



Individual Genome Sequencing in the Clinical Laboratory

Julianne O'Daniel
Genetic Counselor & Clinical Genomics Liaison
Illumina Clinical Services Laboratory

© 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, CPro, 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

Overview

- ▶ Sequencing Technology
- ▶ Sequencing in Clinical Diagnostics
- ▶ Whole Genome Sequencing in a Clinical Laboratory



Sequencing Technology

Concepts and Progress

▶ Sequencing:

- Read off the order of bases in DNA (or RNA)
- Identify variants (mutations) between individuals
- Assess consequence & report

▶ Methods:

- 1977: Sanger Sequencing - gold standard for 30 years
- 1998: Capillary machines – enabled scale up to clinical labs - dominant for 6 years
- 2005: Start of Next-Gen Sequencing

▶ Features:

- All start with single DNA molecules
- Amplified: cloning/PCR (Sanger); beads (454, Solid); clusters (Illumina), balls (Complete)
- Non-amplified (3rd Generation): surface-attached (Helicos), nanowells (PacBio)

▶ Impact:

- Throughput: 300,000 to 3,000,000,000 bases per day
- Cost: \$100M to \$48k per human genome
- Accuracy: ~99% (raw); can be >99.999% (consensus)
- Completeness: ~90% (human genome)

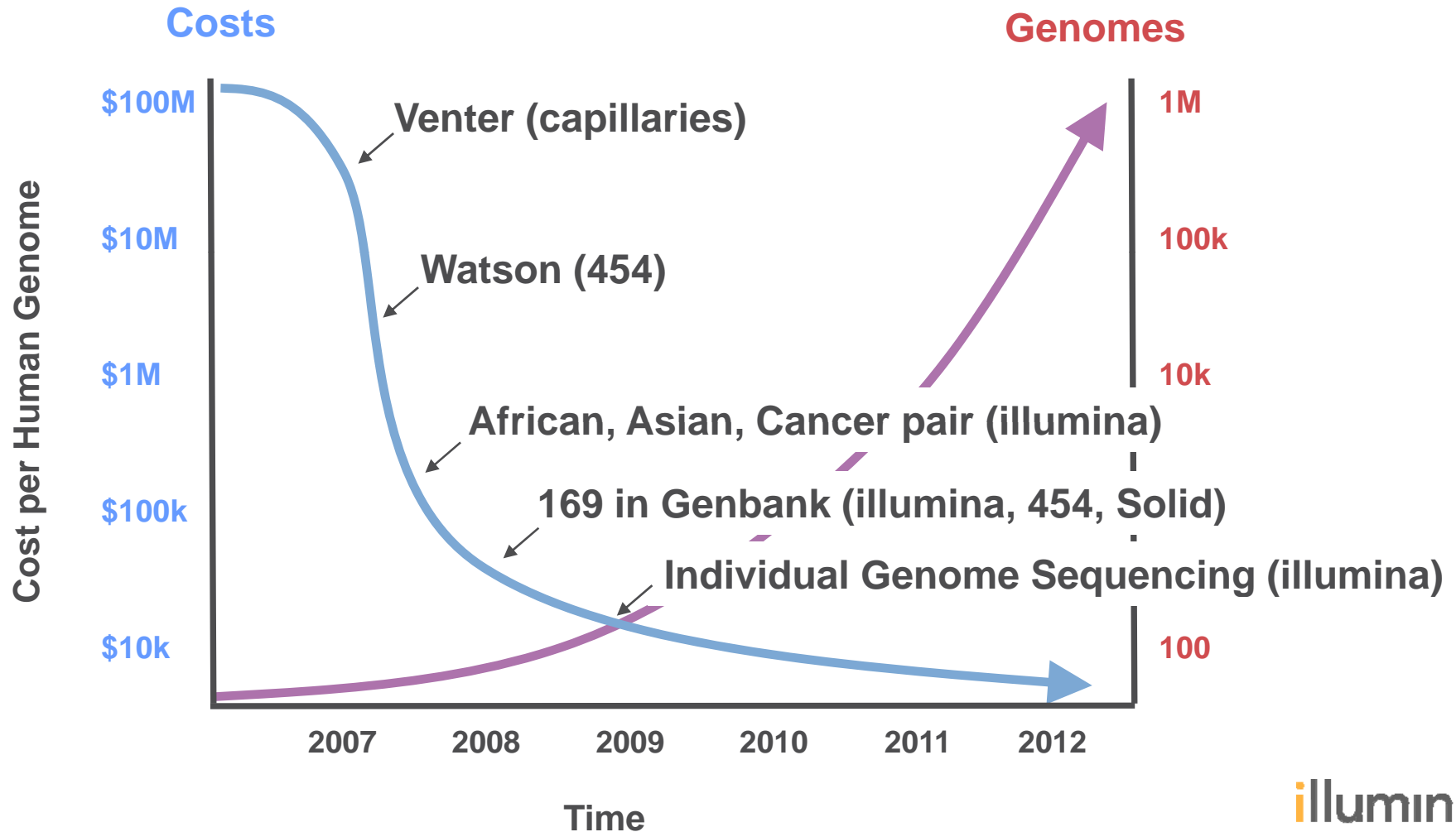
illumina

Illumina Sequencing by the Numbers

- ▶ Each cluster is ~ 200bp
- ▶ 250K clusters per tile
- ▶ 120 tiles per lane
- ▶ 8 lanes per flow cell
 - = 240 million fragments of 200bp
 - = 48 billion bp per flow cell
- ▶ 12x human genome per run
- ▶ Enables scalable genome sequencing
- ▶ Considerations for Clinical use
- ▶ HiSeq will increase capacity even further

illumina

The Tipping Point in Human Genome Sequencing



illumina



Sequencing in Clinical Diagnostics

Genetic Testing Currently in Clinical Laboratories

- ▶ GeneTests.org
 - 1,196 genetic clinics
 - 581 laboratories
 - Genetic testing for 2,070 diseases
 - 1801 Clinical testing
 - 269 Research only testing
- ▶ Most testing performed using:
 - targeted mutation analysis techniques (such as allele-specific PCR)
 - Sanger sequencing to sequence a specific gene region(s)
- ▶ Sanger sequencing
 - Usually bi-directional (2 fold coverage),
 - Accuracy estimated around 98-99% in reports
 - Generally considered “gold standard” for mutation verification

Validity Considerations for Genetic Testing

- ▶ Analytical validity
 - How true is the data?
 - Accuracy
 - Precision
 - Repeatability
 - Reproducibility
- ▶ Clinical validity
 - Does the mutation indicate clinical status? How reliably?
 - Number of reports, case/controls
 - Functional assay
 - Odds ratio
- ▶ Clinical Utility
 - Can the information be used in a medically meaningful way?

CLIA/CAP requirements

- ▶ Analytical validity
 - CLIA and CAP require a laboratory to establish how accurate and precise their testing is
 - CAP runs a proficiency testing program to regularly assess accuracy
 - 2x / year for every test offered

- ▶ Clinical validity
 - CLIA and CAP recommend that a laboratory
 - Maintain and provide documentation around the clinical validity of a test
 - Have information available to physicians to help them assess whether a particular clinical test is appropriate for a particular patient



Whole Genome Sequencing in a Clinical Laboratory

illumina

Establishing Accuracy

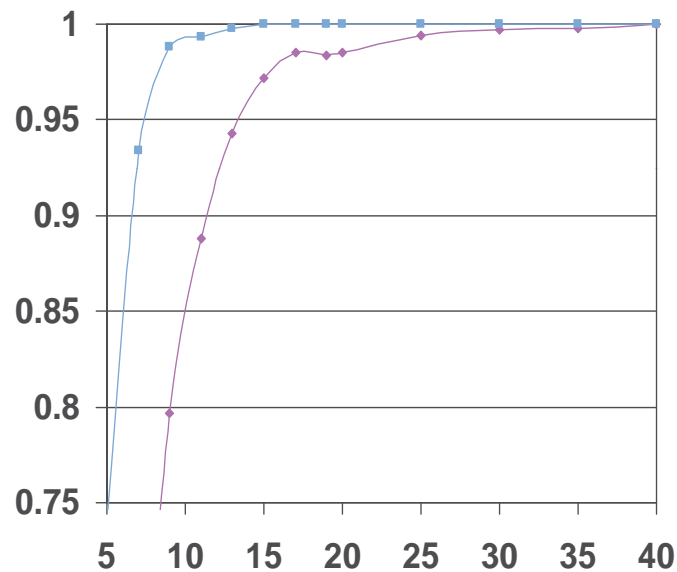
At single base pair level:

- ▶ 14 Coriell samples of known Factor V genotype
- ▶ Sequenced at >6 million-fold depth

= 100% accuracy in genotype calls

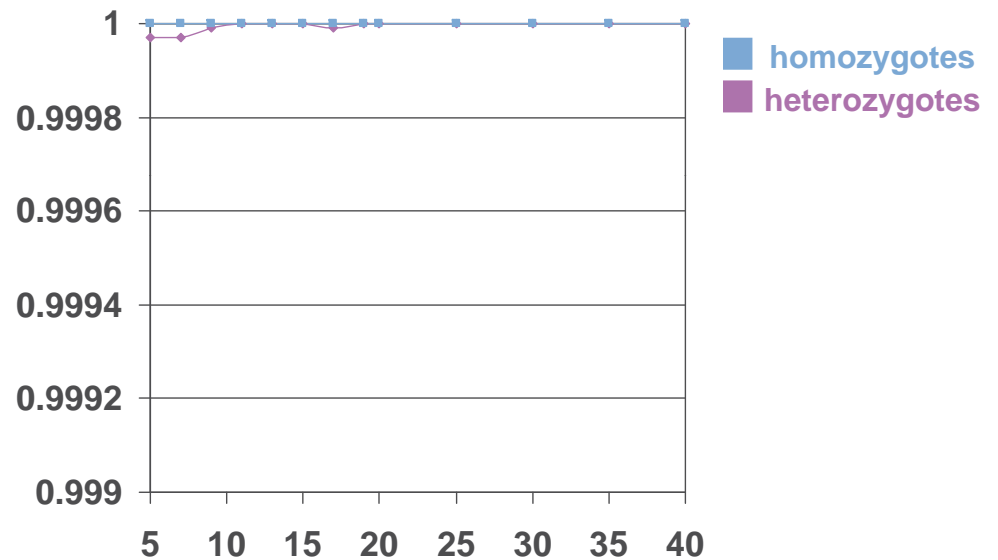
Establishing Accuracy

- Sub-sampled 1,000 times at various folds of coverage



Sensitivity

At average 30 fold coverage:
= 100% for hom
= 99.7% for het*



Specificity

At average 30 fold coverage:
100%

*0.3% heterozygotes called as "n"



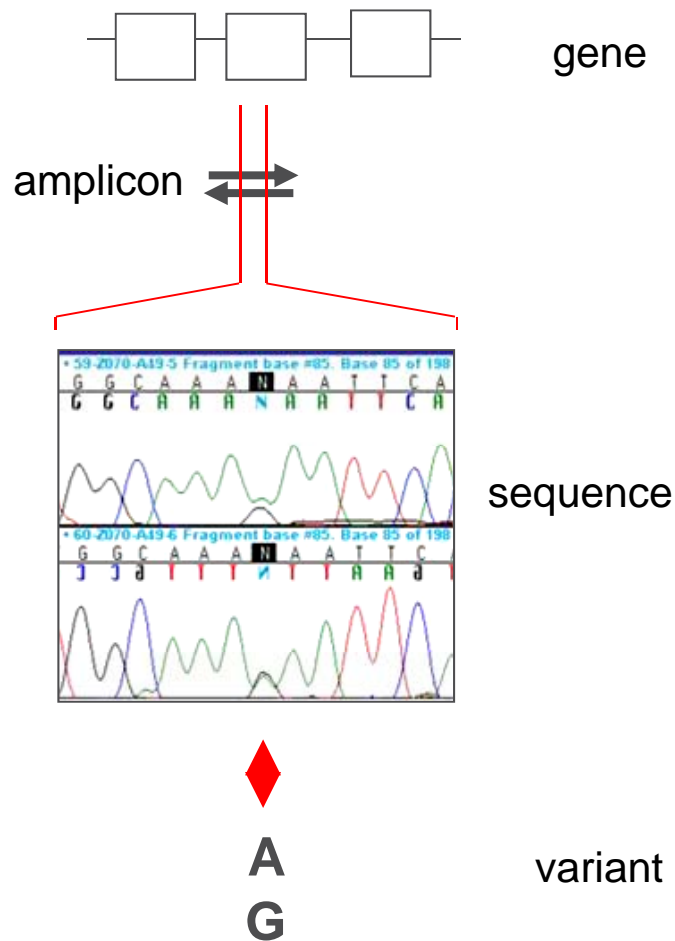
Technical Specifications

- ▶ Paired 100 bp reads from a ~300 bp library
- ▶ Purity filtering
- ▶ Alignment of >130 Gb of sequence to NCBI Human reference
 - ELAND, requires unique alignment of read pairs
- ▶ Quality Metrics
 - raw sequence data,
 - alignments,
 - consensus sequence,
 - coverage of reference,
 - variant calling,
 - concordance with genotype data
- ▶ Method and metrics described in Bentley et al Nature 456: 53-59 (2008)

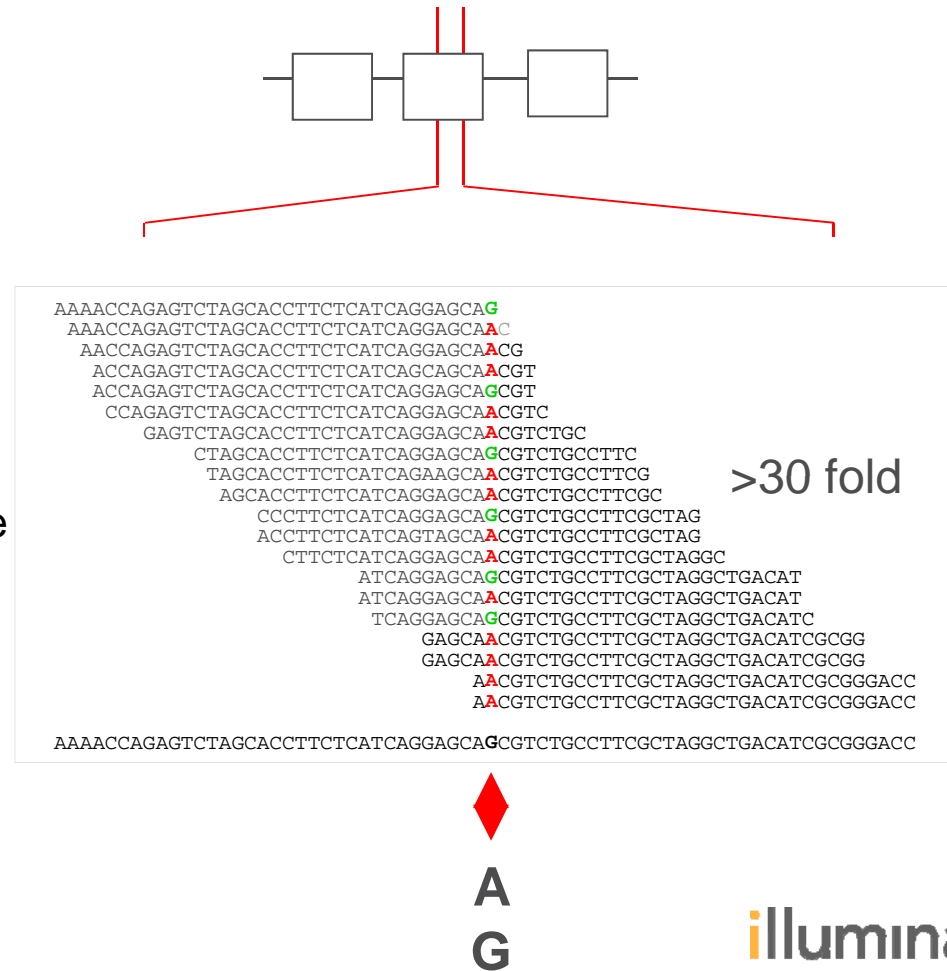
Consensus base call	> three Q30
Average read depth	> 30
Minimum coverage of NCBI36 reference genome by uniquely aligned reads	> 90%

Next Generation Sequencing: Comparing Accuracy

A. Capillary sequencing



B. Illumina IGS sequencing



Precision

Repeatability:

- ▶ 13 sequencing runs over 2-month period
- ▶ Three libraries prepared from the same sample
- ▶ Mitochondrial genome sequences compared
 - >99.99% concordance of base calls
(16,569 of 16,571 bases between 13 different runs and 3 libraries)
 - subsequent runs have been at 100% repeatability

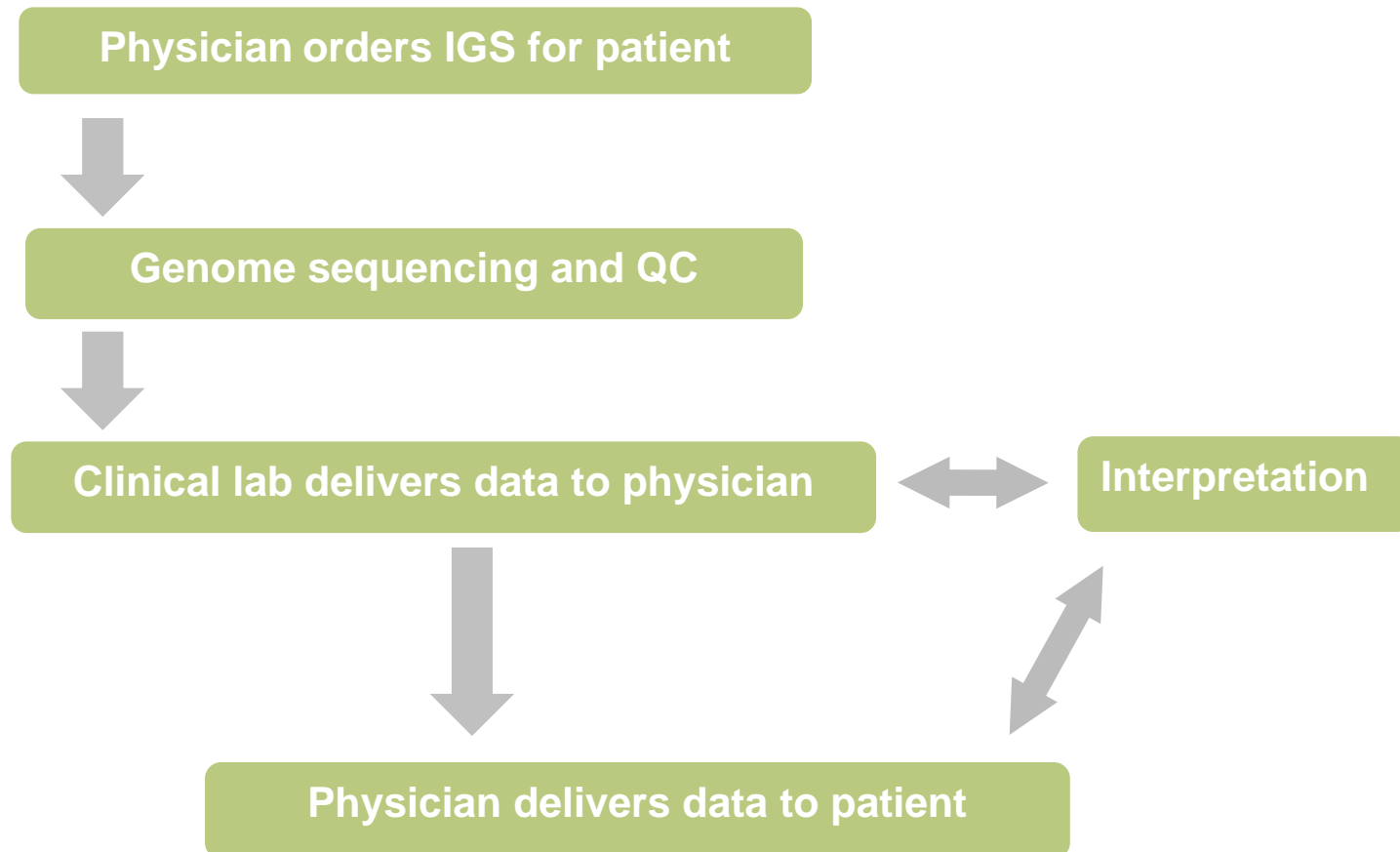
Reproducibility:

- ▶ Two Clinical Laboratory Scientists
- ▶ 3 GAII Sequencers
- ▶ 100% concordance of Phi X genomes between different runs
 - carried out by same or different laboratory scientists

Individual Genome Sequencing in a Clinical Laboratory

- ▶ We established a protocol for sequencing whole genomes based on;
 - Professional and Regulatory agency guidelines
 - Advisement from our external Ethics Advisory Board
 - Guidance from Geneticists, Clinicians, Ethicists and Attorneys
- ▶ Elements of the protocol
 - Process workflow and safety checks
 - Technical validation of the assay
 - Providing a human genome deliverable
- ▶ Protocol and validations were evaluated externally
 - CLIA certified
 - CAP accredited

Individual Genome Sequencing workflow



Individual Genome Sequencing workflow

Physician orders IGS for patient



- Required pre-test discussion/ counseling
- Informed Consent and Service Agreement
- Saliva and blood sample taken
- Cooling-off period (7+ days); order confirmed

Genome sequencing and QC



- Barcode samples for confidentiality
- Saliva and blood genotype for ID match
- Whole genome sequencing of blood DNA
- Check sequence and genotype ID match
- Analyze and QC sequence and called variants
- Archive full dataset

Clinical lab delivers data to physician

illumina

Delivery of Individual Genome Sequence

- ▶ Summary Report
- ▶ Consensus sequence
- ▶ List of SNPs, with dbSNP designation or novel
- ▶ All individual reads, aligned to the human genome reference sequence
- ▶ GenomeStudio genome browser is provided along with annotated variants
- ▶ Encrypted hard drive

illumina

Personal Genome Sequence Report

Dr Naviaux
UCSD School of Medicine
214 Dickinson St., Room C-103,
San Diego, CA 92103-8467
Tel: (619) 543-2904

Physician Name	Dr. Naviaux	
Patient Sex	Male	
Patient Record Number	N/A	
Laboratory Tracking	PG0000001-BLD	
Date Reported		

Specimen type:	Collected	Received
Blood Sample	1/19/09	1/19/09
Saliva Sample		

1 of 7

Services Laboratory
Physician: Dr. Naviaux
No: PG0000001-BLD

605,412,712
1.3%
1,052,543
27,806
3.9
1,240,455,542

DBI published
per chromosome
nt highly repetitive
it have not yet been

Pa in
SNP
1,484
0%

heterozygotes
ozygotes

Novel SNPs	Known in dbSNP				
20	21	22	X	Y	M

one inherited from
2 chromosome as
idual is said to be
to homozygous SNPs

chromosome is

heterozygotes
ozygotes

one inherited from
2 chromosome as
idual is said to be
to homozygous SNPs

heterozygotes
ozygotes

one inherited from
2 chromosome as
idual is said to be
to homozygous SNPs

heterozygotes
ozygotes

one inherited from
2 chromosome as
idual is said to be
to homozygous SNPs

heterozygotes
ozygotes

one inherited from
2 chromosome as
idual is said to be
to homozygous SNPs

heterozygotes
ozygotes



Accuracy at Genomic Level

- ▶ One human genome sequenced to average 33-fold aligned read-depth
 - ▶ aligned to NCBI36 reference
 - ▶ 94% of reference covered
 - ▶ 99.88% sequence identity
 - ▶ 99.72% concordance to genotyping calls using Human1M-duo (across 1,024,890 SNP sites)

Summary Statistics of Pilot Genome

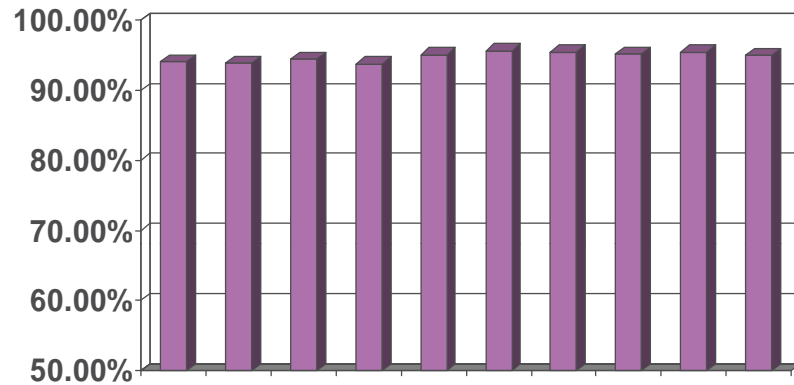
Base pairs reported (consensus)	2,692,877,630
Percentage NCBI36 reference covered (unique aligned)	94.3%
Number of SNPs detected relative to NCBI reference	3,320,918
Number of novel SNPs detected	347,820
Average read depth across reference genome	33
Number of bases used to generate consensus	88,738,112,316

This genome is now publicly available at Genbank

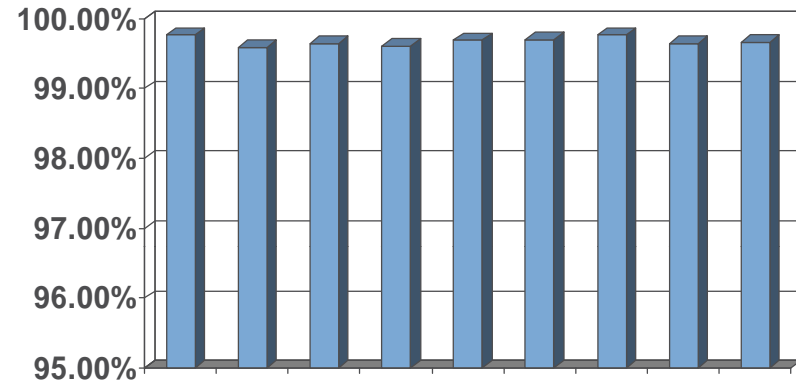


Summary Statistics of subsequent genomes

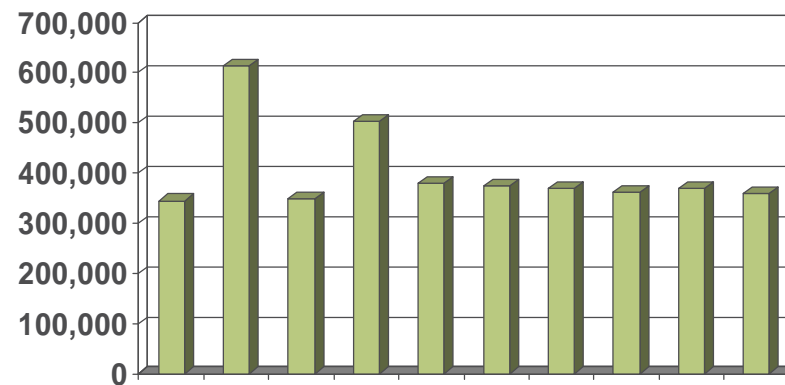
Coverage of NCBI reference



Concordance to Human1M-duo



Novel SNPs detected



Key Features of Individual Genome Sequencing

- ▶ **Physician Oversight:** a physician is an integral part of the process
 - Initial Patient consultations, informed consent
 - Ordering IGS, reviewing results, providing additional counseling/referral
- ▶ **Highly regulated process workflow in the Clinical Services Laboratory**
 - CLIA approved and CAP accredited
 - Advice and regular consultation with external Ethics Advisory Board
- ▶ **High quality IGS deliverable: accurate information but not full interpretation**
 - High quality reads, consensus, and high-confidence genetic variants are provided and all data produced meets a series of defined metrics
- ▶ **Full confidentiality: designed to protect privacy and maintain confidentiality**
 - Samples are positively verified, tracked via anonymous unique identifiers throughout the process
 - No staff have access to both Patient genetic data and Patient ID information
 - Data storage is HIPAA compliant and meets CAP guidelines

Caveats and Considerations

- ▶ Knowledge base
 - What do we know, and what is still uncertain?
 - Are current clinical databases accurate?
 - How do we communicate information and uncertainty?
 - How do we update as understanding is refined?
- ▶ Clinical Aspects
 - What can we say about somebody's health?
 - Carrier status, possible underlying diagnosis, PGx, disease risk associations
 - How can this information aid physicians in treating patients?
 - How can we support clinician education and use?
 - When is the best point during a lifetime to test?
- ▶ Confidentiality
 - What are the risks of having whole genome level information?
 - Who has access to this information?
 - What information should be included in medical records?

Ongoing Developments

- ▶ Physician network (PGNet)
 - Physician education
 - Identify support tools needed
 - Patient and physician referral
- ▶ Engage Professional societies, Clinicians, and Regulatory Agencies
 - Identify specific policy needs
 - Extra sets of eyes
- ▶ Develop additional support tools
 - Interpretation- HGMD, other databases
 - Security options
- ▶ Ongoing evaluations of quality of data
 - Additional validations around
 - Calling novel alleles
 - Sensitivity/specificity across genome
 - Insertions and deletions

Thank You for your attention!

Illumina Clinical Services Laboratory Team:

- ▶ Tina Hambuch
- ▶ Julianne O'Daniel
- ▶ Brad Sickler
- ▶ Marc Laurent
- ▶ Mark Ross
- ▶ Paula Poggio
- ▶ Josh Bernd
- ▶ Suneer Jain
- ▶ Yelena Lyan
- ▶ Phil Cotter
- ▶ David Bentley





jodaniel@illumina.com

illumina

Individual Genome Sequence: Information Workflow

