

Are high-resolution, whole-genome approaches the future of clinical and cancer genetics?

Obviously – Yes!

- Why?
 - The cost reduction of microarrays and whole genome sequencing will be making them more accessible for routine diagnostics
 - They have greater diagnostic power and will give definitive diagnoses in more cases.

E.G. Sequencing

- The Human Genome Project cost
~USD3,000,000,000
 - Illumina now offers a complete genome sequence from USD50,000
 - Complete Genomics will offer a complete genome sequence from USD5,000 in 2010
 - These prices will continue to plummet until they are similar to a single gene test.
 - Using barcoding, single samples can already be processed for under \$400 in sequencing costs
 - There are now an estimated 50 complete human genome sequences

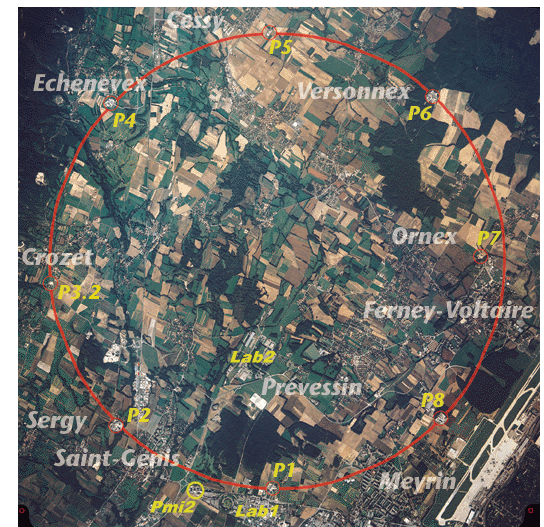
Cost of the Human Genome Project



107 F18 Hornets
- 28M each



19 F22s
- 162M each



1.5 LHC particle accelerator
- 2000M

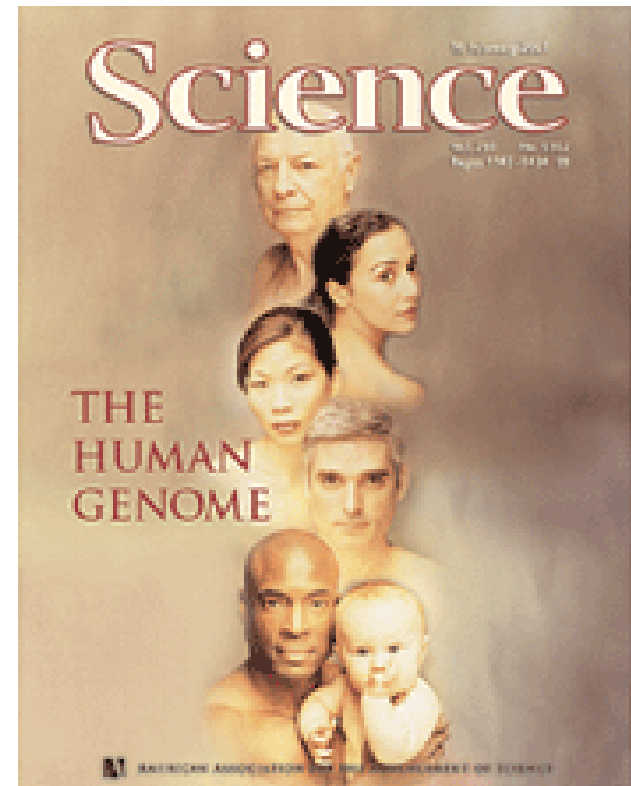
3000M over 15 years
1000M from the DOE
2000m from the NIH

TADA!!

 **2001**

The HGP consortium publishes its working draft in *Nature* (15 February), and Celera publishes its draft in *Science* (16 February).

One composite human genome = USD300 million



A Carrot

- The Archon X PRIZE for genomics – USD10 million
 - 100 human genomes
 - 10 days
 - <10K per genome

The Players

Table 1 Selected next generation sequencing companies

Company	Platform	Developer	Business	Status (price of instrument)
Applied Biosystems	Bead-based massively parallel clonal ligation based DNA sequencing	Agencourt Personal Genomics, Cambridge, Massachusetts	Public company; revenue from instrument and reagent sales	Launched SOLiD in October 2007 (\$591,000)
Complete Genomics	Combinatorial probe-anchor ligation on DNA nanoarrays	Rade Drmanac, Complete Genomics, Mountain View, California	Private company; \$46.5 million raised	Launched as a service company this month (\$5,000/human genome sequence)
Dover, a Danaher Motion Company	Polymerase colony sequencing by ligation	George Church, Harvard University, Cambridge, Massachusetts	Public company; revenue from instrument and reagent sales	Polonator shipped to first users in February (\$150,000)
Helicos	Massively parallel single molecule sequencing by synthesis	Stephen Quake, Stanford University, Stanford, California	Public company; IPO May, 2007	Launched Helicos Genetic Analysis System in February; 2 instruments ordered (\$1.35 million)
Illumina	Sequencing by synthesis	David Bentley, Solexa, UK	Public company; revenue from instrument and reagent sales	Launched IG Genome Analyzer in January 2007; ~200 instruments sold (\$450,000)
Intelligent Bio-Systems	Massively parallel sequencing by synthesis using proprietary reversible fluorescent nucleotide terminators	Jingyue Ju, Columbia University, New York	Private company	Pinpoint Sequencer under development
Pacific BioSciences (formerly Nanofluidics)	Single molecule, real time sequencing by synthesis	Walter Webb and Harold Craighead, Cornell University, Ithaca, New York	Private company; \$178 million raised	SMRT technology under development
Roche	Massively parallel pyrosequencing by synthesis	Jonathan Rothberg, 454 Life Sciences, New Haven, Connecticut	Public company; revenue from instrument and reagent sales	Genome Sequencer (GS) 454 FLX System launched in 2005; ~180 instruments shipped (\$500,000)
VisiGen Biotechnologies	Massively parallel real-time single-molecule sequencing	Susan Hardin, University of Houston, Texas	Private company; ABI and Seqwright, Houston made equity investments	VisiGen sequencing system under development

454 Sequencing Instrument Ease of Use

2. Load PicoTiter plate into instrument

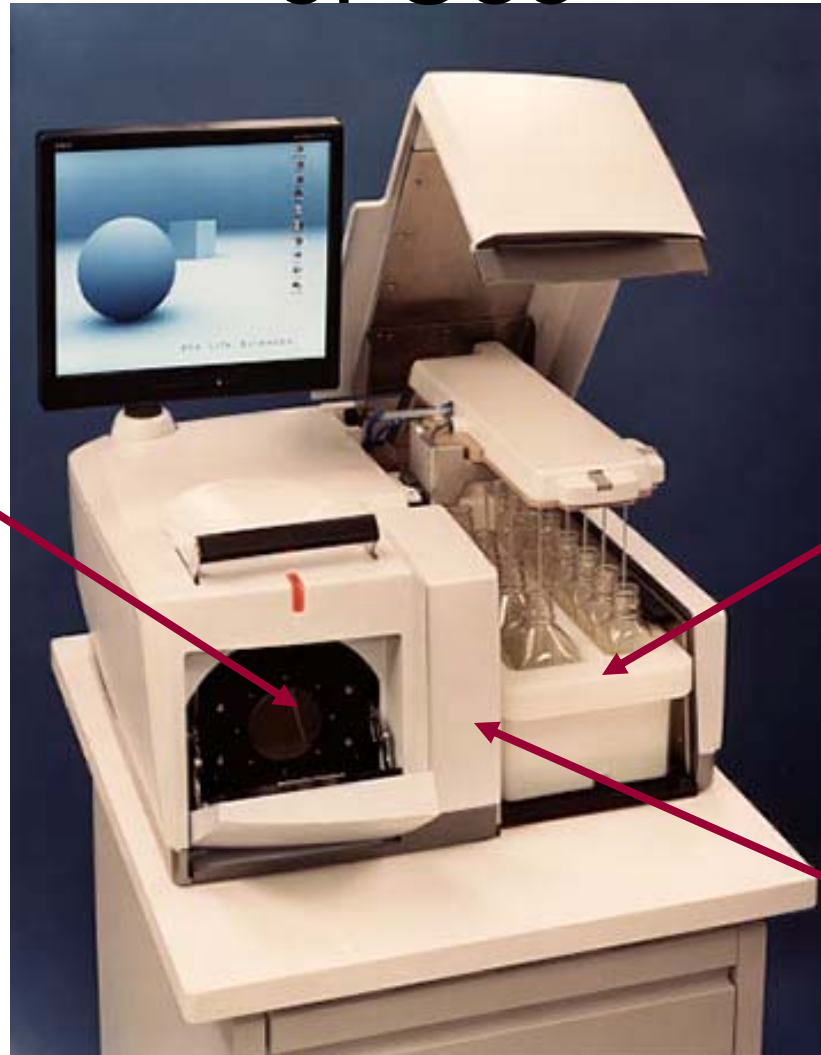
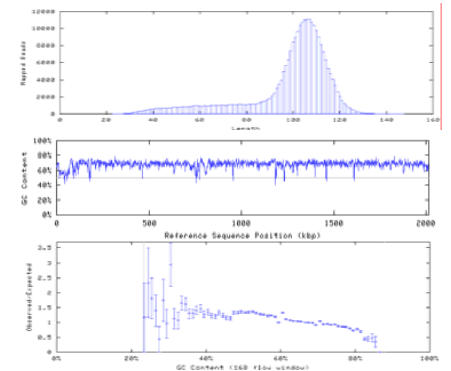


1. Genome is loaded into a PicoTiter™ plate

3. Load Reagents in a single rack



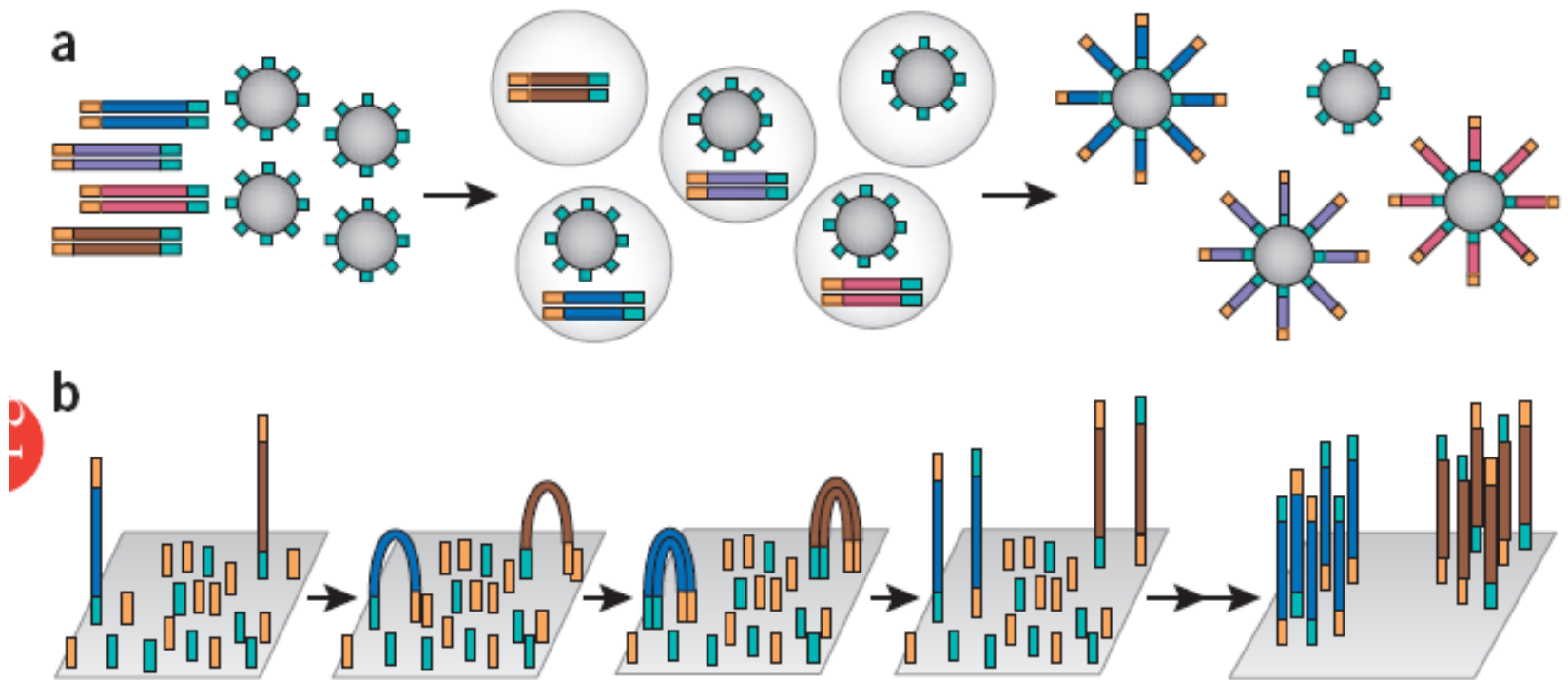
4. Sequence Entire Genome at once, in real-time





What most have in common

- DNA amplification without bacterial cloning



Pacific Biosciences

Single Molecule Real Time (SMRT™) DNA Sequencing
20 zeptoliters (10^{-21} litres) reaction volume

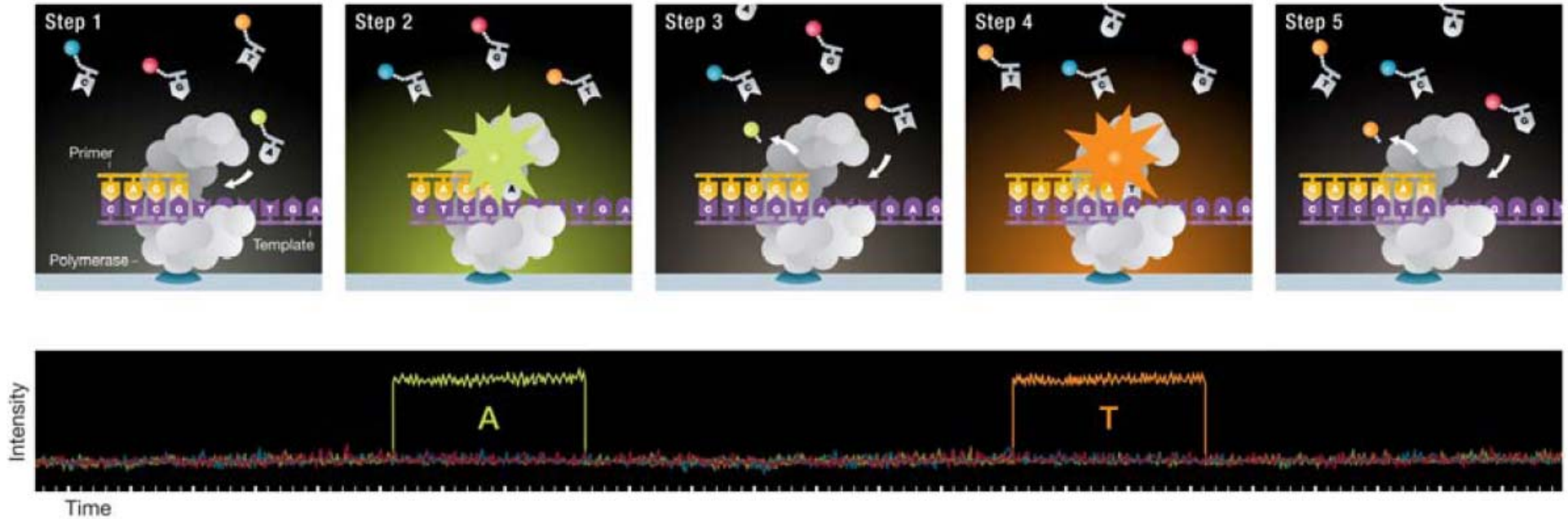


Figure 14. Generation of sequence data.

Enzymatic incorporation of the labeled nucleotide creates a flash of light, which is converted into a base call using optimized algorithms.

Who has been sequenced since

- Craig Venter (3730xl 2007)
- James Watson (454 2008)
- An African (Illumina, 2008)
- An Asian (Illumina, 2008)
- Fibroblasts and blood from an AML patient (Illumina, 2008)

More Genomes coming

- International Cancer Genome Consortium

www.icgc.org

- Australia, Canada, China, India, Singapore, the United Kingdom, the United States (the National Cancer Institute and the National Human Genome Research Institute) and the European Commission
- Coordinate the generation of comprehensive catalogues of genomic abnormalities (somatic mutations) in tumors in 50 different cancer types and/or subtypes which are of clinical and societal importance across the globe.
- a minimum of 500 unique cases for each cancer type or subtype.
- Must have matched normal tissue
- A total of 50,000 genomes
- Funding?

More Genomes coming

- Complete Genomics
 - Complete Genomics is offering a \$5,000 genome as a service from Q2 next year.
 - Over the next five years, Complete Genomics projects that 10 such centers distributed throughout the world will be able to sequence 1 million complete human genomes for <\$1,000 each.
 - Complete Genomics is offering a sequencing service instead of following the traditional instrument sales model, thus relieving its customers of operational, computational and capital purchase burdens, allowing them to focus their resources on scientific discovery
 - Pacific Biosciences anticipates commercializing single molecule a next-generation sequencer by 2010.
 - Pacific Biosciences anticipates having a 15-minute human genome by 2013.

Mol Path in 10 Years

- Routine diagnosis in 10 years time will mainly be performed by high-throughput sequencing
- There will be 4 levels of integrated analysis
 - Expression profiling will be routinely performed
 - Whole genome epigenetic profiling will be feasible
 - Whole germline and somatic genomes will be determined for individual patients for under \$1000
 - Paired end tagged sequencing will partially replace traditional karyotyping
- Single cell and single molecule analyses will be possible

Major Challenges

- Interpretation
- Computing
- Bioinformatics

Why Do Genetic Testing

- Definitive diagnosis
- Prognostic information
- Treatment options
- Family planning

- Increasing complications
 - Phenotypic variability
 - Unknown pathological significance
 - Increased evidence of high frequency of de novo mutations
 - Possible chimerism
 - Possible role of *de novo* mutation

Definitive diagnosis

LAST CHANCE CLINIC

Some diseases defy diagnosis. **Brendan Maher** meets two people who hope that the US National Institutes of Health can help.

Dunham Aurelius is eager to take his shirt off and show his scars. One, a centimetre wide and roughly 20 long runs up his lower back and is from the placement of a steel rod to straighten his spine at the age of 14. Two others, looking like bullet wounds, are above his left buttock. But it's not his scars, nor his barrel-chested physique that have earned him the nickname 'ultimate fighting champion'. His urologist bestowed that title because of the fact that since the age of 22, Aurelius has passed a dozen and a half kidney stones — many, he's proud to say, without assistance. Aurelius is 39, a sculptor and a former triathlete with curly blond locks and a surfer's drawl. His wife, Michelle Barry Aurelius, jokes that he's like a human oyster. But the stones he grows are no smoothed pearls. At the cinema in 2008, Aurelius stepped out to use the bathroom. When he returned, he handed her the four-millimetre wide 'barnacle' of calcium phosphate his body had just expelled. She had noticed he was quiet that evening.

On a February morning this year, Aurelius and Barry are waiting in a hospital room in the sprawling Clinical Center on the campus of the US National Institutes of Health (NIH) in Bethesda, Maryland. They have travelled here from their home in Santa Fe, New Mexico, so that a small team of clini-

"There's nothing so complicated for a patient as not being able to put a name to their disease."

— Carl May

Elsewhere in the Clinical Center, another far-from-average person is awaiting time with Gahl. Sally Massagee, a 54-year-old certified public accountant from Hendersonville, North Carolina, is watching a neuromuscular specialist remove three deep-red slivers of muscle from her bicep. Although she was nervous going into the biopsy, Massagee jokes that she can spare the tissue. A little more than a decade ago, she started putting on weight. By the spring of 2007 she had gained nearly five kilograms on a compact 1.68-metre frame — all of it muscle. People around her thought she was training for competitive bodybuilding, but pain stopped her doing any exercise but tennis. Eventually she outgrew her clothes.

Massagee and Aurelius have few, if any, symptoms in common. What they do both have is a spot in the NIH's Undiagnosed Diseases Program, an effort to identify and characterize previously unknown diseases by drawing on the institution's 6,000 clinical and biomedical experts and the medical technologies at their fingertips. Gahl, a medical geneticist specializing in metabolic disorders, started the programme in May 2008 with \$280,000 in pilot funding from the NIH's Office of Rare Diseases. It received \$1.9 million more in its first year, and has been approved as a fully fledged NIH programme at \$3.5 million per year for the next five. Patients

Next Gen in Diagnosis

enriching for discovery of highly penetrant variants. Here we report on the targeted capture and massively parallel sequencing of the exomes of 12 humans. These include eight HapMap individuals representing three populations⁴, and four unrelated individuals with a rare dominantly inherited disorder, Freeman–Sheldon syndrome (FSS)⁵. We demonstrate the sensitive and specific identification of rare and common variants in over 300 megabases of coding sequence. Using FSS as a proof-of-concept, we show that candidate genes for Mendelian disorders can be identified by exome sequencing of a small number of unrelated, affected individuals. This strategy may be extendable to diseases with more complex genetics through larger sample sizes and appropriate weighting of non-synonymous variants by predicted functional impact.

Figure 08250

ing

Next Gen in Diagnosis

Routine clinical use of massively parallel sequencing will require higher accuracy, better ways to select genomic subsets of interest, and improvements in the functionality, speed, and ease of use of data analysis software. In addition, substantial enhancements in laboratory computer infrastructure, data storage, and data transfer capacity will be needed to handle the extremely large data sets produced. Clinicians and laboratory personnel will require training to use the sequence data effectively, and appropriate methods will need to be developed to deal with the incidental discovery of pathogenic mutations and variants of uncertain clinical significance. Massively parallel sequencing has the potential to transform the practice of medical genetics and related fields, but the vast amount of personal genomic data produced will increase the responsibility of geneticists to ensure that the information obtained is used in a medically and socially responsible manner.

Genomics shifts focus to rare diseases

COLD SPRING HARBOR, NEW YORK

Genome sequencing may finally be living up to its promise of pinpointing genetic mutations that bear on treatment for individual patients. But the breakthroughs are not coming from the DNA analysis of common diseases with complex genetic origins, which has been the obsession of genomics for nearly the past decade. Instead, many genome scientists are turning back to study rare disorders that are traceable to defects in single genes, and whose causes have remained a mystery.

The change is partly a result of frustration with the disappointing results of genome-wide association studies (GWAS). Rather than sequencing whole genomes, GWAS studies examine a subset of DNA variants in thousands of unrelated people with common diseases. Now, however, sequencing costs are dropping, and whole genome sequences can quickly provide in-depth information about individuals, enabling scientists to locate genetic mutations that underlie rare diseases by sequencing a handful of people.

"Years ago, people were using families and mapping approaches to distil down to a region where they thought a causative gene was," says Elaine Mardis, a director of the Genome



Hugh Rienhoff sequenced family transcriptomes to try to diagnose his daughter Bea's genetic disease.

Not Enough or Too much Information?

- Use of these technologies will require changes in the practice of clinical genetics
 - Consent for testing will need to cover the potential of unexpected findings? E.g. a microdeletion causing congenital defects and a potential for cancer predisposition, but this is already done for Down syndrome
 - Consent may have to include that only a small amount of genetic information should be reported on? e.g.. only genes with known single gene disorders, no reporting on complex diseases of little prognostic value

Not Enough or Too much Information?

- Use of these technologies will require a change in the practice of clinical genetics
 - That these tests may result in a lack of definitive diagnostic and prognostic information will need to be carefully explained on consent and test results. Can only report what we know?
 - How does the health care system handle potential increased clinical load?

Summary



LAST CHANCE CLINIC

Some diseases defy diagnosis. **Brendan Maher** meets two people

"It's going to be nice to be able to walk out of here, if not with a diagnosis, with at least the next step." — Dunham Aurelius

Definitive diagnosis and Prognosis?

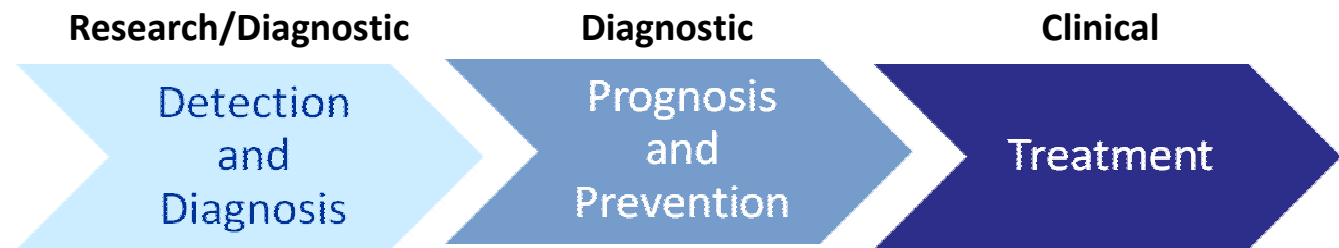
CONCLUSIONS

We have identified recurrent molecular lesions that elude syndromic classification and whose disease manifestations must be considered in a broader context of development as opposed to being assigned to a specific disease. Clinical diagnosis in patients with these lesions may be most readily achieved on the basis of genotype rather than phenotype.

Oncogenomics

- Small DNA sequence abnormality
 - importance of recurrent genetic lesions
- Copy number variation (cytogenetics)
- Structural aberrations (cytogenetics)
- Epigenomic modification
 - Methylation
 - Histone modification
- Transcriptional alterations

Personalized Cancer Medicine



Genotyping Oncomutation Panel

- Identify patterns of genes associated with the development of specific cancers

Methylation EpiPanel

- Study how genetic variations that change protein function cause cells to function abnormally

Gene Expression

- Investigate how those changes are further affected by lifestyle behaviors and environmental factors

Copy Number Analysis

- Develop biomarkers to measure changes in protein and cellular function associated with specific cancers
-
-
-

Monitor cellular function to detect pre-cancerous changes

- Use genetic profiles to identify subsets of cancer types that define prognosis

Choose targeted therapies based on cancer type and the individual biological profile

Use biospecimen tests and imaging techniques to refine treatment

An Integrated Genomic Analysis of Human Glioblastoma Multiforme

D. Williams Parsons,^{1,2*} Siân Jones,^{1*} Xiaosong Zhang,^{1*} Jimmy Cheng-Ho Lin,^{1*} Rebecca J. Leary,^{1*} Philipp Angenendt,^{2*} Parminder Mankoo,³ Hannah Carter,³ I-Mei Siu,⁴ Gary L. Gallia,⁴ Alessandro Olivi,⁴ Roger McLendon,⁵ B. Ahmed Rasheed,⁵ Stephen Keir,⁵ Tatiana Nikolskaya,⁶ Yuri Nikolsky,⁷ Dana A. Busam,⁸ Hanna Tekleab,⁸ Luis A. Diaz Jr.,¹ James Hartigan,⁹ Doug R. Smith,⁹ Robert L. Strausberg,⁸ Suely Kazue Nagahashi Marie,¹⁰ Sueli Mieko Oba Shinjo,¹⁰ Hai Yan,⁵ Gregory J. Riggins,⁴ Darell D. Bigner,⁵ Rachel Karchin,³ Nick Papadopoulos,¹ Giovanni Parmigiani,¹ Bert Vogelstein,^{1†} Victor E. Velculescu,^{1†} Kenneth W. Kinzler^{1†}

Glioblastoma multiforme (GBM) is the most common and lethal type of brain cancer. To identify the genetic alterations in GBMs, we sequenced 20,661 protein coding genes, determined the presence of amplifications and deletions using high-density oligonucleotide arrays, and performed gene expression analyses using next-generation sequencing technologies in 22 human tumor samples. This comprehensive analysis led to the discovery of a variety of genes that were not known to be altered in GBMs. Most notably, we found recurrent mutations in the isocitrate dehydrogenase 1 (*IDH1*) in 12% of GBM patients. Mutations in *IDH1* occurred in young patients and in most patients with secondary GBMs and were associated with overall survival. These studies demonstrate the value of unbiased genomic characterization of human brain cancer and identify a potentially useful gene classification and targeted therapy of GBMs.

Science, 2008

NEJM, 2009

ORIGINAL ARTICLE

IDH1 and *IDH2* Mutations in Gliomas

Hai Yan, M.D., Ph.D., D. Williams Parsons, M.D., Ph.D., Genglin Jin, Ph.D., Roger McLendon, M.D., B. Ahmed Rasheed, Ph.D., Weishi Yuan, Ph.D., Ivan Kos, Ph.D., Ines Batinic-Haberle, Ph.D., Siân Jones, Ph.D., Gregory J. Riggins, M.D., Ph.D., Henry Friedman, M.D., Allan Friedman, M.D., David Reardon, M.D., James Herndon, Ph.D., Kenneth W. Kinzler, Ph.D., Victor E. Velculescu, M.D., Ph.D., Bert Vogelstein, M.D., and Darell D. Bigner, M.D., Ph.D.

ABSTRACT

BACKGROUND

A recent genomewide mutational analysis of glioblastomas (World Health Organization [WHO] grade IV glioma) revealed somatic mutations of the isocitrate dehydrogenase 1 gene (*IDH1*) in a fraction of such tumors, most frequently in tumors that were known to have evolved from lower-grade gliomas (secondary glioblastomas).

Oncogenomics

- Small DNA sequence abnormality
 - importance of recurrent genetic lesions
- Copy number variation (cytogenetics)
- Structural aberrations (cytogenetics)
- Epigenomic modification
 - Methylation
 - Histone modification
- Transcriptional alterations

ARTICLES

**Whole Genome Sequencing
(WGS)****DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome**

Timothy J. Ley^{1,2,3,4*}, Elaine R. Mardis^{2,3*}, Li Ding^{2,3}, Bob Fulton³, Michael D. McLellan³, Ken Chen³, David Dooling³, Brian H. Dunford-Shore³, Sean McGrath³, Matthew Hickenbotham³, Lisa Cook³, Rachel Abbott³, David E. Larson³, Dan C. Koboldt³, Craig Pohl³, Scott Smith³, Amy Hawkins³, Scott Abbott³, Devin Locke³, LaDeana W. Hillier^{3,8}, Tracie Miner³, Lucinda Fulton³, Vincent Magrini^{2,3}, Todd Wylie³, Jarret Glasscock³, Joshua Conyers³, Nathan Sander³, Xiaoqi Shi³, John R. Osborne³, Patrick Minx³, David Gordon⁸, Asif Chinwalla³, Yu Zhao¹, Rhonda E. Ries¹, Jacqueline E. Payton⁵, Peter Westervelt^{1,4}, Michael H. Tomasson^{1,4}, Mark Watson^{3,4,5}, Jack Baty⁶, Jennifer Ivanovich^{4,7}, Sharon Heath^{1,4}, William D. Shannon^{1,4}, Rakesh Nagarajan^{4,5}, Matthew J. Walter^{1,4}, Daniel C. Link^{1,4}, Timothy A. Graubert^{1,4}, John F. DiPersio^{1,4} & Richard K. Wilson^{2,3,4}

Acute myeloid leukaemia is a highly malignant haematopoietic tumour that affects about 13,000 adults in the United States each year. The treatment of this disease has changed little in the past two decades, because most of the genetic events that initiate the disease remain undiscovered. Whole-genome sequencing is now possible at a reasonable cost and timeframe to use this approach for the unbiased discovery of tumour-specific somatic mutations that alter the protein-coding genes. Here we present the results obtained from sequencing a typical acute myeloid leukaemia genome, and its matched normal counterpart obtained from the same patient's skin. We discovered ten genes with acquired mutations; two were previously described mutations that are thought to contribute to tumour progression, and eight were new mutations present in virtually all tumour cells at presentation and relapse, the function of which is not yet known. Our study establishes whole-genome sequencing as an unbiased method for discovering cancer-initiating mutations in previously unidentified genes that may respond to targeted therapies.

Flow Chart of Filters used to identify somatic point mutations in AML tumour genome

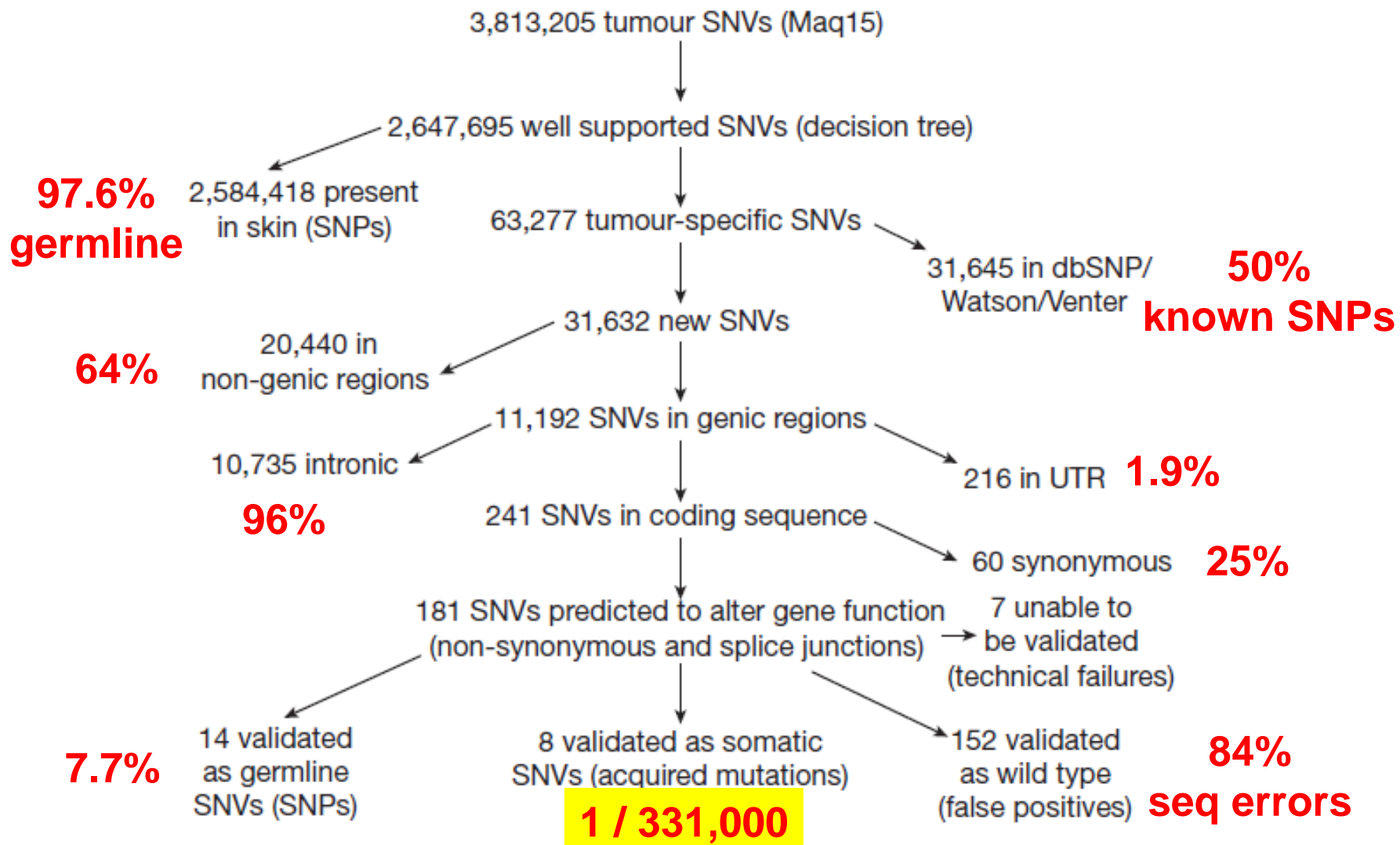


Figure 2 | Filters used to identify somatic point mutations in the tumour genome. See text for details. UTR, untranslated regions.

Table 2 | Non-synonymous somatic mutations detected in the AML sample

Gene	Consequence	Type	Solexa tumour reads WT:variant	Solexa skin reads WT:variant	Conservation score of mutant base	Mutations in other AML cases*
<i>CDH24</i>	Y590X	Nonsense	9:9	16:0	0.998	0/187
<i>SLC15A1</i>	W77X	Nonsense	15:12	19:0	1.000	0/187
<i>KNDC1</i>	L799F	Missense	7:8	20:0	NA	0/187
<i>PTPRT</i>	P1235L	Missense	9:13	16:0	1.000	0/187
<i>GRINL1B</i>	R176H	Missense	15:10	14:0	NA	0/187
<i>GPR123</i>	T38I	Missense	11:11	13:0	NA	0/187
<i>EBI2</i>	A338V	Missense	7:12	18:2	1.000	0/187
<i>PCLKC</i>	P1004L	Missense	19:9	15:1	0.98	0/187
<i>FLT3</i>	ITD	Indel	18:12	8:0	NA	51/185
<i>NPM1</i>	CATG ins	Indel	36:6	33:0	NA	43/180

Gene	Consequence	Mutations in other AML cases*
<i>CDH24</i>	Y590X	0/187
<i>SLC15A1</i>	W77X	0/187
<i>KNDC1</i>	L799F	0/187
<i>PTPRT</i>	P1235L	0/187
<i>GRINL1B</i>	R176H	0/187
<i>GPR123</i>	T38I	0/187
<i>EBI2</i>	A338V	0/187
<i>PCLKC</i>	P1004L	0/187
<i>FLT3</i>	ITD	51/185
<i>NPM1</i>	CATG ins	43/180

Rare
Or
passenger
mutations
(not driver)

ARTICLES

**Whole Genome Sequencing
(WGS)****DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome**

Timothy J. Ley^{1,2,3,4*}, Elaine R. Mardis^{2,3*}, Li Ding^{2,3}, Bob Fulton³, Michael D. McLellan³, Ken Chen³, David Dooling³, Brian H. Dunford-Shore³, Sean McGrath³, Matthew Hickenbotham³, Lisa Cook³, Rachel Abbott³, David E. Larson³, Dan C. Koboldt³, Craig Pohl³, Scott Smith³, Amy Hawkins³, Scott Abbott³, Devin Locke³, LaDeana W. Hillier^{3,8}, Tracie Miner³, Lucinda Fulton³, Vincent Magrini^{2,3}, Todd Wylie³, Jarret Glasscock³, Joshua Conyers³, Nathan Sander³, Xiaoqi Shi³, John R. Osborne³, Patrick Minx³, David Gordon⁸, Asif Chinwalla³, Yu Zhao¹, Rhonda E. Ries¹, Jacqueline E. Payton⁵, Peter Westervelt^{1,4}, Michael H. Tomasson^{1,4}, Mark Watson^{3,4,5}, Jack Baty⁶, Jennifer Ivanovich^{4,7}, Sharon Heath^{1,4}, William D. Shannon^{1,4}, Rakesh Nagarajan^{4,5}, Matthew J. Walter^{1,4}, Daniel C. Link^{1,4}, Timothy A. Graubert^{1,4}, John F. DiPersio^{1,4} & Richard K. Wilson^{2,3,4}

Acute myeloid leukaemia is a highly malignant haematopoietic tumour that affects about 13,000 adults in the United States each year. The treatment of this disease has changed little in the past two decades, because most of the genetic events that initiate the disease remain undiscovered. Whole-genome sequencing is now possible at a reasonable cost and timeframe to use this approach for the unbiased discovery of tumour-specific somatic mutations that alter the protein-coding genes. Here we present the results obtained from sequencing a typical acute myeloid leukaemia genome, and its matched normal counterpart obtained from the same patient's skin. We discovered ten genes with acquired mutations; two were previously described mutations that are thought to contribute to tumour progression, and eight were new mutations present in virtually all tumour cells at presentation and relapse, the function of which is not yet known. Our study establishes whole-genome sequencing as an unbiased method for discovering cancer-initiating mutations in previously unidentified genes that may respond to targeted therapies.

Author Contributions T.J.L. and R.K.W.: project conception and oversight. T.J.L. and E.R.M.: project leaders and analysis coordination. L.D.: supervised variant discovery and characterization, decision tree analysis. D.E.L.: decision tree analysis development. S.S.: automated variant detection by decision tree analysis. B.F.: variant validation oversight. B.F., P.M. and D.G.: Consed multiple sequence viewer development/programming. M.D.M.: auto-analysis and manual review of validation data. K.C.: copy number analysis, variant detection algorithm development. D.C.K.: indel detection algorithm development. K.C. and L.W.H.: indel detection. D.D.: IT and data management, data analysis automation leader. B.H.D.-S.: variant detection algorithm development. S.M. and M.T.: library optimization and construction. L.C.: data generation scheduling and oversight. R.A. and T.M.: variant validation assays. X.S.: variant annotation pipeline development. D.E.L.: variant annotation. J.R.O.: variant data management and pfam analysis. A.H.: validation assay design. C.P.: LIMS (Laboratory Information Management System) oversight. S.A.: LIMS trouble shooting/facilitation of variant detection. D.L.: data analysis. L.F.: production data oversight. T.W. and J.G.: data analysis algorithm development. V.M.: next-generation platform development. J.C. and N.S.: primary next-generation data production. A.C.: analysis oversight for mutation discovery. Y.Z.: manual review of sequence variants. R.E.R. and M.J.W.: comparative genomic hybridization analyses. R.E.R.: cDNA expression analyses. J.E.P.: gene expression array analysis. P.W., M.W., J.I. and S.H.: clinical data and specimen acquisition/processing/management. R.N.: bioinformatic analysis. J.B. and W.D.S.: statistical analysis. P.W., M.H.T., T.A.G., J.F.D. and D.C.L.: study design, execution and analysis. T.J.L., E.R.M., D.D., D.L., L.W.H., P.W., M.H.T., D.C.L., T.A.G., J.F.D. and R.K.W.: manuscript preparation.

Flow Chart of Filters used to identify somatic point mutations in AML tumour genome

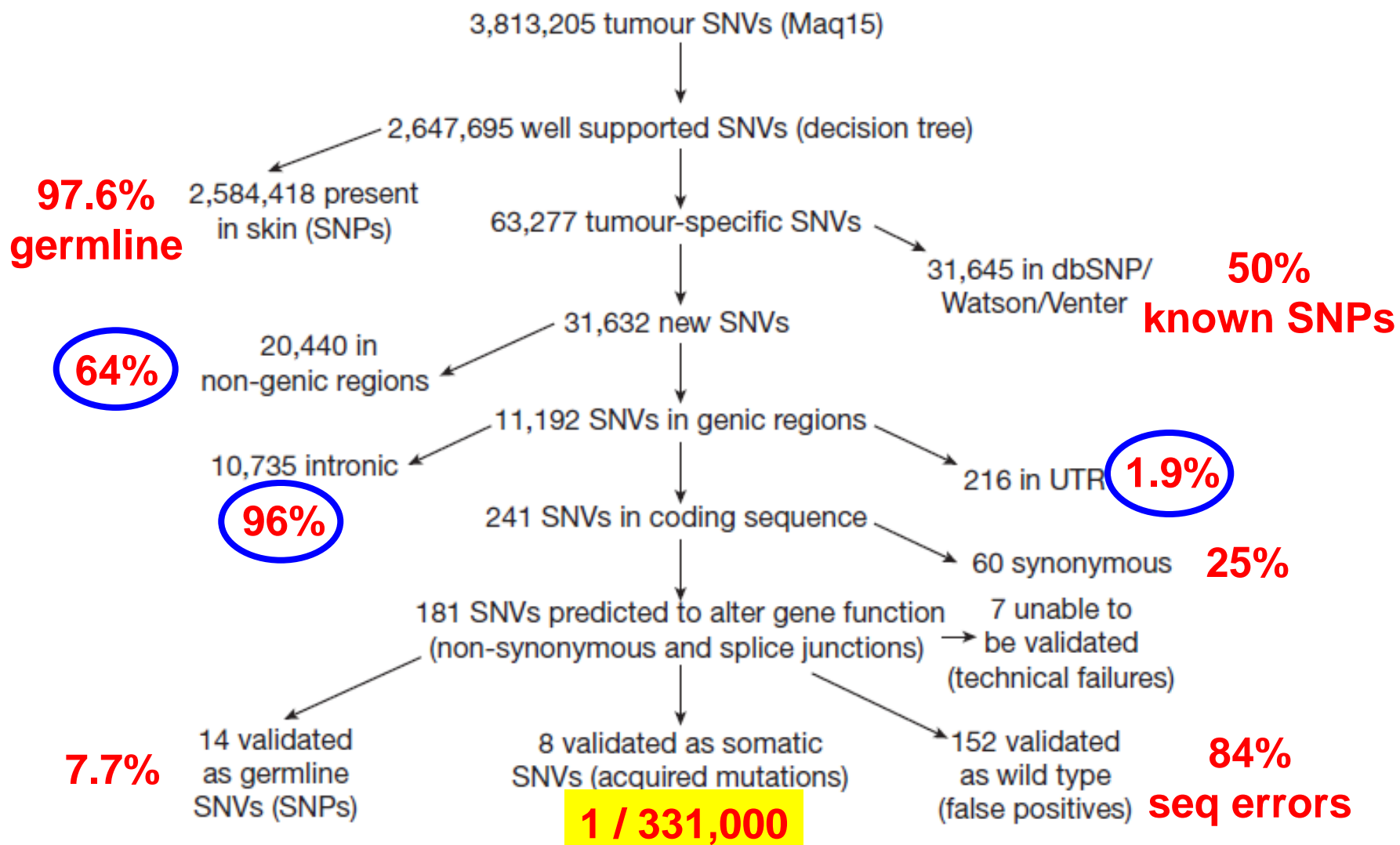


Figure 2 | Filters used to identify somatic point mutations in the tumour genome. See text for details. UTR, untranslated regions.

When shouldn't you have your genome sequenced or... over interpreted?

Table 1 Direct-to-consumer marketers of genetic information

Company	Technology	Cost
23andMe	Illumina HumanHap550+ BeadChip, oligonucleotide bead arrays in microwells on bundled optical fibers for detecting 550,000 SNPs plus 30,000 additional SNPs selected by 23andMe	\$399
deCODEme	Illumina Human1M BeadChip, for detecting >1 million SNPs	\$985 introductory price
Knome	Illumina genome sequencing platform	\$350,000
Life Code	Undisclosed SNP genotyping arrays	€1,200 to €2,400
Navigenics	Affymetrix Genome-Wide Human SNP Array 6.0, a photolithographically synthesized oligonucleotide chip for detecting 600,000 SNPs	\$2,500 plus \$250 per year "for continuous service"

When shouldn't you have your genome sequenced or... over interpreted?

An agenda for personalized medicine

Pauline C. Ng, Sarah S. Murray, Samuel Levy and J. Craig Venter find differences in results from two direct-to-consumer genetics-testing companies. They therefore give nine recommendations to improve predictions.

More than 1,000 DNA variants associated with diseases and traits have been identified^{1,2}. Direct-to-consumer (DTC) companies are harnessing these discoveries by offering DNA tests that provide insights into personal genetic traits and disease risks. Genetic testing can improve lifestyle choices and increase preventive screening³. However, understanding of the genetic contribution to human disease is far from complete.

There is debate in the genetics community as to the usefulness of DTC testing. Therefore,

Two other major concerns are whether the predicted disease risks have any clinical validity, and how well a genetic variant correlates with a specific disease or condition⁴. A few individuals have alluded to getting different predictions from different DTC companies for the same disease^{5,6}. We compared the consistency of disease-risk predictions between the two DTC companies to see where differences may arise (see Table 1).

Both companies report absolute risk, which is the probability that an individual

SUMMARY

- For seven diseases, 50% or less of the predictions of two companies agreed across five individuals
- Companies should communicate high risks better and test for drug response markers
- Community should study markers in all ethnicities and look at behaviour after tests

When shouldn't you have your genome sequenced or... over interpreted?

TABLE 1: PREDICTIONS FOR DISEASE RELATIVE RISKS FOR FIVE INDIVIDUALS					
Disease	Female A	Female B	Female C	Male D	Male E
Breast cancer	↑↑	↑↑	↓↓		
Coeliac disease	↓↓	↓↓	↓↓	↓↓	↓↓
Colon cancer	==	==	=↓	↑↑	=↓
Crohn's disease	↓↑	↓↑	↓↓	↓↓	↓=
Heart attack	↓↓	=↓	=↓	=↓	↑↑
Lupus	↑↓	↓↓	↓↓	↑=	↑=
Macular degeneration	↓↓	↓↓	↑=	↓↓	↓↓
Multiple sclerosis	↑↑		↓↓	↓↓	↓↓
Prostate cancer				↑↑	↓↑
Psoriasis	↓↑		↑↓	↑↑	↓↓
Restless legs syndrome	=↓	↑↑	↓=	↓↑	↑↑
Rheumatoid arthritis	↑↑	↑↑	↓↓	↓↓	↑↑
Type 2 diabetes	↓↓	=↓	↓↓	↑↓	=↓

↑ increased risk (RR > 1.05), ↓ decreased risk (relative risk (RR) < 0.95), = average risk (0.95 ≤ RR ≤ 1.05). First prediction is from 23andMe; second prediction is from Navigenics. Different predictions are highlighted in beige.

When shouldn't you have your genome sequenced or... over interpreted?

Vol 461|8 October 2009|doi:10.1038/nature08494

nature

REVIEWS

Finding the missing heritability of complex diseases

Teri A. Manolio¹, Francis S. Collins², Nancy J. Cox³, David B. Goldstein⁴, Lucia A. Hindorff⁵, David J. Hunter⁶, Mark I. McCarthy⁷, Erin M. Ramos⁵, Lon R. Cardon⁸, Aravinda Chakravarti⁹, Judy H. Cho¹⁰, Alan E. Guttmacher¹, Augustine Kong¹¹, Leonid Kruglyak¹², Elaine Mardis¹³, Charles N. Rotimi¹⁴, Montgomery Slatkin¹⁵, David Valle⁹, Alice S. Whittemore¹⁶, Michael Boehnke¹⁷, Andrew G. Clark¹⁸, Evan E. Eichler¹⁹, Greg Gibson²⁰, Jonathan L. Haines²¹, Trudy F. C. Mackay²², Steven A. McCarroll²³ & Peter M. Visscher²⁴

When shouldn't you have your genome sequenced or... over interpreted?

Table 1 | Estimates of heritability and number of loci for several complex traits

Disease	Number of loci	Proportion of heritability explained	Heritability measure
Age-related macular degeneration ⁷²	5	50%	Sibling recurrence risk
Crohn's disease ²¹	32	20%	Genetic risk (liability)
Systemic lupus erythematosus ⁷³	6	15%	Sibling recurrence risk
Type 2 diabetes ⁷⁴	18	6%	Sibling recurrence risk
HDL cholesterol ⁷⁵	7	5.2%	Residual* phenotypic variance
Height ¹⁵	40	5%	Phenotypic variance
Early onset myocardial infarction ⁷⁶	9	2.8%	Phenotypic variance
Fasting glucose ⁷⁷	4	1.5%	Phenotypic variance

* Residual is after adjustment for age, gender, diabetes.

When shouldn't you have your genome sequenced or... over interpreted?

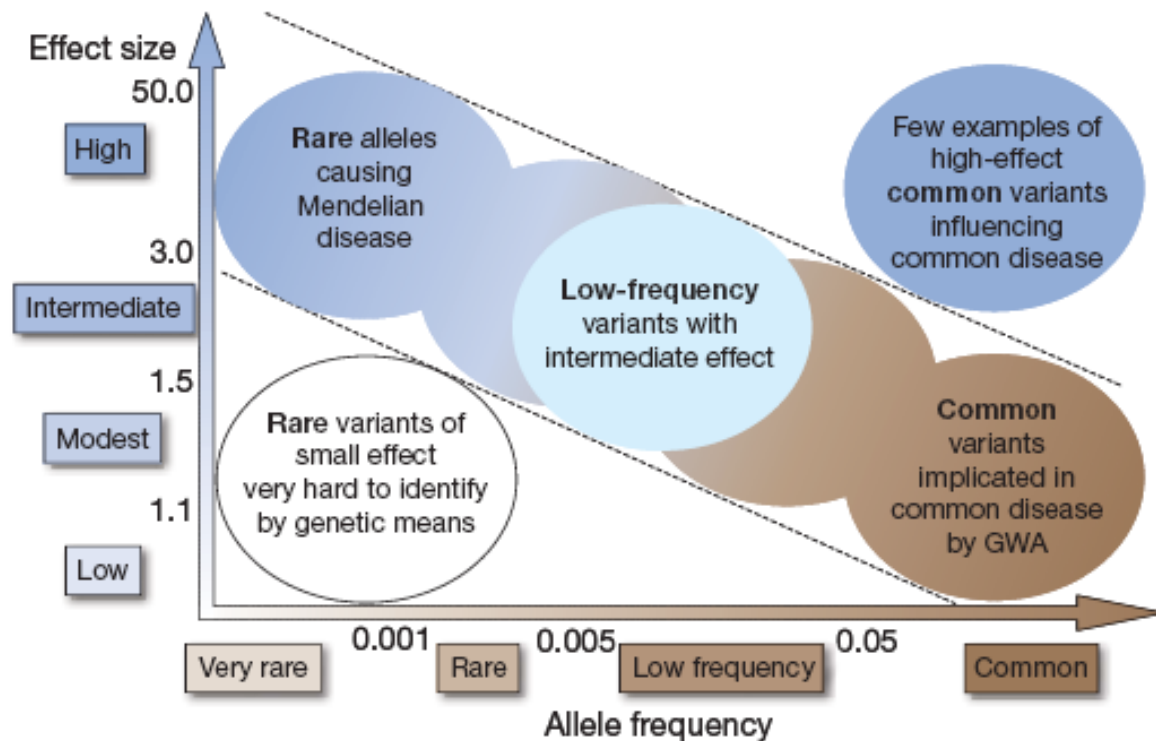


Figure 1 | Feasibility of identifying genetic variants by risk allele frequency and strength of genetic effect (odds ratio). Most emphasis and interest lies in identifying associations with characteristics shown within diagonal dotted lines. Adapted from ref. 42.