Complete genomics

Large-scale whole genome sequencing (WGS) for disease understanding, drug development and genomic medicine

Rade Drmanac, Ph.D., Co-Founder and Chief Scientific Officer

IOM Workshop, March 2012

Safe Harbor Summary

This presentation contains forward-looking statements about Complete Genomics, Inc. Our actual results could differ materially from those discussed due to a number of factors, including but limited to our ability to raise additional equity and debt financing on favorable terms, the ability of our technology to achieve and sustain sufficient market acceptance, our ability to scale and commercialize our genome sequencing service, our ability to hire a sufficient number of qualified individuals to run our service facilities, and our ability to manage our rapid growth. Additional risks and uncertainties are described more fully in our preliminary prospectus and registration statement on Form S-1, as amended, filed with the Securities and Exchange Commission. We are providing this information as of the date of this presentation and do not undertake any obligation to update any forward-looking statements contained in this document as a result of new information, future events or otherwise.

Why is Whole Genome Sequencing Critical?

- **Almost all genetic variants have contextual expression and meaning dependent on other genomic sequences and environment interacting through complex "decision making" regulatory networks**
  - Accurate interpretation and effective use of contextual genetic instructions is impossible by partial access to our genetic code (e.g. targeted genes)

- **Each person has 10,000-100,000 family-specific genetic variants + ~100 de novo personal variants in combination with a few million population variants**
  - No comprehensive predefined genetic variant chips can be designed

- **WGS reveals the ultimate genetic level providing complete genetic information**
  - To gain maximal benefits allowed by genetics; allows standardization

Genomics Opportunities in Drug Discovery and Development: Safety, Efficacy, Success Rate

- Translational Medicine (**active**)
  - Pre-clinical
  - Biomarker Discovery
  - Disease Definition

CALCULATING OVERALL SUCCESS RATES
(simple compounded probabilities)

Phase Transition	Phase Success	Phase LOA
P1 to P2	63%	9%
P2 to P3	33%	15%
P3 to NDA/BLA	55%	44%
NDA/BLA to Approval	80%	80%

.63 x .33 x .55 x .80 = .09

biomed tracker

Bio BIOTECHNOLOGY INDUSTRY ORGANIZATION

http://insidebioia.files.wordpress.com/2011/02/bio-ceo-biomedtracker-bio-study-handout-final-2-15-2011.pdf

- Clinical Trials (**need a push**)
  - Phase I – Focus is on safety in healthy individuals (except oncology)
  - Phase II – Mid-size, multi-end-points (targeting to improved efficacy)
  - Phase III – Large studies, carefully designed end-points
  - Phase IV – Monitoring, Indication Expansion

How Many Human Genomes Do We Need to Sequence?

Number of Sequenced Genomes	Impact
1,000,000s	• Understanding molecular and genetic bases of a) thousands of human diseases b) all other phenotypes • Developing better targeted drugs and other therapies including for disease prevention • Developing personal genome interpretation software
1,000,000,000s	• Genomics-based personalized medicine -Improving health and disease treatments • Genomics-based personalized way of living

Industrialization of Whole Human Genome Sequencing

Novel Sequencing Technology

Designed and Optimized for **Massively Parallel (nanoarray based)** Whole Human Genome Sequencing for **Quality**, **Cost** and **Scale** (*Science* Jan 1, 2010)

Innovative Business Model

A Turnkey Service Enabling Customers to **Outsource** WGS – Samples In, **Research and Clinic Ready Data Out**

>3,000 genomes sequenced in 2011
>10,000 genomes expected in 2012 (just 24 instruments)

Easy to Use Informative Lists of Sequence Variants

Complete genomics

-***Every base*** of the genome is marked as reference, variant (SNPs, indels, multi-base changes, SNVs, SVs) **or no-call**

-A highly informative confidence score for each variant to balance sensitivity vs. specificity

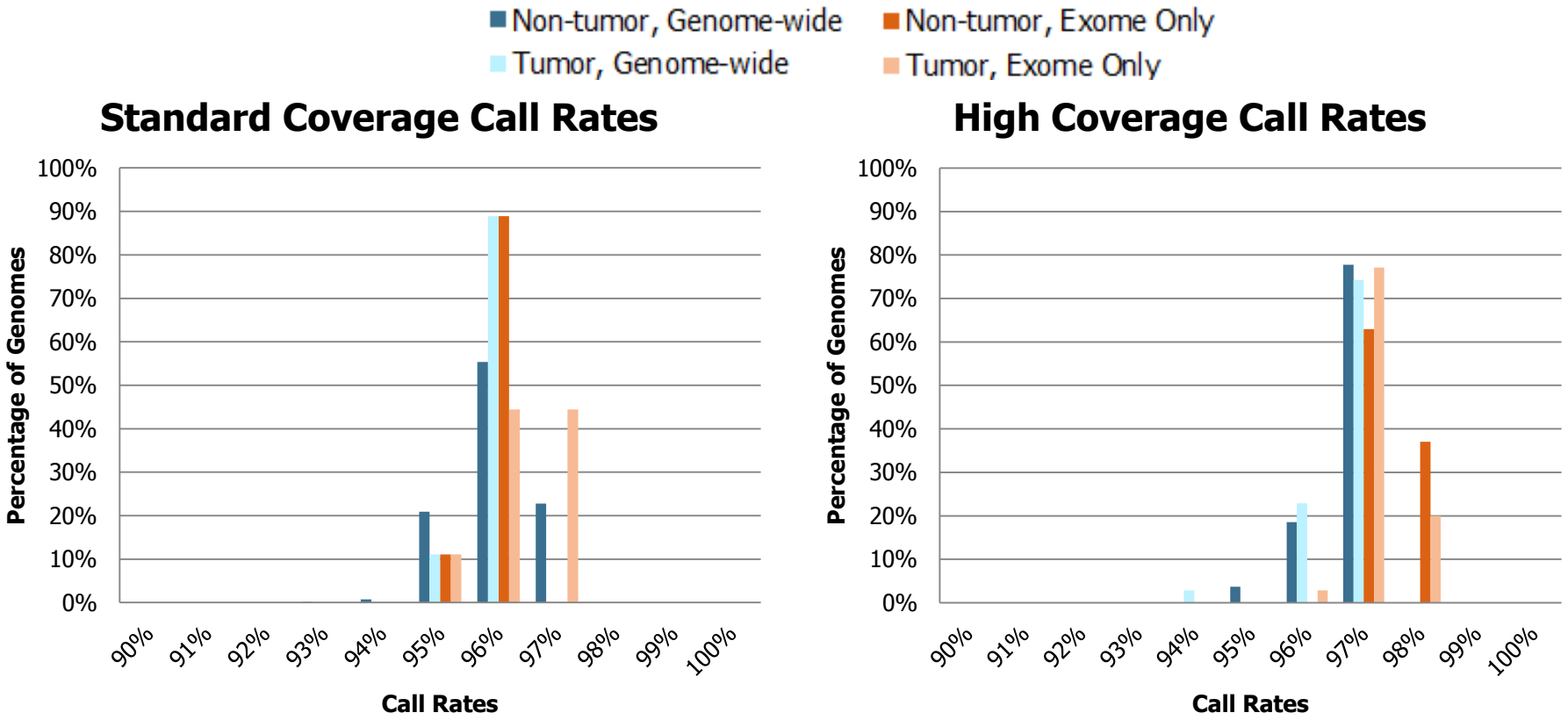
>locus	ploidy	haplotype	chromosome	begin	end	varType	reference	alleleSeq	totalScore	hapLink	xRef
974	2	all	chr1	5099	5126	no-call	=	?			
975	2	all	chr1	5126	5146	ref	=	=			
976	2	1	chr1	5145	5146	snp	G	T	87		dbsnp:806
976	2	2	chr1	5145	5146	snp	G	T	58		dbsnp:806
977	2	all	chr1	5146	5212	ref	=	=			
978	2	1	chr1	5212	5215	ref	GTC	GTC	36		
978	2	2	chr1	5212	5215	no-call-rc	GTC	?T?			

-Annotations of variations in known protein coding (<2% of genome) **and regulatory (more then 10% of genome**) gene sequences

index	locus	Hapl.	Chr.	begin	end	Var Type	Ref.	call	xRef	Gene Id	mrna Acc	proteinAcc	Orient.	Exon Cat.	exon	aa Category	Nucl. Pos	Prot. Pos	Aa Annot	aa Call	aa Ref
66	1269	1	chr1	59315	59316	snp	G	A	dbsnp:rs2854682	79501	NM_001005484.1	NP_001005484.1	+	EXON	1	SYNONYMOUS	362	120	K	K	K
66	1269	2	chr1	59315	59316	ref	G	G		79501	NM_001005484.1	NP_001005484.1	+	EXON	1	NO-CHANGE	362	120	K	K	K
67	1271	1	chr1	59373	59374	snp	A	G	dbsnp:rs2691305	79501	NM_001005484.1	NP_001005484.1	+	EXON	1	MISSENSE	420	140	T	A	T
67	1271	2	chr1	59373	59374	snp	A	G	dbsnp:rs2691305	79501	NM_001005484.1	NP_001005484.1	+	EXON	1	MISSENSE	420	140	T	A	T

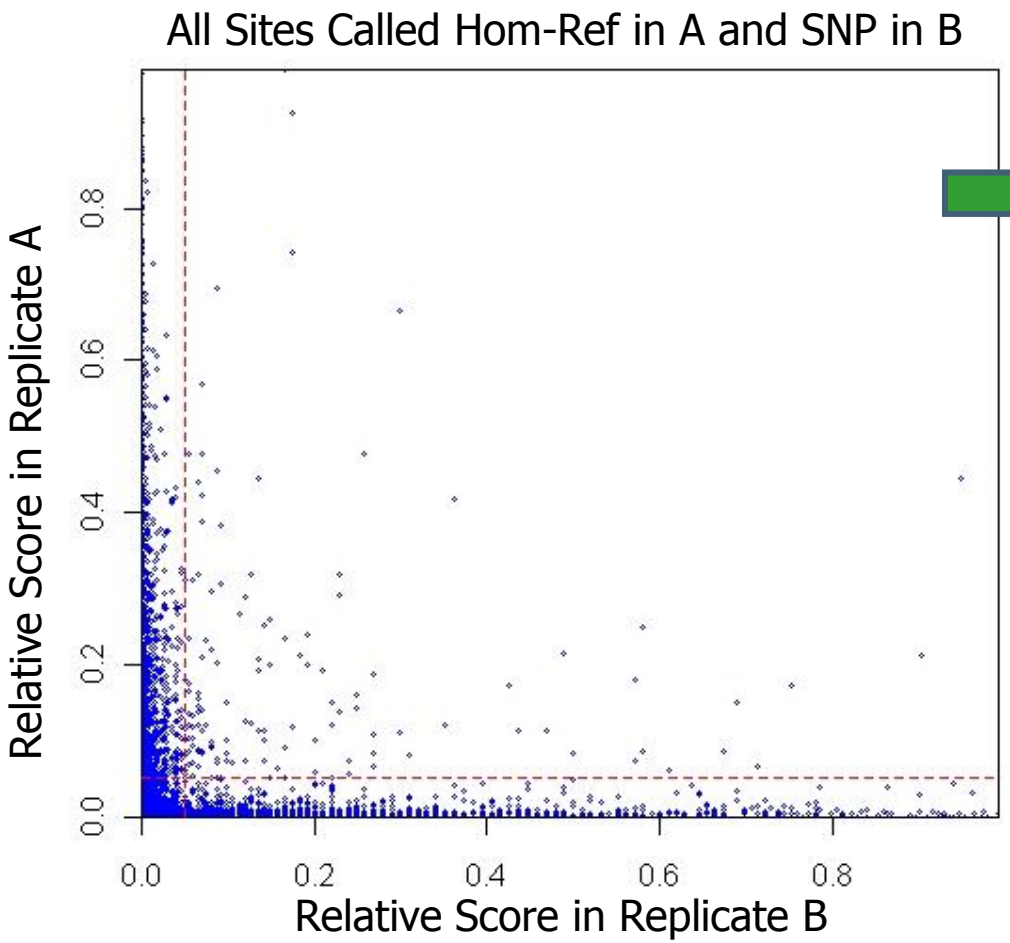
High rates of High-Confidence Base Calls

	Standard Coverage		High Coverage	
	Non-Tumor	Tumor	Non-Tumor	Tumor
Median Genome-wide Call Rate	96.6%	96.5%	97.4%	97.3%
Median Exome Only Call Rate	97.3%	97.0%	97.9%	97.7%



Data for October, November, & December 2011

Sequencing Accuracy in Technical Replicates: Using Variant Confidence Measures



Complete Data

Novel False+ "Somatic SNPs"
= 7,628 (1.10)
= 4,258 (2.0 development pipeline)

Each Such Error Must Be Either:
- False Positive in Replicate B, or
- False Negative in Replicate A

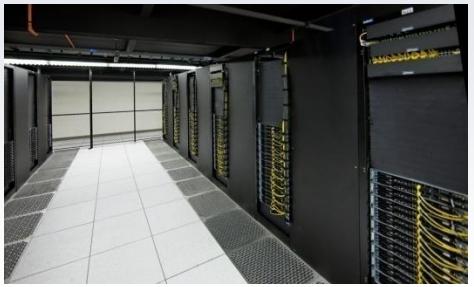
Inside Red Lines

Weakest "Somatic SNPs" Removed
= 5% of True+ (concordant calls)
= 91% of Errors (discordant calls)
= 366 Remaining Errors
= 1 error per 7 called megabases

10x Improvement in Sensitivity/Specificity Trade-Off.

Krishna Pant: NA12878 Replicate Data Analyzed Using Development Pipeline.

Rapid Continued Scaling of Human WGS Service

Samples → DNA Nanoarrays	DNA Nanoarrays → Data	Data → Genomes
		

Our Capacity Expansion

- Up to 2000 genomes per month by end of 2012 (~800 now)
- Expected to have ~5x more powerful instruments in 2012/2013
- Future: Instruments based on 15 fold faster CMOS cameras for sequencing millions of genomes per year

Diverse Base of 100+ Service Users

Complete genomics

Academic/ Government

National Cancer Institute
The University of North Carolina at Chapel Hill
Broad Institute
Stanford University
UCSF
The University of Texas Health Science Center at Houston
Yale University

Medical Research Center

ICR The Institute of Cancer Research
Ernest Gallo Clinic and Research Center at the University of California, San Francisco — New Therapies from Discovery in Molecular Neuroscience™
UCSF
amc Academisch Medisch Centrum Universiteit van Amsterdam
UT Southwestern Medical Center
VIB
Institute for Systems Biology
Harvard Medical School

BioPharma

Lilly
Pfizer
Genentech In Business for Life

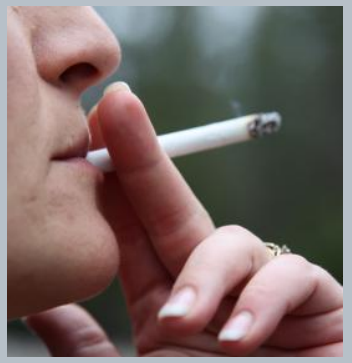
Translational Medicine/Clinical Research

BWH Brigham and Women's Hospital — A Teaching Affiliate of Harvard Medical School
Mayo Clinic
INOVA Health System
The Children's Hospital of Philadelphia®
Children's Hospital Boston
Dana-Farber Cancer Institute
Scripps Translational Science Institute
Erasmus MC

Example WGS Applications

Cancer	• NCI: pediatric cancer (TARGET) **(1,000 genomes)** • Genentech: Hepatitis B Virus (HBV) infection and HCC • AMC (Amsterdam): Neuroblastoma **(Nature 2012 paper; 174 genomes)**
Mendelian Diseases/ De Novo Mutations	• ISB: Miller's Syndrome and disease modifiers (**600 genomes**) • U of Arizona: Infantile Epileptic Encephalopathy • Erasmus: Craniosynostosis
Genomic Variation and Disease	• T2D-GENES: Type II Diabetes risk in families (**700 genomes**) • Stanford: Genetic Determinants of Diabetes Risk • Scripps Health: Wellderly study **(1000 genomes)**
Translational Medicine	• Inova Health System: Pre-term Delivery Study (**1,500 genomes)** • Mayo Clinic: Translational genomics for guiding patient care • USTW: Hypercholesterolimia • Scripps Health: Clinical Annotation of Novel Variants (Cypher)

Accurate WGS are Highly Informative



Detailed Picture of Lung Cancer (Genentech)

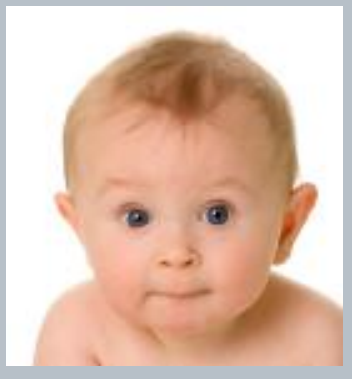
- First Detailed Picture of Cancer Genome Mutations from Primary Non-Small Cell Lung Tumor
- 50,000 New Mutations Indicating Affected Pathways
- **Finding: 8 Mutation per Day of Smoking**
- **Conclusion: Sequence Genome First**

Nature 4-21-10

Discovered Disease-Causing Gene (ISB)

- Miller Syndrome: Formerly Unknown Genetic Cause
- Sequenced 2 Parents and 2 Affected Children
- **Found Causal (Confirmed) and 6 Other -/- Genes**
- **Conclusion: Sequence Genome First**

Science, 3-11-10



Diagnosed Perplexing Case (UTSW)

- 11-Month Old with LDL-C of 837
- Blood Test Failed: Showed Required Protein Not Missing
- **Sequencing Showed Protein Must Be Absent (2 Stop Codons; Confirmed)**; Protein in Blood from Mother's Milk
- **Conclusion: Sequence Genome First**

HMG-Oxford, 8-30-10

WGS Provides Information with Life-saving Treatment Implications

Identification of a Novel *TP53* Cancer Susceptibility Mutation (7kb deletion) Through Whole-Genome Sequencing of a Patient With Therapy-Related AML
Link, D.C. *et al*. *JAMA.* 2011;305(15):1568-1576.

Clinical outcome and benefits: Though patient died, results identified a major risk for her children, who were informed and encouraged to seek genetic testing and counseling

Use of Whole-Genome Sequencing to Diagnose a Cryptic bcr3 PML-RARA Fusion Oncogene
Welch, J.S. *et al*. *JAMA.* 2011;305(15):1577-1584.

Clinical outcome and benefits: Diagnosis and treatment plans were changed on the basis of the identified fusion gene in the course of 7 weeks and patient remains in remission 15 months after her presentation

Whole-Genome Sequencing for Optimized Patient Management
Bainbridge, M.N. *et al*. *Sci. Transl. Med.* **3**, 87re3 (2011).

Clinical outcome and benefits: Identified causal variants (compound mutations in the SPR gene) suggest taking serotonin precursor, 5-hydroxytryptophan, as treatment plan. Patients (paternal twins) are symptom free, one month after treatment.

Golden Age of Disease Research and Translational Medicine

- Expect up to 1 million human genomes sequenced in each of these areas:
  1. **Cancers** (primary, metastatic, CTCs)
  2. Kids and adults with **rare and undiagnosed disease**
  3. Patients with **frequent (chronic) diseases**
  4. **Healthy adults** (e.g. Wellderly, PGP, UK-BioBank, motivated individuals) for longitudinal studies, aging, impact of environment, preventive medicine
  5. Natural or mutagenized or edited iPS and other **stem and stem-like cells** for decoding tissue differentiation and aging (requires haplotyped genomes and complete transcriptomes/epigenomes)
- How quickly can this be done? **5 million genomes from 2012 to 2016?**

  (0.1+0.3+1.0+1.6+2 million genomes each consecutive year)

Genomic Medicine

- **New medicine** based on a) knowing "patient" (and parental) genome sequence, and b) new "genome-specific" remedies

  -WGS as "universal genetic test"; an economical way to use gene sequencing as a screening test without "qualifications"

- This is a predictive, personalized medicine for:
  - **Health improvement (long-term priority)**
  - Disease prevention
  - Treatment improvement
- Improving health and **reducing our unsustainable health care costs**

Genomics and Future Progress in Drug Development

Complete genomics

Expected progress in genomics in the next 3-5 years:

- Exponential growth of WGS capabilities
  - >1 million tests per year capacity
  - ~100% accurate and complete (with haplotypes)
  - Affordable/justifiable price for broad applications
  - Easy lifelong storage of informative variant files (<1Gb/genome)
- Exponential understanding of how human genome works (i.e. how it controls tissue differentiation and functioning)
- Explosive development of sophisticated personal genome interpretation software (accurate and easy to use)
- Synergy with cancer, stem-cell and system biology

Genomics and Future Progress in Drug Development (Cont.)

Impact on Drug Development

- Drug discovery and development standardized on WGS as complete and ultimate genetic information
- Drugs "designed" and clinically tested (with high success rate) for distinct fully characterized "genomic states"
- Only people with personal WGS enrolled (minimize risk, maximize efficacy)
- WGS integrated with complete transciptome and epigenome of relevant tissues for each subject
- WGS as a new foundation that integrates many existing building blocks
- Personal WGS subsumes many companion-diagnostic tests

Possibilities for Future Drugs and Treatments

- Drugs specific for distinct genome states (both efficacy and tolerance)
- Health maintenance drugs (preventing disease development; "supplementing" needs for some genome states)
  - Genome defined age-specific supplementing
  - Living healthier lives in accordance with our genomes
- Targeted "genome-specific" drug delivery (e.g. unique sequences of cancer genomes)
- Nucleic acid (genome-sequence) based drugs
- Cell/tissue therapies; editing genomes in patient's stem cells

*Example: Difficulties in drug treatment of CF in spite of recent progress (Kalydeco form Vertex Pharma.); also, use WGS for healthier reproduction

Complete genomics

Questions, Suggestions

rdrmanac@completegenomics.com