние окна со значением флексибельности больше  $13.7^{\circ}$  называют пиками флексибельности. Нуклеотидная последовательность взята из базы данных GenBank (NT\_010755(37 560 538..3857420)). Анализ способности флексибельных последовательностей формировать устойчивые вторичные структуры проводили с помощью программы Mfold [22]. Для определения зависимости устойчивости вторичной структуры от расположения нуклеотидов в последовательности проводили

Таблица 1. Праймеры и флуоресцентно-меченные пробы для ПЦР в реальном времени

Название гена	Gene ID*	Прямой праймер	Обратный праймер	TaqMan зонд
<i>MEDI</i>	5469	5'-AGGAGAAGAGGGTTGTGATGAG-3'	5'-CTTCTGСААТGTCTCCAААСАG-3'	5'-TCTGGAGGGСАТСААСАТТТGGTCA-3'
<i>STARD3</i>	10948	5'-AGGTTCTCTCCATCCCTACC-3'	5'-TACGGATGCTGTGTGGTGC-3'	5'-GGGGCAGCCCCAGGCCCTTGC-3'
<i>ERBB2</i>	2064	5'-CCCAAGCCCAАААГАААТC-3'	5'-GGGTCTCTCTCTCTCCTCCTA-3'	5'-CGTAAGCCCCCTCCACCCTCTCCT-3'
<i>GRB7</i>	2886	5'-CTGCTTTTGGCCCTTGTCCC-3'	5'-CACTCACTGGGGAGCTCTTG-3'	5'-TGCAGGCTGCTGTGCCCCGTGGGC-3'
<i>IKZF3</i>		5'-GCAGCGGTTTGTGAATGACTAC-3'	5'-ACTACGGCTCACCTCTATGTC-3'	5'-TGGAGAAGGCCAGCCCAATGAA-3'
<i>flankIKZF3-1082</i>	22806	5'-GCACCTGGCАТCTCCTACTCA-3'	5'-GCCACAGTACGGGATATCAА-3'	5'-AGTTGCTCTGCTTGACTCATTTG-3'
<i>flankIKZF3-225</i>		5'-AATCTGTTGGCCCTCCATTC-3'	5'-GGCCTTCTGTGGCTTTTG-3'	5'-AGCGTATGAGGGATGAAATGCGAAGG-3'
<i>GSDML</i>	55876	5'-CCCAААAGGAGACGATGAG-3'	5'-CTGACAGAGCCTGGTTGAATC-3'	5'-CAAAAGGCAGATGGAGAACAGATGGG-3'
<i>PSMD3</i>	5709	5'-CAGTCACTTGGAGGGTACG-3'	5'-AGGTTCACTATTGGCCCTG-3'	5'-AGCCCCAGCCGGGAACGTGCCCGC-3'
<i>THRA</i>	7067	5'-TGCCTCTCCATCACTCATA-3'	5'-GTTCCSAGGCAGGAGTTAC-3'	5'-GCAGGGCTGGAGAGGGCAGGGCGC-3'
<i>CASC3</i>	22794	5'-CCCCCTCCACTGAGGTATGA-3'	5'-CAGAGAAACACACCCAGAAAATGA-3'	5'-CATGCTCCGTCACTGGGCTTCTCCT-3'
<i>flankRARacent</i>		5'-GTGTGCTCTGCATCTTCCA-3'	5'-GTCTCCAGTTCAGGCCCTCTG-3'	5'-TTTAGGGGACCCGGGGAGCC-3'
<i>RARA</i>	5914	5'-GTGTCAACCGGACAAGAACT-3'	5'-CATGTGCACAААСАТССАСА-3'	5'-AACTGAGCGGGACTCGGGCC-3'
<i>flankRARAtel</i>		5'-AGGGTCTCTAGTTCATGTG-3'	5'-ATGGCCTTCCCTTAGAGAA-3'	5'-CCCTTGTCTGGCAGGGCTC-3'
<i>TOP2A</i>	7153	5'-CCCCCTCTGAGTTCACAAC-3'	5'-GTTCCGGCAGCAGCAGTTTG-3'	5'-AGGTGGCCAAAGGATTTGAACGGA-3'
<i>UBC</i>	7316	5'-GGAGCGCAGCAAAATGG-3'	5'-CATTAGCGAAGGCCCTCAAGAC-3'	5'-CGGGCTGTGAGGCTGTGAAACAAG-3'

\*Идентификационный номер гена в базе данных GenBank.



**Рис. 1.** Флексибельность последовательности ДНК и длины ампликонов в районе *MED1-ERBB2-TOP2A* в клетках опухолей РМЖ. Анализ степени флексибельности проведен с помощью программы Ugene («Унипро», Россия). Нуклеотидная последовательность взята из базы данных GenBank (вариант генома GRCh37.p10, NC\_000017.10 (37 560 538.. 38 574 20)). Косыми черными стрелками обозначены пики с максимальным значением флексибельности. Описание пиков представлено в таблице 2. Расположение и ориентация генов хромосоме 17q указаны под графиком флексибельности в виде горизонтальных стрелок. Слева направо: гены *MED1*, *STARD3*, *ERBB2*, *GRB7*, *IKZF3*, *GSDML*, *PSMD3*, *THRA*, *CASC3*, *RARA*, *TOP2A*. Центромерный район располагается слева. Оценка длины ампликонов в ДНК клеток опухолей РМЖ проведена методом ПЦР в реальном времени. Под графиком флексибельности черными прямоугольниками указаны типы и расположение ампликонов относительно анализируемых генов. Пунктирной линией указаны районы с неувеличенной дозой генов в данных образцах (группа III). Справа указаны длины ампликонов (т.п.н.) Слева указано число образцов РМЖ, в которых установлен определенный тип ампликона. В зависимости от типа ампликонов образцы разделены на 4 группы (I–IV). Пояснение см. в тексте. Вертикальными стрелками выделены дистальные и проксимальные границы ампликонов, которые располагаются в области генов *IKZF3*, *RARA* и *CASC3* (подгруппы № 1, № 2 и № 3). Пояснение см. в тексте.

сравнительный анализ значений свободной энергии образования вторичных структур ( $dG$ ) для оригинальной последовательности флексибельного сайта и контрольных последовательностей. В качестве контрольных использованы значения  $dG$  для фланкирующих последовательностей, а также для искусственно сгенерированных последовательностей идентичного нуклеотидного состава ( $p < 0.01$ ).

**Статистическая обработка данных.** Статистическую обработку данных ПЦР проводили по критериям Шапиро–Уилка и Колмогорова–Смирнова при помощи программы STATISTICA 6.0 («Statsoft», Швеция). Значение  $p < 0.01$  использовано как критерий достоверности различий полученных значений.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

### Оценка длины *ERBB2*-содержащих ампликонов в клетках опухолей РМЖ методом ПЦР в реальном времени

Методом ПЦР в реальном времени увеличение дозы гена *ERBB2* выявлено в 49 из 162 образцов

(30.2%) опухолей РМЖ ( $N_{ERBB2} \geq 1.5$ ). Для оценки минимального района амплификации хромосомы вокруг *ERBB2* выбраны смежные гены: *MED1*, *STARD3*, *GRB7*, *IKZF3*, *GSDML*, *PSMD3*, *THRA*, *CASC3*, *RARA*, *TOP2A*, – располагающиеся на расстоянии не более 1015 т.п.н. (рис. 1). В случае увеличенной дозы двух смежных генов считали, что фланкированный ими район также амплифицирован. В качестве границы амплификации района хромосомы принимали точку между смежными генами с увеличенной и неувеличенной дозой. Образцы, в которых наблюдалась амплификация в районе *MED1-ERBB2-TOP2A*, в зависимости от длины ампликонов распределили по четырем группам (рис. 1). В группу I вошли образцы с увеличенной дозой всех генов в районе *MED1-ERBB2-TOP2A* (17 из 49 образцов; 34.7%); в группу II – с увеличенной дозой гена *ERBB2* и одного или более соседних генов (18 из 49 образцов; 36.7%); в группу III – с увеличенной дозой гена *ERBB2* и одного или более генов-мишеней с прерыванием неамплифицированными/делетированными генами-мишенями (12 из 49 образцов;

24.5%); в группу IV – с увеличением дозы только гена *ERBB2* без амплификации окружающих генов (2 из 49 образцов; 4.1%). Наличие лишь двух образцов в группе IV свидетельствует о том, что амплификация единственного гена представляет собой достаточно редкое событие в геноме опухолевых клеток.

Поскольку границы амплификации в образцах группы I выходили за пределы изучаемого района, мы исключили их из дальнейшего анализа. Интересно, что в 12 образцах, входящих в группы II и III, граница ампликонов находилась между генами *IKZF3* и *GSDML*, а в 13 – между генами *CASC3* и *RARA*. Эти образцы составили подгруппу № 1 и подгруппу № 2. Кроме того, в 9 образцах группы III проксимальная граница *TOP2A*-содержащего ампликона также располагалась между генами *CASC3* и *RARA*. Эти образцы объединили в подгруппу № 3. Для изучения особенностей структуры ДНК, которая может оказывать влияние на формирование границ ампликонов, мы исследовали гибкость ДНК в районе *MED1-ERBB2-TOP2A*.

#### **Последовательность ДНК в районе *MED1-ERBB2-TOP2A* содержит сайты высокой гибкости**

ДНК в районе *MED1-ERBB2-TOP2A* содержит 7 участков с высокой гибкостью (значение гибкости > 15.0): в интронных последовательностях генов *ERBB2*, *IKZF3*, *RARA* и в некодирующей последовательности между генами *CASC3* и *RARA* (рис. 1). Для обозначения сайтов высокой гибкости использовали название гена, в котором он обнаружен, с приставкой “*Flex*”. Описание сайтов представлено в таблице 2.

Районы ДНК с высокой гибкостью обладали повышенным содержанием А/Т-нуклеотидов (79% ± 6.7%) по сравнению с прилежащими последовательностями той же длины (55% ± 2.8%,  $p < 0.01$ ). Кроме того, гибкие сайты суще-

ственно обогащены АТ/ТА-динуклеотидами ( $23 \pm 3.3\%$  и  $25 \pm 5.0\%$  соответственно) по сравнению с нефлексибельными районами ( $5.66 \pm 0.49\%$  и  $5.38 \pm 1.56\%$  соответственно,  $p < 0.01$ ).

#### **Гибкие последовательности в генах *IKZF3* и *RARA* способны формировать устойчивые вторичные структуры**

Гибкие сайты *FlexIKZF3-1082*, *FlexIKZF3-225* и *FlexRARA* способны формировать устойчивые вторичные структуры с более низким значением *dG* по сравнению с контрольными значениями (рис. 2). Все проанализированные вторичные структуры характеризовались наличием одной или нескольких удлиненных и сложно-разветвленных шпилек, в образовании которых участвовало более 85%, а в случае *FlexIKZF3-2* – более 95% нуклеотидной последовательности гибкого участка. Именно последовательности *FlexIKZF3-1082*, *FlexIKZF3-225* и *FlexRARA* выбраны нами для дальнейшего исследования в качестве потенциальных границ *ERBB2*-содержащих ампликонов.

#### **Различный уровень амплификации в районе сайтов *FlexIKZF3-1082*, *FlexIKZF3-225* и *FlexRARA***

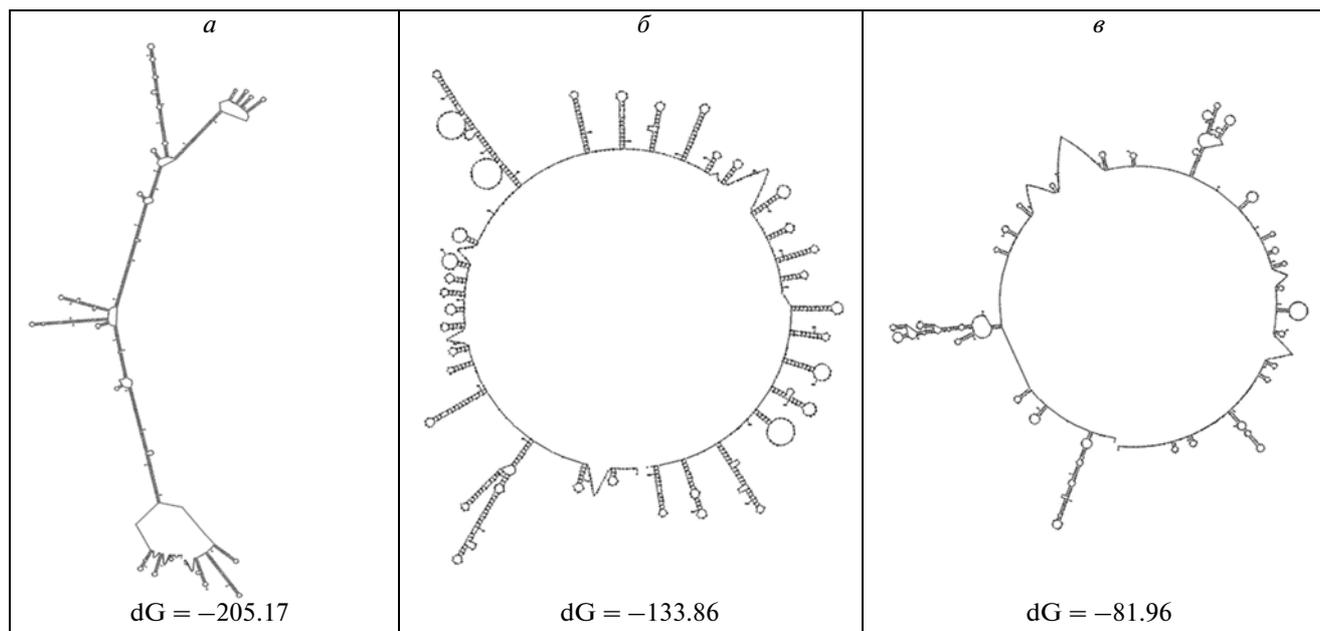
С целью проверить предположение, что сайты *FlexIKZF3-1082*, *FlexIKZF3-225* и *FlexRARA* принимают участие в формировании границ амплификации, мы провели более тонкое картирование границы ампликонов по отношению к этим участкам. Проанализировав уровень амплификации фрагментов ДНК, фланкирующих сайты *FlexIKZF3-1082* и *FlexIKZF3-225* на расстоянии 19 985 п.н. в образцах подгруппы № 1 (рис. 3а), мы обнаружили, что из 12 образцов в десяти (83.3%) граница ампликона располагается в области *FlexIKZF3-1082* и *FlexIKZF3-225*, а в двух (16.7%) – дистальнее сайтов *FlexIKZF3-1082* и *FlexIKZF3-225*.

Анализ дозы фрагментов ДНК, фланкирующих *FlexRARA* на расстоянии 13096 п.н., проведен

**Таблица 2.** Гибкие сайты в районе *MED1-ERBB2-TOP2A*

Название	Локализация	Значение гибкости	Длина, п.н.	Нуклеотидная позиция*
<i>FlexERBB2-1</i>	Ген <i>ERBB2</i>	15.714	147	37852283 .. 37852930
<i>FlexERBB2-2</i>	Ген <i>ERBB2</i>	16.312	180	37855143 .. 37855323
<i>FlexIKZF-1082</i>	Ген <i>IKZF3</i>	16.045	1082	38010948 .. 38012030
<i>FlexIKZF-225</i>	Ген <i>IKZF3</i>	15.267	225	38016857 .. 38017082
<i>Flex_non-cod1</i>	Некод. район	15.262	140	38381472 .. 38381612
<i>Flex_non-cod1</i>	Некод. район	16.052	169	38466714 .. 38466883
<i>FlexRARA</i>	Ген <i>RARA</i>	16.039	166	38476005 .. 38476171

\*Хромосома 17, вариант генома GRCh37.p10, NC\_000017.10.



**Рис. 2.** Сайт *FlexIKZF3-1082* способен формировать устойчивые вторичные структуры. Анализ проведен с помощью программы Mfold. Нуклеотидная последовательность взята из базы данных GenBank (вариант генома GRCh37.p10, NC\_000017.10 (38 010 948..38 012 03)). ДНК сайта *FlexIKZF3-1082* формирует более устойчивые ( $dG = -205.17$ ) вторичные структуры (а) по сравнению с искусственно сгенерированными последовательностями той же длины и нуклеотидного состава ( $dG = -133.86$ ), но с произвольным расположением нуклеотидов (б) и с фланкирующими нефлексибельными последовательностями ( $dG = -81.96$ ) той же длины (в).

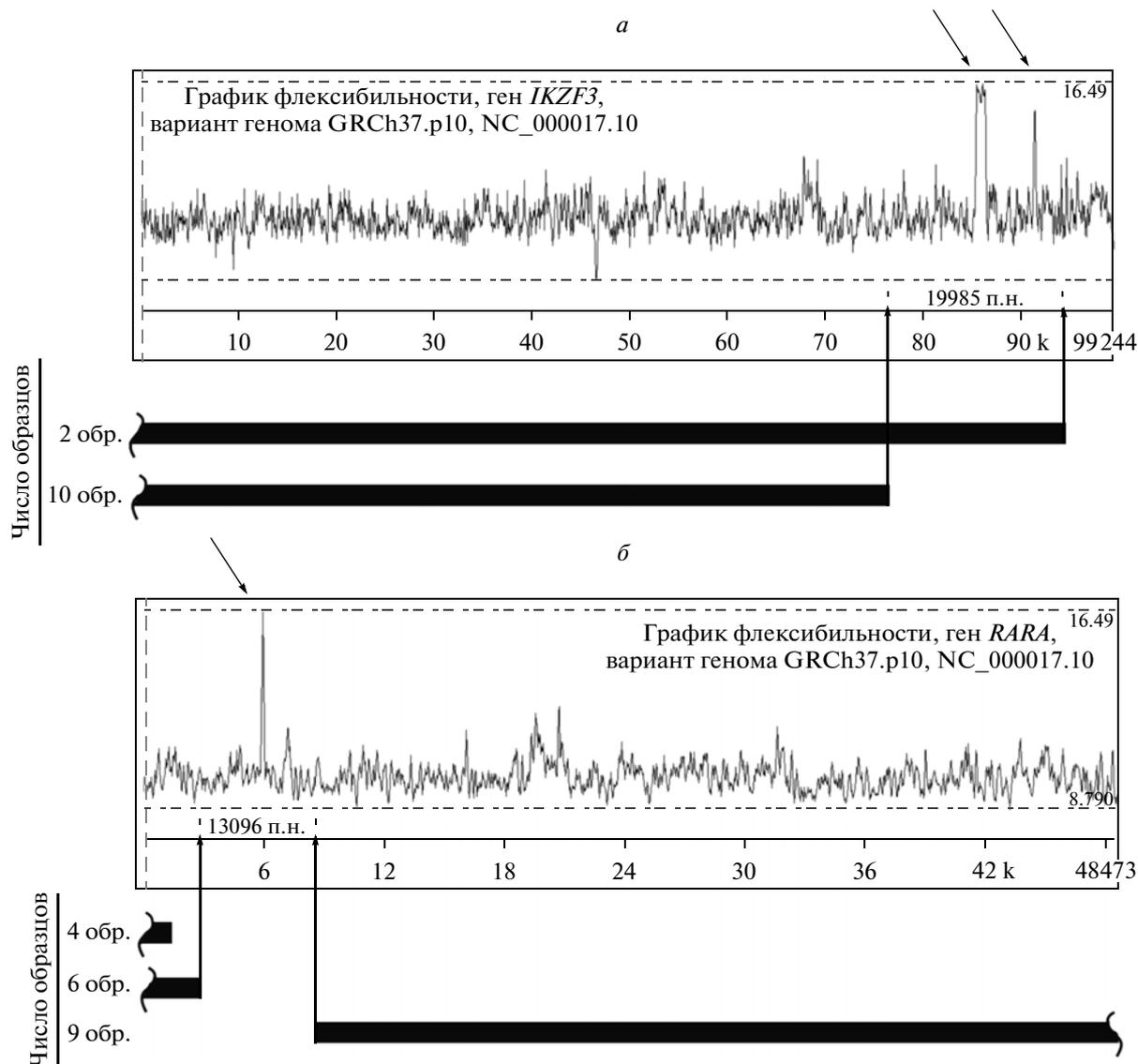
в подгруппах № 2 и № 3 (рис. 3б). Из 13 образцов подгруппы № 2 в шести (46.1%) граница ампликона располагалась в области *FlexRARA*; в четырех (30.7%) образцах амплификация фланкирующих фрагментов ДНК не выявлена; в трех (23.1%) количество ДНК оказалось недостаточным для проведения анализа. Во всех девяти образцах подгруппы № 3 проксимальная граница *TOP2A*-содержащего ампликона располагалась в области *FlexRARA* (рис. 3б).

## ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Амплификация *ERBB2* и ряда смежных генов — достаточно частое событие при заболевании РМЖ [23, 24]. Причины запуска амплификации гена *ERBB2* до сих пор не установлены. В литературе описано несколько моделей механизма амплификации онкогенов [3–6]. Их общая черта — необходимость появления разрыва ДНК с теломерной стороны онкогена в качестве инициаторного события амплификации. Попытки выявить специфические последовательности на границе *ERBB2*-ампликонов не увенчались успехом. Мы предположили, что границы амплифицируемого района определяются не столько специфической последовательностью нуклеотидов, сколько структурными особенностями ДНК, и попытались установить границы *ERBB2*-содержащих ампликонов

при РМЖ и выявить структурные особенности ДНК на их границах.

Длину ампликонов, содержащих ген *ERBB2*, оценивали в клинических образцах опухолей РМЖ, используя метод ПЦР в реальном времени. Возможность использования данного подхода для картирования границ *ERBB2*-ампликонов продемонстрирована в ряде исследований [25, 26]. Мы определили дозу девяти генов, смежных с *ERBB2* и находящихся от него на расстоянии не более 1015 т.п.н., в 162 образцах тканей РМЖ и в 49 из них (30.3%) выявлено наличие ампликонов разной длины (рис. 1). В зависимости от типа и длины ампликонов образцы распределили по четырем группам. Однако, в соответствие с известными моделями амплификации онкогенов, разрыв ДНК с дистальной стороны от онкогена — необходимое условие активации процесса амплификации. С учетом этого факта из групп II и III мы выбрали образцы с одинаковым расположением дистальной границы ампликонов и распределили их по трем подгруппам. В подгруппу № 1 вошло 12 (24.4%) образцов с расположением дистальной границы ампликона между генами *IKZF3* и *GSDML*, в подгруппу № 2 — 13 (26.5%) образцов с расположением дистальной границы ампликона между генами *CASC3* и *RARA*, в подгруппу № 3 — 9 (18.4%) образцов с расположением проксимальной границы ампликона между генами *CASC3* и *RARA*. На основании полученных результатов можно предпола-



**Рис. 3.** Картирование границ амплификации в образцах трех подгрупп относительно сайтов *FlexIKZF3-1082*, *FlexIKZF3-225* (а) и *FlexRARA* (б). В верхней части рисунка (а) приведен график гибкости гена *IKZF3*. Анализ степени гибкости проведен с помощью программы Ugene. Нуклеотидная последовательность взята из базы данных GenBank (варианта генома GRCh37.p10, NC\_000017.10 (37 913 968..38 020 441 complement)). Косыми черными стрелками указаны сайты *FlexIKZF3-1082*, *FlexIKZF3-225*. Вертикальными стрелками указано расположение районов, фланкирующих область высокой гибкости на расстоянии 19 985 п.н., выбранных для тонкого картирования границ ампликонов в образцах подгруппы № 1. Степень амплификации этих районов определяли методом ПЦР в реальном времени. Последовательности пар праймеров *flankIKZF3-1082* и *flankIKZF3-225* приведены в таблице 1. Под графиком гибкости черными прямоугольниками указаны уточненные расположения границ ампликонов относительно сайтов *FlexIKZF3-1082*, *FlexIKZF3-225* в образцах ДНК опухолей РМЖ из подгруппы №1. Слева указано число образцов РМЖ, в которых установлен данный тип расположения границы ампликона. Пояснение см. в тексте. В верхней части рисунка (б) приведен график гибкости гена *RARA*. Нуклеотидная последовательность взята из базы данных GenBank (варианта генома GRCh37.p10, NC\_000017.10 (38 465 423..38 513 895)). Косой черной стрелкой указан сайт *FlexRARA*. Вертикальные стрелки показывают местоположение участков, фланкирующих область высокой гибкости на расстоянии 13 096 п.н., — эти участки выбраны для тонкого картирования границ ампликонов в образцах подгрупп № 1 и № 3. Степень амплификации этих районов определяли методом ПЦР в реальном времени. Последовательности пар праймеров *flankRARAcen* и *flankRARAtel* указаны в таблице 1. Под графиком гибкости черными прямоугольниками указаны уточненные расположения границ ампликонов относительно сайта *FlexRARA* в образцах ДНК опухолей РМЖ из подгрупп № 2 и № 3. Слева указано число образцов РМЖ, в которых установлен данный тип расположения границы ампликона. Пояснение см. в тексте.

гать, что ДНК в этих областях обладает повышенной чувствительностью к появлению разрывов, формирующих границы ампликонов. Какие же особенности ДНК могут обуславливать ее чувствительность к появлению разрывов?

В геноме человека описаны сайты ломкости — специфические районы ДНК, которые обладают повышенной чувствительностью к появлению разрывов за счет особенностей нуклеотидной структуры. Коквелле (Coquelle) с соавт. [14] показали, что экспрессия сайтов ломкости ведет к образованию двухцепочечного разрыва ДНК и может инициировать процесс амплификации. Последовательная индукция экспрессии двух фланкирующих сайтов ломкости может определять размер ампликона и расположение клона внутри амплифицированной области [13]. Изучение причин повышенной чувствительности ДНК к разрывам привело к открытию группы часто встречающихся сайтов ломкости (common fragile sites), в состав которых входят АТ/ТА-обогащенные участки ДНК с высокой степенью торсионной напряженности. Такие фрагменты внутри сайтов ломкости названы гибкими сайтами. Гибкие сайты формируют устойчивые вторичные структуры, которые затрудняют работу репликативного комплекса и приводят к возникновению нестабильности и повышенной ломкости ДНК в этих районах [14–16, 27]. Следует отметить, что в литературе нет данных о наличии подобных сайтов ломкости в области гена *ERBB2*. Мы предположили, что именно гибкость становится причиной появления разрывов ДНК и последующих хромосомных перестроек в этих районах. В связи с этим возникновение *ERBB2*-содержащих ампликонов в геноме опухолевых клеток РМЖ может быть связано с возможным присутствием в этом районе сайтов высокой гибкости. Биоинформатическое исследование структурных особенностей ДНК в районе *MED1–ERBB2–TOP2A* показало наличие семи сайтов с высокой степенью гибкости (рис. 1), три из которых (*FlexIKZF3-1082*, *FlexIKZF3-225* и *FlexRARA*) способны формировать устойчивые вторичные структуры (рис. 2). Сопоставление данных *in silico* по расположению сайтов высокой гибкости в районе *MED1–ERBB2–TOP2A* и результатов ПЦР в режиме реального времени по установлению границ *ERBB2*-амплификации в образцах РМЖ показало, что в формировании границ амплификации могут принимать участие сайты *FlexIKZF3-1082*, *FlexIKZF3-225* и *FlexRARA* (рис. 1).

С целью уточнить место расположения дистальной границы района амплификации в образцах подгруппы № 1 проведен дополнительный анализ амплификации фрагментов ДНК, максимально близко фланкирующих *FlexIKZF3-1082* и *FlexIKZF3-225* (рис. 3а). В 10 из 12 (83.3%) образцов граница ампликона располагалась

внутри фланкируемой области в районе *FlexIKZF3-1082* и *FlexIKZF3-225*. Это свидетельствует о том, что эти сайты гибкости принимают участие в формировании границ амплификации. В 6 из 13 (46.1%) образцов подгруппы № 2 уровень амплификации районов, максимально близко фланкирующих *FlexRARA*, увеличен для проксимального фрагмента; при этом амплификация дистального фрагмента отсутствует (рис. 3б). Это свидетельствует о том, что сайт *FlexRARA* принимает участие в формировании дистальной границы *ERBB2*-содержащих ампликонов. Отсутствие амплификации обоих фланкирующих районов в четырех (30.7%) образцах подгруппы № 2 может быть связано с влиянием сайта гибкости, располагающегося проксимальнее гена *RARA* в некодирующем районе ДНК (рис. 1). Кроме того, для девяти образцов группы III обнаружено, что сайт *FlexRARA* принимает участие в формировании проксимальной границы *ERBB2*-независимых ампликонов, в состав которых входит ген *TOP2A*. Полученные данные свидетельствуют о том, что гибкие сайты ДНК могут определять границы *ERBB2*-содержащих ампликонов в клетках опухолей РМЖ и быть независимыми маркерами участков ДНК с повышенной чувствительностью к появлению разрывов и последующими хромосомными перестройками. Благодаря развитию современных технологий, позволяющих исследовать ДНК на нуклеотидном уровне, стало возможно изучать роль отдельных участков ДНК в механизмах внутривхромосомной амплификации. В 2012 г. Мароттаб (Marottab) с соавт. [28] описали длинный (>300 т.п.н.) район, состоящий из ряда дупликаций генов *KRTAP*, со сложной структурной организацией и присутствием однонуклеотидных полиморфизмов, который располагается на 1500 т.п.н. дистальнее гена *ERBB2*. Авторы предположили, что неординарная структура этой области может приводить к формированию границ амплификации гена *ERBB2*. Полученные нами результаты также подтверждают, что область хромосомы вокруг гена *ERBB2* имеет сложную организацию. Наличие сложных структурных перестроек, ампликонов разной длины дает основание предполагать, что существует целый комплекс причин, вызывающих амплификацию гена *ERBB2* в опухолях; при этом важнейшую роль в инициации процесса амплификации играют структурные особенности ДНК. На основании результатов проведенного нами исследования можно предполагать, что существенную роль в инициации амплификации играют участки ДНК с высокой торсионной напряженностью (гибкие участки), в которых формируются характерные вторичные структуры. Детальное исследование таких районов как регуляторных последовательностей позволит внести существенный вклад в понимание молекулярных механизмов онкогенеза.

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации (Госконтракт № 14.740.12.0819) в рамках Федеральной Целевой Программы “Научные и научно-педагогические кадры инновационной России” на 2009–2013 годы.

Работа поддержана Российским фондом фундаментальных исследований (проект № 12-04-32216).

## СПИСОК ЛИТЕРАТУРЫ

- Jacobson K.K., Morrison L.E., Henderson B.T., Blondin B.A., Wilber K.A., Legator M.S., O'Hare A., Van Stedum S.C., Proffitt J.H., Seelig S.A., Coon J.S. 2004. Gene copy mapping of the *ERBB2/TOP2A* region in breast cancer. *Genes Chromosomes Cancer*. **40**, 19–31.
- Поляновский О.Л., Лебеденко Е.Н., Деев С.М. 2012. ERBB-онкогены – мишени моноклональных антител. *Биохимия*. **77**, 289–289.
- McClintock B. 1951. Chromosome organization and genic expression. *Cold Spring Harbor Symp. Quant. Biol.* **6**, 13–47.
- Singer M.J., Mesner L.D., Friedman C.L., Trask B.J., Hamlin J.L. 2000. Amplification of the human dihydrofolate reductase gene via double minutes is initiated by chromosome breaks. *Proc. Natl. Acad. Sci. USA*. **97**, 7921–7926.
- Kuwahara Y. 2004. Alternative mechanisms of gene amplification in human cancers. *Genes Chromosomes Cancer*. **41**, 125–132.
- Копнин Б.П. 2007. Нестабильность генома и онкогенез. *Молекуляр. биология*. **41**, 369–380.
- Zlotorynski E., Rahat A., Skaug J., Ben-Porat N., Ozeri E., Hersberg R. 2003. Molecular basis for expression of common and rare fragile sites. *Mol. Cell Biol.* **2**, 7143–7151.
- Glover T., Stein C. 1987. Induction of sister chromatid exchanges at common fragile sites. *Am. J. Hum. Genet.* **41**, 882–890.
- Glover T., Stein C. 1988. Chromosome breakage and recombination at fragile sites. *Am. J. Hum. Genet.* **43**, 265–273.
- Wang N., Testa J., Smith D. 1993. Determination of the specificity of aphidicolin-induced breakage of the human 3p14.2 fragile site. *Genomics*. **17**, 341–347.
- Rassool F., McKeithan T., Neilly M., van Melle E., Espinosa III R., Le Beau M. 1991. Preferential integration of marker DNA into the chromosomal fragile site at 3p14: an approach to cloning fragile sites. *Proc. Natl. Acad. Sci. USA*. **88**, 6657–6661.
- Hellman A., Zlotorynski E., Scherer S.W., Cheung J., Vincent J.B., Smith D.I. 2002. A role for common fragile site induction in amplification of human oncogenes. *Cancer Cell*. **1**, 89–97.
- Miller C. 2006. Genomic amplification of MET with boundaries within fragile site FRA7G and upregulation of MET pathways in esophageal adenocarcinoma. *Oncogene*. **25**, 409–418.
- Coquelle A. 1997. Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons. *Cell*. **89**, 215–225.
- Fry M., Loeb L.A. 1994. The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. *Proc. Natl. Acad. Sci. USA*. **91**, 4950–4954.
- Gacy A.M., Goellner G., Juranic N., Macura S., McMurray C.T. 1995. Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell*. **81**, 533–540.
- Pearson C.E., Eichler E.E., Lorenzetti D., Kramer S.F., Zoghbi H.Y. 1998. Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand DNA (S-DNA) formation. *Biochemistry*. **37**, 2701–2708.
- Cao W., Hashibe M., Rao J. 2003. Comparison of methods for DNA extraction from paraffin-embedded tissues and buccal cells. *Cancer Detect. Prev.* **27**, 397–404.
- Pfaffl M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucl. Acids Res.* **29**, 2002–2007.
- Persson K., Pandis N., Mertens F., Borg A., Baldetorp B., Killander D. 1999. Chromosomal aberrations in breast cancer: a comparison between cytogenetics and comparative genomic hybridization. *Genes Chromosomes Cancer*. **25**, 115–122.
- Okonechnikov K., Golosova O., Fursov M. 2012. Uni-pro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. **28**, 1166–1167.
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.* **31**, 3406–3415.
- Edgren H., Murumagi A., Kangaspeka S., Nicoric D., Hongisto V., Kleivi K., Rye I.H., Nyberg S., Wolf M., Borresen-Dale A.L., Kallioniemi O. 2011. Identification of fusion genes in breast cancer by paired-end RNA-sequencing. *Genome Biol.* **12**, 345–358.
- Hyman E., Kauraniemi P., Hautaniemi S., Wolf M., Mousses S., Rozenblum E. 2002. Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res.* **62**, 6240–6245.
- Arriola E., Marchio C., Tan D.S., Drury S.C., Lambros M.B., Natrajan R., Rodriguez-Pinilla S.M., Mackay A., Tamber N., Fenwick K., Jones C., Dowsett M., Ashworth A., Reis-Filho J.S. 2008. Genomic analysis of the HER2/TOP2A amplicon in breast cancer and breast cancer cell lines. *J. Tech. Meth. Pathol.* **88**, 491–503.
- Sircoulomb F., Bekhouche I., Finetti P., Adelaide J., Ben Hamida A., Bonansea J., Raynaud S., Innocenti C., Charafe-Jauffret E., Tarpin C., Ben Ayed F., Viens P., Jacquemier J., Bertucci F., Birnbaum D., Chaffanet M. 2010. Genome profiling of ERBB2-amplified breast cancers. *BMC Cancer*. **10**, 539–557.
- Mishmar D., Rahat A., Scherer S.W., Nyakatura G., Hinzmann B. Molecular characterization of a common fragile site (FRA7H) on human chromosome 7 by the cloning of a simian virus 40 integration site. 1998. *Proc. Natl. Acad. Sci. USA*. **95**, 8141–8146.
- Marottab M., Chen X., Inoshita A., Stephens R., Thomas Budd G., Crowe J.P., Lyons J., Kondratova A., Tubbs R., Tanaka H. 2012. A common copy-number breakpoint of ERBB2 amplification in breast cancer colocalizes with a complex block of segmental duplications. *Breast Cancer Res.* **14**, R150.