PRECISION MEDICINE IS COMING

PATIENTS ARE WAITING, WHAT DO WE NEED TO DO TO GIVE IT TO THEM?

WE'VE BARELY BEGUN TO SEE WHAT'S POSSIBLE
Jonathan Bingham talks us through the Google Genomics vision for future research.

NGS: HOW AND WHY WE GOT HERE
Explore the young, but rich, history of Next Generation Sequencing, and know which option works best for you.

THE VIEW FROM THE TOP
We speak with Jean-Claude Marshall, Director of Pfizer's Clinical Pharmacogenomics Laboratory, about genomics in Big Pharma and FDA regulations.
We never stop seeking.

We are driven to know more—to relentlessly search for the answers that will advance the understanding of genomics to improve human health. And we realize we can’t do it alone. We’re counting on the next generation of scientific minds to help us keep up the momentum. As the leading developer of life science technologies and services, we help accelerate genetic research and its use in the fields of cancer, hereditary disease, reproductive health, infectious disease, forensics, and agriculture. Together, we’ll realize the promise of personalized medicine.

www.Illumina.com/vision
Here have been some big changes since we launched the magazine in December. It’s always staggering to witness how quickly things progress in and around genomics. To mention a couple of major milestones: there was the mapping of the epigenome, Obama’s State of The Union Address, mitochondrial donation approval in the UK, 23andMe opening up their database to pharma and receiving FDA approval, the 100,000 Genomes Project entered its main phase... A lot, that may hold a significant place in history when we look back in years to come.

That’s not to say we can start popping the champagne just yet. There is still a lot of work needed on standards. I think this is something that most people recognise, and work is obviously being done. But it is still something that is going to require a lot of input, and a strong driving force, to keep pushing for something actionable. Hopefully this gets some much needed attention.

With genomics constantly creeping closer into the patient domain, there is remarkably little going on to help support and develop genetic counsellors. As a profession they have an absolutely essential role in ensuring that patients understand and support genomic research.

We’ve been overwhelmed by the positive feedback you’ve given us following last year’s launch. Thank you. It’s been very useful in helping to define our future direction. All of us here at FLG are working on projects we have a real passion for, so hearing directly from those we aim to serve is always beneficial and a real pleasure.

In this issue, we have great contributor pieces from Shea Robinson and Richard Wintle on how epigenetics is changing our understanding of cancer and how sequencing technology continues to develop for the field. We also have some fantastic interviews with Gholson Lyon, David Stone, Jean-Claude Marshall, Jonathan Bingham and Alaistair Kent OBE. As a collection of interviews, these cover a great deal. Everything from life in Big Pharma, the need for standardisation in research, integrating genomics into healthcare, and what the future of genomics holds. I hope you enjoy reading them as much as I enjoyed putting them together!

There’s a lot more for us to do here at FLG, but we can’t do it without you. We’re going to be looking at what we can to do to really help drive our social mission in the coming months. There are some very passionate people out there doing their best to help deliver the benefits of genomics to patients faster. And that’s whom we really want to help. If you’d like to get involved, I would love to hear from you.

All that remains to be said is – I hope you enjoy reading our first issue of 2015!

---

**About Us**

Our mission is to help bring the benefits of genomics to patients faster.

To achieve this, we work with the smartest people and organizations to produce a series of events and a free-to-access web portal & magazine for the genomics community.

Our products are designed to support scientists, clinicians, business/research leaders and officials, from academia, research institutes, industry, healthcare and government organizations.
WELCOME
Managing Editor, Carl Smith, reflects on the major developments between issues and outlines key areas of focus at FLG.

MAKING THE MAGAZINE
Find out who helped put the magazine together and who contributed to this issue.

GENOMIC LENS
This issue’s photo comes from Genetic Alliance UK’s Patient Charter launch at the Wellcome Trust in London.

ROUND UP
We look at President Obama’s State Of The Union Address and the Precision Medicine Initiative.

ROUND UP BUSINESS & FUNDING
Assessing 23andMe’s industry partnerships and recent FDA approval.

ROUND UP SCIENCE RESEARCH
Looking at the successful mapping of the epigenome, and mitochondrial donation laws passed in the UK.

OUR PICK FROM THE WEB
Reviewing the timely interview of the Personalized Medicine Coalition’s Amy Miller on The Bio Report.

INTERVIEW
The View From The Top – We interview Jean-Claude Marshall, Director of Pfizer’s Clinical Pharmacogenomics Laboratory. He outlines the current state of pharmacogenomics and assesses the impending FDA regulations on LDTs.

COMMENT
The Next Frontier For Cancer Research – With the newly mapped epigenome and new research being published, our understanding of cancer could increase dramatically in the near future.

INTERVIEW
Cut The Hype. Accuracy And Standards Come First – Gholson Lyon, gives us his view on the dangers of political hype and the benefits of developing a broad research background. He also rallies for much needed standards and improved accuracy in research.
**INTERVIEW**

From High School Biology To Genomic Pioneer – Merck’s David Stone, shares his remarkable journey from high school biology teacher to playing a pivotal role in big pharma.

**INTERVIEW**

Giving Patients A Voice – As large scale sequencing projects progress, the role of ‘patient-power’ is beginning to be recognised. But what is patient-power and how is it being used to effect change?

**EVENTS**

A preview of what to expect at The Festival of Genomics this summer.

**INTERVIEW**

We’ve Barely Begun To See What’s Possible - Google has had a massive impact on how we do business and how we live our lives. Google Genomics is aiming to do the same for the field of genomics, and help us ask much bigger questions.

**REVIEW**

Genetics Goes Mainstream – 1997’s GATTACA is a cult classic in modern film history, but is now more relevant than ever after recent developments in healthcare.

**FIRESTARTER**

The Genetic Portrait Project – Genetic information can illicit an emotional response due to the personal information involved. Imaging artist, and former Biologist, Stefan Petranek explains The Genetic Portrait Project.

**COMMENT**

NGS: How And Why We Got Here – Next Generation Sequencing technology has advanced at a startling pace over the last few years. What options are available today and which one should you choose?

Sign up and become a Front Line Genomics member. It’s completely free of charge. Receive exclusive content and discounts on event registration.

For full list of benefits and sign up, visit www.frontlinegenomics.com/become-member
This is our first print issue of the magazine. It added another aspect to the design process: the look and feel of the magazine. Getting the design element for our first issue was critical, so we went back to many of the same sources to try make sure that the physical version of the magazine was consistent with the look and feel that digital version has. Suddenly terms such as ‘perfect bound’ and ‘saddle stitched’ were being thrown around the office, as well as talk of the ideal GSM for the pages.

The content itself was once again a collaborative effort from various people. We are very much indebted to all the people who have been kind enough to talk to us on the phone, step up to write and join in for interviews. Above all, we are very grateful to our advisory board who have all been inspiring, patient and incredibly helpful as always.

**ADVISORY BOARD**

Jean-Claude Marshall  
Director, Clinical Pharmacogenomics Lab  
Pfizer

David Smith  
Professor of Laboratory Medicine & Pathology and Chair of Technology Assessment Committee for the Center for Individualized Medicine  
Mayo Clinic

Gholson Lyon  
Assistant Professor  
Cold Spring Harbor Laboratory

**CONTRIBUTORS**

Dr Carl Smith / Managing Editor  
Carl looks at the growing social implications of preventative medicine in this issue’s movie review of the 90’s cult classic GATTACA. For this issue’s pick of the web, he also reviews Amy Miller’s guest appearance on The Bio Report, giving the Personalized Medicine Coalition perspective on the Precision Medicine Initiative.

Shea Robison / The Nexus Of Epigenetics  
Shea considers the limitations of genomic information for cancer research. With new advances in epigenetics, and the mapping of the epigenome, new information could be set to greatly improve our understanding of this most complex group of diseases. Shea takes a look at some of these developments in his critical look at cancer epigenetics.

Richard Wintle / The Center For Applied Genomics  
As Assistant Director at the The Center For Applied Genomics, Richard comes into contact with sequencing technology everyday. In this issue he takes us through the intriguing story of how cancer research has developed along with sequencing technology, and guides you through what to bear in mind as you consider which NGS platform to use for your own cancer projects.

Laura Rae / Festival Director  
Our Festival Director, Laura, interviewed Alastair Kent, Director of Genetic Alliance UK for this issue. They talk about the recently launched Patient Charter, the importance of including patients in major policy discussions, direct-to-consumer testing, and the changing healthcare landscape in the UK.
On the 11th of February, Genetic Alliance UK launched their Patient Charter for the integration of genomic sequencing into healthcare. Genetic Alliance UK, Director, Alastair Kent OBE (right) and Minister for Life Sciences, George Freeman MP delivered the opening addresses at the event held at the Wellcome Trust in London.
GENETIC ALLIANCE UK LAUNCH PATIENT CHARTER ON GENOME SEQUENCING

In February, Genetic Alliance UK launched their Patient Charter- a series of recommendations to the NHS on how to integrate genomics into public healthcare. The charter was compiled following extensive research with patients and their families.

Minister for Life Sciences, George Freeman MP, gave his support to the charter and the Genetic Alliance UK at the launch event. He reiterated the government’s commitment to genomic medicine and the importance of giving patients an active voice in this period of pioneering innovation.

The 15 recommendations are broken down into 4 main areas of focus: 1) Patients want the option to receive as much information about their health as possible; 2) Patients value genetic counselling and are keen for the support of genetic counsellors before and after genome sequencing; 3) Patients welcome the sharing of their genomic data for research purposes; and 4) Patients think that the NHS needs to make more progress towards preparing for the integration of genome sequencing into clinical practice.

PRESIDENT OBAMA UNVEILS GENOMIC PLANS IN STATE OF THE UNION ADDRESS

President Barack Obama’s State Of The Union Address, outlined the new Precision Medicine Initiative in January. Stating that his administration is committed to increasing the use of genetic information to help treat complex diseases like cancer. He strongly urged Congress to review research funding to support precision medicine.

“I want the country that eliminated polio and mapped the human genome to lead a new era of medicine – one that delivers the right treatment at the right time...”

The immediate goal of the Precision Medicine Initiative will be to greatly increase research efforts into developing treatments and preventions for more cancers. In collaboration with pharmaceutical companies, the project will support clinical trials to expand our understanding in these areas making use of latest technologies.
23ANDME PARTNER WITH DRUG DEVELOPERS AND GET FDA APPROVAL

3andMe have been making news for a variety of reasons since our last issue. Back in December, we reported that the direct-to-consumer testing company opened for business in the UK. In the past couple of months, the big news has been back in the USA. On the direct-to-consumer side, they received a huge boost with FDA approval for their Bloom Syndrome test.

The real story here isn’t so much 23andMe, but the FDA re-evaluating how it intends to regulate new types of genetic tests. Illumina, Thermo-fisher, Pacific Biosciences, and Oxford Nanopore are just some of the potential beneficiaries here. In short, the potential for higher-quality tests with faster approvals just increased significantly.

The other half of the 23andMe story over the past months is in their dealings with drug developers. Having built up a considerable database of genomic information, the company is now selling access to their anonymised data for research purposes. Pfizer and Genentech were the first two partners announced, with a total of 10 rumoured to follow shortly.

If 23andMe get back up to full scale direct-to-consumer testing on both sides of the Atlantic, there could be some very interesting times ahead for the company. Should Pfizer and Genentech make significant progress through their Research Portal, it might raise 23andMe to a very prominent position in the wider genomics community.

Anne Wojcicki, CEO & Co-Founder, 23andMe

Anne Wojcicki graduated from Yale University with a BSc in Biology. She developed her interest in molecular biology research at the NIH and UC San Diego. Following a 10-year spell as a healthcare investment analyst, she co-founded 23andMe.
UK PASSES LAWS ALLOWING MITOCHONDRIAL DONATION

The House of Commons and House of Lords have approved laws allowing mitochondrial donation. This makes the UK the first country in the world to legalise ‘three-person babies’. The fertility regulator will now decide how to license the procedure. This is a significant step forward in biomedical ethics and policy. The move gives hope to many families at risk of certain inherited diseases.

Despite vocal opposition on contentious ethical grounds, the vote passed with a strong majority both times.

DNA PHENOTYPING USED BY POLICE

Police in Columbia, S.C. used DNA phenotyping to issue a police sketch in an ongoing murder case. This is not the first instance of DNA phenotyping being used in a police case, but is one of the highest in profile to date. It is a technique that is growing in popularity, along with other applications of forensic genomics. The Toronto Police Service took advantage of this to gain new insights into several ‘dead end’ cases thanks to the technology provided by Identitas and their partner AKESOgen.

“LOOKING AT THE CHEMICAL SIGNATURES RESULTING FROM VARYING GENE EXPRESSION, CAN HELP INFORM HOW THE GENOME AND EPIGENOME MIGHT LEAD TO CONDITIONS SUCH AS ALZHEIMER’S DISEASE, CANCER, ASTHMA, AND FOETAL GROWTH ABNORMALITIES”

EPIGENOME MAPPED FOR OVER 100 CELL AND TISSUE TYPES

Researchers, supported by the Roadmap Epigenomics Program, have mapped the epigenomes of more than 100 types of cells and tissues. This has the potential to provide valuable insights into how some diseases develop.

The description of the maps were published in Nature, with over 20 additional papers published on how they can be used in further research. In particular, looking at the chemical signatures resulting from varying gene expression, can help inform how the genome and epigenome might lead to conditions such as Alzheimer’s disease, cancer, asthma, and foetal growth abnormalities.

“This represents a major advance in the ongoing effort to understand how the 3 billion letters of an individual’s DNA instruction book are able to instruct vastly different molecular activities, depending on the cellular context,” said NIH Director Francis Collins, M.D., Ph.D. “This outpouring of data-rich publications, produced by a remarkable team of creative scientists, provides powerful momentum for the rapidly growing field of epigenomics.”

“The NIH Roadmap Epigenomics Mapping Consortium is committed to producing a public resource of human epigenomic data to support basic biology and disease-oriented research. The Consortium is also developing and disseminating standardised protocols, reagents and analytical tools to help researchers integrate this body of data into their work.
WILL FUNDING FOR OBAMA’S PRECISION MEDICINE INITIATIVE MATCH ITS AMBITION?

Our favourite podcast for this issue comes from ‘The Bio Report with Daniel Levine’. Shortly after President Obama unveiled the details of his $215 million Precision Medicine Initiative, Daniel spoke with Amy Miller, Executive Vice President of The Personalized Medicine Coalition. The President’s announcement was important news for many of us, but in particular for those at the Personalized Medicine Coalition. Launched in 2004, the Coalition is a non-profit education and advocacy organisation, which educates the public and policy makers. Given the focus on patient-power, and building standards and regulatory frameworks, Dr Miller had some very relevant views to share.

The main point of conversation was around the level of funding being provided for such a big project. Given the scale and ambition of what is being proposed, the question is whether the funding will be enough. “I don’t know exactly what it’s going to take. It’s important to remember this is for one budget year – 2016. So it’s a very good start to get the planning underway for such a momentous project”, explained Dr Miller. Picking up on this, Mr Levine pointed out that there are commercial interests here too. So there might be a potential scenario where industry begins to spearhead this kind of work, leaving behind some of the benefits promised by the US Government. Dr Miller confirmed her faith in NIH Director, Francis Collins, citing his track record in major projects such as these that have faced competition from industry. She does concede that it’s still not quite known how partnerships with industry might work. But being able to pool together data and research participants, would not only help get to the 1 million sequences mark, but also help make that data more accessible for researchers.

Looking at the project’s funding in broader terms, the NCI funding came under scrutiny. Dr Miller felt that the NCI would be in a good position to make sure that some sub-types of cancer get more attention where there may not be such an immediate pay-off for commercial interests. She also strongly supported the portion of funding being given to the FDA. Looking at the bigger picture, she says that “…research is the first step…” regulation being the second, and reimbursement being the third. While a lot of progress has been made, she recognises that there is still a long way to go.

The conversation then turned to privacy matters, in particular how much of a barrier it might be in getting 1 million participants. Dr Miller opened, “There are a number of different ways to look at privacy. One – we know from a community of concerned activists, that when they are sick, or their child’s sick, or their loved one is sick, they want answers. They are willing to take their data and give it to those who can help solve their problems.” She then went on to explain that there are questions that we need to ask ourselves about the nature of our data. The Genetic Information Nondiscrimination Act of 2008 was also mentioned.

A great interview and a timely topic. For more, The Bio Report is free on iTunes and worth a listen!
Developing personalised medicines is one of the major applications of genomic research. Pharmacogenomics is the study of the role that genes play in drug response. Traditionally, this has focused on linking single nucleotide polymorphisms with a drug response. Now, with the birth and proliferation of whole genome sequencing, the scope of what pharmacogenomics can, and will, encompass has grown significantly.

Bringing this to fruition will require attention in certain areas, particularly in the recording of phenotypic data through large scale clinical trials. Communication within the research community and regulatory authorities is crucial to keep developing useful and relevant information. One thing is certain - pharmacogenomics laboratories are going to be even more invaluable in modern drug development.

**FLG:** What does the term ‘pharmacogenomics’ actually mean today? Now that you’re not just dealing with SNPs, it seems that the definition has expanded to encompass a lot more.

**JCM:** The term pharmacogenomics has indeed significantly expanded in meaning from the linkage of a single nucleotide polymorphism (SNP) with a drug response, to now encompass any genetic change related to both drug response and toxicity. This change has largely been driven by the increased use of next generation sequencing, particularly whole exome and whole genome sequencing of patients from clinical trials.

**FLG:** NGS is probably the technology that has really opened things up quite a lot. How big of an impact has it had on pharmacogenomics up to this point?

**JCM:** I would characterize NGS as just beginning to have significant impacts on the field of Pharmacogenomics. We have begun to see case reports of rare variants being described in the literature with clinical impact. Where NGS has already changed the field significantly is in our understanding of what constitutes a reference genome versus true variants. By that I mean we have much better understanding of variants at a population level than we did even five to ten years ago.

**FLG:** Are we still in a position where ‘more is better’ when it comes to producing data?

**JCM:** I think that the prevailing thought in the field is that more data is better, as long as you have a clear indication of what you are going to do with it. The generation of large amounts of sequencing data without any associated clinical data is much less useful. For example, there is currently a significant limitation on making phenotype associations with NGS data, compared to the progress that has been made in linking single SNPs or haplotypes with phenotype.

**FLG:** With more and more data being produced, the need for common standards grows. How tough is it to harmonize data across such a quickly growing field?

**JCM:** The rapid expansion of NGS has led to multiple different approaches for data analysis pipelines, some of which are slowly beginning to become standardized now. The major challenge that we still face is the integration of this genomic data into meaningful data sets, such as electronic health records. Even more important is to break down the data silo problem which has evolved across academic institutions, government and industry as a whole.
I would characterize NGS as just beginning to have significant impacts on the field of Pharmacogenomics.

March 2015 / Frontline Genomics Magazine / 11

PFIZER

Pfizer was founded by German-American cousins Charles Pfizer and Charles Erhart in 1849 as a chemicals business. Based out of Williamsburg, Brooklyn, they had immediate success producing an antiparasitic. The company experienced its first big period of growth through their production of citric acid.

Following a shortage of calcium citrate after World War I, Pfizer chemists developed expertise in fermentation to find an alternative way to produce citric acid. This new expertise was applied to penicillin production to support US and Allied soldiers. In the 1940’s the price of penicillin dropped dramatically. This drove the company to discover and commercialise Terramycin in 1950. This was the pivotal moment that changed Pfizer into a research-based pharmaceutical company. Today it is one of the world’s largest pharmaceutical companies by revenues.

“My hope is that we will begin to truly see the development of precision medicine, translating to direct improvements for patients across not only oncology but a range of disease areas such as cardiology and pain.”

“I would characterize NGS as just beginning to have significant impacts on the field of Pharmacogenomics.”
I think that we’ve finally begun to understand the fact that there aren’t just a few types of cancer, there are hundreds, perhaps thousands of types.

Jean-Claude Marshall
Director, Clinical Pharmacogenomics Lab
Pfizer

Jean-Claude currently serves as Director of the Pfizer Clinical Pharmacogenomics Laboratory, which is a CLIA certified lab working with preclinical and clinical trial samples. He also serves as subject matter expert on CLIA regulations. He has extensive experience with the use of next generation sequencing, genotyping and microarray platforms in personalised medicine for oncology patients. Prior to his work with Pfizer, Jean-Claude was Cancer Research Fellow for the NCI, and Research and CLIA lab manager at the Catholic Health Initiative. He completed his PhD at McGill University on Ophthalmic Pathology.

FLG: A lot of drugs fail in development due to safety issues. We’re starting to see some interesting work being published that could potentially help bring that failure rate down by expanding the predictive capabilities of data mining. Is bringing the cost of development down the big win for drug developers, or is there a bigger win out there than increasing efficiency?

JCM: Anything that increases our success rate in later stage clinical development will have a significant impact across all pharmaceutical companies. If there’s a way to predict for significant toxicity events, such as liver or kidney toxicities, before we reach first ‘in human’ studies, that would certainly be considered a significant advance. In addition, we need to remember that many drugs fail for lack of demonstrated efficacy or lack of improvement over current therapy. The promise of pharmacogenomics is to bring greater stratification of patients, creating enhanced efficacy or reduced toxicity, what we refer to as ‘Precision Medicine’.

FLG: We hear a lot of great things about the power of genomics, but progress can be frustratingly slow. Is there a risk that the clinical environment is too cautious to allow the field to really flourish?

JCM: Advances in scientific research are almost always a few steps ahead of clinical implementation. The cautious approach clinically is justifiably tied to a desire on the part of clinicians to base their decisions off of a solid base of scientific and clinical evidence. The competing drive between clinical caution and scientific drive for advancement has generally played out well. The area that may lead to significant delays in genomic personalized medicine may be the current uncertainty in the regulatory environment around these genetic tests. I would note however, that in the field of oncology, incorporation of pharmacogenomics has led to rapid demonstration of efficacy and drug approvals.

FLG: Regulators are also in a difficult position. While their purpose is to ensure patient safety, they do seem to create a challenging environment for innovation at times. The FDA’s LDT plans are proving less than popular at the time being. Is there anything regulators could realistically do to help support innovation as well as protect consumer safety?

JCM: The question around the FDA’s draft guidance for LDTs is not an easy one. On one hand you have companies who are offering direct to consumer testing that does not necessarily require significant clinical validation. On the other, you have laboratories that are able to respond to scientific advances and offer testing rapidly due to the current CLIA oversight regime, such as the rapid deployment of Ebola testing. Striking a balance between the two may require repeated rounds of talks between the FDA and the clinical laboratories they are proposing to regulate. Neither side has historically interacted with the other and that may prove to be just as challenging as forging a balanced regulatory approach.

FLG: Who do you think is likely to feel LDT regulations most?

JCM: Clinical laboratories who fill niche markets, those who make early testing available in response to scientific advances, such as those offering NGS testing right now, and small clinical labs who don’t necessarily have the capability to interact with the FDA will feel these proposed regulations the most. That isn’t to say that the current draft guidance by the FDA won’t affect others. In fact, I think it’s clear that this draft
guidance, if implemented in its current shape, may have broad reaching implications for the entire clinical testing field within the United States.

**FLG:** As well as being an expert in setting up regulated genomics laboratories, you also specialise in cancer. ‘Curing cancer’ is one of the real holy grails out there at the moment. With emerging genomic technology and its applications, are we in a position to realistically say that we have the tools to launch a devastating attack on cancer?

**JCM:** Every five or so years someone in the field comes forward and makes a bold declaration that they expect “cures” for cancer within the next few years. I think that we’ve finally begun to understand the fact that there aren’t just a few types of cancer, there are hundreds, perhaps thousands of types. There’s clear evidence that we are making progress on attacking cancer using these understanding, for example immune therapies and antibody drug conjugates. There’s still a lot of work to be done though, and I do have hope that we are making progress.

**FLG:** What’s the most exciting part of working at one of the world’s biggest pharmaceutical companies?

**JCM:** The breadth and width of the clinical trials that my laboratory supports is something that continually excites me. Being involved in trials from multiple disease areas, and being asked to support new clinical targets is something that is both energizing, and challenging. Being able to interact with people who are all experts in their realms and contributing to those teams is equally rewarding.

**FLG:** Prior to your work with Pfizer, you were at the CHI Center for Translation Research. Part of your work there was to engage with physicians to help drive genetic testing for solid tumour samples. What were the biggest obstacles to over come in getting front line practitioners bought in?

**JCM:** The education of clinicians around what molecular testing was, and how it could directly impact their patient treatments was always a challenge. There are multiple, sometimes contradicting, demands on a clinician’s time and resources which only adds to that challenge.

**FLG:** How do you think drug development and healthcare might change over the next few years?

**JCM:** I think that we are working during a time of great change in both the drug development and healthcare fields. The FDA regulatory oversight of LDTs will most probably continue to play out over the next few years, along with rapid advances in our understanding of genetic variants and the utility of whole exome and whole genome sequencing. There are many exciting opportunities available to us, although certainly those don’t come without risks. My hope is that we will begin to truly see the development of precision medicine, translating to direct improvements for patients across not only oncology but a range of disease areas such as cardiology and pain.

**FLG:** Is there any advice you would give to people just starting out in their scientific careers?

**JCM:** This is something that is often said, but I think bears repeating: do what excites you and what you love. Find the area of industry, government, academia or a hybrid of all of those that challenges you and forces you to expand your thought processes, and go after that. Even failing at the right challenge, while painful initially, can be rewarding.

"THE TERM PHARMACOGENOMICS HAS INDEED SIGNIFICANTLY EXPANDED IN MEANING FROM THE LINKAGE OF A SINGLE NUCLEOTIDE POLYMORPHISM (SNP) WITH A DRUG RESPONSE, TO NOW ENCOMPASS ANY GENETIC CHANGE RELATED TO BOTH DRUG RESPONSE AND TOXICITY“
Epigenetics deals with the biochemical processes ‘above’ or ‘before’ the genes which regulate the expression of genes in the genome, usually in response to influences in the immediate environment. What is particularly intriguing about epigenetics is how much it blurs the traditional boundaries we have erected between our genes and our environments. The novel complications introduced by epigenetics suggest—if not require—commensurately novel ways of conceptualizing our relationships with ourselves and with our environments that transcend this conventional dichotomization.

The unique nexus of genetics and the environment presented by epigenetics is of considerable practical relevance for diseases such as cancer which have as yet defied understanding via existing approaches that dichotomize genetics versus the environment (or vice versa). In this context, one major purpose of this article is to provide a brief survey of how the conventional emphasis on genetics in cancer research is being extended and empowered by epigenetics to perhaps finally realize the much-anticipated promise of cancer genomics.

CANCER, GENES AND “BAD LUCK”
The unique perspective from epigenetics in cancer as compared to the more conventional dichotomy of genes versus the environment are particularly noticeable in the discussions of a recent paper by Cristian Tomasetti and Bert Vogelstein on the causes of cancer which has already generated a significant amount of controversy.

In this paper the main question the authors attempt to answer is why there is such a disparity in the incidence of cancer between different kinds of tissues—e.g., as the authors note, “the lifetime risk of being diagnosed with cancer is 6.9% for lung, 1.08% for thyroid, 0.6% for brain and the rest of the nervous system, 0.003% for pelvic bone and 0.00072% for laryngeal cartilage,” just as “cancer risk in tissues within the alimentary tract can differ by as much as a factor of 24 [esophagus (0.51%), large intestine (4.82%), small intestine (0.20%) and stomach (0.86%)].” These disparities in risk of cancer between tissue types has been recognized for more than a century, but have not yet been reducible to either hereditary or environmental factors, which until now have been the only ways to parse the causes of cancer.

Building on Vogelstein’s previous pioneering work in somatic mutation, or mutational changes in cells’ DNA that are not passed along via the germ line but which occur during a person’s life, Tomasetti and Vogelstein hypothesized that the relative incidences of cancer in different kinds of tissues could be
caused by random mistakes when DNA is copied during cell division. In other words, the more times cells in a particular tissue type divide, the more opportunities for such copying errors to occur, the greater the risk of cancer.

However, to test this idea Tomasetti and Vogelstein needed a way to assess the rates of cell division of different kinds of tissues. Because only stem cells (versus differentiated cells) live long enough to initiate a tumor, Tomasetti and Vogelstein plotted the rates of stem cell divisions of the 31 tissue types for which the rates of stem cell divisions are known against the lifetime risk for cancer for each type of tissue on a log-log axis, predicting that “there should be a strong, quantitative correlation between the lifetime number of divisions among a particular class of cells within each organ (stem cells) and the lifetime risk of cancer arising in that organ.”

As shown in the figure below, there is a clearly noticeable relationship between these two very different measures. Tomasetti and Vogelstein report a strong positive correlation (0.80) between the lifetime risk of cancer and the number of stem cell divisions for a particular tissue type. From this correlation, the authors thereby conclude that around two-thirds of the variation in cancer risk between tissue types can be explained by the total number of stem cell divisions unique to that tissue.

To distinguish this stochastic cell division from external environmental and heredity causes, Tomasetti and Vogelstein construct an “extra risk score” (ERS) as a function of lifetime risk and the total number of cell divisions (log10 values). Utilizing machine learning methods and unsupervised classification, the 31 cancers clustered into two groups, high ERS (9) and low ERS (22): the higher the ERS (basically, the higher the risk of cancer relative to the number of stem cell divisions), the more likely are external environment factors to play a role. The authors found that the high ERS cancers were those with known links to specific environmental or hereditary risk factors, with the low ERS cancers being more likely to be caused by these stochastic errors during DNA replication.

These findings are particularly noteworthy for a couple of reasons. First, because before now the term “environmental” in cancer epidemiology has been used to denote anything not hereditary, such that these kinds of developmental processes had been “grouped with external environmental influences in an uninformative way.” Now these stochastic errors in DNA replication can be distinguished from external environmental factors. Second, because these non-hereditary genetic causes were found to contribute more to cancer risk than either hereditary or external environmental factors. This is important because, as reiterated by Tomasetti in a follow-up interview with Science, “if you go to the American Cancer Society website and you check what are the causes...”
of cancer, you will find a list of either inherited or environmental things. We are saying two-thirds is neither of them.”

SO WHAT?
What are the implications of this identification of a third way by Tomasetti and Vogelstein, and how is it related to epigenetics?

To the first point, as explained by one of the reviewers of the Tomasetti and Vogelstein paper, it is—or should be—common knowledge that even though the somatic mutations identified by Tomasetti and Vogelstein are legitimately genetic phenomena, they “are not in the germ line...are not transmitted from parents to offspring...don't generate family risk correlations [and therefore] can't be found by GWAS or other studies based on sequencing inherited genomes.” This reviewer also describes how it is—or should be—common knowledge that “environmental or life-history risk factors, like diet or reproductive history and so on,” can affect the risk of mutations identified by Tomasetti and Vogelstein, but that because this exposure “has to affect a cell in a given tissue and in a particular relevant gene being used by that tissue,” the net effect of these mutagens, and hence their predictability, is usually very small. In the end, for this reviewer the Tomasetti and Vogelstein paper uses new data but doesn't show much that wasn't already understood; perhaps the most salient point of this paper is how it demonstrates that “the love affair with inherited genotypes, enabled, encouraged, and funded by a variety of enthusiasms, opportunities, and vested interests, has distracted attention from working from what we knew.”

However, this point about the effects of age on the rate of somatic mutation is what opens the door for an epigenetic explanation of Tomasetti and Vogelstein’s results. Although Tomasetti and
epigenetic clock based on DNA methylation age as a measure of the cumulative effect of an epigenetic maintenance system which predicts not the age of the cells but of the person the cells inhabit. The median error of this clock is 3.6 years, which means it can predict the age of half the donors to within 43 months for a broad selection of tissues. Horvath also analyzed 6,000 cancer samples of 20 cancer types, all of which showed significant age acceleration, except for “a significant negative relationship between age acceleration and the number of somatic mutations.” Subsequent studies have also found an advanced methylation aging rate in tumor tissue, and that DNA methylation-derived measures of accelerated aging predict mortality independently of health status, lifestyle factors, and known genetic factors.

That epigenetics could be playing such a significant role in this longstanding puzzle about the disparity between the cancer risks of different tissue types, is intriguing. These results are preliminary at best, but quite suggestive of the profound role of epigenetics in cancer. Tomasetti and Vogelstein provided one important piece by identifying the role of stem cell divisions in risk of cancer. The next step is suggested by the connection between DNA methylation, somatic mutation, aging and cancer. The next step remains to be seen, but with the recent release of the first full mapping of the human epigenome, new developments are likely to come even more frequently.

Tomasetti and Vogelstein do not explicitly identify the epigenetic components of their findings as such, the copying errors which are such an important component of their model likely have epigenetic causes. This oversight is more than a little curious as Vogelstein has been a central figure in cancer epigenetics from its very beginning.

To explain how this might work, the reviewer from before goes on to identify a very clear environmental factor related to cancer risk not addressed by Tomasetti and Vogelstein in their model: “If mutations arising by chance during cell division ultimately lead to transforming genotypes in some cells, the longer one lives the more likely such changes are likely to arise in at least one such cell in the person. This is generally why most cancer rates rise with age in ways correlated with rates of cell division...That is environmental causation, even if indirect!” This oversight about the causal influence of age, “though it won’t change the empirical fact that neither inherited genotypes nor most environmental exposures do not have highly predictive effects,” suggests that Tomasetti and Vogelstein missed something important.

There are a number of recent papers published on the connections between DNA methylation and aging which have relevance for this proposed connection between somatic mutations and cancer. In particular, a 2013 paper by Steve Horvath describes his discovery of a highly accurate epigenetic clock based on DNA methylation age as a measure of the cumulative effect of an epigenetic maintenance system which predicts not the age of the cells but of the person the cells inhabit. The median error of this clock is 3.6 years, which means it can predict the age of half the donors to within 43 months for a broad selection of tissues. Horvath also analyzed 6,000 cancer samples of 20 cancer types, all of which showed significant age acceleration, except for “a significant negative relationship between age acceleration and the number of somatic mutations.” Subsequent studies have also found an advanced methylation aging rate in tumor tissue, and that DNA methylation-derived measures of accelerated aging predict mortality independently of health status, lifestyle factors, and known genetic factors.

That epigenetics could be playing such a significant role in this longstanding puzzle about the disparity between the cancer risks of different tissue types, is intriguing. These results are preliminary at best, but quite suggestive of the profound role of epigenetics in cancer. Tomasetti and Vogelstein provided one important piece by identifying the role of stem cell divisions in risk of cancer. The next step is suggested by the connection between DNA methylation, somatic mutation, aging and cancer. The next step remains to be seen, but with the recent release of the first full mapping of the human epigenome, new developments are likely to come even more frequently.
Over 200 Peer-Reviewed Droplet Digital PCR (ddPCR™) Publications*

From detection of rare mutations and cancer biomarkers to quantification of gene expression and miniscule viral loads, the QX100™ and QX200™ Droplet Digital PCR Systems have been used to redefine the limits of absolute nucleic acid quantification. With over 200 peer-reviewed publications, ddPCR platforms have outperformed other digital PCR systems by several orders of magnitude. The third generation QX200™ AutoDG™ System now brings automation and scalability to digital PCR.

Visit bio-rad.com/info/FLG for the publication list and to learn more.

*Based on PubMed data, February 2015.
Genomics is growing in just about every way imaginable. Advances in sequencing technology and cloud computing are making genomic analysis more accessible to researchers; more and more genomic data is being produced; and the political backing is also bringing more funding. It is a very exciting time to be involved in genomics at any level, but we can't get too excited just yet.

Managing stakeholder expectations is a constant battle. Especially when it comes to ensuring that robust and accurate science remain a priority over short term goals. Hype is great, in that it shows people are excited, but there's still a long way to go.

**FLG:** The political arena on both sides of the Atlantic has been very vocal about their support and backing of Genomics and Precision Medicine in recent weeks. Is there a risk in overhyping the public?

**GL:** Yes, of course. There seems to be an enormous amount of hype in the field of human genomics, and we must constantly remind people that the path to new drugs and/or the prevention of disease can take on the order of decades.

**FLG:** There's a sense of a growing disconnect between the growing political element and those doing the actual research. Is that funding being focused in the right areas of genomic research at the moment?

**GL:** More funding should be directed toward developing more accurate and faster sequencing methods, along with engaging much more with software engineers and cloud-based computing. We have to collectively develop ways to store and share millions of genomes going forward, and this is what various scientists and members of the Global Alliance for Genomics and Health are working on. From my perspective, the emphasis should be on further technology development and the implementation of highly accurate genome sequencing. This is part of what we have been working on in collaboration with Michael Schatz and others at Cold Spring Harbor Laboratory and elsewhere. Using whole genome datasets from 10 members of one family, my graduate students, Jason O’Rawe, Han Fang, and Yiyang Wu, showed that one can increase the reliability of the biological inferences with an integrative bioinformatics pipeline, including a new algorithm, Scalpel, developed by Giuseppe Narzisi and Michael Schatz, for more accurate identification of indels. We find a 2 to 5-fold difference in the number of variants detected as being relevant for various disease models when using different sets of sequencing data and analysis pipelines, and we derive greater accuracy when more pipelines are used in conjunction with data encompassing a larger portion of the family. We also collaborated with Min (Max) He and Kai Wang on the development of SeqHBase, a big data-based toolset for analysing family-based sequencing data, and we demonstrated SeqHBase’s high efficiency and scalability on several disorders, including a new syndrome, which we are currently calling RykDax Syndrome, where we identified a maternally inherited missense variant in an X-chromosomal gene, TAF1. A "genotype-first" approach led us to other families with variants in TAF1 and containing individuals having a remarkably similar clinical presentation.

**FLG:** Genomic technology has, and continues to develop at a rapid pace. At the center of this was the race to the "$1000 dollar genome". In pursuing faster and cheaper sequencing options, accuracy suffered. It has resulted in a drastic growth in the rate at which sequencing data is being produced. Was the compromise towards speed and affordability worth it?
This $1000 dollar genome is still a myth, as you are quite correct that such a thing is not accurate. We published papers in 2013 and 2014 showing that the accuracy of “whole genome sequencing” is far from ideal, and we have recommended that 60x coverage is needed to detect >95% of indels, if one uses 100 base pair reads from Illumina. This mythical $1000 dollar genome is only at 30x coverage and this is the amortized price on the current Illumina X10 instrument, which assumes running the instrument constantly to churn out 18,000 such genomes per year. I have heard of some genome centers charging $1500 for this 30x coverage, although the real price is closer to $2,000, I am told, and this does not include any of the costs for analysis. Therefore, for a 60x coverage Illumina genome, the current cost is $3,000, and even this sort of genome still lacks many regions that can only be filled in with much longer reads. There is certainly ongoing improvement and innovation, including from other sequencing companies, so that hopefully in a few years’ time, the accuracy will be much improved. However, there is still plenty of work to be done.

Are people sequencing too much at the moment? It seems that every week we see a newly completed genome announced. Is this an indication of the potential power of genomics or are we just in a period in which sequencing technology is being used simply because it has become more accessible?

About four years ago, I started broadly calling for better standards in terms of exome sequencing. However, the response basically was that implementing better standards makes the cost of research too much. So, we find ourselves in the situation of having had tens of thousands of exomes sequenced in research environments, where the only variants that can be returned to the research participants are ones that get Sanger validation (or some other validation) in a clinical environment. This was mentioned and discussed yet again at the Precision Medicine Workshop held recently by the NIH, with various people lamenting the fact that there is such a huge divide between the “research” and “clinical” worlds, and the response once again concerned mostly cost. So, from this perspective, the main work right now should focus on getting the cost of sequencing down much further, along with getting higher accuracy with longer reads, so that eventually people can get their whole genomes sequenced in clinical environments, where the chances for sample-swaps and other inaccuracies will be less. I do applaud that the FDA has finally approved a direct to consumer genetic test for Bloom Syndrome by 23andMe, but I certainly hope that the pace and scope of such approvals will dramatically increase. We need to get to a world of highly accurate and relatively cheap (~$100 genomes), so that it is then cost-effective to sequence millions of people, and then collect and analyse these data in aggregate to begin to understand how any particular genotypes express themselves among many different genetic backgrounds. Such things will only be possible with broad data sharing, including on the level of phenotype data. One can see that this very broad sharing of data, including pictures, is possible, as demonstrated by innovative companies like Facebook, 23andMe, PatientsLikeMe, and Ancestry.com, although the privacy concerns and issues with genetic data are a big issue that must be carefully considered. There have also been some recent innovations in face recognition and image processing, where one can begin to classify genetic syndromes based on photographs.

In terms of technology coming through today, what do you feel is going to make the most useful impact?

The developments going on right now at Pacific Biosciences and Oxford Nanopore are very promising in terms of longer reads, for sure. Researchers at Cold Spring Harbor Laboratory, including...
Dick McCombie and Michael Schatz, have been working hard on assessing and pushing forward these technologies from Pacific Biosciences and Oxford Nanopore, and it seems that we might only be a few years away from highly accurate and relatively inexpensive human whole genomes. It is also useful that people at the FDA seem to be beginning to engage more with how to regulate direct to consumer genetic testing in broader fashion. One can also see that companies like Google are getting interested in this sector, based on their cloud computing capabilities, so one could imagine that this could further enable broader data sharing.

**FLG:** The road to the promised ‘Genomic Revolution’ seems to keep stretching out. One of the first major milestones was the completion of the Human Genome Project. Back then, President Clinton suggested that “our children’s children will know the term ‘cancer’ as a constellation of stars.” The miracle cures that the public were hoping for, never arrived. It was, however, the starting point for a lot of great work. Now with large scale sequencing projects taking place, it feels like the expectation is that a whole bunch of variants are going to be found that we can drug and eradicate disease. The reality is that it is very unlikely that a single gene holds the answer. So the next step will be to take all of that data and turn it into useful information. Is there enough focus on funding the kinds of rigorous functional studies that will help deliver something tangible to patients?

**GL:** Although it seems that most people focus on the development of drugs to treat illness after it has started, I personally feel that large-scale sequencing might actually be more effective in terms of early detection and prevention of disease. This remains to be proven, of course, but there are early indications that screening of family members (known as cascade carrier screening) can help to identify other members of the family carrying particular mutations and thus at increased risk. Such people can at least know about their elevated risk, and there are some instances in which people can take action. Perhaps the most famous example of this is in women with prominent family histories of breast and ovarian cancer, who also carry mutations in the BRCA1 gene. Such women can undergo more intensive screenings, or in some cases, made famous by Angelina Jolie, elect to undergo interventions to reduce their risk, such as with mastectomies. Of course, the decision to undergo such a radical intervention particularly when the expression of this phenotype is quite variable is something that needs to undergo much more study, but on an individual level, some people are acting on their genetic information already.

**FLG:** You’ve had some experience on the front lines as a clinician yourself. Here in the UK, one of the biggest challenges the 100,000 Genomes Project is trying to address is how to integrate genomics into the NHS. This will have to look at work force planning, in particular around how to use and support Genetic Counsellors. How difficult can it be to introduce something as big as genomics into regular clinical practice?

**GL:** This is indeed an enormous challenge, particularly in the United States, where we do not have a national health care system. In addition, the number of genetic counsellors in America is on the order of only a few thousand, for a population numbering well over 300 million people. There is also a pressing need to educate health care professionals much more about genetics. I am constantly reminding people in the genomics world that we live in a tiny bubble in comparison to the vast landscape of healthcare in America, and we have to prove clinical validity and utility in order to get any sort of wider adoption of exome or whole genome sequencing. Such things will be enabled by better technology, higher standards, lower costs, and broader data sharing.

**FLG:** Your research is focused on the genetics of neuropsychiatric illnesses at Cold Spring Harbor. What drew you to that particular field?

**GL:** The short answer is that these illnesses are among the most fascinating in all of medicine. The long answer concerns the broad training that I have had, which requires me to explain my background a bit. I conducted some of my training in the Cornell/Rockefeller/Sloan-Kettering M.D./Ph.D. program. Just prior to that, I had spent one year as a Rotary Scholar at the University of Cambridge, England, working toward a Master’s degree in Genetics at the Wellcome CRC Institute, with Martin Evans as my research mentor. By the time I finished my Ph.D. at Rockefeller with Tom Muir and Richard Novick and returned to medical school, I had been exposed to a variety of research experiences. I had conducted research in various laboratories at Dartmouth College, the NIH, the University of Cambridge and the Cornell/Rockefeller/Sloan-Kettering M.D./Ph.D. program. Some of my research interests had included by then: thyroid hormone and its effects on the brain during development; cancer research with a focus on chemoprevention; the creation and phenotypic characterization of mouse models of human disease; the structure and function of proteins, including the use of chemistry to synthesize unusual protein variants and to analyze complex mixtures; and the development of novel anti-infectives for Staphylococcus aureus and other bacterial infections. It is safe to say that I had, and continue to have, wide-ranging interests in biological chemistry, which broadly defined encompasses the targeted use of chemistry to elucidate biological processes and vice-versa.

---

**Gholson Lyon**

**Assistant Professor**

**Cold Spring Harbor Laboratory**

Gholson Lyon is on faculty at Cold Spring Harbor Laboratory and is a research scientist at the Utah Foundation for Biomedical Research. He is also a board-certified child, adolescent and adult psychiatrist. He earned an M.Phil. in Genetics at the University of Cambridge, then received a Ph.D. and M.D. through the combined Cornell/Sloan-Kettering/Rockefeller University training program. He started his independent research career in 2009, after finishing clinical residencies in child, adolescent and adult psychiatry. In addition to his research on the genetics of neuropsychiatric illnesses, Gholson is focusing on the genetic basis of rare Mendelian diseases and the development of clinical-grade exome and whole genome sequencing.

**“ABOUT FOUR YEARS AGO, I STARTED BROADLY CALLING FOR BETTER STANDARDS IN TERMS OF EXOME SEQUENCING. HOWEVER, THE RESPONSE BASICALLY WAS THAT IMPLEMENTING BETTER STANDARDS MAKES THE COST OF RESEARCH TOO MUCH”**
Upon returning to medical school and rotating through the many different clinical rotations, I was struck by the fact that the most fascinating and challenging clinical cases for me involved illnesses affecting the brain and mind, due mostly to our relative lack of knowledge of the complexities of psychiatric and neurologic illnesses. I realized that much remains to be discovered and that one could fill an entire career focusing on the clinical and basic science aspects of diseases like intellectual disability, autism, schizophrenia, obsessive-compulsive disorder, Tourette syndrome, or other brain-based illnesses. In order to make substantive contributions in this area, I felt that I needed extensive clinical exposure to these illnesses, so I decided in the year 2003 to pursue clinical residency training, first as a psychiatry resident at Columbia and the New York State Psychiatric Institute, followed by additional training in child and adolescent psychiatry at New York University, Bellevue Hospital and Rockland Children's State Psychiatric Hospital. Upon completion of 5 years of clinical training, I started my independent career in 2009 in the state of Utah, and I was recruited in 2012 to establish an active laboratory at Cold Spring Harbor. We focus on the discovery of families with rare diseases and/or increased prevalence for syndromes such as intellectual disability, autism, Tourette syndrome and schizophrenia. Once we identify mutations that likely contribute to a disease, we undertake detailed functional studies of these mutations and the biological processes affected. Proving the biological relevance for newly discovered mutations is the major problem, so having access to research participants and derived tissues is critically important, hence the need to engage directly with families.

**FLG:** You split your time as a researcher at the Utah Foundation for Biomedical Research (UFBR) as well. What kind of projects are you working on at the moment?

**GL:** We study the breadth and depth of genetic variants in Utah, where there is a large founding population, large family structures and good genealogical records, which enables well powered family-based genetic studies for rare diseases. We use exome and whole genome sequencing (WGS) to identify mutations that segregate with various idiopathic syndromes, and we undertake comprehensive functional studies of many of the newly identified mutations. This has led to the discovery of many new genetic syndromes, including Ogden Syndrome, RBCK1 Syndrome, and most recently RykDax Syndrome. This latter syndrome presents with severe intellectual disability (ID), a characteristic intergluteal crease, and very distinctive facial features. We continue to advocate for more comprehensive and accurate whole genome analyses in large pedigrees, and we have collected ~2000 DNA samples to date from >100 families in Utah, including detailed phenotyping information. Some of these samples have undergone exome or whole genome sequencing, and we are currently analyzing these data. This includes the ongoing analysis of whole genomes from 3 families with singleton cases of autism, and an analysis of nine whole genomes from a pedigree with Prader–Willi Syndrome (PWS), Hereditary Hemochromatosis, Familial Dysautonomia (FD), and Tourette Syndrome.

**FLG:** We interviewed Michael Vellard, from Ultragenyx last December. He had family reasons for wanting to develop treatments for rare diseases, and was very passionate about the potential relevance of his research to other, more common, indications. What attracted you to start researching into rare diseases?

**GL:** When I finished my M.D.-Ph.D. training in 2004, I was full of ideas about what research I might do in my future academic career. However, I decided to broaden my training by undertaking the clinical residency. Now, many years later, I am so glad that I undertook clinical training, as my eyes are now fully open to the complexity and nuance of the human condition, which cannot possibly be understood fully by studying the outcome of mutations only in mice or other lower animal models. Due mostly to my clinical training, my outlook has broadened to include a focus on rare diseases with very strong phenotypes, as these provide a window into very interesting and important biology. I believe that my background in genetics, chemistry, pharmacology, and medicine allows me to interface with basic biology.
Scientists and clinicians in the discovery and characterization of new genetic syndromes. It is critical for suitably trained physician-scientists or other broadly trained individuals to be involved with careful phenotyping and collection of human pedigrees with particular disorders, followed by a well-thought-out experimental design in terms of whole genome sequencing and follow-up experiments.

**FLG:** Going back to the NIH Precision Medicine Workshop, do you think more people should try to get a broad experience across research and clinical practice to try eliminate the present divide between the two?

**GL:** Yes, definitely. There is a重大 dearth of physician-scientists and other broadly trained scientists right now in America, and there is also a hyper-specialization that has occurred partly due to the way that NIH funding is determined, including evaluation of research grants by hyper-specialized study sections. I have been constantly amazed that there is very little reward in the current system for people with broad, interdisciplinary training, and in fact, I have received grant evaluations somehow lamenting the fact that I am not “focused enough” on one particular topic. I have heard that the NIH is trying to figure out ways to support broadly trained individuals, and I would certainly support such efforts. We definitely need to figure out ways to bridge this substantial divide between research and clinical practice, and this can only be done with the aid of people trained in both areas. I have been incredibly lucky at Cold Spring Harbor Laboratory that the president, Bruce Stillman, and the chancellor emeritus, Jim Watson, have been so supportive of my work.

**FLG:** As genomics develops as a field over the next 15-20 years, what would you like to have achieved through your work?

**GL:** I have been advocating for more rigorous standards regarding the collection of human genetic data, including with the accuracy of variant calls. Instead of compartmentalizing research and medicine, the two should be integrated so that physicians who are most familiar with human “phenotypes,” can inform the other arms of science. This is certainly beginning to happen, and I am very much helping to push this forward.

**FLG:** For students at the start of a potential career in genomic research, would you have any advice or recommendations about how they should look to develop their areas of focus?

**GL:** Get a broad training! For me, science is all about trying to understand biology using whatever tools I can bring to the table. Therefore, genomics is just one of many tools out there. Over the course of my career, I have used many other tools, including cell culture, genome-edited mice, peptide chemistry, protein expression, mass spectrometry, and yeast genetics. So, I tend to pick biological questions and then figure out the best tools and techniques that I might need to answer the questions.

**FLG:** Since you joined Twitter in 2011, you have been steadily developing an impressive following. It seems that twitter is the social media platform of choice for a lot of researchers at the moment. As a frequent tweeter, you give a lot to the community. What's your main motivation for maintaining your online presence? Have you had anything out of the ordinary happen to you online yet?

**GL:** I use Twitter as a way to communicate with other scientists and the general public. It is a great way to keep up to date on what is happening in science in general, along with also seeing how the blogosphere reacts to overly hyped papers. Most people do not want to take the time to criticize papers, either at all or in various snail-mail venues such as “letters to the editor”, whereas a few scientists are at least willing to send out a tweet or post something on a blog to call out various papers. I do think that such analyses help to alert people regarding recidivist behaviours on the part of some scientists who tend to overly hype their results. This is particularly prominent for some people who tend to issue overly dramatic (and sometimes misleading) press releases about their work. These things take time, but it is my hope that the younger generation of scientists will learn from such things on Twitter that it is actually damaging to your career and reputation to engage in so much hype and spin. People may or may not directly call out such behaviour publicly, but they certainly talk about these things at meetings and in other venues, and a poor reputation ultimately leads to less funding and support for your work. In regards to your other question, luckily, I have had mostly positive interactions on Twitter to date.

**FLG:** Is there anything else you would like to mention to our readers?

**GL:** The incentive structure in academic science is really skewed in favour of publications. This results in the churning out of many substandard papers, all due to the fact that each person traditionally been held to the standard that they must be first or last author on some decent number of publications in order to be considered for certain grant monies. This dis-incentivizes collaboration, and I have personally witnessed behaviours involving withholding genetic data and pedigrees due to the fact that some group demands that they absolutely must be the sole first and/or last author on some paper. I have tried to counteract such behaviour by contributing to and helping to promote the BioRxiv preprint server, which was started by Cold Spring Harbor Laboratory Press as a way to encourage the open sharing of results. I am also on the editorial board of a new journal from CSHL Press, called Molecular Case Studies, which aims to present genomic and molecular analyses of individuals or cohorts alongside their clinical presentations and phenotypic information. The plan is to have a rapid peer-review process that is based on technical evaluation of the analyses performed, not the novelty of findings, and offers a swift, clear path to publication.

**FLG:** Thank you very much for your time, and good luck with your research!

**GL:** Thank you!
Stefan Petranek took images of attendees at last year’s ASHG meeting, posing them the question “How do you think genetic research will affect the future?”
**HOW DO YOU THINK GENETIC RESEARCH WILL AFFECT THE FUTURE?**

Whether you’re a scientist, or not, you will have an opinion on the implications of genetic research. It is a deeply emotive issue because it affects all of us and raises some interesting ethical dilemmas. Image-based artist, Stefan Petranek, is looking to you to find the answers through the Genetic Portrait Project.

**STEFAN PETRANEK**

Assistant Professor of Photography and Intermedia, Herron School of Art & Design, Indiana University

Petranek is an emerging visual artist. He studied Biology at Bowdoin College, before moving on to Rochester Institute of Technology to pursue Imaging Arts. Recently, he exhibited his Genetic Portrait Project at Rochester Contemporary in Rochester, New York. Petranek spoke at the 2012 Society of Photographic Education’s Mid-West Conference on his work dealing with genetics. In 2008 he spoke at SPE’s national conference on the Subject of Science in Contemporary Art. He resides in Indianapolis where he is an Assistant Professor of Photography and Intermedia at Herron School of Art and Design.

**FLG:** What is the Genetic Portrait Project?

**SP:** The Genetic Portrait Project is a photographic based documentary that examines individuals’ current understanding and attitudes towards genetic research through an artistic lens. From the food we consume to the medical advice we receive, advances in genetics and genomics are being translated into tangible technologies that influence a multitude of practical, social, and personal choices we make. This photographic project gives voice to how individuals from all walks of life perceive these rapidly expanding fields, which by many accounts is on the cusp of changing how we all think about life and our power to influence it in the most fundamental way. To participate, individuals are photographed alongside their response to the question: “How do you think genetic research will affect the future?” To date, over 375 individuals have participated, and although it primarily reflects American sentiment, the project aims for a globally diverse response.

**FLG:** What made you decide to make the switch from Biology to pursuing a career in imaging arts?

**SP:** I started out thinking I wanted to be a biological scientist in high school, but shortly after completing a Fulbright Research Grant in Chile after undergrad I realized one could not be a generalist in science. I began to find scientific methodology too reductive in its stance. While I have very high respect for scientific researchers, I recognized that I personally did not possess the patience for the rigors of scientific methodology. Art, like science, is interested in deeper truths but it allows for a much more liberal approach to research practices that I find fulfilling.

**FLG:** Why is visual art a good medium to explore the social impact of genomics?

**SP:** Hopefully it provides an intuitive entry point to consider the social and psychological implications of what advances in genetics means to the average person. Science and the specialized vocabulary that surrounds it, is intimidating for much of the population and so visual art has an opportunity to move past these language barriers and deal with the biology in a direct way that doesn’t require a sophisticated understanding.

**FLG:** What’s next for the project?

**SP:** The general goal is to find new audiences with differing views on genetic advances to participate in the project, giving voice to a diverse set of individuals and mind sets. Currently, the next planned stop for the Genetic Portrait Project will be Front Line Genomics’ Festival of Genomics in Boston this June. There I will be working with attendees to photograph them alongside their own response to how genetics will affect the future.

**FLG:** What are you ultimately hoping to achieve with the project?

**SP:** In the near future, it is my hope that I can create an online portal so anyone can post their own photograph of themselves alongside their response to increase the awareness of how thinking about science’s impact on society is important. Long term, I would like to have a book published which chronicles the project from its humble beginnings of me with a camera to a project in which anyone has that opportunity to participate.

**FLG:** How can people get involved?

**SP:** Short term, you can head to the Festival of Genomics in June. You can also see links to the project’s Facebook Page (facebook.com/geneticportraitproject) or visit my personal website www.stefanpetranek.com.

March 2015 / Frontline Genomics Magazine / 25
BETTER TOXICITY ASSAYS IN 3D

Organoid, Spheroid & TriCulture Models
Toxicity Assessment
Organ models; skin, liver, neurons, kidney & others
Oncology Drug Profiling
Stem Cell and Regenerative Medicine
Mitochondrial Function & Toxicity
IHC, FACS, Genomic & Proteomic readouts

Quantitative • Reproducible • Easy to use

info@amsbio.com • www.amsbio.com DE +49 (0)69 77 90 99 • UK +44 (0)1235 232100
US/CA: +1 (800) 987-0985 / +1 (617) 945-5033 • CH +41 (0)91 604 55 22
There are so many variables in the world, that the most mundane of occurrences can have a significant knock on effect. Often the difference between ourselves and those above us, can come down to simply ‘right place, right time’. David Stone’s career includes an enviable list of firsts, and is currently involved in pioneering research at one of the world’s most important pharmaceutical companies. For a career that has already contributed significantly to improving human health, it almost didn’t happen at all.

We were fortunate enough to get an opportunity to hear the full story, as David reflects on his career to date and life in Big Pharma.

**FLG:** Most people in genomics seem to be working towards a common goal of improving the lives of patients. That being said, people have found their way into the field in a variety of different ways. What first attracted you to a life of research and the commitment of undertaking a PhD way back in the early 90’s at University of Pittsburgh?

**DS:** Oddly enough, it was light traffic on a Thursday afternoon. I was teaching high school biology a couple of years after completing college, and working on a master’s degree in education. I had left a little early, and with no traffic I ended up at the university about half an hour before class began. I wandered into the anthropology department and struck up a conversation with a couple people, one of whom would later be my PhD advisor.

I have to give a lot of the credit for my success and drive to work on human health to my postdoc advisor, Caleb Finch. My dissertation work was very much on the theoretical side of things, and I’m still not certain what made him take a chance on me for a postdoc in molecular neuroscience.

“I THINK THE SCIENTIFIC RIGOR IS FAR GREATER IN INDUSTRY THAN IN ACADEMIA, AND I SUPPOSE THIS MAKES SENSE CONSIDERING THE RISKS”
We hit it off famously; he is a guy who never allows himself to be pigeonholed, and follows the questions wherever they lead, always with the end goal of improving human health. I decided that was the kind of scientist I wanted to be, and I try to let that guide what I do.

**FLG:** As part of your impressive journey to Merck and Co, you were at Harvard Medical School in the late 90s where you set up the first molecular biology lab at the Harvard Brain Tissue Resource Center. How did the opportunity come about?

**DS:** Again, I was unbelievably lucky. The head of the HBTRC (Francine Benes, also I believe the first woman to make full professor at Harvard Medical School) recognized that they needed to bring the current techniques in-house to stay on the front edge of research; I was fortunate enough to have the right skills at the right time. I was extremely close with the people from my postdoc, but they practically threw me out when they heard I had a shot at a Jr. faculty role at Harvard Medical School. They said most people never get a chance like that, and I had to take it.

**FLG:** Harvard Medical School comes with a very big reputation. How does it feel to be part of that history?

**DS:** People do notice that on your resume, and without question there are a lot of great minds there. I’m certainly happy to have been there, but I have to say I’ve met brilliant people from almost every university, country, and walk of life.

“I’VE SEEN GOOD PROJECTS GET DERAILED BECAUSE OF SITE CLOSURES AND LAYOFFS. OBVIOUSLY ACADEMIC RESEARCH IS AFFECTED BY ECONOMIC FORCES AS WELL, BUT I THINK TO A LESSER EXTENT THAN FOR INDUSTRY.”

**MERCK**

Merck began when Jacob Merck purchased a drug store in Darmstadt, Germany, in 1668. In the 1800’s Emanuel Merck and his successors developed the enterprise into a chemical-pharmaceutical factory. 1891 saw the first expansion into the USA as George Merck emigrated his family to set up Merck & Co as the US arm of Merck KGaA.

Merck & Co was confiscated in 1917 and re-established as an independent American company in keeping with wartime policy. Merck KGaA retain the naming rights outside of North America.

In 1953, Merck & Co merged with Philadelphia-based Sharp & Dohme, Inc. This move made them the largest US drug maker, combining Merck’s research and chemical manufacturing expertise with Sharp & Dohme’s strength in sales, distribution and marketing. Today Merck & Co (known as Merck Sharp & Dohme outside of North America) is the world’s seventh largest pharmaceutical company by market capitalisation and revenue.
**FLG:** The next step was what took you into the world of drug development. You moved to CuraGen where you co-led the team that demonstrated for the first time that drug efficacy could be predicted from gene-expression profiling. Were you ever in any doubt that you were going to be able to do it?

**DS:** I’ll admit it was a very pleasant surprise when it worked the first time! I had suspected it was possible in one way or another, but had never thought of attempting it. Then I was talking with a co-worker, Erik Gunther, and found out he had put a huge amount of thought into this problem, and finding a “kindred spirit” was enough for us to try to prove it in the wet lab. But we knew the devil was in the details; even if it would work in theory there were dozens of things that could go wrong in practice—so there was a LOT of thought and planning that went into the experiments. After we had published the work it was picked up as an “Editor’s Choice” in both Nature and Science—which again was nice considering how many people told us it would never work!

**FLG:** You’ve been at Merck and Co now for over ten years. With the explosion in genomic technology in that time, it must be an exciting time for you. One of the highlights was completing the first genome-wide siRNA screen for a neurological indication. Your focus these days is on using genetics and genomics to enable drug discovery. How different is the ‘Big Pharma’ environment to academia?

**DS:** In some ways the research environment is better in big pharma; in other ways it’s better in academia. I think the scientific rigor is far greater in industry than in academia, and I suppose this makes sense considering the risks. In academia if your research is wrong you risk having your paper refuted by a colleague; but in industry if you are wrong you risk literally millions of dollars on a failed trial, or worse yet harming a patient. I think most people are surprised when they find out how carefully research is done in big pharma, and how critically results are reviewed.

On the other hand, big pharma is corporate America. I’ve seen good projects get derailed because of site closures and layoffs. Obviously academic research is affected by economic forces as well (e.g. government funding), but I think to a lesser extent than for industry.

**FLG:** What projects are you working on at the moment?

**DS:** I’m using a lot of the newer sequencing technologies to look at neurodegenerative diseases.

**FLG:** Despite having accomplished so much already, you’re showing no sign of letting up. Is there anything in particular that keeps you motivated to keep innovating?

**DS:** It always comes back to improving human health. As long as you keep that as your primary goal, there’s never a question as to why you are doing this.

**FLG:** Have there been any moments in particular that you think you will always remember?

**DS:** An unfortunate incidence at the AD/PD conference in Salzburg 2007. I was given the honor of choosing the bar we would attend after dinner, and led about a dozen researchers into a hole-in-the-wall establishment in an old building (I chose this place on the strict criteria that I thought the name sounded funny and it was close by). As we sat there drinking beer with the locals, I noticed the young man I was sitting next to had some racist tattoos—it turns out I had quite unwittingly led everyone into a pub that was frequented by the local neo-Nazis. For some reason I decided it was a good idea to explain to him that the genetic evidence refuted his world view, and as you can imagine this conversation got heated...
and loud rather quickly. Long story short we got out of there without injury or anyone spending the night in an Austrian prison, but I was told I was not allowed to choose the bar the team would go to at the Society for Neuroscience meeting that year.

**FLG:** You'll be taking to the stage at the Festival of Genomics this summer. For the crowd at the Pharmacogenomics stage, you'll be presenting 'Connecting opposite ends of the pipeline - how clinical trial pharmacogenetics can enable target ID in drug discovery'. That's quite a lot to pack into 30 minutes. What are they main points you're hoping to get across to crowd?

**DS:** I think most scientists are beginning to realize that the information generated at each end of the pipeline can, in many instances, be of great use to the other end – especially in the case of genetic information. But not all opportunities are equal. If we are going to try to reduce this to practice at this time we have to focus on the right areas. I'm hoping to give people a view into how Merck is trying to do this.

**FLG:** It looks like pharmaceutical companies are starting to increase their investment in genetics and incorporating genomic information into drug discovery and development. Improving success rate and making clinical trials much more efficient should help reduce the cost of drug development significantly. This is something that is already starting to take place, but there is still a long way to go before it is a consistent reality. From a drug development perspective, what would you say is the biggest challenge facing those who are looking to use a genomic approach?

**DS:** On a practical level, dealing with the sheer volume of information is a major concern. If you want to have a well-powered genetic study you need thousands of individuals, and whole genome sequencing data on that many people dwarfs what we could imagine even a few years ago. But big data are here to stay, and I think we are already getting a handle on this. On another level, the question of where to focus internal research and what to leave to the academic world isn't quite clear. What will have the biggest payoff- PheWAS? Mendelian Randomization? WGS of large populations? Family studies on extreme phenotypes? We all have opinions on this, but we really don't know which will result in the most therapeutics in the coming years.

**FLG:** As more and more ethics studies come out, it's becoming clearer that patients support the use of their genotypic data for research purposes. Which is great, but there still seem to be some misgivings around allowing commercial organisations access. What can drug developers do to try help patients feel more comfortable in supporting their research?

**DS:** I think we have to make it clear that the end goal of the pharmaceutical industry is the same of that of the practitioners and patients: to develop therapeutics and get them to the people who need them. The more we understand the genetic factors in disease, the more we can focus our research on that area. In some ways this gets at the idea of personalized medicine; not everyone with a given disorder has exactly the same for of the disease, and they will not all respond to the same treatment. The best way to ensure that we are working on your form of the disease is to make sure you are in the genetic studies. And of course we need to let people know that they are anonymous in these studies- we NEVER want to connect names or any person information to genetic data.

**FLG:** One of the comments I hear a lot in relation to the use of genetics and genomics from non-scientists is “If there was anything to it, drug companies would already be doing it”. It’s infuriating, but also shows how little people see beyond the political hype. It also suggests that there is a fundamental lack of understanding of the drug development process out here. Associating a genetic variant to a particular indication takes a lot of work to produce a lot of data. There is also no guarantee that it will be drugable. To help all of us who deal with this kind of question, what should we be telling people about how pharmaceutical companies are putting genetic and genomic data to use?

**DS:** I think we need to be clear about how much evidence there is that genetics can enable drug discovery (for example, the fact that the targets of successful drugs are quite frequently genes which have been associated with human disease or clinical phenotypes). When you pair this with the fact that we still have a vast amount to learn in human genetics and its relationship to disease, I think it is clear that this is the best way forward. Our best hope of finding therapeutics for autism, ALS, schizophrenia, cancer and a host of other diseases really is understand the disease pathophysiology and design treatments that get at the heart of the process.

**FLG:** The Festival line up was announced a few weeks back for the first time. Who are you most excited to see up on stage in Boston?

**DS:** Honestly it looks like a great line up. When I read through it I was surprised how many of the talks I wanted to go to!

**FLG:** We look forward to seeing you there. Thank you very much for sharing your thoughts with us!
Millions of Droplets Matter

RainDrop™ Digital PCR System. Precise detection of mutation frequencies as low as .000001%. See the proof now at RainDanceTech.com/RainDrop

RainDance Products are for Research Use Only. Not for use in diagnostic procedures.
See the Genome in a New Light with PacBio® Sequencing

Revolutionizing genomics with 20,000 bp reads

PacBio® RS II Sequencer

- Industry’s Longest Read Lengths
- Achieves > 99.999% Consensus Accuracy (QV 50)
- Direct Sequencing of DNA
- Unbiased Genome Coverage

www.pacb.com

For Research Use Only. Not for use in diagnostic procedures.
NGS: HOW AND WHY WE GOT HERE

In the early part of the twenty-first century, we were treated to the completion of the first drafts of the human genome sequence. In the ensuing decade, surprising new technologies for DNA sequencing have been developed. Today's highest throughput sequencers can generate over 20,000 times as much data in a single run as those used in the Human Genome Project, making routine, cost-effective genome sequencing from individuals realistic. Here, we will discuss where these sequencing technologies came from and what they've been used for. But even by 2015, a number of “next-generation” sequencing approaches have come and gone, falling by the wayside and leaving only a very few real genome sequencing contenders left.

In 2015, it is reasonable to say that we are living in what could be described as yet another “Golden Age” of genomics. Technologies now exist that allow for sequencing of complete human genomes in a matter of days. Coupled to modern, high-performance computing and a host of bioinformatics tools that have been developed to analyze the data, whole genome sequences from individual humans can now be generated at costs approximating that of single DNA laboratory tests. Collectively referred to as “next-generation sequencing”, or NGS, these technologies all share several characteristics: massive parallelization of chemical sequencing reactions, micro- to nano-scale reaction volumes, and a hefty dose of computational power to capture the raw data and process it to formats interpretable by analysis software running on external computers.

Since the first publication of a human genome sequence using NGS, that of Nobel Laureate James Watson in 2008, there have been numerous studies of single genomes, using various NGS approaches (for relatively early examples, see references 2 and 3). More recently, we have begun to see larger-scale studies, applying NGS to the whole-genome analysis of larger patient and family cohorts (for example, references 4 and 5). Not surprisingly, cancer biologists have embraced these new technologies: the field is replete with examples, including major international efforts such as the International Cancer Genome Consortium (https://icgc.org), which aims to thoroughly characterize fifty different tumour types, including whole-genome sequence analysis. Indeed, these are exciting times for genome scientists, far from the early days of automated DNA sequencing.

GOOD OLD DAYS - THE HUMAN GENOME PROJECT

In many ways, the future of NGS was determined by a roadmap laid out during the Human Genome Project (HGP). Both the public- and privately-funded human genome reference sequences were created mainly on fluorescent, automated capillary sequencers. Developed initially as early as 1990, capillary sequencing ultimately took conventional Sanger
sequencing and parallelized it into instruments that could run 96 or 384 reactions at once in microplate format. Along the way, radioisotope labels were abandoned in favour of fluorescent detection, polyacrylamide slab gels exchanged for polymer-filled capillaries, and X-ray film done away with by laser fluorescent detection systems. The development of this technology was an inexorable result of the massive sequencing capacity required to complete the human genome.

Nowadays, HGP-era sequencing looks distinctly low-throughput, and rooms full of hundreds of capillary sequencers have given way to institutes running comparatively small numbers of modern NGS instruments. The last major human genome sequencing paper using capillary sequencing was published in 2007, and even that study used mainly existing data. While this study and subsequent analysis of the sequence remain the benchmark of thorough analysis of an individual genome, even in 2007 the newer technologies were beginning to take over. A veritable earthquake was beginning to shake the sequencing field.

“NEXT GENERATION” SEQUENCING
While Levy and colleagues were re-assembling the first individual, diploid human genome sequence, others were applying NGS approaches, most notably to produce a genome sequence from the DNA of James Watson, the first NGS-sequenced human genome. This study used instruments from 454 Life Sciences, an offspring from Curagen that was later acquired by Roche Life Sciences in 2007. If that sounds a bit convoluted, it is: the short history of NGS is remarkably messy, replete with acquisitions, hostile takeovers, and mergers, and littered with the skeletons of dead-end technologies. In this fast-moving field, by 2015 we have already seen numerous instruments reaching end of life, and others upgraded almost beyond recognition. Still other promising approaches have never made it to market and been abandoned. One or two that had been available for sale never worked as advertised, and have also subsided into the increasingly-murky history of the field.

The Roche technology used for Dr. Watson’s genome was the first NGS technology to market, with the now long-deceased GS20, so named because it was a Genome Sequencer capable of producing a massive 20 million bases of data. Compared with the most popular of the capillary sequencers, Applied Biosystems’ 3730xl, it was a monster, producing over 250 times as much data in a single run. Its latest version, the GS-FLX, increased this output another 35-fold to 700 million bases, but even this was not enough – Roche announced in 2013 that it was shutting down its 454 sequencing division, and that it would discontinue support in 2016. In the meantime, the company had other sequencing goals in sight: they attempted a failed hostile takeover of industry leader Illumina in 2012, and forged an alliance with nanopore sequencing pioneer Pacific Biosciences in 2013.

Other players in the field have followed similarly convoluted evolutions. Genotyping heavyweight Illumina began their foray into sequencing by acquiring Solexa, a spin-off company founded by scientists from Cambridge University. Late in 2006, Illumina announced the acquisition of Solexa in a takeover that had been a closely-guarded secret, immediately catapulting the former genotyping company into the forefront of the NGS field.
The Solexa 1G, rapidly re-branded as the Illumina Genome Analyzer, promised up to 1 Gigabase of sequence per run, a vast increase over that of the Roche instruments. By using four-colour fluorescent sequencing-by-synthesis (i.e., the measurement of different coloured DNA bases as they are incorporated into a growing chain), the instruments also had a familiarity about them that appealed to those used to four-colour capillary sequencing. Later versions, the Genome Analyzer II, GAIIx, and the HiSeq series, have increased this output remarkably, to the just-released spec of the HiSeq 4000, a single instrument reputed to produce up to 1.5 Terabases\textsuperscript{17} of sequence in a single run. Bundled into packages of 5 or 10, the HiSeq X instruments provide additional capacity, up to a reported 1.8 Terabases per run. For those keeping score, that’s about 90 thousand times as much sequence as the venerable 454 GS20 could manage.

The third big player, at least initially, was Applied Biosystems. Long the market leader in capillary sequencing, the company bought 454 Life Sciences, developers of a sequencing-by-ligation technology termed SOLID.\textsuperscript{18} SOLID instruments went through a similar number of iterations as the Illumina’s have done, culminating in the SOLID 5500xl W, still available in 2015 with an output of up to 360 Gigabases, depending on the specific application.\textsuperscript{19} In parallel, having become part of Life Technologies along the way, which in turn is now all part of the lab supply and instrumentation juggernaut Thermo Fisher, in August of 2010 the company hedged its bets by announcing the acquisition of Ion Torrent, developer of a completely different sequencing technology.\textsuperscript{20}

The current Ion Torrent Personal Genome Machine (PGM) and Proton instruments form the strongest competition to Illumina at the present time. The somewhat confusingly-named PGM, can generate up to 2 Gigabases of sequence per run, or only about two-thirds coverage of a human genome.\textsuperscript{21} The Proton, while a capable workhorse for exome sequencing, is only specified to produce 10 Gigabases of sequence per run, although in reality, 15 Gb is routinely achievable.\textsuperscript{22} While enough to redundantly cover the roughly 3-billion base human genome five times or so, this is not nearly enough throughput in a single run to achieve the depth of coverage required for robust whole-genome re-sequencing.

OTHER TECHNOLOGIES
Along the way, there have been many other sequencing technologies. One interesting example is the “Polonator”, initially distributed by Dover Systems, with a claimed output of 8-10 Gigabases per run\textsuperscript{23}, and an open architecture model for both the instrument, and sequencing applications and chemistries for it. How successful this has been is somewhat difficult to determine, and it appears most recently to have been taken over by Azco Biotech. There are few, if any, currently in operation.

Other companies have taken the approach of using single molecules as sequencing templates, thereby avoiding any biases that might be introduced through PCR amplification of the templates. Notable among these, the Helicos Biosciences Heliscope certainly had the capability to sequence whole genomes.\textsuperscript{24} The instrument, though, was both huge and expensive. Ultimately Helicos, after a series of press releases promising upcoming sales, filed for bankruptcy.\textsuperscript{25} More successful single-molecule technologies include those from Pacific Biosciences (another one-time rumoured acquisition target of Applied Biosystems) and Oxford Nanopore Technology. These rely on sequencing by detecting molecules as they interact with nanopores in a membrane (see reference \textsuperscript{26} for a recent review). Other companies have either fallen by the wayside, or changed focus, one example being BioNanomatrix, now Bio Nano Genomics, which focuses on genome mapping rather than sequencing per se.\textsuperscript{27}
One other important option is provided by Complete Genomics, a company built entirely on a service model and using proprietary technology based on DNA “nanoballs.” Complete Genomics was founded initially to focus only on human, whole-genome sequencing, but importantly for cancer biologists, it integrated matched tumor:normal sample sequencing from very early in its existence. Now owned by Chinese sequencing company BGI, it has shifted its focus to clinical projects, but is still apparently completing some legacy research projects as well. Persistent rumours suggest that Complete Genomics’ instruments might find their way into the market at some point, but concrete evidence is lacking.

Of all of these and many others, only Pacific Biosciences and Oxford Nanopore Technologies appear to be well established, with instruments well suited for niche sequencing (targeted gene panels or regions, small genomes, pathogens). No other player is taking a serious run at whole genome sequencing, and with serious questions about Ion Torrent ever achieving instrument outputs that can realistically sequence whole human genomes, that leaves really only one player selling instruments that can: Illumina.

CURRENT STATUS
So where does this leave us when faced with the problem of sequencing cancer samples today?

Sequencing a human genome and comparing it to the existing reference sequence, a process usually referred to as “re-sequencing,” requires redundant coverage in order to accurately map and overlap all of the individual sequence reads. This, in turn, allows for confident “calling” of variants in the test genome as compared with reference. The scientific community seems to have settled on roughly 30 to 40-fold sequence “depth” as a standard required for a robust, constitutional genome sequence. This means that a minimum of 90 Gigabases of sequence are required, given that the human genome is about 3 billion or so bases in total length.

Tumour sequencing generally requires much greater depth of sequencing, as tumour samples can have mixes of cell types, with only some cells harbouring disease-relevant somatic mutations. How much, exactly, appears to be a somewhat unresolved question, with greater depth revealing rarer mutations. A recent white paper by Illumina suggests that in fairly pure tumour samples (those with >80% cancer cells), good detection of high-resolution somatic events can be achieved with 60x genome coverage.

So where does this leave prospective genome scientists today? For sequencing of the roughly 1.5% of the genome that contains the coding regions of the genes (exome sequencing), there are instruments from Illumina and Ion Torrent, which despite a great deal of misinformation in the marketplace, work more or less equivalently well for this application. For whole genome analysis, if running such experiments in house is desirable, purchase of Illumina’s instruments is the only real option. However, outsourcing is still a viable alternative, with services from Illumina itself, Illumina-based service providers such as South Korea’s Macrogen, or Complete Genomics, offering WGS. Specialized instruments such as PacBio’s RS II and the Oxford Nanopore MinION seem useful for targeted analysis of difficult genomic regions, haplotyping (i.e., determining the arrangement of markers on the same chromosome, either the paternally- or maternally-inherited homologue) in individual genes, or the analysis of small genomes such as those of viruses. Realistically, only ten years since 454 rattled the sequencing world with its GS20, Illumina has now essentially cornered the market for whole-genome sequencing, and appears that only a technological revolution as big as NGS itself might change this.

REFERENCES
Your partner for international laboratory services and products

Supplying the global healthcare, life science research and bio pharma markets

Sample preparation
- PCR purification
- DNA/RNA extraction
- Cell culture products
- Transfection reagents

Gene expression
- Pre-made clones
- Gene expression analysis services

Sequencing
- Sanger sequencing
- Next generation sequencing

Detection
- Genotyping
- Protein detection
  - Antibodies

Bio Storage
- Blood and cell lines
- DNA and stem cells
- Clinical trial material
- and much more...

Talk to the premium global genomic services provider:
email: sales@sourcebioscience.com
To find out more visit: lifesciences.sourcebioscience.com
On both sides of the Atlantic, large scale genome sequencing projects are underway. Both are focused on empowering patients to lead the integration of genomic sequencing into healthcare. Genetic Alliance UK have captured the patient voice in their Patient Charter on Genome Sequencing - A document that makes 15 recommendations for the incorporation of genomic sequencing into the NHS.

It is hoped that the Charter will influence policy and decision-makers to encourage more in-depth and creative public and patient involvement in health care decision-making. As a document it is remarkably powerful as a representation of the patient voice.

FLG: You’ve been Director of Genetic Alliance UK for over 20 years now. A lot has happened in that time. Arguably, now might be the most exciting time for genetics and genomics in healthcare. What have been the biggest milestones for you in your time as Director?

AK: What has been really exciting is seeing the kind of progress towards generating understanding of the biology of some of these horrendous conditions, and in some cases moving towards the point of which we can intervene to change the path, to improve quality and quantity of life for patients and for families who are affected by previously incurable conditions. That steady kind of progress over what’s really a very short timescale has been hugely exciting.

FLG: The patient voice is one that, at times, remains unheard. In the case of the genomic revolution, it is getting harder to ignore the growing calls that patients want to be part of it rather than sit back and let it happen to them. Here in the UK there is a real drive from Genomics England to integrate genomics into routine healthcare. How important is it that patient voice to be able to do this?

AK: The patient voice is absolutely essential; patients are uniquely qualified to add insight into the way in which genetic diseases, rare diseases, affect their lives and the impact on the quality and quantity of life that they can expect to enjoy. It’s also the case that active participation by patients, and by families, as partners in the research and development process increases both the quality and effectiveness of the research. It allows for the identification of targets that are really important to those affected and also for the creation of methods for investigation that patients and families are able more easily to support and participate in. So it has a benefit both for the planning and delivery of research but also in terms of the effectiveness of the outcomes and the opportunity to deliver real benefit.

FLG: In your recently launched Patient Charter on genome sequencing, there was a lot of support and demand for genetic counsellors to be an integral part of the genomic journey for patients. I think it may have surprised some, just how much importance patients place on their profession. Does this show that we need more genetic counsellors, or is it indicative of the lack of a particular skill set in today’s front line practitioners?

AK: It’s an indication of both of those. What patients in this situation want is the opportunity to understand the implications of insights that are revealed through analysis of their genome. That needs a framework within which you can discuss what’s known, what its implications might be and how you are going to handle it in terms of your own life and your family circumstance. For that, genetic counselling is crucial to create the relationship which makes this possible. So clearly there needs to be an increase in the number of qualified genetic counsellors who have gone through proper professional training to build this relationship - but also other healthcare professionals, clinicians, non-geneticists-

GIVING PATIENTS A VOICE

Alastair Kent OBE, Director, Genetic Alliance UK

BOTH THE USA AND UK GOVERNMENTS HAVE COMMITTED TO DEVELOPING PRECISION MEDICINE, THROUGH LARGE SCALE SEQUENCING PROJECTS. THE KEY STAKEHOLDER AND PRIMARY BENEFICIARY IN THIS WILL BE PATIENTS AND THEIR FAMILIES. FOLLOWING THE LAUNCH OF THE GENETIC ALLIANCE UK’S PATIENT CHARTER, ALASTAIR KENT TALKS TO US ABOUT PATIENT-POWER.
nurses, midwives etc. need to understand how to incorporate communications around genetically based insight into their daily practice. It's not just something you can say “oh this is genetics I need to shove it off to a genetic counsellor” and ignore it. It needs to be incorporated into the way in which you build the relationship with patient and family such that you are able to modify your professional practice in the light of emerging knowledge from Genomics England and the 100,000 Genomes Project.

**FLG:** We are often told that we are on the cusp of a new age in healthcare, with the boom of genomic technology and information. Data has had a major impact on how we interact with the world around us. The rise of the smart phone has been the big enabler of the data-centric life style many of us live. Is NGS the big innovation that makes genomics part of everyday life, or is that big ‘smart phone’ moment yet to come for genomics in healthcare?

**AK:** I think it will. But I think we need to make sure that we create the infrastructure, the framework that will allow it to influence the way in which we think about planning and delivery of healthcare. Also the way in which we incorporate insights from genomics into maximising the opportunities for people to enjoy a healthy lifestyle or a healthy life even. But that’s not going to happen without some strategic thinking about how to do this, and without some
investment in creating the insights, and the tools, and the opportunities for that new knowledge to be translated into things that are directly applicable by people in their everyday lives.

**FLG:** Education and engagement is an important aspect in the successful transitioning of genomics into routine healthcare. There has been a lot of research carried out recently into the ethics of patient consent. On the one hand, you have patients themselves telling you that they want as much information as they can get, and support the use of their data for research. On the other hand, you get some in the field who suggest patients are not well informed enough to fully understand what they are consenting to. What is the real risk a patient might run by being involved in a genomic clinical trial that they do not understand fully?

**AK:** Well I guess the primary risk is that they will be disappointed. If they don’t, as it were, form a realistic understanding of what the likely outcomes might be for them (and how long it might take to bring these about), then the biggest risk is that they will be disappointed. They will be frustrated and the future sustainability of the research effort will be compromised. I think it’s probably a throwback to a more paternalistic history that people think that patients are not capable of understanding complex issues incorporating a high level of uncertainty as to the robustness of the knowledge on which they’re being asked to make decisions. I mean, everyday life is full of uncertain decisions - if we never took a decision where there was an element of risk we would still be squatting in caves waiting for the sabre tooth tigers to eat us because we were frightened to pick up the burning brand and scare them off.

**FLG:** Is there a risk of giving too much power over to patients?

**AK:** I would say no. Patients are absolutely the ultimate beneficiaries of the appropriate application of this new knowledge and they should be integral to the process by which the research and development is guided and developed. The rhetoric of “nothing about us without us” actually should be central to everything.

**FLG:** With so much hype around genomics at the moment coming from the very top of politics, is there a danger of over-hype?

**AK:** Oh yes, yes. I don’t think this is unique to genomics, it’s a feature of an awful lot of everyday life that we seem to have lost the understanding of how long things take. That’s not to say that we can just kind of take our time and let things roll on under their own momentum and be laid back. We need to keep the pressure on but it takes time for ideas to mature, to be tried, to be translated into applicable interventions that people can use. We must be sensible in the way in which we expect the benefits to be rolled out. Part of the problem is that we have got so many conflicting time scales - politicians want massive results before the next elections; scientists need a reasonable degree of expectation that if they deliver outcomes that have quality, and are sensible, then their research funding will be renewed; industry needs to think about patent license and what have you. All of these conflicting time scales tend to point us towards short term outcomes, rather than recognition of the need to accumulate quality insight and translate them sensibly into changes in the way in which we do things.

**FLG:** Recently, 23andMe launched in the UK. There are concerns around the legitimacy surrounding reported results, and the potential distress they might cause. A counter argument is that it might be a good way to introduce core concepts to the public and grow familiarity with genetic information. Ancestry tests have already proved quite popular. Should we be doing more to encourage people to explore their genetics on their own terms?
**AK:** I think any offer of access to genetic or genomic information about yourself in a healthcare context should be absolutely supported by comprehensive and comprehensible information delivered in an environment where there are opportunities to ask questions. The thing about genetic information is that it’s a one-way street. You can’t un-know something about a current, or a future, health risk once you’ve got it. So, therefore there needs to be the opportunity for someone to think “do I want this information and how am I going to handle it if I do”. There’s a world of difference between discovering that ostensibly you are descended from a pharaoh and discovering that you may have a predisposition to a life limiting condition. One is entertaining, the other is potentially life changing but not necessarily in a good way unless it comes with the knowledge that allows you to take actions that might mitigate that risk.

**FLG:** With your third Patient Charter launched, what’s next for you and Genetic Alliance UK?

**AK:** We’ll continue to make sure that the voice of those affected by life-limiting conditions and their families is heard in the policy and the practical process for supporting high quality sustainable biomedical research and for the development of timely and appropriate services and support for those affected or at risk. There’s also a lot more to be done in terms of continuing the conversation with patients about genome sequencing as this technology makes its way into mainstream clinical care.

**FLG:** What keeps you motivated to keep fighting for patients?

**AK:** I guess it’s the sense that it’s real. If we can build that critical mass, if we can articulate the patient, the family expectations, which by and large are reasonable and rational and not unrealistic. If we can build that and present it to those people with the opportunity and resources to bring about change than it is likely quality and quantity of life for those with currently intractable genetic diseases will be improved, and it will be improved more quickly because of that input than might otherwise be the case. That, frankly, seems an issue that’s worth getting out of bed in the morning for.

**FLG:** For patients with rare disease one of the most difficult things to deal with can be the delay in getting a diagnosis. What will the recent advances in genomics mean for the rare disease diagnostic odyssey?

**AK:** Potentially they will have a major impact. With increased insight into what whole genome sequencing can tell us about current and future health then this will potentially shorten the time to an accurate diagnosis very significantly. It won’t happen immediately – it’s important to realise that just because we can sequence a genome, doesn’t mean that we understand what that sequence means in all its detail. Building the insight to allow the interpretation is currently the rate determining step. Obviously there are significant efforts being made to try and shorten that process. The clinical interpretation partnerships that Genomics England are setting up is one tool that will hopefully link what we can see or predict in our bodies with what we can determine from our genomes much more quickly. And that linking of societal, of clinical, of observable information with insight into the genome will help us to build that accurate diagnostic framework that will allow patients and families to learn more quickly and more accurately what the situation is in which they find themselves and what they can do about it.

**FLG:** What do you think healthcare will look like in the UK in 15 years time?

**AK:** Who knows! In terms of technological possibility, we’ve seen progress happening much more quickly than we would have thought even a few years ago. If someone had said in 2000 when the first draft of the human genome was produced, after hundreds of thousands of scientific years of effort and billions of dollars of investment, that in 15 years you would be capable of sequencing a genome for under a $1000, people would have said “get real, which planet are you on!!?” So I think the opportunity that scientific progress is creating is huge. But of course healthcare systems are hugely complex organisations which have got a kind of momentum all of their own, so changing the way in which we do things in healthcare to adapt to these new insights into disease and its management will require consistent and sustained pressure and the ability to make tough decisions about allocating and reallocating resources in order to build a framework which will allow this new knowledge to be translated into better outcomes for patients.

**FLG:** How can people help to get us there?

**AK:** By engaging with the issues. By thinking about how the new knowledge will transform their understanding of situations and by being creative in their willingness to embrace change and new ways of working with patients, with families and across professional and institutional boundaries so that bright ideas can be translated into patient benefits in a way that is timely, that is user friendly and that is sustainable.

AK: I think any offer of access to genetic or genomic information about yourself in a healthcare context should be absolutely supported by comprehensive and comprehensible information delivered in an environment where there are opportunities to ask questions. The thing about genetic information is that it’s a one-way street. You can’t un-know something about a current, or a future, health risk once you’ve got it. So, therefore there needs to be the opportunity for someone to think “do I want this information and how am I going to handle it if I do”. There’s a world of difference between discovering that ostensibly you are descended from a pharaoh and discovering that you may have a predisposition to a life limiting condition. One is entertaining, the other is potentially life changing but not necessarily in a good way unless it comes with the knowledge that allows you to take actions that might mitigate that risk.

**FLG:** With your third Patient Charter launched, what’s next for you and Genetic Alliance UK?

**AK:** We’ll continue to make sure that the voice of those affected by life-limiting conditions and their families is heard in the policy and the practical process for supporting high quality sustainable biomedical research and for the development of timely and appropriate services and support for those affected or at risk. There’s also a lot more to be done in terms of continuing the conversation with patients about genome sequencing as this technology makes its way into mainstream clinical care.

**FLG:** What keeps you motivated to keep fighting for patients?

**AK:** I guess it’s the sense that it’s real. If we can build that critical mass, if we can articulate the patient, the family expectations, which by and large are reasonable and rational and not unrealistic. If we can build that and present it to those people with the opportunity and resources to bring about change than it is likely quality and quantity of life for those with currently intractable genetic diseases will be improved, and it will be improved more quickly because of that input than might otherwise be the case. That, frankly, seems an issue that’s worth getting out of bed in the morning for.

**FLG:** For patients with rare disease one of the most difficult things to deal with can be the delay in getting a diagnosis. What will the recent advances in genomics mean for the rare disease diagnostic odyssey?

**AK:** Potentially they will have a major impact. With increased insight into what whole genome sequencing can tell us about current and future health then this will potentially shorten the time to an accurate diagnosis very significantly. It won’t happen immediately – it’s important to realise that just because we can sequence a genome, doesn’t mean that we understand what that sequence means in all its detail. Building the insight to allow the interpretation is currently the rate determining step. Obviously there are significant efforts being made to try and shorten that process. The clinical interpretation partnerships that Genomics England are setting up is one tool that will hopefully link what we can see or predict in our bodies with what we can determine from our genomes much more quickly. And that linking of societal, of clinical, of observable information with insight into the genome will help us to build that accurate diagnostic framework that will allow patients and families to learn more quickly and more accurately what the situation is in which they find themselves and what they can do about it.

**FLG:** What do you think healthcare will look like in the UK in 15 years time?

**AK:** Who knows! In terms of technological possibility, we’ve seen progress happening much more quickly than we would have thought even a few years ago. If someone had said in 2000 when the first draft of the human genome was produced, after hundreds of thousands of scientific years of effort and billions of dollars of investment, that in 15 years you would be capable of sequencing a genome for under a $1000, people would have said “get real, which planet are you on!!?” So I think the opportunity that scientific progress is creating is huge. But of course healthcare systems are hugely complex organisations which have got a kind of momentum all of their own, so changing the way in which we do things in healthcare to adapt to these new insights into disease and its management will require consistent and sustained pressure and the ability to make tough decisions about allocating and reallocating resources in order to build a framework which will allow this new knowledge to be translated into better outcomes for patients.

**FLG:** How can people help to get us there?

**AK:** By engaging with the issues. By thinking about how the new knowledge will transform their understanding of situations and by being creative in their willingness to embrace change and new ways of working with patients, with families and across professional and institutional boundaries so that bright ideas can be translated into patient benefits in a way that is timely, that is user friendly and that is sustainable.
Helping You Deliver the Benefits

EXPERIENCE THE MOST EXCITING TOPICS IN GENOMICS WITH SOME OF THE BIGGEST LUMINARIES AND INNOVATORS IN THE FIELD.

Focused, purposeful content and visionary sessions in a fun, engaging and vibrant setting. Come be part of something extraordinary and help deliver on the promise of genomics for patients.

1000+ Attendees
50+ Exhibitors
100+ Speakers
70+ Sessions

PRICES FROM $150

Exclusive $100 ticket price for Front Line Genomics Magazine readers.* Enter code FLGReader when you register at festivalofgenomics.com

FOR MORE INFORMATION ON PARTNERING:
Contact our commercial director Freddy White:
T: +44 (0) 207 384 8155
E: Freddy@frontlinegenomics.com

FOR FULL FESTIVAL OF GENOMICS DETAILS AND UPDATES:
Visit our website:
www.festivalofgenomics.com

*Not Applicable to Vendor Festival Passes

Register now at www.festivalofgenomics.com
DON'T MISS PLENARY SESSIONS BY THESE INDUSTRY LEADERS

GEORGE M. CHURCH
Professor of Genetics
Harvard Medical School

ERIC D. GREEN
Director
National Human Genome Research Institute

C. JIMMY LIN
President
Rare Genomics Institute

JAMES R. LUPSKI
Professor of Pediatrics
Baylor College of Medicine

ROBERT C. GREEN
Associate Professor of Medicine
Harvard Medical School

J. CRAIG VENTER
Founder, Chairman
J. Craig Venter Institute (JCVI)

JONATHAN BINGHAM
Product Manager
Google Genomics

of Genomics to Patients, Faster.

Tech Forum Stage
Poster Area
The Main Stage
TED-Style Plenary Session

The Blue Stage
The Main Stage
The Play

Exhibition Open
23-24 June

SPONSORS:

@FLGenomics #GenomicsFest
The age we live in is increasingly data centric. Businesses are constantly looking for new ways to create and leverage information; while we are quick to adopt new tools and applications that make our personal lives easier. Central to all of this is Google. Be it, Google Analytics, Google Maps, or the Google search engine, the company has changed how we look and interact with the world around us.

As the field of genomics hits its stride, there is a swelling volume of sequencing data being amassed. So how will Google Genomics help us do things we can’t even imagine right now?

**FLG:** As the field of genomics is a rapidly growing and very exciting field to be a part of, what initially led you to be part of this community?

**JB:** Ever since I was a student, my interests have always been interdisciplinary. My practical side was drawn to computer science. My heart was drawn to applications far away from the typical software companies and corporate IT - including simulation, modelling, social science, the humanities. Bioinformatics was a new and exciting field, and I wanted to see where it would lead.

**FLG:** As the field of genomics is a rapidly growing and very exciting field to be a part of, what initially led you to be part of this community?

**JB:** Ever since I was a student, my interests have always been interdisciplinary. My practical side was drawn to computer science. My heart was drawn to applications far away from the typical software companies and corporate IT - including simulation, modelling, social science, the humanities. Bioinformatics was a new and exciting field, and I wanted to see where it would lead.

**FLG:** What made you decide to make the move from PacBio over to Google?

**JB:** From my time at PacBio, I knew that the health care system was about to be swept up in an information tsunami from ubiquitous DNA sequencing as well as imaging and sensors. When I learned of the potential to bring Google’s scale and technology to bear on these challenges, of course I wanted to be a part of it.

**FLG:** Google has developed a reputation for being able to do just about anything with sufficient data. There’s a lot of interest around what Google will be able to do for the field of genomics. Does that add any extra pressure?

**JB:** From chairman Eric Schmidt, who sits on the board of Mayo Clinic and Broad Institute, to founder Sergey Brin, who was married to the founder of 23andMe, to senior fellow Jeff Dean, who was an intern at the Center for Disease Control, to many other software engineers with backgrounds or friends who work in the field, there’s an appreciation for the challenges facing the genomics field. There’s also some healthy realism about what is and isn’t possible with data analysis. I’d say there’s more curiosity, and optimism, than pressure.

**FLG:** How is the amount of data produced within a project such as veteran affairs million genome project, comparable to the size of data Google handles on a daily basis?

**JB:** To put it in perspective, users upload over 300 hours of video per minute to YouTube. At typical smartphone video resolution, that’s like loading and indexing the raw sequence information from 30 whole human genomes every single minute. The Google Search index itself is over 100 petabytes in size. That’s like storing the genetic variant calls from 100 million people, and being able to look up results in a quarter of a second.

**FLG:** What makes genomics a market of interest for Google? Between Google Genomics, the Baseline project and other moonshots, it seems like you guys are investing quite seriously in healthcare in general?
“IMAGINE IF YOU NEVER HAD TO DOWNLOAD DATA AGAIN; IF YOU DIDN’T HAVE TO MANAGE A COMPUTER CLUSTER, EVEN A VIRTUAL ONE IN THE CLOUD; IF YOU COULD USE STATISTICAL TOOLS, AND YOU NEVER HAD TO THINK ABOUT RUNNING OUT OF COMPUTER MEMORY”

Google Genomics got started up from the grassroots efforts of software engineers, in their 20% time, because they were interested in the field.

GOOGLE GENOMICS

Google Genomics provides an API to store, process, explore and share DNA sequence reads, reference-based alignments, and variant calls, using Google’s cloud infrastructure. It offers users data storage, processing power and allows them to explore and share their data more effectively.
**JB:** Given the growing importance of information science within the healthcare system, and the role that Google has played in developing technologies for large-scale analysis, it only makes sense for Google to get involved to help tackle some of the data challenges in genomics. In the broader health space, we can bring technology to bear as somewhat of an R&D lab for our life sciences partners, helping them answer big questions more quickly and efficiently.

**FLG:** How was Google Genomics started?

**JB:** Google Genomics got started up from the grassroots efforts of software engineers, in their 20% time, because they were interested in the field. They had the idea that Google Cloud Platform, and a technology called BigQuery, which we use extensively to interactively explore trillions of log entries, might be a natural fit for mining sequence data. Other software engineers had thoughts about how to compress and index DNA sequences. A Google VP with a deep personal interest in cancer genomics saw this activity, and helped catalyze the formation of a full-time team. That’s how it began.

**FLG:** Cloud technology is quickly becoming an indispensable tool for business and personal life. You’re collaborating with Institute for Systems Biology on the NCI Cloud Pilot program. What kind of impact do you think cloud technology is going to have for genomics research? What will it enable us to do that we couldn’t do before?

**JB:** Cloud technology makes a lot of sense in general. Most companies today don’t even think about building their own power generating plants. Why would they build their own data centers?

In the first wave, many bioinformaticians think about cloud technology the same way they think about university compute clusters. They imagine doing the same things, but not having to deal with an IT department or manage the hardware themselves. And they’re saving money and moving faster with that approach, so it’s a great beginning.

The real power will come when we embrace the fact that cloud computing brings entirely new capabilities. Imagine if you never had to download data again; if you didn’t have to manage a computer cluster, even a virtual one in the cloud; if you could use statistical tools, and you never had to think about running out of computer memory; if you didn’t have to downsample to analyze your complete cohort; if you could have an idea, and test it right away, without having to wait weeks or months to write complicated code and manage servers or even virtual machines. Imagine if you could do this from anywhere in the world, without needing to be at a well-funded university with a large IT budget, without even needing more than a tablet or phone. We’ll have high school science fair participants in developing countries asking questions that faculty at our top universities have never thought to ask, and making discoveries that matter for basic science and for human health.

**FLG:** As genomic database tools grow and improve, what do you see as being the greatest benefit for drug developers and what will this mean for patients in the near future?

**JB:** For drug developers, there’s long been the idea of rational drug design, and of personalized medicine. Having more complete genomic databases will help, especially when combined with other kinds of clinical information. The benefits will come from getting to a scale where machine learning and sophisticated analytical methods become possible. That means there’s going to need to be more collaboration and data sharing.

Google Genomics is designed to facilitate that kind of collaboration by building on the open standards developed by the Global Alliance for Genomics and Health. A scalable, interoperable genomics platform will make possible entirely new applications. Drug developers and patients will benefit from the combined contributions of the community.

**FLG:** What’s your view on open source data sharing i.e. personal genome project, opensnp.org?

**JB:** Google Genomics supports sharing genomic data as widely or as narrowly as the institutional review board and the patient consent forms allow. If a study is limited to one set of researchers for one use only, we support setting restrictive access, and in fact that’s the default. If a study is open to collaborators, that’s supported as well. If it’s open to researchers for other qualified use through an application process, that’s supported. And if the project is truly open to anyone, like Personal Genome Project and opensnp.org, that’s supported too. Our belief is that sharing data more widely will benefit more people. Bartha Knoppers and others at the Global Alliance for Genomics and Health are framing the issues in a way that we support.

**FLG:** How easy is it to integrate a project into the cloud once it’s already begun?

**JB:** There are multiple ways to make the transition, some of them are really simple. To get started with Google Genomics, you can take the variant calls from your DNA sequencing experiments and load them up to the cloud, and begin exploring trends and patterns right away. In other words, you can focus on adding new capabilities, rather than replicating what you’re already doing on local compute resources. You can get started in an afternoon.

You can also take your entire IT infrastructure and move it to Google Cloud Platform, using Google Compute Engine virtual machines, Google Cloud Storage for files, Google Cloud SQL as a relational database, Google Cloud Datastore as a NoSQL alternative. Anything you’re running locally, you can probably get it set up in the cloud, without too many changes.

Longer term, you’ll get the most benefit from thinking about infrastructure differently.

**FLG:** At the moment these tools seem to largely be targeted to bioinformaticians, or at least researchers with a familiarity with analytical data outputs – How user friendly can you realistically make things without sacrificing depth?

**JB:** We’re starting with the applications where Google can bring the most unique value. We have a lot of great technologies for working with large data sets, and bioinformaticians are excited to get access to those tools.

Institute for Systems Biology is building a more biologist-friendly interface for the Cancer Cloud Pilot. Autism Speaks is building a
researcher portal for the MSSNG project. Over time, the trend is toward friendlier interfaces, making advanced tools more widely available.

**FLG:** What potential risks do you foresee for patient privacy, how do you think these can be handled?

**JB:** Google Cloud Platform supports HIPAA covered entities by signing Business Associates Agreements. We encrypt all information where it's stored and when it's transferred over the network. There are many security and compliance certifications that provide assurance that information is shared only as planned.

**FLG:** What have been the highlights for you since Google genomics first began? Atul Butte has compared the work at Google genomics to how travel agents felt when they saw expedia.

**JB:** So far a few of the highlights have been the positive community response to the announcement that Google Genomics had formed and joined the Global Alliance for Genomics and Health; the ICGC DREAM somatic mutation calling challenge; the NCI Cancer Cloud Pilot with Institute for Systems Biology; the launch of the MSSNG database for autism research; and the first peer-reviewed publication to mention Google Genomics. This is just the beginning. Over the course of 2015 and the coming years, the momentum will increase.

**FLG:** In 10 years’ time, will the term ‘Googling’ have taken on a life of its own in genomics?

**JB:** I’ve been struck by how often I hear that clinicians and researchers find out more about a genetic variant by starting with a Google Search. Results today are based on web pages. Google Genomics is building optimized storage, processing, and exploration that’s adapted to the domain. We’ve barely begun to see what’s possible. I think we’ll all be surprised by how genomics as a field looks in 10 years.

**FLG:** What do you think the next big story in genomics will be?

**JB:** The small stories are easier to predict – the incremental improvements, as technologies that exist today are applied to genomics. The big news stories that excite me most will be the unexpected breakthroughs and insights. These are exciting times for genomics.

**FLG:** We hear you’ve got some good restaurant recommendations for people travelling into the Bay Area. What are your top five picks?

**JB:** When not studying DNA, we should all make a point to eat foods filled with plant and animal DNA. In Berkeley, Gather offers some of the most inventive farm-to-table food. In San Francisco, Lers Ros is my favorite Thai restaurant. Orenchi in Santa Clara is a top spot for ramen. Madras Cafe in Sunnyvale serves excellent South Indian. Koi Palace in Daly City has amazing dim sum. Is that five already?

**FLG:** Any last questions, comments for readers?

**JB:** Thanks for reading about Google Genomics. We’re thrilled to collaborate with the global genomics community and make an impact together.
In 1997, we were living in the era of ‘The Human Genome Project’. Unlocking our genetic code would seemingly answer many of life’s questions and lead to the eradication of diseases like cancer. Amidst all of this, a particularly interesting film emerged from writer/director Andrew Niccol: 1997’s GATTACA.

Now in 2015, GATTACA is more relevant than ever. In the film’s version of the world, Galton’s ideas on eugenics have resurfaced and taken hold. Parents choose their optimal gametes for reproduction, to produce the healthiest offspring they can. As this approach becomes widespread, it has knock-on effects on society. In particular for physically demanding roles, that rely on genetic profiling as their interview process.

As a thoughtful piece of entertainment, GATTACA is great. Strong performances from the lead cast of Ethan Hawke, Jude Law and Uma Thurman, bring this dystopian vision to life. The science isn’t over played, as with modern sci-fi offerings like Interstellar. It’s there to give context and reason for an enjoyable drama to unfold.

GATTACA raises interesting questions about where we set boundaries. At what point do we go from saving lives, to giving people a chance at a normal life? How do we determine what normal looks like? What’s does a healthy/optimal genome look like? Google’s Baseline study is trying to figure that out as we speak.

In a hypothetical world, where we have a better understanding of the genomic basis of disease and tools to act on it, how far do we go? What is an acceptable level of health risk we are willing to live with? As large-scale genome sequencing projects come to fruition, we should know a lot more about the diseases that affect us. With developing technologies, we might soon be in a position to ask ourselves some challenging ethical questions.

Fortunately, the law has kept up with developments in the field thus far. A GATTACA style future is out of the question while laws like the Genetic Information Nondiscrimination Act (USA) of 2008 is in place. But many employers already require psychometric and aptitude tests to predict the suitability of candidates. Would genetic profiling, be any more of a form of discrimination than today’s tests if they were relevant to the job? You’re still looking at probabilities after all.

You can go around in circles debating the ethical landscape of genetics. As fun as it can be, one does get the feeling that a lot of these hypothetical scenarios might not be too far away anymore. While I love GATTACA, I think a more interesting story would be the political and social drama that leads to the presented vision of the future. That being said, GATTACA is a great modern day parable. A modern ‘tortoise and hare’ story that champions the rewards of hard work and determination over natural talent.

**VERDICT: GREAT**

Great acting, great pacing, great cinematography… there’s not a lot not to like. Raises interesting and relevant questions, while keeping you entertained. Worth watching, just so you can answer the inevitable ‘GATTACA’ questions you’ll get asked, by non-scientists, about genomics.

**PROS**
- Thought provoking, and still highly relevant
- Great for introducing complex ideas to the uninitiated
- Stands up in its own right as a thriller/drama

**CONS**
- Genetic health seems to be directly proportional to amount of hair gel used
- Outside of the main cast, the world and characters can seem a little underdeveloped

**RATING** 8
ACCELERATING PROGRESS IS IN OUR GENES

You see the world as it is: a complex system of fundamental elements. So do we. Whether your focus is discovery or clinical research, we help you see deeper.

See Deeper. Reach Further.

Innovations that will redefine Genomics:

- **OneSeq**: Detect genome-wide CNVs, SNPs, indels and LOH in one assay
- **HaloPlex**: Enables detection of high sensitivity target enrichment using molecular barcodes
- **SureSelect FFPE Solutions**: From sample to data analysis
- **SureSelect Clinical Research Exome**: Exome solution with deeper coverage of disease-associated regions

DISCOVER MORE: www.genomics.agilent.com
seqcap EZ choice
Target Enrichment System

Sequence What Matters

Whole Genome Sequencing can be difficult. Your target enrichment choice is EZ... SeqCap EZ Choice.

- Increase the efficiency of targeted capture — benefit from 2.1 million overlapping probes.
- Discover more variants while optimizing sequencing efficiency — multiplex up to 12 samples in a lane and achieve higher capture uniformity, minimizing your sequencing.
- New! Overnight hybridization protocol increases your speed to discovery.

For more information, visit www.nimblegen.com/seqcap or contact your Roche representative.

For life science research only. Not for use in diagnostic procedures.
NIMBLEGEN and SEQCAP are trademarks of Roche. © 2015 Roche 408-60986-0115