

Секвенирование нового поколения

Демидова И.А.

Лаборатория молекулярной биологии МГОБ 62, Москва

*In the beginning there was nothing.
God said : 'Let there be light !'
There was still nothing, but now
you could see it. (Dave Thomas)*

В начале была пустота. Господь
сказал: «Да будет свет!»
Пустота осталась, но теперь ее
стало видно (Дейв Томас)

Немного истории



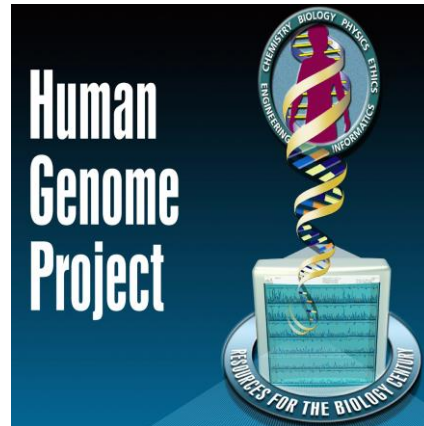
Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

Немного истории

- 1953 : Discovery of DNA structure by Watson and Crick
- 1973 : First sequence of 24 bp published
- 1977 : Sanger sequencing method published
- 1980 : Nobel Prize Wally Gilbert and Fred Sanger
- 1982 : Genbank started
- 1983 : Development of PCR
- 1987 : 1st automated sequencer : Applied Biosystems Prism 373
- 1996 : Capillary sequencer : ABI 310
- 1998 : Genome of *Caenorhabditis elegans* sequenced
- 2000 : Human genome sequenced
- 2005 : 1st 454 Life Sciences Next Generation Sequencing system : GS 20 System
- 2006 : 1st Solexa Next Generation Sequencer : Genome Analyzer
- 2007 : 1st Applied Biosystems Next Generation Sequencer : SOLiD
- 2009 : 1st Helicos **single molecule** sequencer : Helicos Genetic Analyser System
- 2011 : 1st Ion Torrent Next Generation Sequencer : PGM
- 2011 : 1st Pacific Biosciences **single molecule** sequencer : PacBio RS Systems
- 2012 : Oxford Nanopore Technologies demonstrates ultra long **single molecule** reads

Немного истории Проект «Геном человека»

- Длительность проекта – 10 (15) лет
- Стоимость – 3 миллиарда долларов
- Участвовали многие страны, для исследования генов использовался метод секвенирования по Сэнгеру

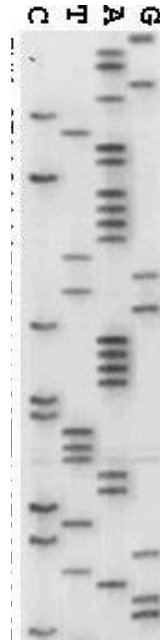


Alan Guttamacher, NHGRI NIH, August 16 2000

Страны, принимавшие участие в проекте

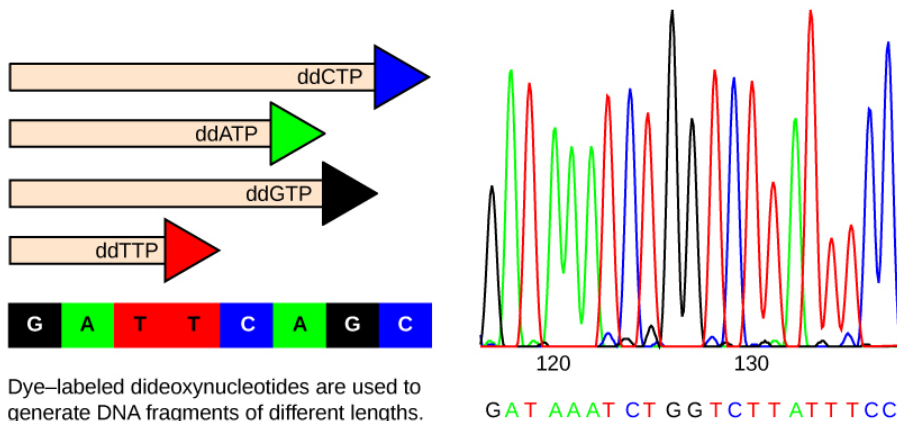


www.sciencemuseum.org



Фредерик Сэнгер (р. 1918) Дважды лауреат Нобелевской премии, за определение последовательности инсулина (1958) и за метод секвенирования ДНК (1980)

Принцип секвенирования по Сэнгеру



<https://www.boundless.com/biology/biotechnology-and-genomics>

Возможности секвенирования по Сэнгеру

- В качестве матрицы возможно использование только небольшого участка гена (желательно не более 1000 п.н.)
- Процесс секвенирования включает этапы постановки ПЦР, очистки ПЦР-продукта, постановки сиквенс-реакции, проведение секвенирования, оценки полученных результатов методом сравнения с имеющимися базами данных
- Итого процесс секвенирования полной (не только кодирующей) последовательности одного гена может занимать несколько недель

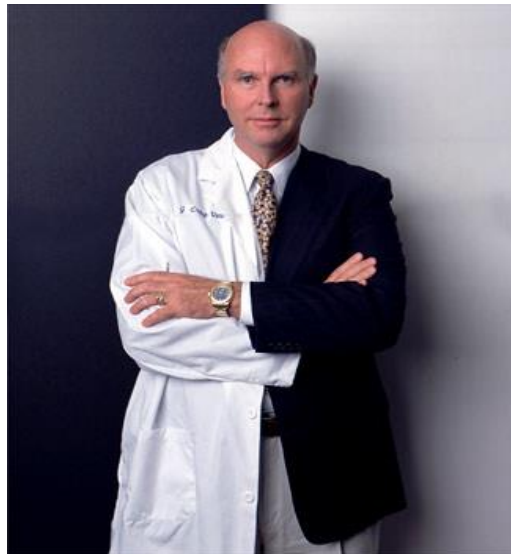
Преимущества и недостатки метода

- Позволяет точно идентифицировать изменения в изучаемом участке
- До сих пор является «золотым стандартом» генетических исследований
- Обладает невысокой чувствительностью (15-20%)
- Исследование полной последовательности гена очень дорого и долго

Что мы получили по завершении проекта?

Итог – «осознание того, что мы получили алфавит, сам по себе крайне необходимый для жизни, но мало что дающий в понимании процессов, происходящих в каждом отдельно взятом организме, а тем более – в опухоли»

Alan Guttmacher, NHGRI NIH, August 16 2000



Крейг Вентер (р.1946). Ученый и бизнесмен, талант и авантюрист.

Методы секвенирования нового поколения

- Основная идея секвенирования нового поколения: потенциальная возможность одновременного исследования последовательностей любого количества генетического материала
- Основное преимущество: значительно сокращает время и стоимость массивных исследований
- Основной недостаток: требует сложной биоинформатической обработки, в ряде случаев – сложнейшей биоинформатической обработки

Геном Джеймса Уотсона

doi:10.1038/nature06894 | Vol 452 | 7 April 2008

The complete genome of an individual by massively parallel DNA sequencing

David A. Wheeler^{1*}, Maitreyan Srinivasan^{2*}, Michael Egholm^{3*}, Yufeng Shen^{1*}, Lei Chen¹, Amy McGuire², Wen He², Yi-Ju Chen², Vinod Makhijani², G. Thomas Roth², Xavier Gomes², Karrie Tartaro², Fahsem Niaz², Cynthia L. Turcotte², Gerard P. Izzyk², James R. Lupski^{4,5,6}, Craig Chiu^{7,8}, Xing-zhi Song¹, Yue Liu¹, Ye Yuan¹, Lynne Nazareth¹, Xiang Qin¹, Donna M. Muzny¹, Marcel Margulies², George M. Weinstock^{1,4}, Richard A. Gibbs^{1,4} & Jonathan M. Rothberg^{1,2}

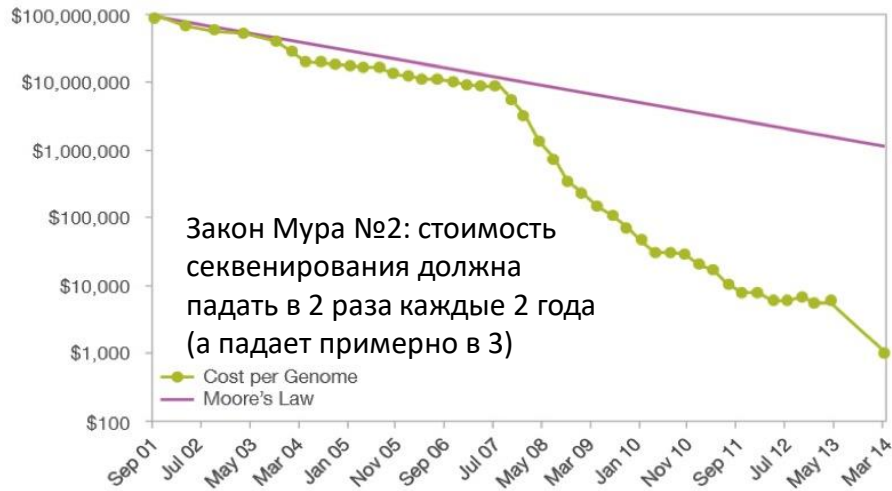
- 1st genome sequenced by a next generation sequencing method (*in 2 months!*)
- \$1,000,000 (1/100th the cost of traditional methods)
- 3,300,000 SNPs relative to HGP reference genome (82% previously known)
- Most SNPs presumed to be neutral, however, 11,000 (85% of those previously known) altered proteins
- ~23 CNVs ranging from 26 kb to 1.6 Mb (9 gains, 14 losses)
- 1 region of homozygous loss!
- Sequence typical for human genetic variation
- *Currently, extremely difficult to extract medically relevant inferences from individual genomes*
- *Couldn't even make a rough prediction of Watson's height from his genome sequence!!!*

From Olsen, M (2008) Dr. Watson's base pairs. Nature 452. 819



NUI Galway
OE Gaillimh

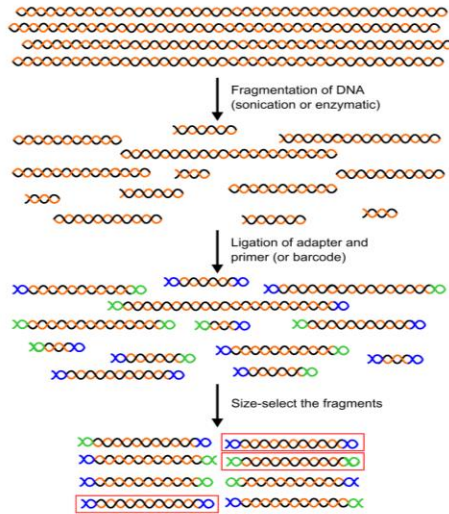
Драматическое снижение стоимости секвенирования генома



Принципы работы основных платформ

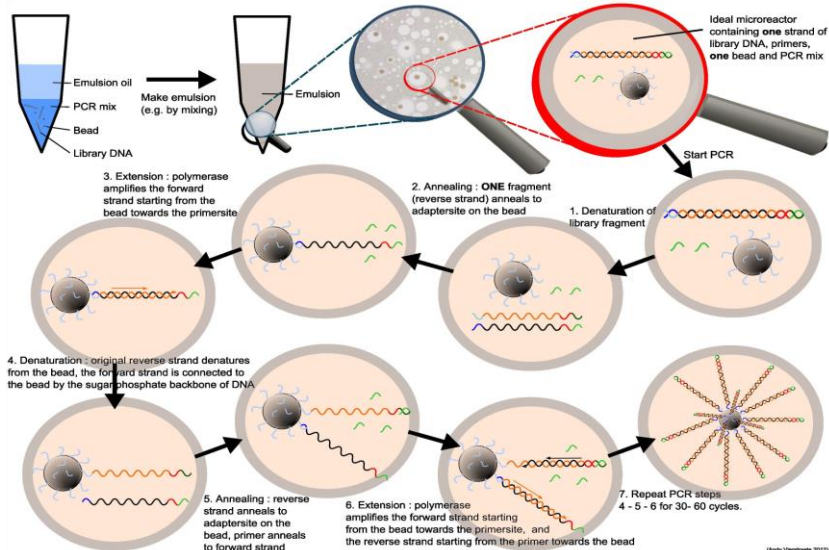
- **454 Sequencing / Roche**
 -
 -
 -
 - **Illumina**
 -
 -
 -
 - **Ion Torrent - Life Technologies**
- } Пиросеквенирование
- } Секвенирование методом синтеза
- } Полупроводниковое секвенирование

Первый этап – приготовление геномной библиотеки (на примере полногеномного секвенирования)

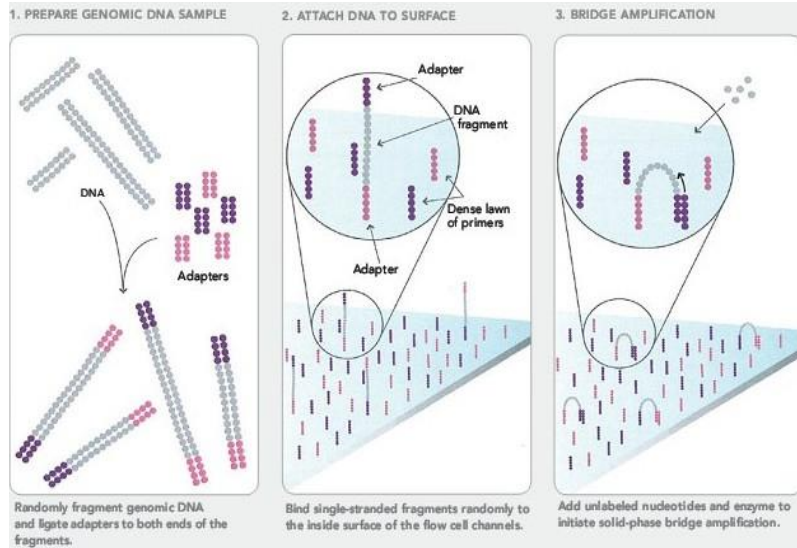


- Первый этап – фрагментирование ДНК (ультразвуковое или ферментативное)
- Второй этап – пришивание адаптеров, праймеров, и баркодов
- Третий этап – отбор полноразмерных фрагментов

Наработка материала- эмульсионный ПЦР (пиросеквенирование, полупроводниковое секвенирование)

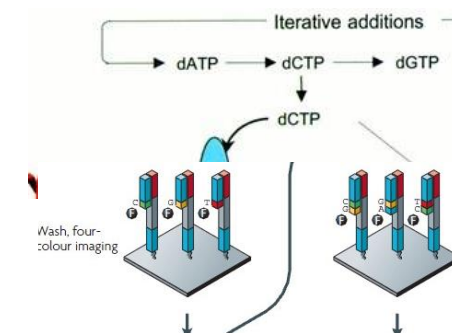


Наработка материала – на твердой поверхности (Bridge PCR, Illumina и проч)



Следующий этап – сам процесс секвенирования

- Пиросеквенирование (с использованием фермента люциферазы)
- Синтез (присоединение каждого нуклеотида в 1 реакцию)
- Полупроводниковое секвенирование (освобождение молекул водорода)



n Sequencing forms

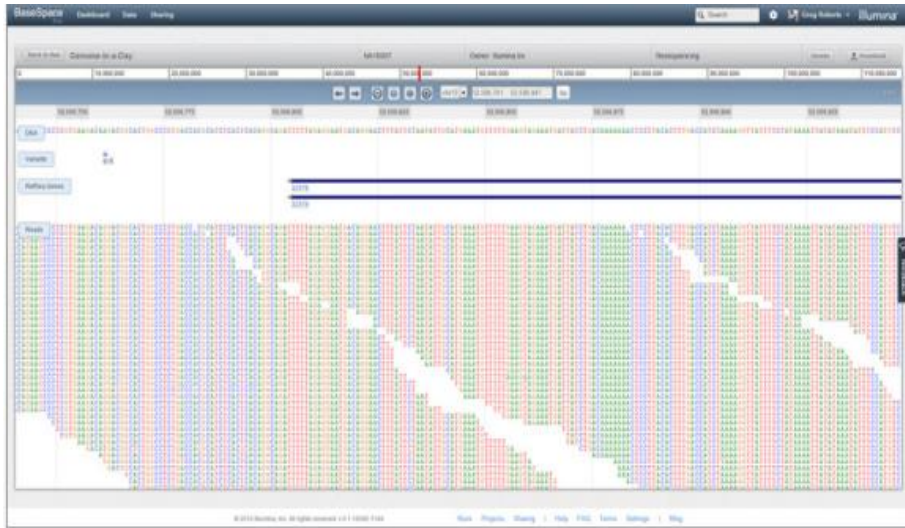
n Sequencing : Amplified Single Molecule Sequencing

Andy Vanraaij
Department of Biology
Ghent University, June 2012

Average Corrected Ionogram



Следующий этап – сборка данных



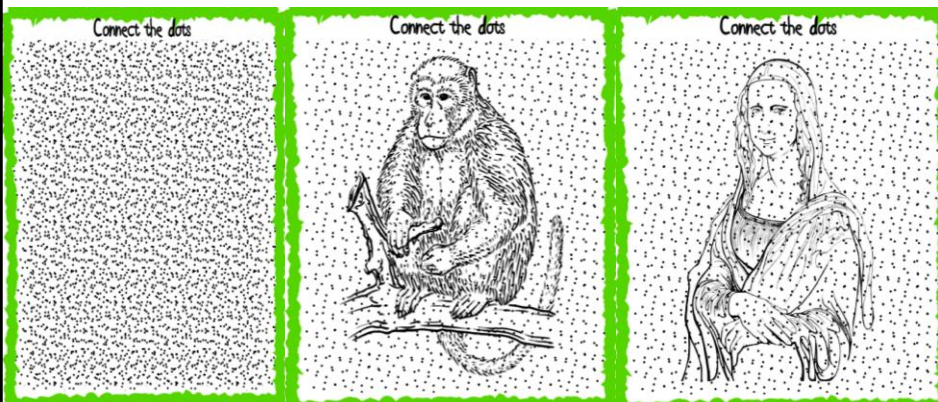
Биоинформатическая обработка

- Оценка качества прочтения (QS, количество прочтений и т.д.)
- Первичная обработка с отсечением «шума» и выравниванием (+ оценка variant frequency)
- Получение полноценных файлов в формате FASTQ
- Сравнение с нормальным образцом пациента, вычитание SNP
- Выделение aberrантных последовательностей, аннотация по базам данных
- Проверка впервые обнаруженных мутаций стандартными методами
- Оценка онкогенного потенциала обнаруженных de novo нарушений

Количество биоинформатических инструментов огромно...количество специалистов в России пока ничтожно



Важность правильной биоинформатической обработки



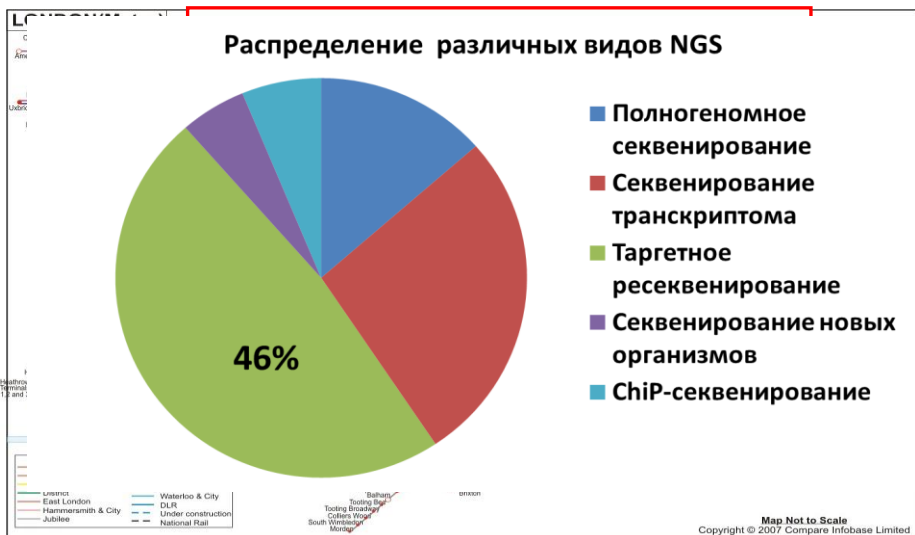
Где же истина?-только в международной единой системе биоинформатической обработки

Виды NGS

Точка приложения	Виды	Возможности и недостатки
Геном	Полногеномное секвенирование	<p>Возможности: исследования генетической предрасположенности и особенностей метаболизма пациента, максимальная характеристика опухоли – применение научно-исследовательское</p>
	Секвенирование экзома	
	Таргетное ресеквенирование	
Транскриптом	Секвенирование транскриптома	<p>Возможности – исследование генов-целей для препаратов. Возможно практическое применение</p>
Эпигеном	Бисульфитное секвенирование	<p>Возможности: исследование экспрессии генов и ее регуляции. Применение, как правило, научно-исследовательское</p>

Dienstmann et al, JCO 2013; 31: 1874-1884

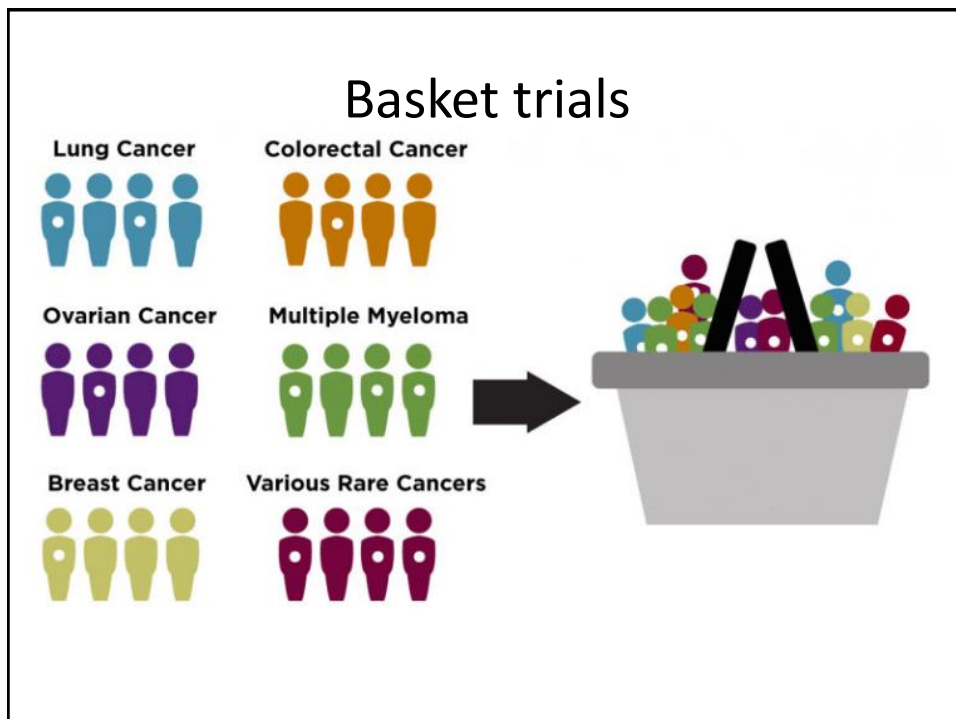
Насколько мы готовы использовать метод NGS в клинической практике?



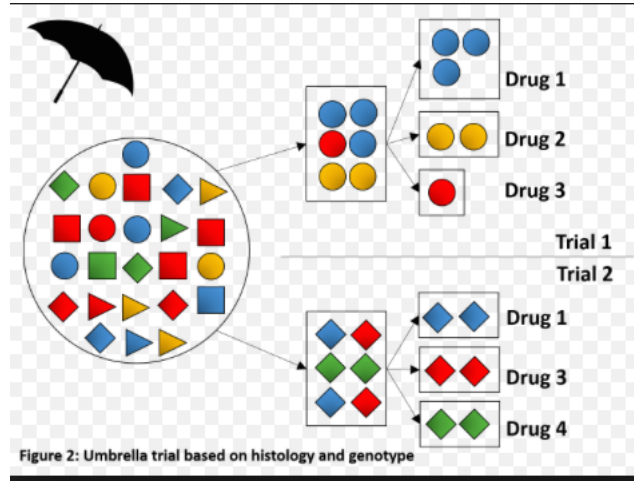
<http://www.pancreapedia.org/pathways/egfr>

Каковы основные применения NGS в онкологии в настоящий момент

- Оценка факторов риска онкозаболеваний (несколько довольно проблемных панелей, частичное экзомное секвенирование с оценкой SNP)
- Оценка риска развития наследственных онкозаболеваний (включают ряд ассоциированных с наследственными раками генов, основные панели несколько разнятся, обязательно включение BRCA1 и BRCA2, TP53)
- Секвенирование генов, являющихся целями для таргетной терапии (таргетное ресеквенирование)



Umbrella trials



За и против

Table 2. Preliminary Best Response According to Cohort.*

Variable	NSCLC (N=20)	Colorectal Cancer		Cholangio- carcinoma (N=8)	ECD or LCH (N=18)	Anaplastic Thyroid Cancer (N=7)
		Vemurafenib (N=10)	Vemurafenib + Cetuximab (N= 27)			
Patients with ≥1 postbaseline assessment — no.	19	10	26	8	14	7
Complete response — no. (%)	0	0	0	0	1 (7)	1 (14)
Partial response — no. (%)	8 (42)	0	1 (4)	1 (12)	5 (36)	1 (14)
Stable disease — no. (%)	8 (42)	5 (50)	18 (69)	4 (50)	8 (57)	0
Progressive disease — no. (%)	2 (11)	5 (50)	7 (27)	3 (38)	0	4 (57)
Missing data — no. (%)†	1 (5)	0	0	0	0	1 (14)
Overall response — no. (%) [95% CI]	8 (42) [20–67]	0	1 (4) [<1–20]	1 (12) [<1–53]	6 (43) [18–71]	2 (29) [4–71]

*The denominator for patients with a complete or partial response, stable disease, or progressive disease is the number of patients with a postbaseline assessment or early withdrawal. Of the 19 patients in the NSCLC cohort, 1 patient withdrew before the assessment of response but was included in the denominator for the efficacy assessment (as having had no response).
†All patients with missing data withdrew early.

Hyman et al N Engl J Med 2015; 373:726-736

За и против



Прежде чем начать работу, подумай...

- Решит ли NGS поставленные задачи?
- Какой вид NGS необходимо применить? (геном, транскриптом, экзом, микроРНК и тд)
- Какую платформу оптимально использовать для данной задачи?
- Продуман ли дизайн эксперимента?
- Где найти биоинформатиков?
- Как определить, насколько корректны результаты?
- Как быть со статистикой?
- Garbage in – garbage out, или что посеешь, то и пожнешь (аккуратный и продуманный подход к материалу)
- Если ты не знаешь – спроси того, кто знает (зачастую, к сожалению, не бесплатно)

Спасибо за внимание!

CTAGGCTAGCTAGTCG
GCTLIFECISGATAG
C4-LETTERWORDT
GCTATATCGTAGCTG