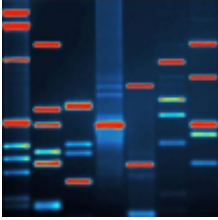


Personal Genome Revolution

Melanie Swan, MS Futures Group, m@melanieswan.com, 415-505-4426
June 14, 2009; http://melanieswan.com/documents/Genomics_Revolution.pdf



Genomics is making faster progress than any other technology field in recent history. Usually the vista from any point on an exponential curve looks flat to the experimenter but not so with consumer genomics, the field is exponentiating from any vantage point. Genomics scientific research and commercialization issues were discussed with excitement at the first-ever [consumer genomics conference](#) in Boston, June 9-11, 2009.

Image credit: <http://www.geekologie.com/2008/01/28/dna-art.jpg>

Summary

1. Advent of whole human genome sequencing: Automatic whole human genome sequencing of all individuals could be a reality within the next few years.
2. Medically actionable now: Genetic data is medically actionable now and becoming increasingly so, particularly in routing higher-risk individuals into earlier screening. It is estimated that each individual is in the upper 5% risk tier for at least one chronic disease.
3. New era of information and communications technology: Genomic data requires a significant new level of information processing, storage and transfer. One whole human genome can range from 6GB-8TB in terms of the data currently transferred between researchers. Genomic data is growing at 10x per year vs. Moore's Law's 1.5x per year.
4. Social inevitability: Widespread genomic sequencing appears to be inevitable which has great benefits together with social challenges such as revealing non-paternity ([commonly thought to be 10%, but may have a median of 3.7%](#)), terminal disease conditions and reproductive issues (e.g.; recessive carrier status).
5. Heightened role of the consumer: Consumers will have unprecedented access to health information about themselves and could take a much more active and self-directed role in their health management, more likely responding favorably than being consumed with their '[incidentalome](#).'

Genetic tests – what is now available

Physician-ordered tests (generally insurance-reimbursed)

For some time, physicians have been ordering any number of one-off genetic tests for specific conditions such as Cystic Fibrosis, Huntington's Disease, breast cancer (mutations in the BRCA1 and BRCA2 genes) and other conditions. Physicians can also order any of the below tests for patients.

Consumer-ordered tests (no doctor-order required, unreimbursed)

For about the last year and half, consumers have been able to order their own genetic tests directly without a doctor's order and several thousand have done so from the offerings below.

- Single condition tests ([DNA Direct](#), \$200-1,000)
- SNP ([Single-nucleotide polymorphism](#) – an allele variation from others) risk assessment tests ([23andme](#) (\$399, down from \$1,000), [DeCODEme](#) (\$985), [Navigenics](#) (\$2,499))
- Whole genome scan ([Knome](#) (\$99,500)) or whole [exome](#) scan ([Knome](#) (\$24,500)) [The price just decreased from \$350,000 to \$99,000, but a few more zeros might drop off within months from this provider or others]
- [Personal Genome Project](#) (PGP), Harvard Medical School, genome sequencing for free in exchange for open data publishing; now expanding from ten subjects to 100,000
- Family planning genetic screening: [Counsyl](#)
- Mate compatibility analysis based on immune system variation: [ScientificMatch](#), [GenePartner](#) (One next obvious step would be including recessive disease carrier status, for example in the back-end matching algorithms of dating services)

**Out of work due to technological advance:
elevator operator, stock broker, physician(?)**



Images: <http://www.discoversandiego.com/features/anderson/images/2003/08/movieland09.jpg>
<http://www.russiablog.org/StockBroker-MosNews.jpg>
<http://blog.soliant.com/wp-content/uploads/doctor-physician.jpg>

Current genomic testing issues: validity and utility

Validity

There are differing levels of data validity depending on which chip array and methodology is used to sequence the genomic data. [Illumina](#) reports being at two 9s now (e.g.; 99.99% error free; experiencing one error per 1,000 reads) and is hoping to move to four and then six 9s of quality. Sequencing is executed at different levels of coverage ranging from 1x to 30x coverage, meaning how many times a sequence is read; 30x coverage is the most accurate and highest industry standard at present.

A few people who have tried multiple DTC (direct-to-consumer) SNP chip offerings have found consistent genotyping data (e.g.; having a ‘CT’ at a certain SNP), but different interpretations in lifetime risk probabilities as diverse markers are evaluated and aggregated into risk assessments across the companies. There is significant risk of false negatives and false positives.

**Direct-to-consumer genomic testing companies:
Heterogeneous breast cancer markers assessed**

Navigenics	DeCODEme	23andme
chr8.128424800 [rs13281615, rs2981582 and rs3803662]	16q12 / rs3803662	rs3803662
TNRC9 [rs3803662]		
chr2.217614077 [rs13387042]	2q35 / rs13387042	
MAP3K1 [rs889312]	5q11 / rs889312	
	10q26 / rs1219648	rs1219648
FGFR2 [rs2981582]		
CASP8 [D302H variant (rs1045485) and the -652 6N ins/del promoter variant (rs3834129)]		
LSP1		
	5p12 / rs4415084	
	8q24 / rs13281615	
	11p15 / rs3817198	
		185delAG BRCA1 mutation
		5382insC BRCA1 mutation
		6174delT BRCA2 mutation

Sources: [Navigenics](#), [DeCODEme](#), [23andme](#)

Not only do different services map different markers to meta conditions like cardiovascular disease, but the most relevant medical SNPs are often not included in DTC SNP chips, possibly due to patent and cost issues. A notable example is [Myriad](#), which owns patents on the breast cancer-related BRCA1 and BRCA2 genes. This has become the focus of a timely [lawsuit brought by the ACLU](#) regarding the patentability of natural materials such as genes and industry norms of how genes are licensed for diagnosis and therapy.

Whole human genome sequencing renders the patented-gene issue moot as anyone having access to their raw data could look up their genotypes for particular SNP/rsid

numbers such as those corresponding to the BRCA1 and BRCA2 genes. ([Knome](#) customers can do this now). There will be a need for interpretation tools appropriately aggregating multiple risk alleles. Fee-based or open source genomic data interpretation tools like the [SNPedia's Promethease report](#) could proliferate.

Utility

People would like to know definitively if they are going to have a disease but aside from monogenic conditions (for example, Muscular Dystrophy, Huntington's Disease, sickle cell disease and Cystic Fibrosis), most chronic diseases are polygenic and influenced by many factors. The current genetic testing for these conditions does not deliver a simple Yes/No, but rather assesses the lifetime risk probability for an individual and whether the individual is at higher or lower risk than the average.

There is ample room for risk interpretation mechanisms for polygenic conditions to become more sophisticated, right now the practice is a multiplicative technique, taking the risk value for each genotyped allele associated with the condition and multiplying them together; weighting and cluster-evaluation could be refinements that research may support over time.

Genetic variation and disease causality

[NHGRI](#) (National Human Genome Research Institute) and other [GWAS](#) (genome-wide association studies) researchers find that genes, as they have been studied so far, only account for a small percent of explaining disease. However, studies have been preliminary, the 1,000 genomes studied may not be enough for complete understanding, for example, about 35 common diseases have been found to have widely replicated common variants. One next step targeted by the NHGRI is to look at rare variants, low-frequency (e.g.; 1-2%) GWAS variants with intermediate penetrance, to possibly explain a larger percentage of disease causality. Simultaneously, our systemic understanding of biology is slowly improving, it seems that in many disease cases it may not be the gene or genotype, but rather the number of copies of the same gene ([CNVs](#)), translocations, inversions, and other problems with gene expression and DNA repair that are responsible for disease.

Knowledge gap

Genomic technology has been moving so fast that at present, most physicians do not have genetic training. The genetics community is the primary party helping to generate, interpret, present and monitor genomic data. Over time, other communities such as physicians and genetic counselors (one of the world's fastest-growing job categories) will hopefully become helpful in interpreting data together with patients. Genetic training is a key target area of CME (continuing medical education), for example the [National Coalition for Professional Education in Genetics' "Genetics Education for Health Professionals: What are the Key Messages? How do we deliver them?"](#) (Sep 2009) and Harvard Medical School's "[What the Primary Care Provider needs to know about the Genetic Basic of Adult Medicine](#)" (Oct 2009).

Medical relevancy

That disease has a molecular basis is now undisputed and medicine is slowly shifting to

reorganize around this. Presently, about 1,400 genes can be tested to inform various clinical decisions and 225 are deemed clinically significant. 100 new tests are being added annually. In some cases, medical information exists but is not being used, for example a straightforward marker for poor drug metabolizers, [CYP2D6](#). About 10% of Caucasians are poor metabolizers however this is not routinely tested for ahead of time (nor in the DTC SNP chip tests mentioned above) and the same drugs are given to all patients in a trial and error process, [sometimes in lower doses \(e.g.: warfarin\) due to fear of overdosing those for whom it could be harmful](#).

Another example of medical relevancy in genomic testing is the NHGRI's GWAS study finding of the first nine genetic risk variants for type 2 diabetes: TCF7L2, IGF2BP2, CDKN2A/B, FTO, CDKAL1, KCNJ11, HHEX/IDE, SLC30A8 and PPARG; particularly the first one, [TCF7L2](#). Higher-risk individuals identified early in life could receive targeted healthcare.

Additive statistical approach

So far, general genomic testing suggests that on average, each patient is in the upper 5% risk tier for at least one chronic disease (e.g.; cancer, cardiovascular disease, myocardial infarction, etc.) and that there is value in understanding genomic risk factors earlier in life. Whole human genome sequencing automatically at birth could mean a lifetime of personally relevant healthcare.

Although genomic tests do not predict polygenic disease definitively, they are medically actionable in taking conventional risk percentages (e.g.; American female lifetime breast cancer risk = 12%; American male lifetime prostate cancer risk = 16%) and layering on the specific genetic risk of the individual to route higher-risk individuals to screening and therapeutics earlier. Several researchers estimate that the earlier identification of higher risk patients could reduce overall healthcare costs by about ~\$100,000 per person per condition.

Patient behavior: a key component of medical actionability

Although there is no known cure for Alzheimer's Disease, and even a firm diagnosis can only be made at autopsy, [Boston University's REVEAL study](#) has shown that people change their behavior after receiving a positive diagnosis for Alzheimer's Disease (mainly through purchasing supplements and some increase in exercise). It is also known that mid-life cholesterol levels correlate with Alzheimer's Disease, so the highly actionable behavior for someone with an [APO E4 positive allele](#) could be more closely managing cholesterol intake.

Family history

The role of family history is another important component of disease prevention, diagnosis and management, and there are starting to be helpful web-based tools for consumers to assemble, manage and access family history data such as [My Family Health Portrait](#).

Technology status

Technology advance has been the key enabler of the genomics revolution. The first genome sequencing project, completed in 2003 cost \$3b. Now, the cost of genetic sequencing is dropping to the point where a \$100 whole human genome may be available in the next few years, in 2010 according to [Pacific Biosciences](#). There are several next-generation sequencing platforms in process now that could supersede the current array-based method.

Next-generation sequencing platforms

Next-generation genomic sequencing platforms are generally falling into two categories, those using synthesis (specifically, multiplex cyclic sequencing by synthesis) and those not using synthesis. Some of the most interesting next-generation companies using synthesis are [Pacific Biosciences](#), [Ion Torrent Systems](#) and [RainDance Technologies](#). Some of the most exciting non-synthesis-based next-generation sequencing companies are [Oxford Nanopore Technologies](#), and [NABsys](#) and [Halcyon](#). NABsys and Halcyon are electromagnetically-based rather than optically-based which means they are not dependent on light or fluorescence so the cameras can go much faster, perhaps 10,000 frames per second. Harvard Medical School maintains a nice [overview of current and emerging gene sequencing technologies](#).

Transcriptome, proteome, metabolome, microbiome...

In addition to improving the cost and speed of existing genomic scanning, sequencing advances could open up the way to the eventual characterization of the whole cell and its interactions through the sequencing of the [transcriptome](#), the [proteome](#), the [metabolome](#), the [microbiome](#) and other biological features. In the farther future, histone modification sequencing, DNA methylation, acetylation and phosphorylation are other characterization processes of interest that could be included.

Petabyte data era: processing, storage and transfer challenges

The biggest challenge consuming national genomic research labs at present is data processing and network communications. Genomic data is growing at 10x per year (vs. Moore's Law growing at 1.5x per year). Research labs have problems with data storage, mapping and access, together with intra-site data transfer and external transfer. Shipping terabyte drives via fedex is the best current data transfer method, and at least one lab finds resequencing data cheaper than storing it.

The raw data of the 6b base pair whole human genome is 6GB, not challenging to store, but challenging to work with, it is not like just opening up and manipulating a word document. New data processing algorithms will need to be developed to interact with whole genome data, link it to reference tools and make it searchable and meaningful. Whole businesses can be formed to focus on genomic data curation alone (another wind for Google?).

Even though the most basic raw data version of the whole human genome is 6GB, the full collection of files in use by researchers for one whole human genome may reach 8TB. The large dataset may include an intensity file, a BAN file (binary), a SAN file (searchable) and other files with coordinates, variations and other details. Part of the

challenge is that appropriate data abstractions from the raw sequencing output are not yet known so all of the data is kept. There is not yet a good reference model. Apparently, the [Archeron X-Prize for genomics](#) (sequencing 100 human genomes within 10 days or less at a maximum cost of \$10,000 per genome) remains outstanding not because it cannot be done, but because the results cannot be recapitulated.

Testing inevitability and social implications

It seems quite possible that initial and ongoing whole human genome sequencing (and eventually, on-demand proteome, metabolome, microbiome, etc. sequencing) would be a routine component of everyone's EHR (electronic health record) available to both patients and physicians for ongoing predictive, preventive healthcare modeling and monitoring. There are some important social implications of widespread whole human genome testing, for example:

Non-paternity

One genetic issue is non-paternity ([commonly thought to be 10%, but may have a median of 3.7%](#)). In the era of whole human genome sequencing, paternity would be quite easy to trace. One possible impact is that the divorce rate could increase and single mothers could be stratified into lower economic tiers.

Right not to know

Another genetic issue is that of a [person's right not to know about their medical situation](#). With improving remedies, the right not to know could decrease in importance. It may be straightforward for practitioners to deliver healthcare without breaching the patient's right not to know their genetic information as they do currently. With more actionable treatments and increasing health literacy, it could become the social norm to know one's genetic profile, to learn about potential conditions and work collaboratively with others with similar conditions in attempts to mobilize long-tail medicine, as [PatientsLikeMe](#) health social network participants are doing to [run their own clinical trials](#).

Discrimination

[GINA, the Genetic Information Nondiscrimination Act of 2008](#), protects U.S. citizens from discrimination by employers and insurance companies. It is a step in the right direction, but many are not reassured. The law has holes, such as not covering long-term care providers, and will have to be strengthened via interpretation as real-life cases arise.

DNA Forensics – Gattaca?

In an age of inexpensive genomic testing, the on-demand testing of other people (such as a prospective mate, business partner, supervisor or tenant), as portrayed in the movie [Gattaca](#), could easily occur; one such example provided [decisive evidence in a recent divorce case](#). DNA privacy would become impossible as a practical matter. However, precisely because everyone would be subject to genetic openness and since the present world is not one of scarcity and control as the dystopian Gattaca, it may be that DNA testing and knowledge would not be a substantive issue. Already, several individuals in support of hastened scientific advance and open medicine have [open-sourced their genomic data on the SNPedia](#) or via the [Personal Genome Project](#).

Venture capital investment opportunities

There are many exciting potential opportunities for venture capitalists, entrepreneurs and researchers in helping to realize the genomics revolution. The money is already arriving before the physicians as companies, backed by varying degrees of research, seek to monetize genetic risk. The potential demand for personal genomic products and services could be enormous, for example, the market for weight-loss products is \$40b/year. Here are some potential opportunities:

- Personalized genetic testing, counseling, supplements and other action programs and remedies, for example, [Inherent Health's Weight Management, Heart Health and other tests](#), and the [APO E Gene Diet](#).
- More DTC (direct-to-consumer) genetic testing and interpretation offerings stratified towards differing enduser tiers (e.g.; the aggressive early adopter, the lay person, the Boomer, the Gen Y'er)
- A line of genomic testing services to be offered by spas, concierge doctors, private clinics and early adopter physicians; possibly positioned as a luxury item vs. a medical necessity to accelerate adoption
- Next-generation sequencing, and next-next- generation sequencing, innovating the technology and the applications to commercialize the technology
- Web-based tools for integrating medical records, family history and genomic data, facilitating data collection, entry and access
- Genetic literacy products and services for physicians and consumers
- Web-based tools to appropriately and dynamically aggregate multiple risk alleles into chronic disease meta conditions such as cancer and cardiovascular disease
- Fee-based genomic data interpretation tools like the [SNPedia's Promethease](#)
- Data processing algorithms to interact with whole genome data, making it searchable and meaningful with links to external reference databases
- Genomic data curation and cloud computing for genomic data analysis
- Health social networks or other tools for deep longitudinal monitoring by consumers/patients over time of many complex health factors

Conclusion

As our molecular understanding of disease progresses and genomic testing continues to decrease in cost and become increasingly medically relevant, adoption could become extremely widespread almost overnight. Physicians could start to see the additive, precise information conferred by genomic testing as a means of improving the care they now deliver, finding themselves initially encouraged and eventually forced into the genomic revolution. Pharmaceutical companies could start to use genomic testing as a means of improving efficacy in drug discovery and delivery, providing much-needed assistance to their ailing cost models. Consumers could be radically empowered to become curious about and responsible for self-managing their health with automated easy-to-use tools. Genomics as an enhanced approach to healthcare could transform the quality of life worldwide for all humanity.

Melanie Swan, the principal of [MS Futures Group](#), is a science and technology researcher and hedge fund manager based in Palo Alto, CA.