



What's New in Sequencing: A Sequencing System adapted to your Needs

Marco Cappelletti
Product Marketing Manager - Europe

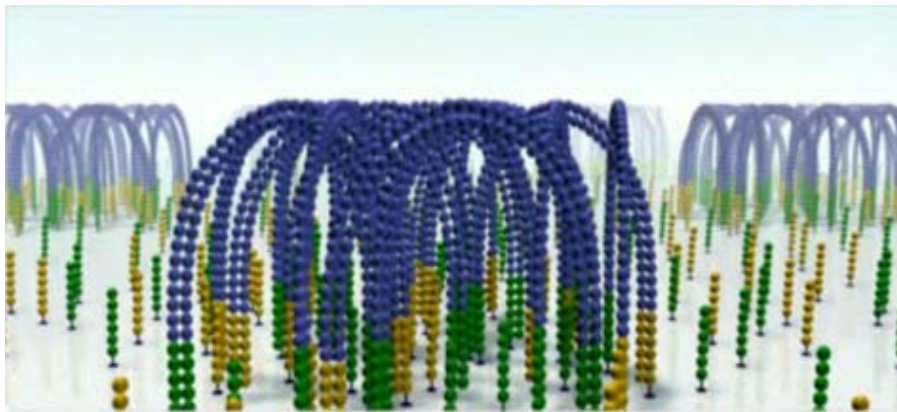
© 2009 Illumina, Inc. All rights reserved.

Illumina, illuminaDx, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, GoldenGate Indexing, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, CSPRO, and GenomeStudio are registered trademarks or trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.

illumina[®]

Agenda

- ▶ Illumina Sequencing Technology
- ▶ Sequencing Workflow
- ▶ Applications



illumina®

HiSeq 2000

Redefining the trajectory of sequencing

HIGHEST OUTPUT

Initially capable of up to 200 Gb per run

FASTEST DATA RATE

~25 Gb/day

7-8 days for 2 x 100 bp

HIGHEST NUMBER OF READS

One billion single-end reads*

Two billion paired-end reads*



*Based on one billion clusters passing filter

illumina®

An Illumina Sequencer for Every need

Next Generation Sequencing made accessible



Genome Analyzer_{IIe}

- 40 Gb/Run
- 260M Reads
- Read Length:150bp

Most widely adopted NGS platform



Genome Analyzer_{IIx}

- 50 Gb/Run
- 500M Reads
- Read Length:100bp

Unique combination of sequencing & arrays



HiScanSQ

- 50 Gb/Run
- 500M Reads
- Read Length:100bp

Redefining the trajectory of sequencing



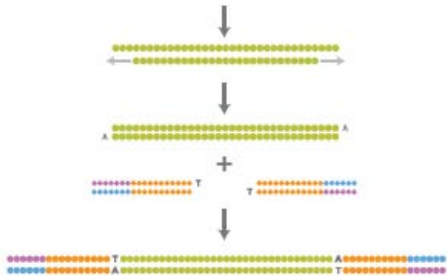
HiSeq 2000

- 200 Gb/Run
- 2B Reads
- Read Length:100bp

illumina[®]

Simplest Sequencing Workflow

SIMPLIFIED SAMPLE PREP



Parallel sample processing

cBot CLUSTER GENERATION



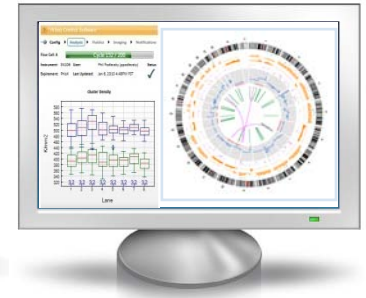
Automated cluster generation

HiSeq 2000 SEQUENCING



Automated sequencing

DATA PROCESSING & ANALYSIS



Simple, efficient data analysis

illumina®

New Simplified Sample Prep Kits*

Multiple DNA and mRNA samples for sequencing application

HIGHLIGHTS

Streamlined Workflow: Pre-mixed reagents reduce hands-on time when preparing multiple DNA or mRNA samples

Faster Throughput: Simultaneous preparation of 96 DNA or mRNA samples increases the number of samples processed/day

Automation-Friendly Process: Master-mixed reagents are compatible with liquid-handlers

Cost-Effective Solution: Significant reduction in sample preparation costs



illumina®

*Expected availability Q2 2010

Simplified Sample Preparation

Multiple DNA and mRNA samples for sequencing applications

STREAMLINED WORKFLOW

Fast method for preparing gDNA or total RNA samples

Designed to process 12, 96, or more samples

Optimal for High-throughput studies

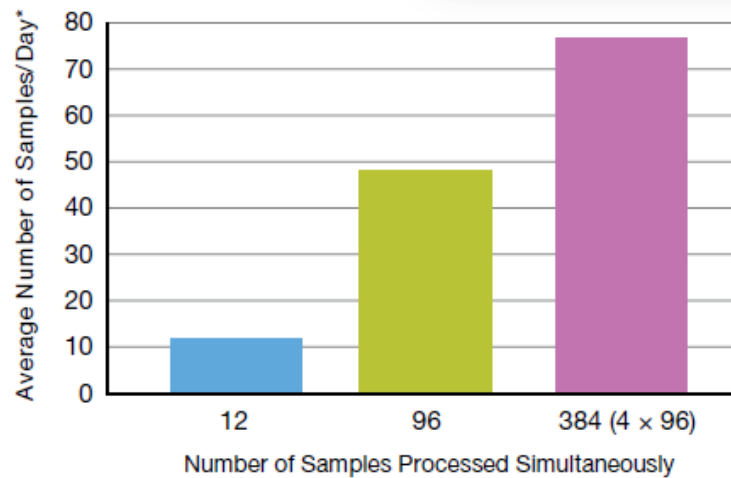
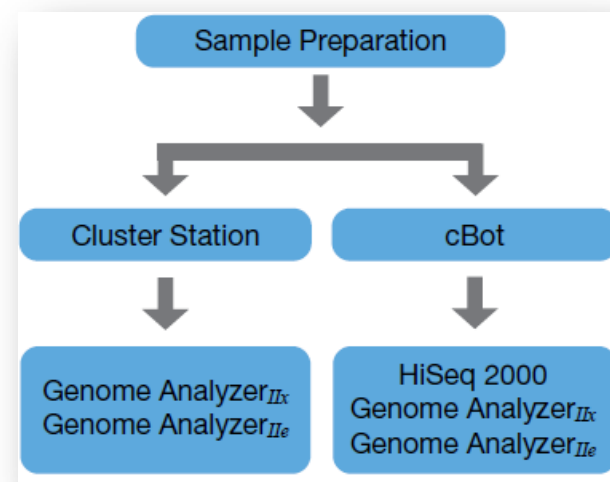
Universal Design, no need for multiple application kits

New Simplified Sample Multiplexing

Built-in Quality Control

Designed Around Automation

Part of the Illumina Solution



illumina®

cBot System – Cluster Generation Workflow

- ▶ Breakthrough workflow automation system
 - Clonal amplification of DNA for Illumina sequencing



illumina®

The flow cell design

5'-PS-TTTTTTTTAAATGATACGGGACCACCGAGAUCTACAC-3'

5'-PS-TTTTTTTTCAAGCAGAGAGCGGCATACGAG0x0AT-3'

Surface of flow cell coated with a lawn of oligo pairs



8 channels

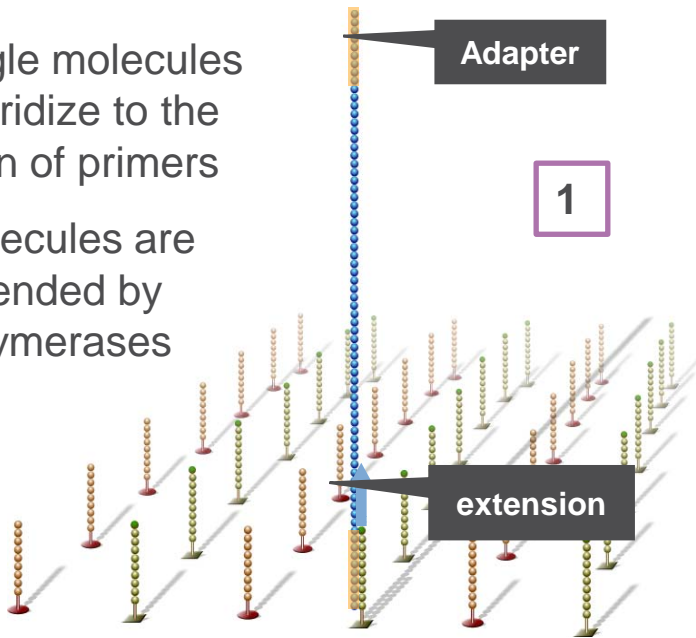
- ▶ Contained environment
- ▶ No need for clean rooms
- ▶ Sequencing performed inside the flow cell

illumina®

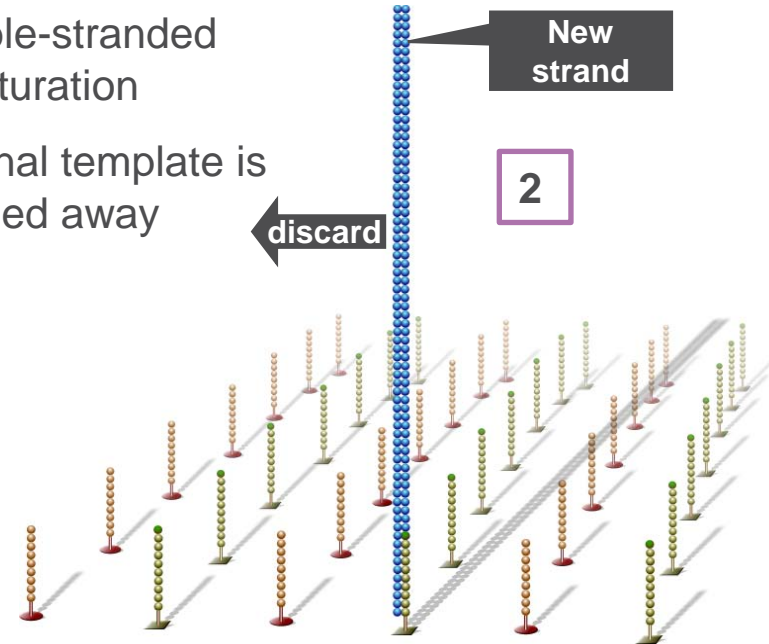
Cluster Generation

Hybridize Fragment & Extend

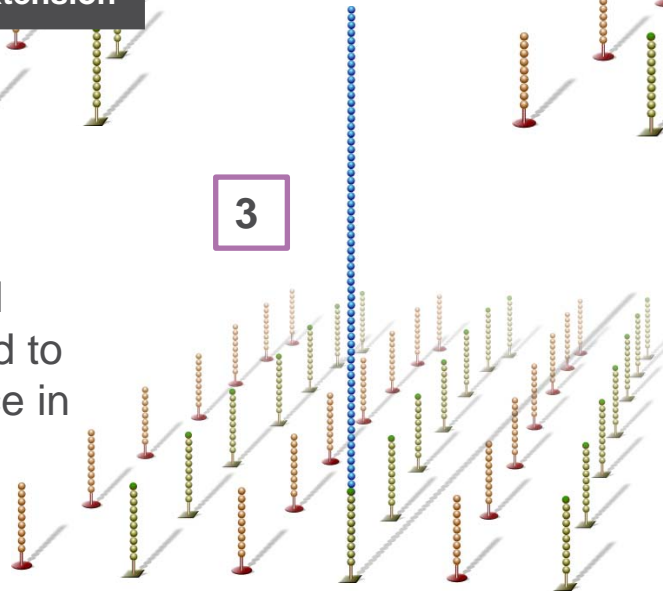
- ▶ Single molecules hybridize to the lawn of primers
- ▶ Molecules are extended by polymerases



- ▶ Double-stranded denaturation
- ▶ Original template is washed away

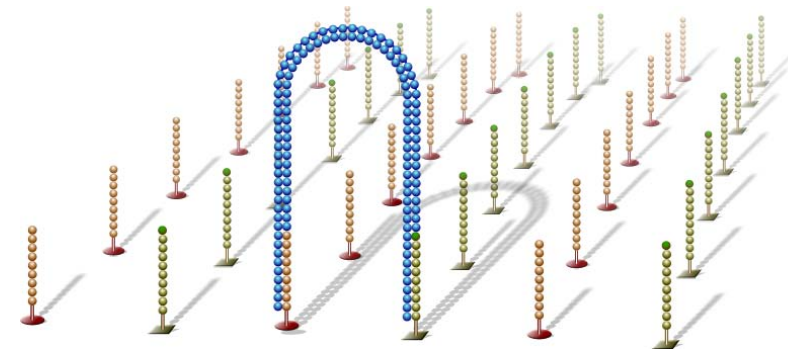
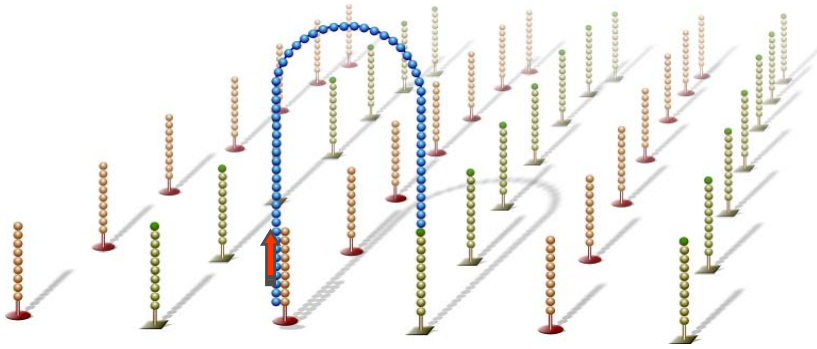


- ▶ Newly synthesized covalently attached to the flow cell surface in a random pattern

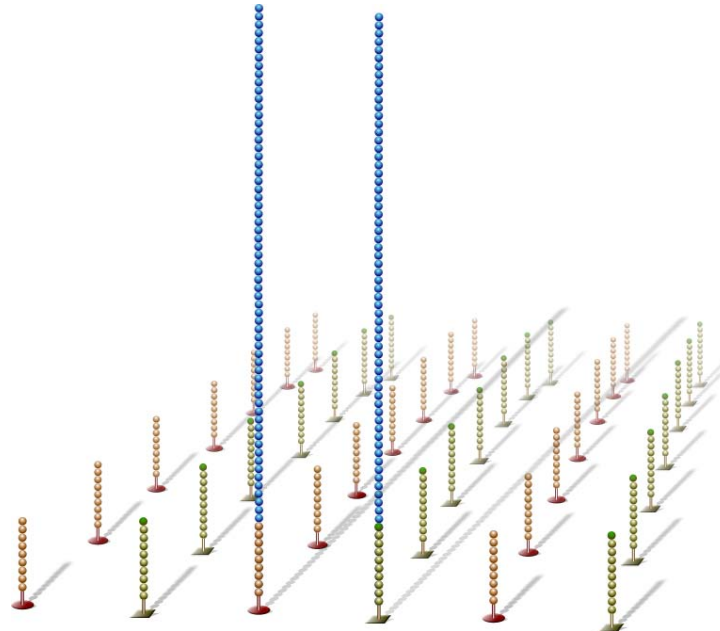


Cluster Generation

Bridge Amplification



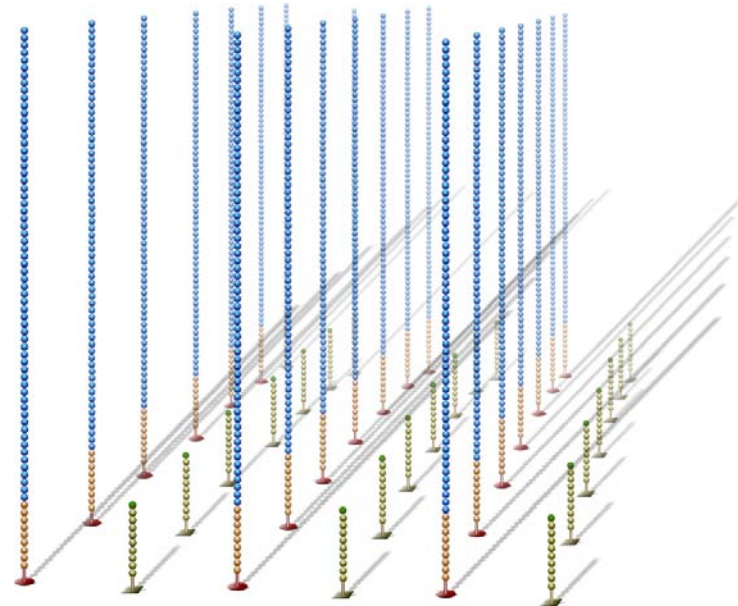
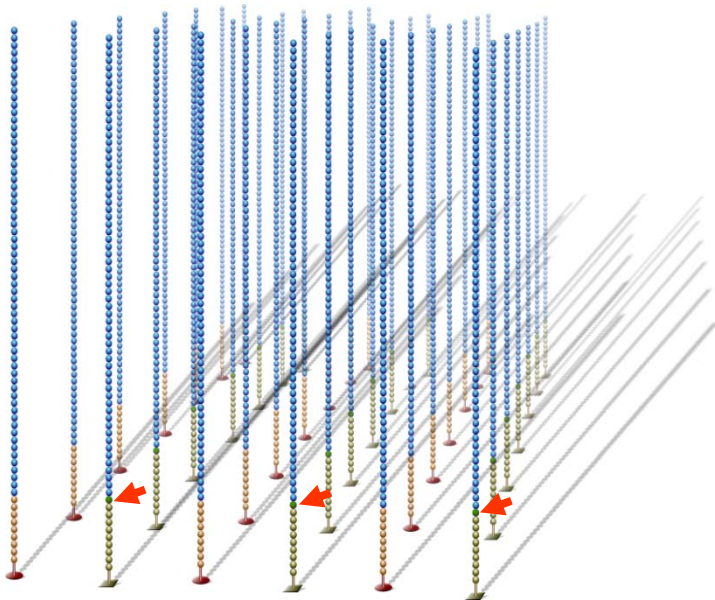
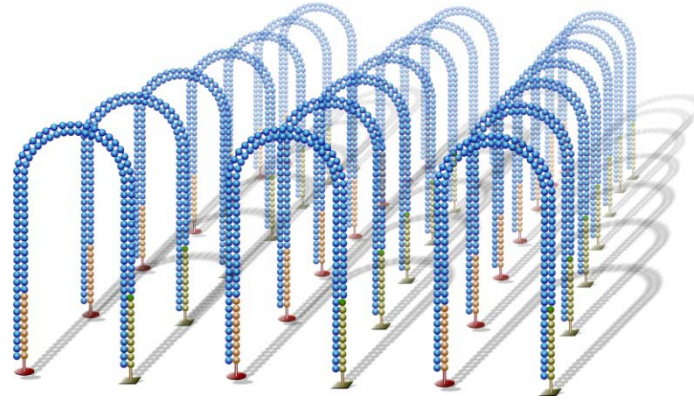
- ▶ Single-strand flips over to hybridize to adjacent primers to form a bridge
- ▶ Hybridized primer is extended by polymerases
- ▶ Bridge is denatured



Cluster Generation

Bridge Amplification

- ▶ Bridge amplification cycle repeated until multiple bridges are formed
- ▶ Bridges denaturation
- ▶ Reverse strands cleaved and washed away



cBot Performance Specifications

- ▶ Compatible with GA & iScan Seq Ext
- ▶ User installable
- ▶ Total hands-on time: <10 min
- ▶ Total run time: ~4 hrs
- ▶ Integrated barcode scanner
- ▶ Integrated touch-screen monitor



illumina®

Sequencing by Synthesis Chemistry

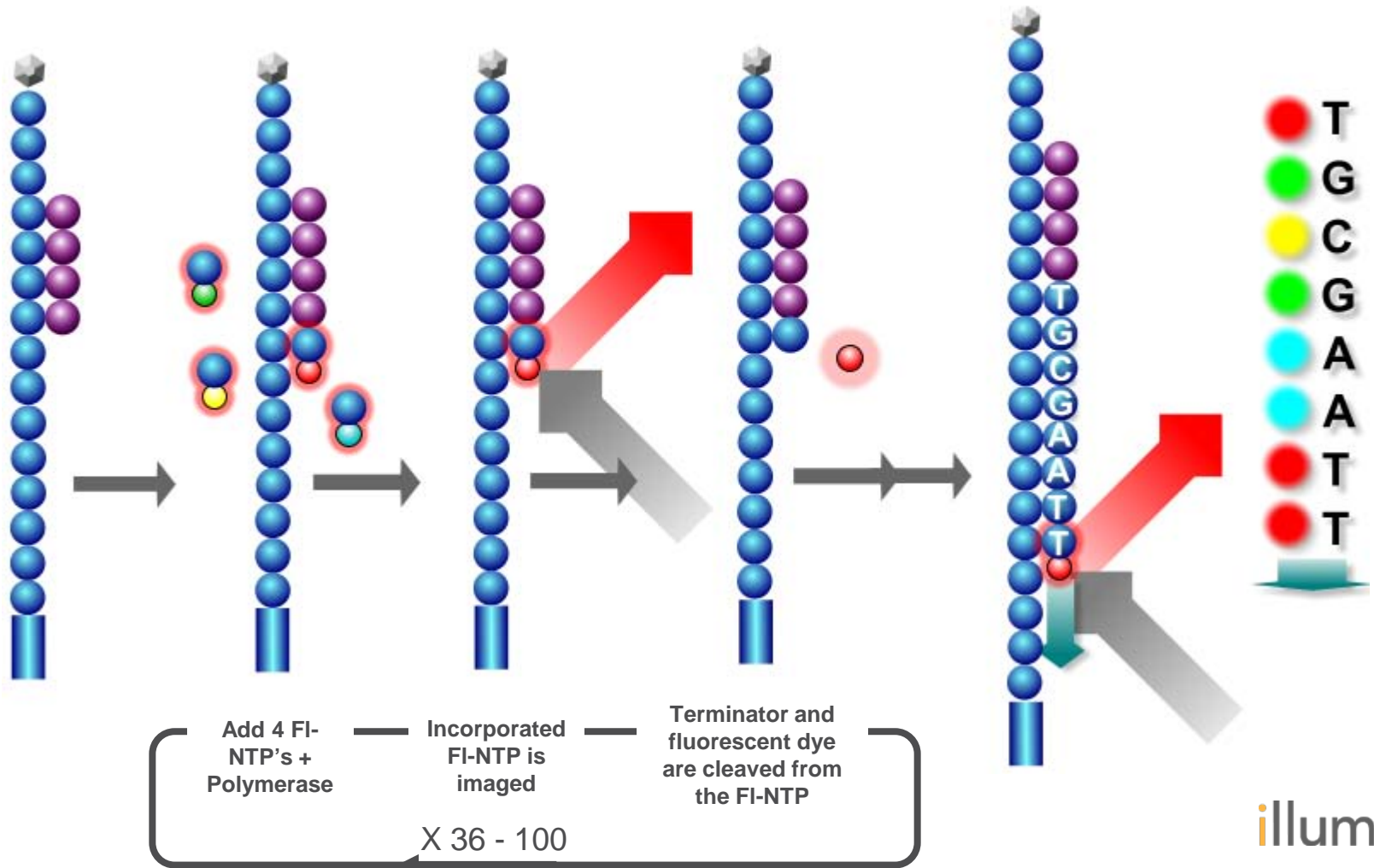
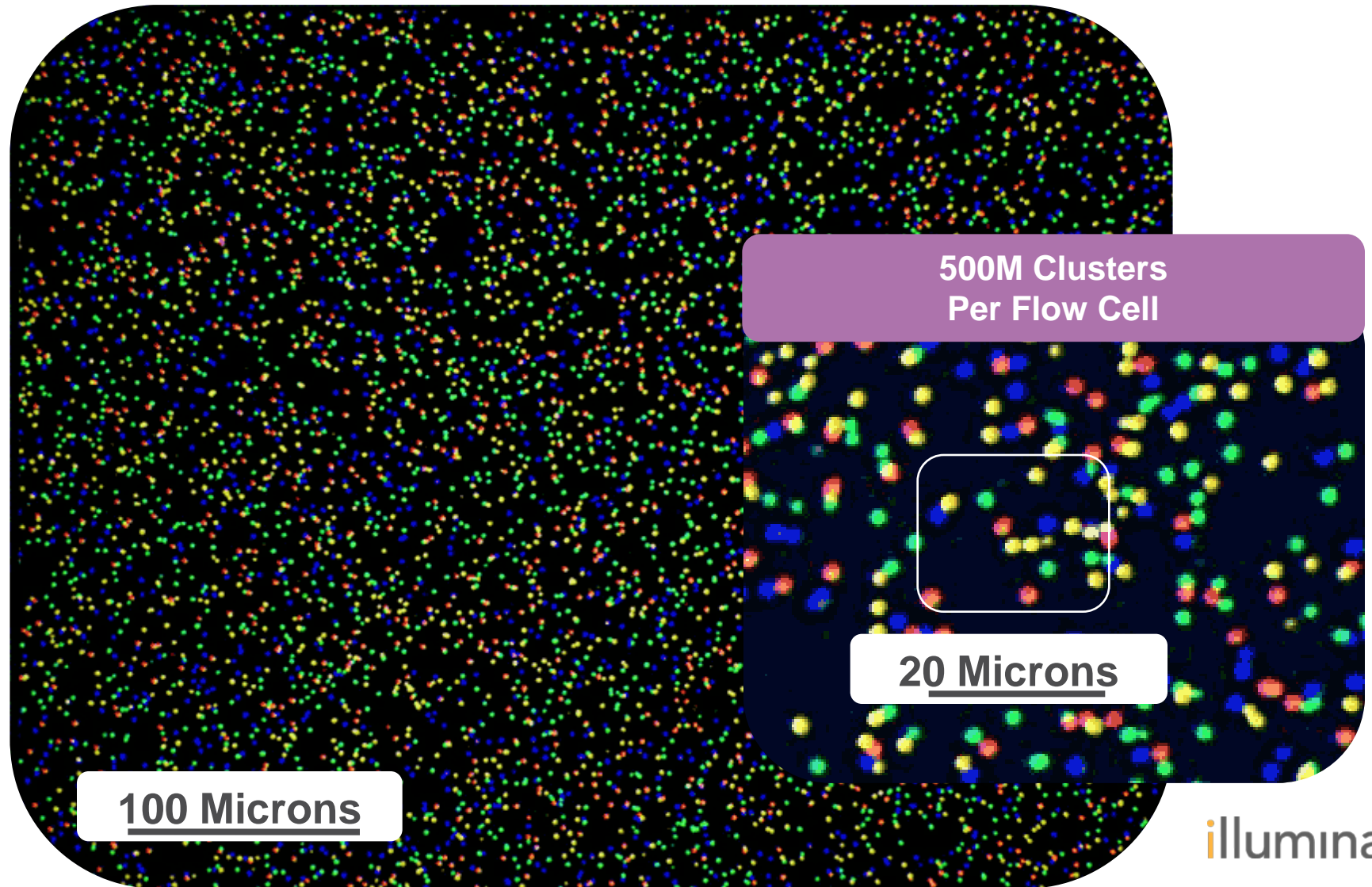


Image acquisition



Sequencing Data Analysis

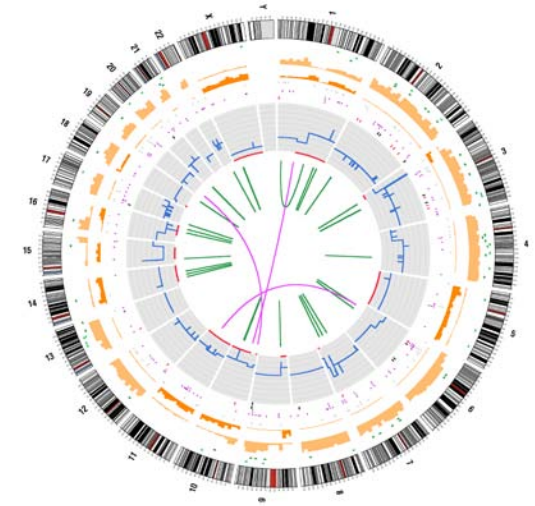
A simple, familiar workflow



**SEQUENCING
CONTROL
SOFTWARE**
Read Generation



CASAVA
Alignments,
variations, builds



VISUALIZATION
GenomeStudio, or
favorite browser

illumina®

Broadest range of applications

Optimized, streamlined and easy-to-use reagent solutions

Sample Prep

Genome

- WG Reseq.
- De-novo
- Targeted

Transcriptome

- RNA-Seq
- Small RNA
- miRNA
- Directional Seq

Regulation

- Methylation
- ChIP-Seq

Automated Cluster Generation



Sequencing



illumina®

Published Clinical Applications

- ▶ Copy Number Variation Analysis
- ▶ Pre-natal Diagnosis
- ▶ Metagenomics
- ▶ Expression Analysis from FFPET
- ▶ Epigenomics

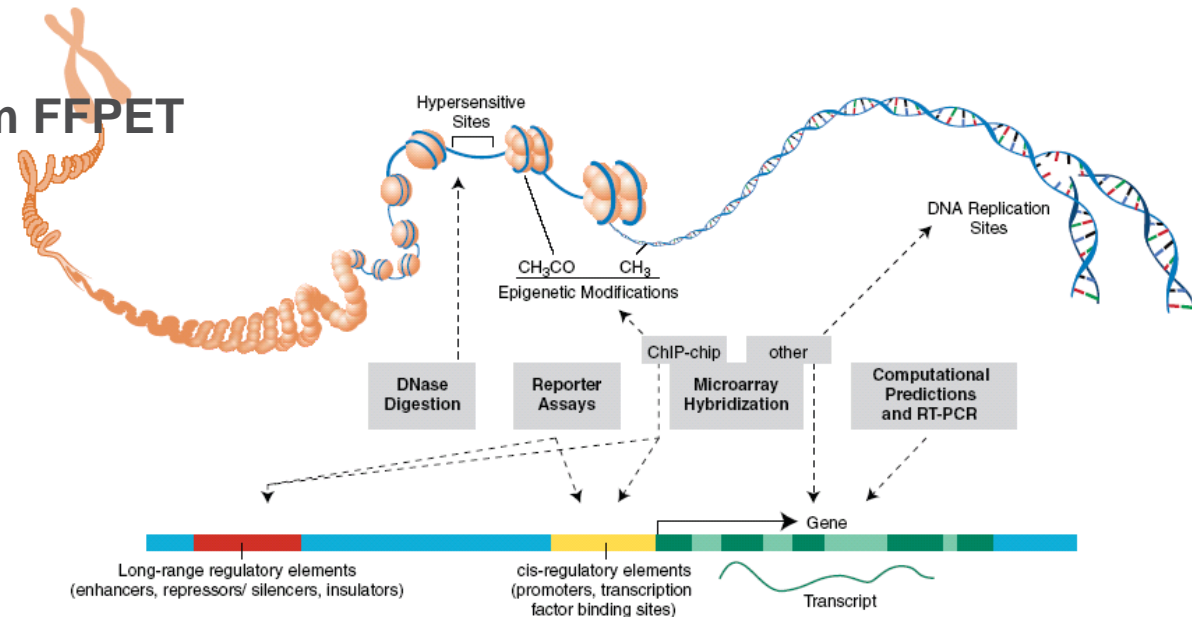


Fig. 1. Functional genomic elements being identified by the ENCODE pilot phase. The indicated methods are being used to identify different types of functional elements in the human genome.

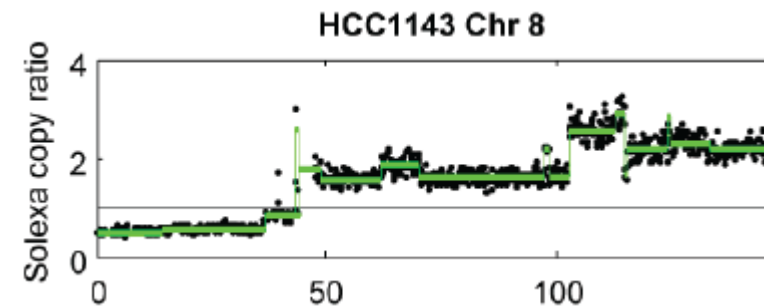
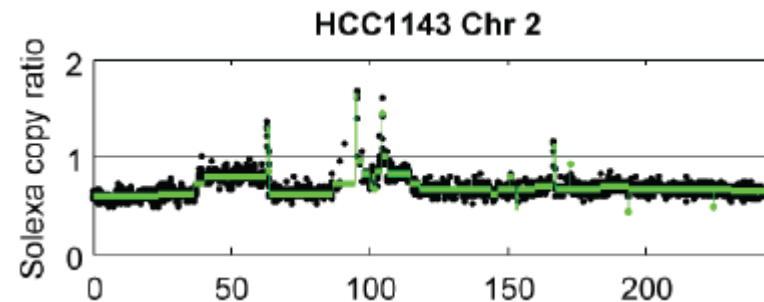
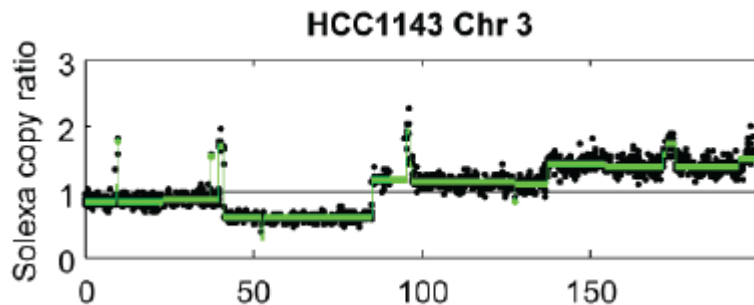
www.sciencemag.org SCIENCE VOL 306 22 OCTOBER 2004

illumina®

Identify copy-number alterations

Cancer High-resolution mapping of copy-number alterations with massively parallel sequencing Nat Methods. 2009 Jan;6(1):99-103

- ▶ Cancer results from somatic alterations in key genes, including point mutations, copy-number alterations and structural rearrangements.
- ▶ The collection of million aligned sequence reads has comparable power to detect events as the current generation of DNA microarrays and has over twofold better precision for localizing breakpoints



Prenatal Diagnostics by Deep Sequencing

(PNAS December 23, 2008 vol. 105 no. 51)

- ▶ “Fetal DNA has been found in maternal plasma but exists as a minor fraction among a high background of maternal DNA”...quantitative perturbations caused by an aneuploid chromosome in the fetal genome would be small”
- ▶ “The use of a locus-independent method would greatly increase the number of target molecules from the aneuploid chromosome that could be analyzed within the same fixed volume of plasma
- ▶ Sequencing is the clear way to do non-invasive prenatal testing. ... existing noninvasive Down syndrome tests are not very informative and provide variable results depending on the ethnicity of those taking the test.”

Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood

H. Christina Fan^a, Yair J. Blumenfeld^a, Usha Chitkara^a, Louanne Hudgins^a, and Stephen R. Quake^{a,b}

^aDepartment of Bioengineering, Stanford University and Howard Hughes Medical Institute, 318 Campus Drive, Clark Center, Room E300, Stanford, CA 94305; ^bDivision of Maternal-Fetal Medicine, Department of Obstetrics and Gynecology, Stanford University, 300 Pasteur Drive, Room HH333, Stanford, CA 94305; and ^cDivision of Medical Genetics, Department of Pediatrics, Stanford University, 300 Pasteur Drive, Stanford, CA 94305

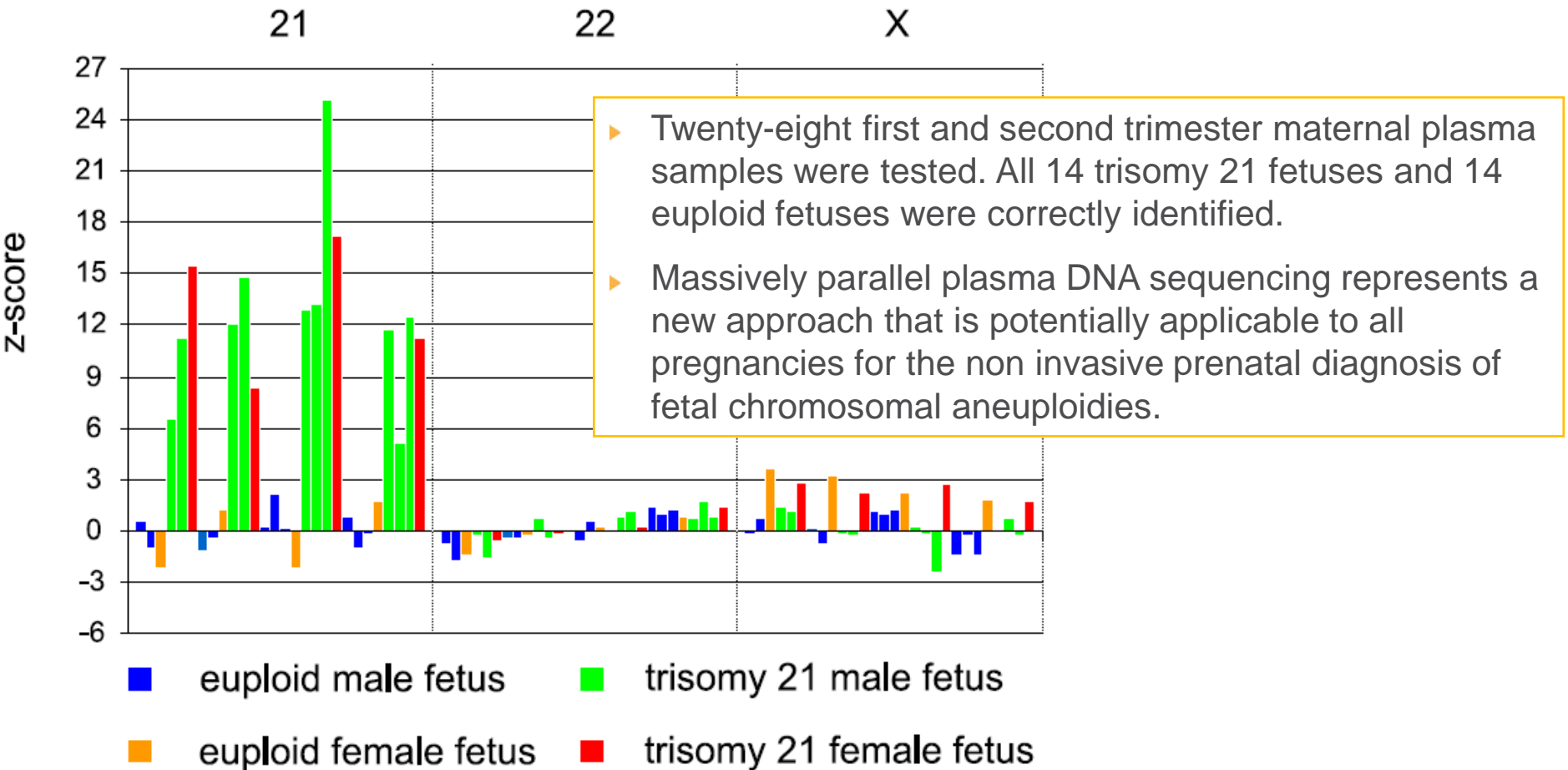
Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma

Rossa W. K. Chiu^{a,b}, K. C. Allen Chan^{a,b}, Yuan Gao^{c,d}, Virginia Y. M. Lau^{a,b}, Wenli Zheng^{a,b}, Tak Y. Leung^a, Chris H. F. Foo^e, Bin Xie^e, Nancy B. Y. Tsui^{a,b}, Fiona M. F. Lun^{a,b}, Benny C. Y. Zee^f, Tze K. Lau^g, Charles R. Cantor^{h,i}, and Y. M. Dennis Lo^{a,b,h,1}

^aCentre for Research into Circulating Fetal Nucleic Acids, Li Ka Shing Institute of Health Sciences, Departments of ^bChemical Pathology and ^cObstetrics and Gynaecology, and ^dCentre for Clinical Trials, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China; ^eCenter for the Study of Biological Complexity and ^fDepartment of Computer Science, Virginia Commonwealth University, Richmond, VA 23284; and ^gSequenom, Inc., San Diego, CA 92121

Prenatal Diagnostics by Deep Sequencing

(PNAS December 23, 2008 vol. 105 no. 51)

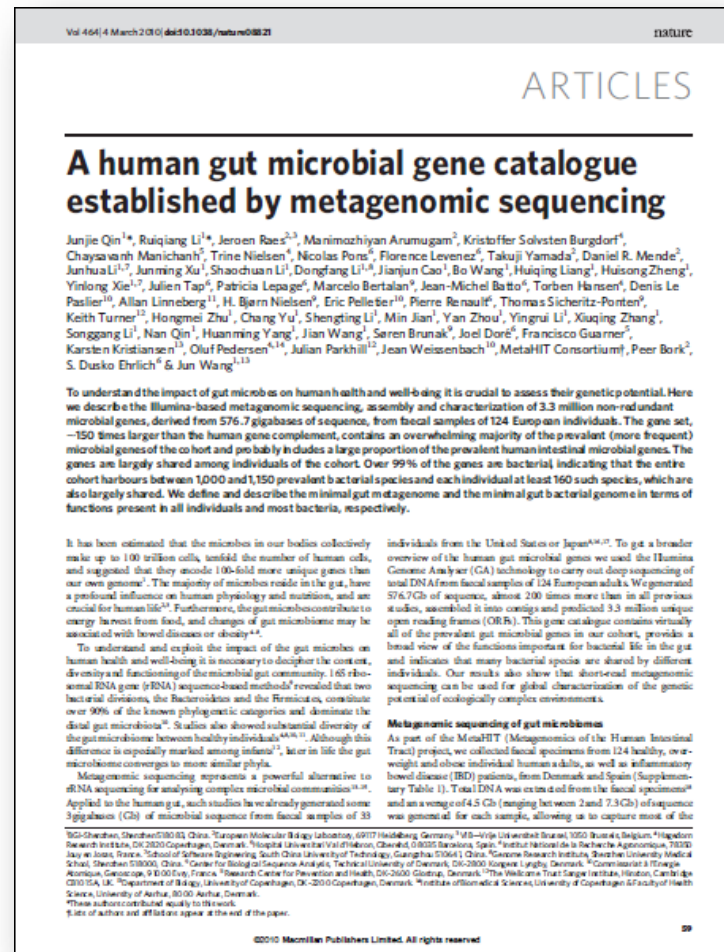


illumina®

A human gut microbial gene catalogue established by metagenomic sequencing

Nature (2010), Vol 464|4 March 2010| doi:10.1038/nature08821

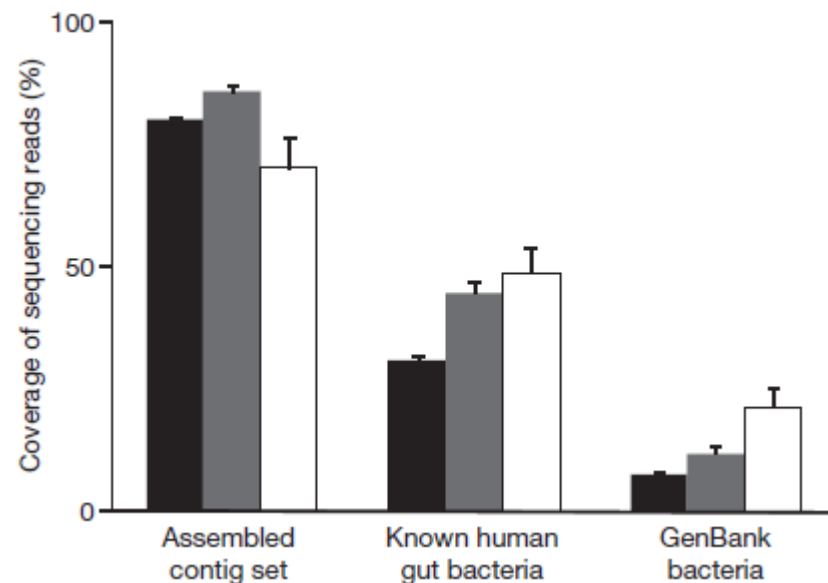
- ▶ To understand the impact of gut microbes on human health and well-being it is crucial to decipher the content, diversity and functioning of the microbial gut community
- ▶ Metagenomic sequencing, assembly and characterization of 3.3 million non-redundant microbial genes from human faecal samples (576.7 Gb of sequence, 124 European individuals)
- ▶ The gene set contains an overwhelming majority of the prevalent microbial genes..



A human gut microbial gene catalogue established by metagenomic sequencing

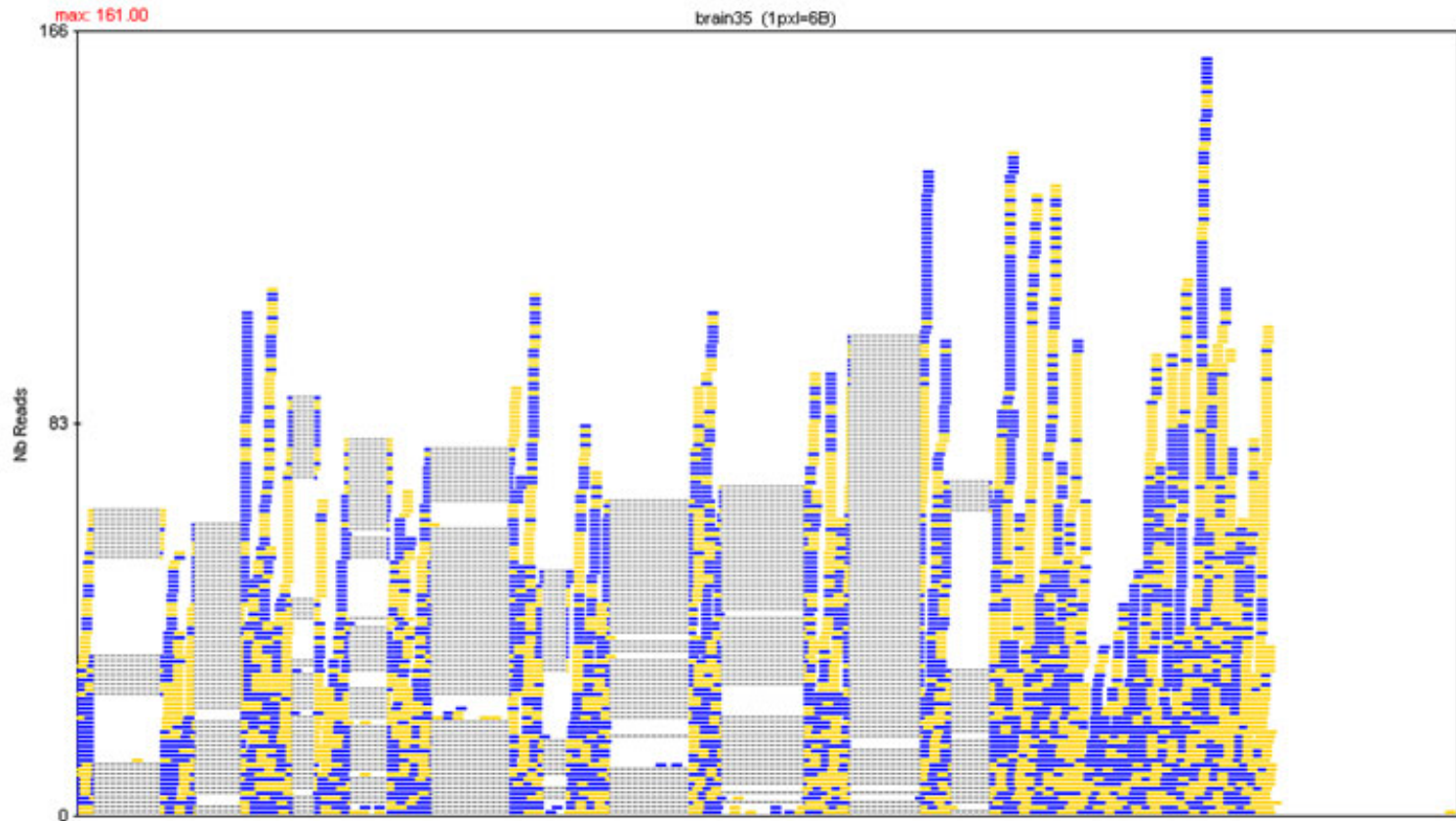
Nature (2010), Vol 464|4 March 2010| doi:10.1038/nature08821

- ▶ Over 99% of the genes are bacterial
- ▶ The entire cohort harbours between 1,000 and 1,150 prevalent bacterial species
- ▶ Each individual at least 160 such species, which are also largely shared

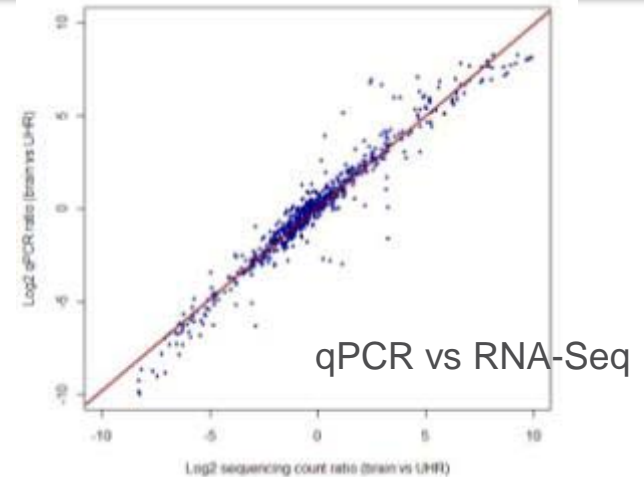
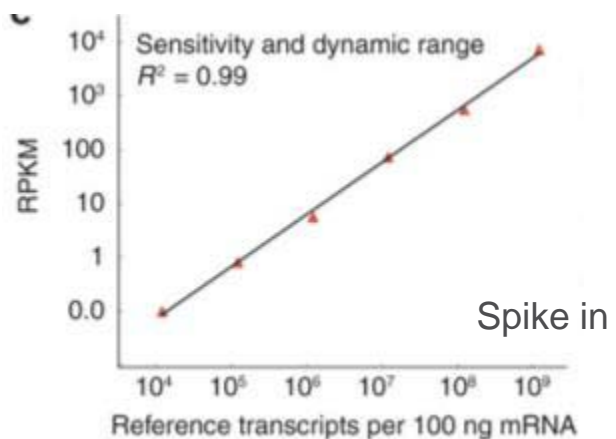
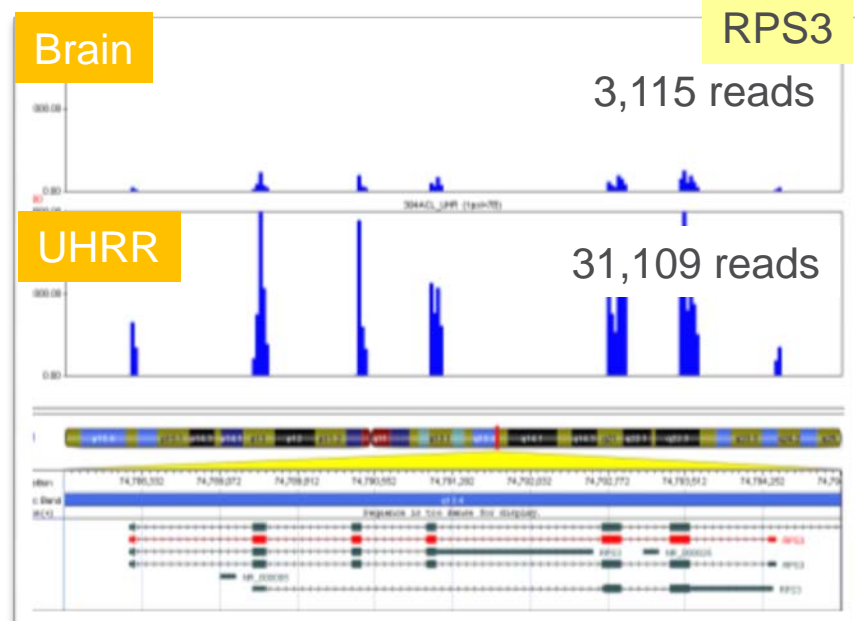
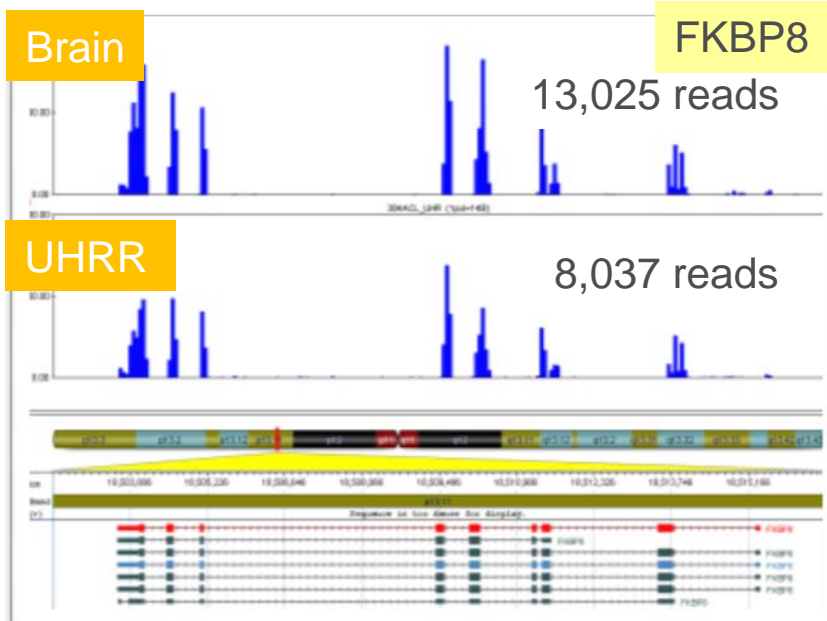


- ▶ **85.7% and 69.5% of our contigs were not covered by the reads from the two previous studies highlighting the novelty we captured**

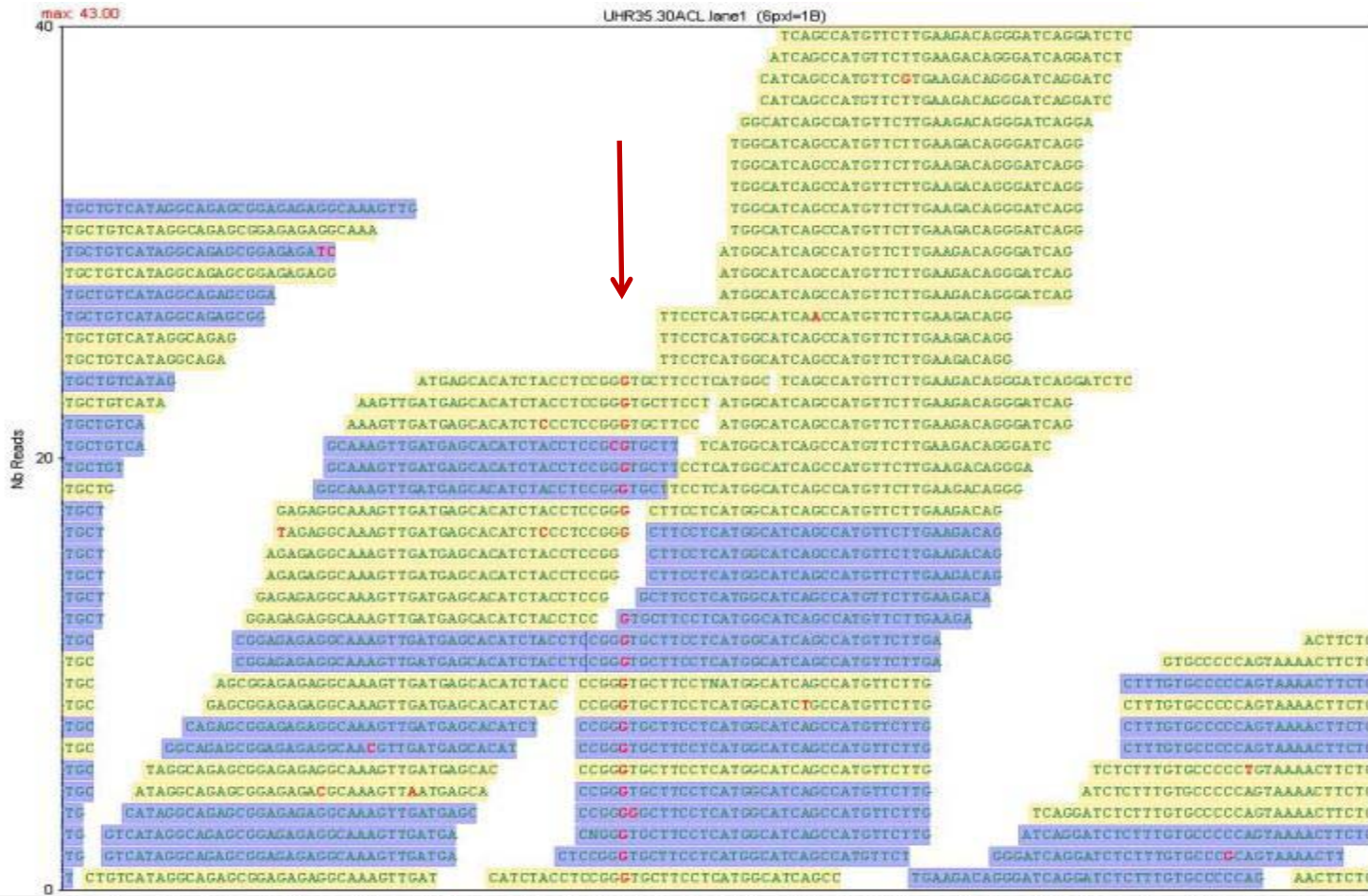
Transcriptomics



Quantitative Measure of Gene Expression



Visualization of Polymorphisms in the Transcriptome



SNP Call made in GenomeStudio

3'-End Sequencing for Expression Quantification (3SEQ) from Archival Tumor Samples

January 2010 | Volume 5 | Issue 1 | e8768

- ▶ A novel procedure (3SEQ) for gene expression profiling from FFPET using NGS.
- ▶ GEX profiling by 3SEQ and microarray on both frozen tissue and FFPET from two soft tissue tumors (desmoid type fibromatosis (DTF) and solitary fibrous tumor (SFT))
- ▶ Analysis of 3SEQ data revealed many genes differentially expressed between the tumor types on both the frozen tissue (~9.6K genes) and FFPET (~8.1K genes)
- ▶ Found Biological pathways known to be important in DTF and SFT pathogenesis

OPEN ACCESS Freely available online

PLOS ONE

3'-End Sequencing for Expression Quantification (3SEQ) from Archival Tumor Samples

Andrew H. Beck^{1,2}, Ziming Weng^{1,2,3}, Daniela M. Witten⁴, Shirley Zhu¹, Joseph W. Foley^{1,2}, Phil Lacroute^{1,2}, Cheryl L. Smith^{1,2}, Robert Tibshirani^{2,4}, Matt van de Rijn^{1,2}, Arend Sidow^{1,2,5}, Robert B. West^{1,2,6*}

1 Department of Pathology, Stanford University Medical Center, Stanford, California, United States of America, **2** Department of Genetics, Stanford University Medical Center, Stanford, California, United States of America, **3** Department of Statistics, Stanford University, Stanford, California, United States of America, **4** Department of Health Research and Policy, Stanford University Medical Center, Stanford, California, United States of America, **5** Pathology and Laboratory Service, Palo Alto Veterans Affairs Health Care System, Palo Alto, California, United States of America

Abstract

Gene expression microarrays are the most widely used technique for genome-wide expression profiling. However, microarrays do not perform well on formalin fixed paraffin embedded tissue (FFPET). Consequently, microarrays cannot be effectively utilized to perform gene expression profiling on the vast majority of archival tumor samples. To address this limitation of gene expression microarrays, we designed a novel procedure (3'-end sequencing for expression quantification (3SEQ)) for gene expression profiling from FFPET using next-generation sequencing. We performed gene expression profiling by 3SEQ and microarray on both frozen tissue and FFPET from two soft tissue tumors (desmoid type fibromatosis (DTF) and solitary fibrous tumor (SFT)) (total n = 23 samples, which were each profiled by at least one of the four platform-tissue preparation combinations). Analysis of 3SEQ data revealed many genes differentially expressed between the tumor types (DTF < SFT) on both the frozen tissue (~9.6K genes) and FFPET (~8.1K genes). Analysis of microarray data from frozen tissue revealed fewer differentially expressed genes (~4,64K), and analysis of microarray data on FFPET revealed very few (65) differentially expressed genes. Functional gene set analysis of 3SEQ data from both frozen tissue and FFPET identified biological pathways known to be important in DTF and SFT pathogenesis and suggested several additional candidate oncogenic pathways in these tumors. These findings demonstrate that 3SEQ is an effective technique for gene expression profiling from archival tumor samples and may facilitate significant advances in translational cancer research.

Citation: Beck AH, Weng Z, Witten DM, Zhu S, Foley JW, et al. (2010) 3'-End Sequencing for Expression Quantification (3SEQ) from Archival Tumor Samples. PLOS ONE 5(1): e8768. doi:10.1371/journal.pone.0087688

Editor: Irene Q. Lin Ng, The University of Hong Kong, HONG KONG

Received: September 5, 2009; **Accepted:** December 21, 2009; **Published:** January 19, 2010

Copyright: © 2010 Beck et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Supported by grants from the NIH, CA12627 (RWB), CA12230 (RWB), the Desmoid Tumor Research Foundation, Inc (MVR), and the Stanford Department of Pathology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: awest@stanford.edu (R.B.W.)
* These authors contributed equally to this work.

Introduction

The development of gene expression microarrays in the mid-1990s represented a significant technical achievement that, for the first time, permitted the systematic genome-wide evaluation of gene expression [1,2]. Since their introduction, these technologies have been widely used for gene expression profiling of cancer samples, leading to the identification of gene expression patterns that predict the biological and clinical features of a wide range of human malignancies [3–16].

Despite the large numbers of gene expression profiling experiments performed on human cancers, the full potential of these technologies for impacting the clinical management of cancer patients has not yet been realized [17–21]. A major limitation of gene expression microarrays for translational cancer research is that they rely on the availability of fresh frozen tissue and show inconsistent performance on formalin fixed paraffin embedded tissue (FFPET) [22–27]. Consequently, gene microarrays cannot be used effectively on the vast majority of tumor specimens, since few samples are stored frozen. In contrast,

essentially all tumor samples are stored as FFPE in pathology laboratories around the world [28]. In an attempt to utilize this rich source of human tumor samples, investigators have resorted to measuring the expression of relatively small numbers of known transcripts from FFPE through the use of a variety of targeted approaches, including reverse transcription-polymerase chain reaction (RT-PCR) [29,30] and cDNA-mediated amplification, selection, extension and ligation (ASL) [31,32]. No technique currently exists for accurate quantitative genome-wide expression profiling from FFPE.

In the past several years, there have been major advances in sequencing technologies, resulting in the development of ultra high-throughput sequencing (UHTS) platforms that have allowed significant increases in sequencing throughput and decreases in sequencing cost [33,34]. There is considerable hope in the scientific community that UHTS will overcome the major limitations of microarray technology and revolutionize the field of functional genomics [35,36]. UHTS has been developed in several platforms, including Roche 454, Illumina Genome Analyzer, and ABI SOLiD. These technologies have been used

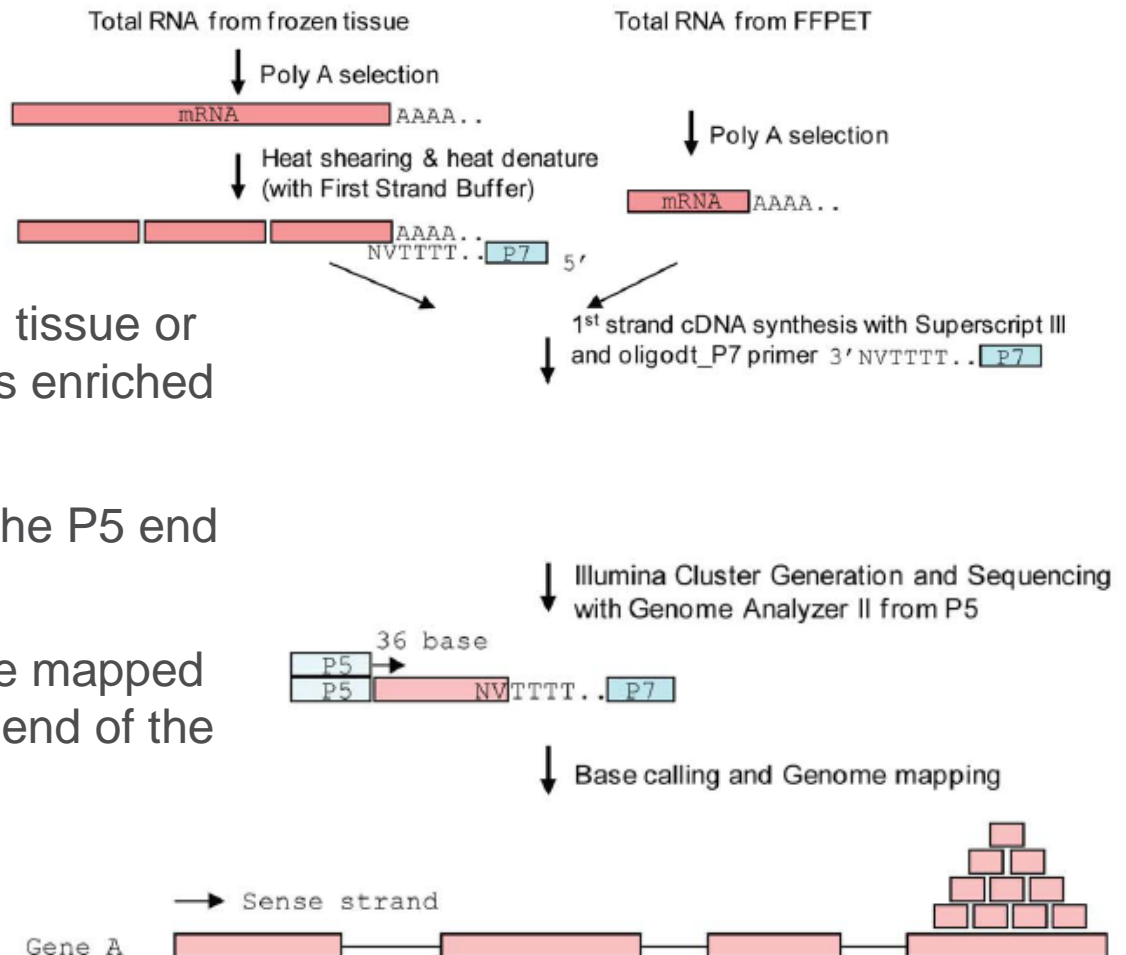
PLOS ONE | www.plosone.org

January 2010 | Volume 5 | Issue 1 | e8768

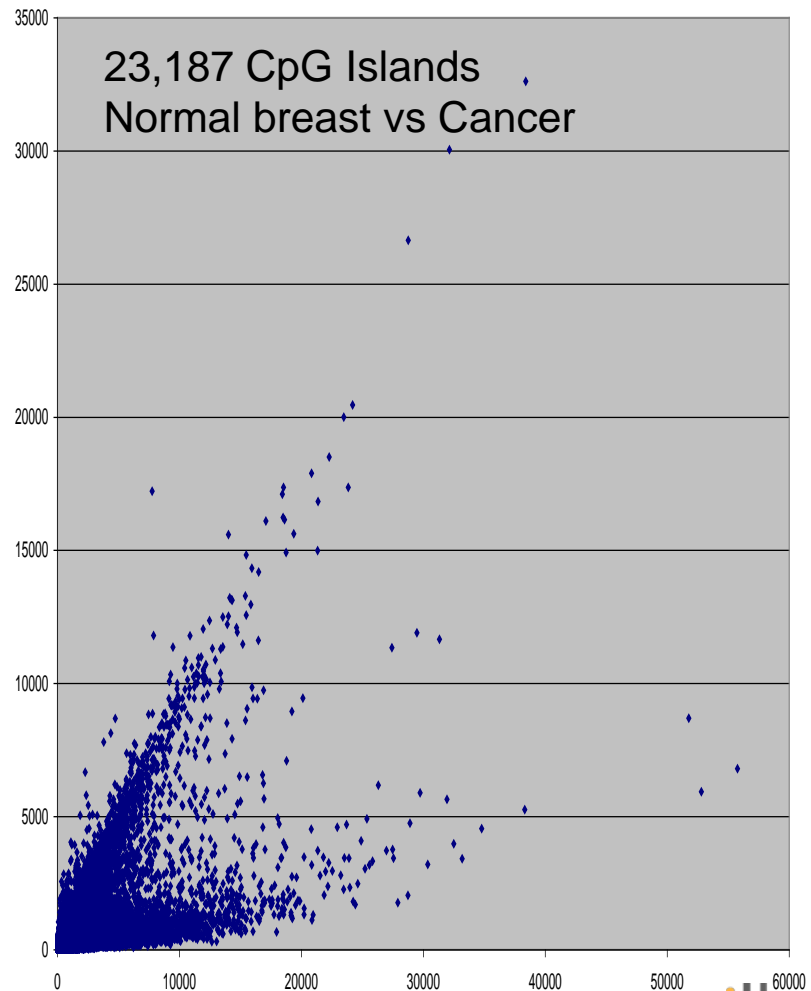
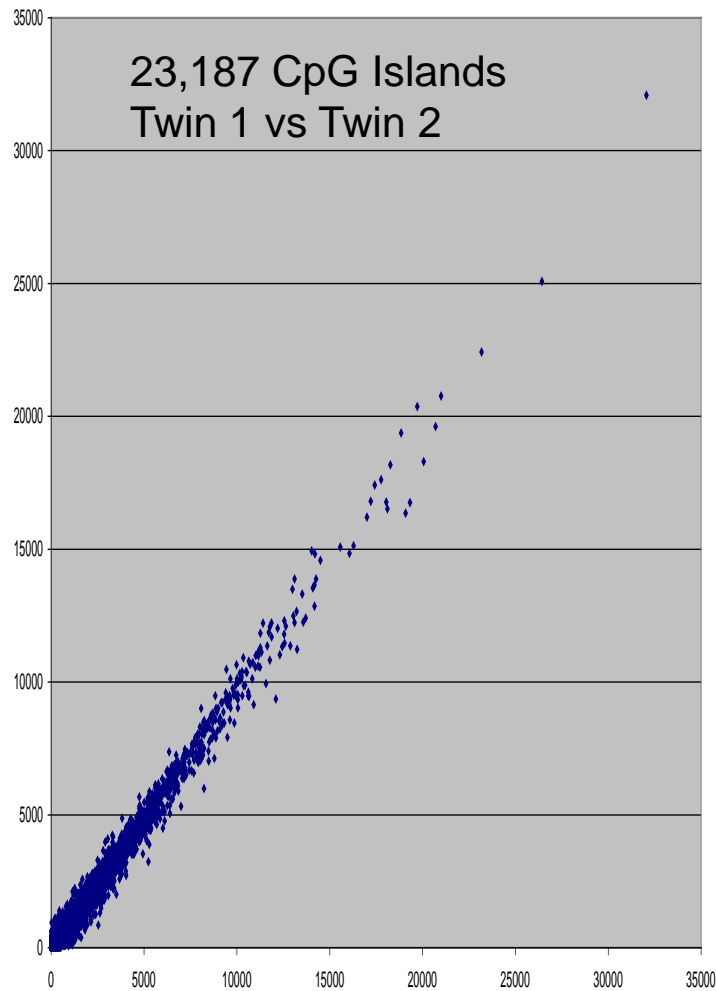
3'-End Sequencing for Expression Quantification (3SEQ) from Archival Tumor Samples

January 2010 | Volume 5 | Issue 1 | e8768

- ▶ Either intact mRNA from frozen tissue or degraded mRNA from FFPE is enriched by poly-A selection
- ▶ The library is sequenced from the P5 end to generate 36 bp reads
- ▶ These reads are expected to be mapped towards to the 3' UTR or the 3' end of the expressed genes



Average CpG Island Methylation



What is Targeted Resequencing?

- ▶ **What is targeted resequencing?**
 - Target Enrichment, Sequence Capture, Genome Partitioning, Genomic Capture, Target Capture, Targeted Pullout
- ▶ **Focus on subset of the genome**
 - Remaining genomic material discarded
- ▶ **Up to several Mb size candidate genomic regions**
- ▶ **Find exact disease causing genes and mutations**
- ▶ **Understand complex human traits**



Genetic diagnosis by whole exome capture and massively parallel DNA sequencing www.pnas.org/cgi/doi/10.1073/pnas.091067210

- ▶ Whole-exome sequencing coupling whole exome arrays to the Illumina DNA sequencing platform.
- ▶ The ability to capture approximately 95% of the targeted coding sequences with high sensitivity and specificity for detection of homozygous and heterozygous variants
- ▶ We illustrate the utility of this approach by identification of a rare mutation in a patient that led to an unexpected clinical diagnosis of the **Barter syndrome**
- ▶ “Because the cost of sequencing on the Illumina platform is potentially considerably lower, we adapted hybrid capture using the NimbleGen 2.1M Human Exome Array to the Illumina DNA sequencing platform”

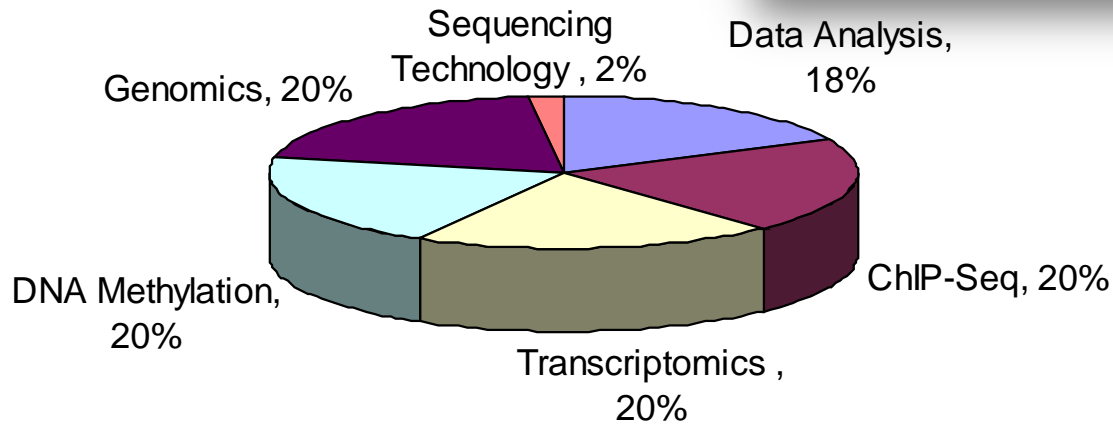
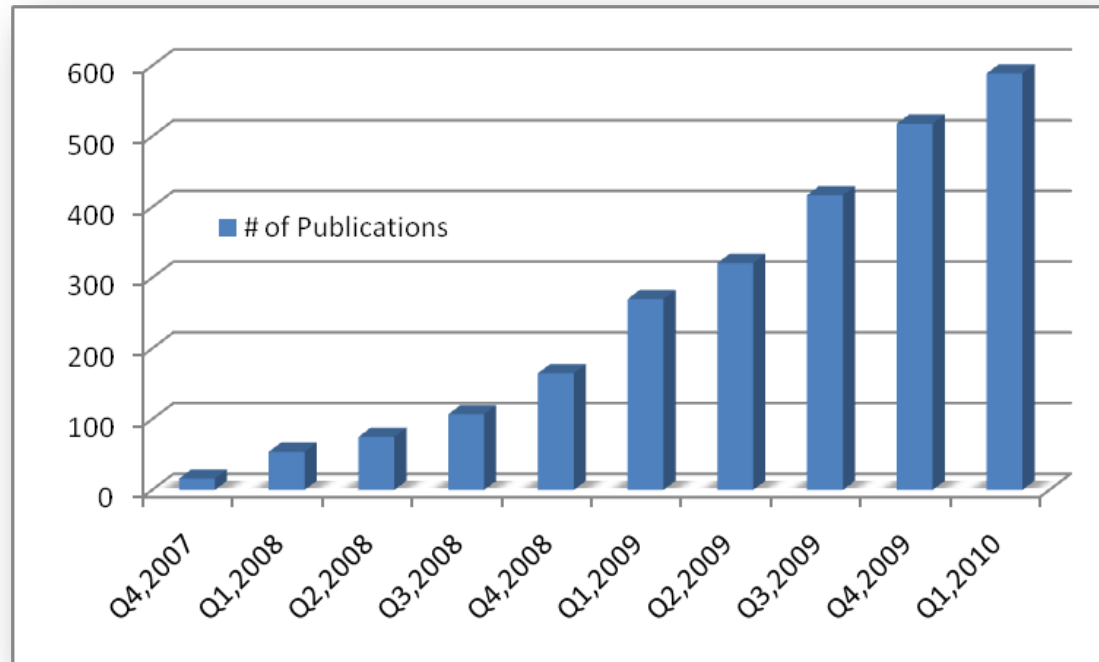
Reference	P L N I E V P K I S L H S L I L [*] F S A V S F L D V S S V R G L K
GIT 264-1	P L N I E V P K I S L H S L I L N F S A V S F L D V S S V R G L K
Sense	5' - CCTCTCAACATTGAGTCCCAAAATCAGCCTCCACAGCCTCATTCTCGACTTTTCAGCAGTGTCCCTTCTTGATGTTTCTTCAGTGAGGGCCCTTAAA - 3'
Antisense	3' - GGAGAGTTGTAACCTCCAGGGGTTTTAGTCGGAGGTGTCGGAGTAAGAGCTGAAAAGTCGTCACAGGAAAGAATCAAAAGAAGTCACTCCCGGAATTT - 5'

3' - GGAGCGTTGTAACCTCCAGGGGTTTTAGTCGGAGGTGTCGGAGTAAGAGT^{*} - 5'
 3' - GTTGTAACTCCAGGGGTTTTAGTCGGAGGTGTCGGAGTAAGAGTTGAAA - 5'
 3' - AACTCCAGGGTTTTTCGTCGGAGGGTTCGGAGTAAGAGTTGAAAAGTCGT - 5'
 5' - ctccaggggttttagtcggaggtgctggagtaagagttgaaaagtcgtca - 3'
 3' - CCAGGGGTTTTAGTCGGAGGTGTCGGAGTAAGAGTTGAAAAGTCGTCACA - 5'
 5' - ggggttttagtcggaggtgctggagtaagagttgaaaagtcgtcacagga - 3'
 3' - TTTTGGTGGGAGGTGTCGGAGTAAGAGTTGAAAAGTCGTCACAGGAAAG - 5'
 3' - TTTAGTCGGAGGTGTCGGAGTAAGAGTTGAAAAGTCGTCACAGGAAAGAA - 5'
 3' - GTCGGAGGCGTCGGAGTAAGAGTTGAAAAGTCGTCACAGGAAAGAATAC - 5'
 5' - cggaggtgctggagtaagagttgaaaagtcgtcacaggaagaactacaa - 3'
 3' - GGGGGGTCGGAGTAAGAGTTGAAAAGTCGTCACAGGAAAGAATCAAAA - 5'
 5' - gaggtgctggagtaagagttgaaaagtcgtcacaggaagaactacaaag - 3'
 3' - GGGTCGGAGTAAGAGTTGAAAAGTCGTCACAGGAAAGAATCAAAAGAAG - 5'
 5' - tcggagtaagagttgaaaagtcgtcacaggaagaactacaaagaagtca - 3'
 3' - GAGTAAGAGTAGAAAAGTCGTCACAGGAAAGAATCAAAAGAAGTCACTC - 5'
 5' - agagttgaaaagtcgtcacaggaagaactacaaagaagtcactccccgg - 3'
 3' - GTTGAAGTCGTCACAGGAAAGAATCAAAAGAAGTCACTCCCGGAAT - 5'



Illumina Sequencing Publications:

nature	93
Science	25



An Illumina Sequencer for Every need

Next Generation Sequencing made accessible



Genome Analyzer *IIe*

- 40 Gb/Run
- 130M Reads
- Read Length: 150bp

Unique combination of sequencing & arrays



HiScanSQ

- 50 Gb/Run
- 250M Reads
- Read Length: 100bp

Most widely adopted NGS platform



Genome Analyzer *IIx*

- 50 Gb/Run
- 250M Reads
- Read Length: 100bp

Redefining the trajectory of sequencing



HiSeq 2000

- 200 Gb/Run
- 1B Reads
- Read Length: 100bp

illumina[®]



Go where the biology takes you

illumina®

Thank You

replay

illumina at ASHG

Workshops, product announcements, expert sessions and more...



Publications

Our researchers publish at an astounding rate



cBot

Breakthrough automated cluster generation gives you sequencing simplified



Discover genetic energy

It's more than genetic analysis. It's the unstoppable drive to know more.

