2008 Presidential Address: *Principia Genetica*: Our Future Science

Aravinda Chakravarti[1,*]

Aravinda Chakravarti

Good afternoon, and welcome to the 58th annual meeting of The American Society of Human Genetics in historic Philadelphia. I am Aravinda Chakravarti, your Society's President, and I am honored and proud to have represented you this year. I want to spend my time with you today reflecting not on what we have already accomplished, which is substantial, but on the future of our science: our drive toward a *Principia Genetica*.[1]

We live in very historic times. Each generation believes this dearly, but our times are special indeed. As all of you know, we had a national election in the United States last week. The campaigning is finally over; Americans have overwhelmingly elected Barack Obama as President, and we stand many inches taller in having overcome the prejudices of our recent past to elect a person of color to the highest office in the land. The President-elect, alluding to the momentous changes of this day and our times, said:[1]

> "A man touched down on the moon, a wall came down in Berlin, a world was connected by our own science and imagination."

I consider it telling that the new president considered the importance of connecting science with our imagination as a central agent of change in our world, and so I have hope that science will return to its central position in the life of this great country. For us as human geneticists, this annual meeting is *our* forum to connect science with our imagination, thereby creating new possibilities.

The American Society of Human Genetics

Ours is a young Society at 60 years, but this is quite a long haul in the history of genetics. Our Society was born on September 11th, 1948, starting with the presidency of the great geneticist Hermann Joseph Muller.[2] We are fortunate to have a well-documented birth, and the papers incorporating the organization with its list of members are available within our archives.[2] These members could not have envisioned the future progress in our science, the open culture of our work, or the borderless collaborations that have ensued. Today, The American Society of Human Genetics can no longer afford to be cast in terms of its geography but needs to be cast by its science and its future. Full membership is now open to all individuals around the world based on their professional interest. They can, and we expect them to, participate in every aspect of the academic life of this Society and engage in ventures that use the creativity among all our members. I take our Society's motto of "Discover-Educate-Advocate" seriously: consequently, we need to take leadership in promoting human genetics worldwide. As a start, our Society is contemplating meetings beyond our national assemblage and beyond North American shores, including smaller meetings on focused topics in partnerships with other organizations. As an inaugural, in March 2009, The American Society of Human Genetics will organize and host, in collaboration with *HUGO* and the Nature Publishing Group, an international meeting in Singapore entitled *Genetics and Genomics of Infectious Diseases*.[3] I hope that many of you will attend, send your trainees, or recommend the meeting to your colleagues and collaborators. I also hope that these meetings will bring in new members and cement the connections between old members, here and around the world.

[1]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA
*Correspondence: aravinda@jhmi.edu
DOI 10.1016/j.ajhg.2010.01.005.

The Logic of Human Genetics

The theme of my talk is "Our future science," and I have called this *Principia Genetica*, in flattery and tribute to the *Principia Mathematica*, published as three books in the early part of the twentieth century.[4] These were monumental works, compiled, written, sweated, and argued over by Alfred North Whitehead and Bertrand Russell. Such was their effort that a fourth book on geometry never materialized because they were intellectually exhausted. They succeeded in showing that all pure mathematics followed from purely logical premises, using only concepts defined in logical terms. These books are important and legendary not only in mathematics, but stand as one of the pinnacles of human creativity and logic.

Today, we are at a new dawn of human genetics. The last many decades of work, by the many members of this Society and others, were largely descriptive, yet teased from them a set of principles of human inheritance, many unique, such as imprinting or expansion mutations. However, we have much hard work ahead to uncover the full logic of human genetic inheritance. Indeed, we need our *Principia Genetica* with a logical foundation such that any new inherited phenomena can be described in these terms. In this new phase in human genetics, we need to expand our understanding of genetic biology and increasingly focus on its mechanisms. I strongly believe that only from such very fundamental understanding will appear the compelling "translations of knowledge" for tomorrow. And that is my main message to you today. In my talk today, I will focus on the practice of genetics in the genomic age; genetic disease and personalized medicine; and my own opinions and thoughts on the nature of the gene and human diversity.

Genetics by Sequence—Human Genetics in the Genomic Age

We no longer do human genetics by breeding but through the DNA sequence. The classical disadvantages of working with humans as a genetic system, small numbers of offspring and the inability to perform controlled crosses, have been demolished in our genome world. From 1993 to the present, international collaborative projects such as the Human Genome Project and the HapMap, ENCODE, Structural Variation, and 1000 Genomes projects have altered our ability to probe the human genome efficiently. I have no doubt that many more such projects will follow. But what did we seek to understand from all of this data? What new principles did we learn?

There is absolutely no doubt that we have profoundly altered our understanding of the human genome and how altering its functions leads to variation in phenotypes. First, we now know that humans, indeed all mammals, have a small repertoire of genes whose molecular functions can be modulated in numerous ways. Second, the structural and functional diversity of RNA is entirely new and astounding. Third, and an unanticipated genomic feature, the widespread conservation of noncoding DNA in amounts greater than that in coding DNA is both real and largely unexplained. Fourth, we now know of pervasive transcription across the genome whose meaning is elusive. Finally, there is abundant polymorphism, in SNPs (single nucleotide polymorphisms) and CNVs (copy number polymorphisms), to convince even the skeptic of their widespread genetic impact. Indeed, we have been surprised and astounded by almost every feature of our genome's structure and function. And there is no doubt in my mind that we are in store for many more surprises. Sequence-based biology and genetics is only beginning, and in the years ahead, according to Eric Lander,[5] we will wish to learn *all* sequence in the human genome, *all* human genetic variation and its relationship to disease, *all* functional elements in the genome, and *all* signatures of cellular response. To these aims I would add: know how to modulate *all* genes and how to *predict* functional effects. These are grand and difficult feats to be sure, but they are problems worth every ounce of our attention and effort. Understanding the central logic of how information is stored in our genomes and how it plays out its role in normal physiology and disease *is* our future science.

Genetic Disease and Personalized Medicine

The central feature of genetics is the ability to predict phenotype based on genotype, once the details are understood. The promise of individualized therapies relies on this edict. Let's consider two successes, one old and one new, to see how our discoveries evolve into practice, in both cases depending crucially on basic genetics knowledge.

In human genetics, the ABO blood group system has played a major and classical role in our understanding of many genetic principles, from early studies of its polymorphisms and inheritance, to genes that modify cell surface expression, to the biochemical basis of red cell antigens. In medicine, the ABO types have been critical to blood transfusions ever since this therapy became widespread during World War II. ABO genotyping is global, largely used without a geneticist's help, and yet, it's the most widely used genetic test. Intriguingly, it's been widely used, despite variable frequencies of the ABO alleles across human groups and without regard to the "race" of either the donor or the recipient. This is personalized medicine at its very best, but newer, deeper understanding promises more exciting possibilities for the future. Studies by Clausen and colleagues[6] have shown the practical possibility of enzymatically reducing all red cell types to "O," promising an eventual common therapy for all individuals without regard to their genotype.

I use the ABO example to argue that the broader medical community will utilize genetics whenever we can provide an effective solution to a dire need. I personally found this out, recently, in a very different part of the world. In trying to deal with an episode of malaria in my mother in India, I discovered that local physicians test patients for *G6PD* deficiency prior to beginning primaquine therapy. This test is

being widely used because of need, but not one of the physicians using it thought of it as a genetic test. Thus, a compelling need with a clear solution is the sine qua non of adopting personalized medicine.

My second example is from work by two of my colleagues whose contributions, decades apart, exemplify what the recent future is likely to be. Victor McKusick used Marfan syndrome to highlight the many genetic and pleiotropic features of heritable disorders of connective tissues.[7] Hal Dietz, in collaboration with McKusick, identified the gene as fibrillin 1 in 1991 and subsequently showed the Marfan phenotype to arise from molecular *deficiency* of fibrillin 1.[8] However, a possible therapy for Marfan syndrome required a second crucial genetic insight; namely, that the phenotype arose from a consequent molecular TGFβ *activation*. This sleuthing led Hal Dietz to suggest a commonly available angiotensin II receptor antagonist drug (Losartan) as likely to reduce, or even reverse, the life-threatening aortic root enlargement, as his recent work suggests.[9] This finding may have a much broader use in nongenetic forms of aortic aneurysms, suggesting personalized medicine is not restricted to diagnosis but widely applicable to therapy, once we understand the molecular basis of the pathophysiology.

Human geneticists have a critical role to play if we are to see many more successes such as these. Our community has been involved in the mapping and elucidation of the molecular basis of close to 4000 disorders. But there is much unfinished business. We still fail to understand the molecular basis of the thousands of remaining Mendelian disorders: each of them is an unlearned lesson that we could use for gaining insight into new treatments. There is both a need, since there are still too many childhood developmental and cognitive disorders of unknown origin, and an opportunity, since we have the genomic tools, to complete this task rapidly.

The Riddle of Complex Disease Inheritance

There are other classical unsolved problems as well, particularly one that has gnawed at us for the last 100 years and ever since the birth of our discipline. The genetic architecture of complex traits is not well understood even today, and is a puzzle that Francis Galton, Karl Pearson, and William Bateson argued over in very strong, sometimes personal, terms in the 1910s (Mendelian-Biometrical debate).[10] The vigor of their arguments challenged the basic understanding of the nature of inheritance, as to whether the logic of inheritance was primarily from rare mutations of large effects (Mendelian) or whether small additive effects of many genes (Biometrical) was the norm. Then, biometricians thought rare mutations to be inconsequential in evolution while Mendelians argued the small effects to be of environmental origin and of no importance to heredity. Although the early geneticists understood the nature of the problem, they clearly did not have the tools to crack it open or resolve these questions.[11] But, it is now time to focus intensely on this task, because its solution is central to our understanding of how information is encoded in our genomes and how it is compromised in human disease. Today, once again, we appear to have two extreme notions of human inheritance: one that of a single gene and the other that of an infinite number of genes. We need solutions to the vast discontinuity that lies in between these two "models" since, in my view, that understanding will be central to how we might approach therapies for multifactorial defects.

Over the past two years we have made some progress in identifying many common polymorphisms that influence the risk of common disease. Some will disagree with me and claim dramatic progress, but this depends on whether we look back at the start or look forward to our destination. Nevertheless, there are intense debates, much like the ones Galton-Pearson-Bateson had, on the importance of the findings and their meaning for common diseases. Current research on mapping common disorder genes by genome-wide association studies (GWAS) have had five major findings: (1) each complex trait or disease has contributions from many genes with highly polymorphic effects; (2) there are often multiple, independent effects at each locus; (3) genetic effects are mostly small; (4) there is little evidence of locus interactions; and (5) there is a significant role for noncoding variants. These results raise intriguing questions: Is the widespread human polymorphism the basis for these traits? How much of this variation is under natural selection? What is the diversity of sequence variation and their effects at any one such locus, and does it involve only one gene? Does the high fraction of noncoding variants suggest mostly regulatory than structural variation? How do so many variants across the genome collaborate to affect a phenotype? Admittedly, the knowledge at this stage is incomplete. Some have objected to the utility of GWAS since they cannot adequately explain anywhere near 100% of the trait variation or disease risk. This remains an important conundrum; however, this is not an excuse for not forging ahead with understanding disease pathophysiology that each new mapped locus provides. As you will experience at this meeting, there is a rise in our understanding of complex diseases. However, we do need to fill the big gap between mapping and identifying the gene and its disease mechanism: this remains the central task for human genetics to prosper.

The challenge in understanding mechanisms in complex diseases, in my opinion, revolves around the hypothesis that humans have a much greater disease burden that arises from regulatory than structural variation. In other words, our disorders are those of protein quantity than protein quality. Compounding this challenge is the enigma of how small the genetic effects can be. The smallness of the effect is constrained by development and natural selection but also does not distinguish between a smaller effect in everyone versus a larger effect in a subset of individuals. These two contrasting scenarios suggest different mechanisms and different implications for

eventual therapies. The smallness of the effect may make predictions of disease risk poor but is no guide to the utility of any therapy we can fashion from the identified target. So, it is armed with this knowledge that many scientists, as well as the lay public, have gotten interested in disease phenotype prediction in individual genomes.

Personal Genomics and Phenotype Prediction

The first two recognized individuals to have their genomes sequenced are J. Craig Venter[12] and James D. Watson.[13] Their genomes are publicly available, have been subjected to many analyses, and prompt us to ask: What have we learnt from their genomes? Maynard V. Olson[14] wrote, on the completion of the Watson genome, that: "The application of new technology to sequence the genome of an individual yields few biological insights. Nonetheless, the feat heralds an era of 'personal genomics' based on cheap sequencing." I am sure he would have added the caveat "so far." We, as geneticists, have been in two minds on the value of personal genomes for individualized medicine: we know that the codes to our phenotypic information must be embedded in these sequences, but we are still poor at identifying them, obtaining biological insights, or predicting disease. This duality leads to caution on our part but does not deter all, nor should it. Misha Angrist, a human geneticist and one of ten participants in George Church's Personal Genome Project (PGP), has written[15] that "…personal genomics is criticized both for the paucity of 'real' information it delivers and for the possibility of tragic consequences resulting from people receiving 'life-altering' genotype and sequence data in an unmediated way…by pretending personal genomics is still years away, we deny people's agency and autonomy, we ignore reality, and we do so at our peril." This is a clear call to improve the prediction of phenotypes from individual genomes, but how good might this be?

The primary challenge in predicting phenotype from genotype is not sequencing technology, which is improving very rapidly, but in *understanding* the myriad variations in our genomes and their effects. The success of individualized medicine is currently limited by our understanding of the meaning of most sequence changes, most of which are rare, have not been seen earlier, yet many are not without phenotypic consequence. For common variants, coding or noncoding, repeated observation in different individuals allows us to both compile them and test their effects on human phenotypes in a rigorous manner. Witness the rise in GWAS. However, we cannot database and test all variants this way, and this is impractical for most human populations. So, how does one attach meaning to DNA variants generally?

Our current attempts at assessing the biological consequences of a variant are crude and, primarily, involve segregation analysis and phenotypic correlations, identifying de novo mutations, testing differences in cases and controls, searching for evolutionary conservation, and, for some classes of proteins, looking at measured or modeled activity. But we need to improve on these methods vastly. It is not simply a matter of collecting more data and "databasing" but, rather, there is a need for a new *theory* that can reliably predict what every sequence change in the genome can lead to. In other words, we need to understand the numerous biological codes in genomes and how they can be compromised in disease.

One aspect of the solution is to investigate how often any nucleotide in the genome changes and how long it survives; i.e., estimation of the human mutation rate. This is one of the most basic parameters of our genome, and yet all current estimates are over four decades old. Although we are making some headway in understanding somatic mutation and its rates, from sequencing cancer genomes, the data on germline mutation rates, thought to be ~2–5 $\times$ 10^{-5} per gene per generation, are ancient, indirect, and nonrepresentative. The primary reason for assessing nucleotide turnover is that extant sequence variation is confounded by past and present demography. We need to untie this knot to illuminate both human biology and human history. We need to know the mutation rate at the sequence level, how it varies across the genome and across individuals, and, importantly, which variants are kept and which discarded and why.

The second aspect is to improve the theory that allows better recognition of genomic codes. There has been some exciting progress in this direction. At least for proteins, Rama Ranganathan and his colleagues at UT Southwestern have developed "statistical coupling analysis" and experimental tests to identify specific residues in a protein that provide critical function with a resolution far above standard conservation analysis.[16] Future developments in this arena can be crucial for identifying protein-encoding mutations in human genomes in a de novo manner. But the problem goes beyond understanding proteins. We have widespread conservation in noncoding DNA, and we need to understand those bases as well. Groundbreaking work by Eran Segal and his colleagues at the Weizmann Institute in Israel have shown how mathematical modeling on the *Drosophila* genome sequence, coupled with accurate experimental data on transcription factor abundance and sequence recognition, can accurately predict the expression pattern of the majority of segmentation genes during fly embryogenesis.[17] It is quite possible that these kinds of analyses can uncover the bases where genetic changes are likely to disturb gene expression or its regulation. I expect that these theoretical-cum-experimental approaches will be absolutely necessary for understanding the functions of the human genome and which sequence variations can compromise its biology and genetics.

The Nature of the Gene and Human Diversity

The changing notions of functions, how they are coded and how they might affect phenotypes, are likely to lead to a great revision in our thinking of the "gene." The notion of the gene, in the minds of the public and to

many scientists, is that of an inflexible machine with deterministic outcomes. In this view, specific genotypes lead to predetermined, fixed, and specific phenotypes. Our community's choice of the Mendelian rather than the biometrical path, in our early history, has led to remarkable achievements in understanding, but relied largely on the biology uncovered by single gene mutations of qualitative effect. This is the basis of our belief in prediction and the source of medical genetic interventions. But we know, and have known for a long time, that phenotypes do not necessarily obey these neat patterns. In fact, the vast majority of human phenotypes do not Mendelize: Mendelian inheritance of *phenotypes* is the exception rather than the rule. Most phenotypes show "complex inheritance" and, despite being heritable and genotype dependent, are subject to many factors that blur the direct effect of genotypes. The multifactorial model is adequate for explaining many genetic features of the *population* being a statistical model but is poor at predicting effects in *individuals* since it is not mechanistic. We need to understand the molecular reasons for this less-than-perfect correlation. First, we need to appreciate the large effect that simple stochastic variation in biological processes can induce on phenotypes. One needs only to observe a pair of identical twins to appreciate how *different* they can be! Second, we need to understand *how* environmental differences can yield phenotypic differences through molecular means. Third, as Emma Whitelaw and her colleagues have shown,[18] isogenic agouti mice can have markedly different coat colors arising from methylation of a repeat element outside the gene. We need to assess the importance of such *heritable* epigenetic effects, beyond those modulated by chromatin, and how they contribute to inheritance.[19] All of these features argue that most phenotypes are not discrete and inviolate given a genotype but are canalized with a range of variation modulated by sequence-dependent (mutation, polymorphism) and sequence-independent (stochastic, epigenetic) effects.[20] No wonder, then, that most phenotypic inheritance is non-Mendelian, since the effect of genes may be both modifiable and dynamic. If generally true, then we have a very different genetic lesson to convey to the public: the gene as an adaptable and dynamic machine.

Our thinking of the nature of the gene significantly affects our thinking of human phenotypes and how they aggregate in families and populations. The reason why the general view of humanity, and its myriad groups, is still stereotypical is because we view, and associate, individual groups by their geography *and* phenotypes. This is true for geneticists, biologists, and the lay public alike. Many recent population genetic studies, in which a thousand or more individuals have been examined for half a million or more SNPs, have reinforced a "genes as proxies of geography" scenario. Humans can now be placed into their continental origins as well as, sometimes, smaller geographic regions.[21] This determination of "ancestry," an industry of its own, is, first, statistical (not foolproof) and, second, depends on an admittedly artificial view (a model) of humanity being composed of distinct and recognizable homogeneous groups, with some exceptions from genetic admixture. No doubt reality is more complex, and so we need our descriptions and population genetic analyses to be more nuanced and our definitions of "ancestry" contingent. Our Society has produced a task force report on this issue to clarify both the meanings of the term "ancestry" and what answers genetics can and cannot provide.[22] Importantly, genetic admixture is a universal phenomenon recognizable in our genes when the admixture events are recent and between individuals from groups isolated from one another for a long time. But, not all populations are of that sort, and human groups lie along a spectrum from recent to ancient admixture, usually between neighboring groups where genetic differentiation is usually low. In other words, almost all humans are likely "admixed" with ancestry from different "populations." "Same" and "different" can be mathematically modeled and culturally imposed, but whether genetic variation follows this edict is still, I contend, poorly known.

Consequently, we have much to learn from a wider sampling of humanity across the globe and not only from its peripheries of variation (continental groups). These future studies can better define what we mean by "ancestry" and at what time in the past this ancestry refers to. Each of us has innumerable ancestors: in my case, recent ancestry from India, ancient ancestry from Africa, and many others from geographies and times in between. So, what do homogeneity, admixture, and ancestry mean in this sense, and what notions do they convey to us as individuals and to others? Many ongoing genetic studies search to find and exclude individuals in that study that are "different" from the others based on their genetic variation data, since otherwise their inclusion may induce errors in the analysis. I believe that alternative methods that explicitly allow for "ancestry" differences are required. Consider that in the United States 68% of the population classified as "non-Hispanic White" in 2008 is expected to decline to 46% by 2050. The current majority is giving way to a new, evolving, highly heterogeneous and admixed majority. This trend is worldwide. We can no longer bend to the current assumptions of population genetic analysis but invent new theories and methods that can study all peoples.

There is also a major social component to studying diversity as it appears. As geneticists, genetic and phenotypic diversity is our currency. But, outside our research, we have much work to do to show respect for the diversity on which our daily research depends. Donna Nelson, at The University of Oklahoma, has done extensive work on the academic status of women and other underrepresented minorities in US university faculties.[23] In the top 50 biological departments, which are representative of where you and I work, she has demonstrated that women are 25% of the faculty, but this figure decreases from 35%

for Assistant Professor, to 30% for Associate Professor, to 17% for Professorial ranks. The corresponding figures for underrepresented minorities are 7%, 4%, and 3%, respectively (4% overall); for Asians, these figures are 21%, 15%, and 8%, respectively (13% overall). The Nelson report does not have separate figures for human genetics, but my personal experience and anecdotal data suggest that it is perhaps not much better. The reasons for the inequality of rank and number in academic departments, with regards to women and other minorities, are complex indeed and not of singular cause. Nevertheless, the situation has changed very slowly over time and remains an impediment to attracting the best and the brightest persons to our faculties. Each of us needs to make progress to remove the remaining obstacles, whatever they may be; the numbers speak for themselves that obstacles remain. As a member of a minority group myself, I find, despite my personal success, the status quo hurtful and best articulated in the words of others:[24]

> Many rivers to cross
> But I can't seem to find my way over
> Wandering I am lost
> As I travel along the white cliffs of Dover

I do not want to end on a somber note. We have an extremely bright future, with success well within our control. There are many different kinds of science that we can and will do. I am also confident that our work will lead to an improved logic in human genetics. But, we also need to make our science more open and convey our progress and possibilities to a much wider audience in the public.

Two years ago, I came across a photograph that showed two people, whom I normally would not associate with one another, one a geneticist and the other not, at a special celebratory dinner hosted by *TIME* magazine. The familiar figure was James Watson, molecular biologist and geneticist. There is no doubt that, despite some of his utterly negative public comments on human variation and its consequence, Jim Watson is a central figure who has made genetics vibrant, stronger, and whose considerable efforts launched the genome project. Watson, unlike most scientists, is universally recognized by others, including the lay public, and probably by several million people. The second person, unknown to most of us, was Aishwarya Rai, a Bollywood actress from India. She is considered by some to be the world's most, and second most, beautiful woman! She is easily recognized in South Asia, as well as in many countries across the Middle East, the Far East, Africa, and South America. By one report, she is recognized by half a billion people! This exaggerated, even crass, difference in popularity underscores who the broader public is, to whom we need to speak and who the ultimate consumers of the fruits of our imagination are. Our charge, then, is to try to educate everyone into the grandness of our field, what our future science can do to help humanity, but, first and foremost, to improve the fundamentals of our science and our *Principia Genetica* that can lead us there.

Acknowledgments

Aravinda Chakravarti a paid member of the Scientific Advisory Board of Affymetrix. This potential conflict of interest is managed by the policies of Johns Hopkins University School of Medicine.

References

1. Obama, Barack. "Election Night Victory Speech". Grant Park, Chicago, IL, USA. November 5, 2008.
2. Muller, H.J. (1950). Our load of mutations., Presidential address, Second Annual Meeting of The American Society of Human Genetics. December 29, 1949. Am. J. Hum. Genet. *2*, 111–176.
3. *Genetics and Genomics of Infectious Disease*, March 21–24, 2009, Singapore. http://www.ashg.org/meetings/meetings_previous.shtml.
4. Whitehead, A.N., and Russell, B. *Principia Mathematica*, 3 volumes (Cambridge, UK: Cambridge University Press) 1910, 1912, and 1913. Second edition: 1925 (V.1), 1927 (V.2,3).
5. Lander, E. S. Keynote address, International Congress of Genetics. July 12–17, 2008. Berlin, Germany.
6. Liu, Q.P., Sulzenbacher, G., Yuan, H., Bennett, E.P., Pietz, G., Saunders, K., Spence, J., Nudelman, E., Levery, S.B., White, T., et al. (2007). Bacterial glycosidases for the production of universal red blood cells. Nat. Biotechnol. *25*, 454–464.
7. McKusick, V.A. (1955). Heritable disorders of connective tissue. III. The Marfan syndrome. J. Chronic Dis. *2*, 609–644.
8. Dietz, H.C., Cutting, G.R., Pyeritz, R.E., Maslen, C.L., Sakai, L.Y., Corson, G.M., Puffenberger, E.G., Hamosh, A., Nanthakumar, E.J., Curristin, S.M., et al. (1991). Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. Nature *352*, 337–339.
9. Brooke, B.S., Habashi, J.P., Judge, D.P., Patel, N., Loeys, B., and Dietz, H.C., 3rd. (2008). Angiotensin II blockade and aortic-root dilation in Marfan's syndrome. N. Engl. J. Med. *358*, 2787–2795.
10. Provine, W.B. (1971). The Origins of Theoretical Population Genetics (Chicago, IL, USA: University of Chicago Press).
11. Altenburg, E., and Muller, H.J. (1920). The genetic basis of truncate wing,-an inconstant and modifiable character in Drosophila. Genetics *5*, 1–59.
12. Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G., et al. (2007). The diploid genome sequence of an individual human. PLoS Biol. *5*, e254.
13. Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T., et al. (2008). The complete genome of an individual by massively parallel DNA sequencing. Nature *452*, 872–876.
14. Olson, M.V. (2008). Human genetics: Dr Watson's base pairs. Nature *452*, 819–820.
15. Angrist, M. (2008) The Stewardess is Flying the Plane: Genomic Determinism and Personal Genomics Gone Wild! Cold Spring Harbor Symposium: Personal Genomes. http://www.personalgenomes.org/pgp10.html.
16. Socolich, M., Lockless, S.W., Russ, W.P., Lee, H., Gardner, K.H., and Ranganathan, R. (2005). Evolutionary information for specifying a protein fold. Nature *437*, 512–518.

17. Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U., and Gaul, U. (2008). Predicting expression patterns from regulatory sequence in Drosophila segmentation. Nature *451*, 535–540.
18. Morgan, H.D., Sutherland, H.G., Martin, D.I., and Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. Nat. Genet. *23*, 314–318.
19. Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635–638.
20. Waddington, C.H. (1957). The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology (London: Allen & Unwin).
21. Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A.R., Auton, A., Indap, A., King, K.S., Bergmann, S., Nelson, M.R., et al. (2008). Genes mirror geography within Europe. Nature *456*, 98–101.
22. Royal, C., Novembre, J., Fullerton, S.M., Goldstein, D.B., Long, J.C., Bamshad, M.J., and Clark, A.G. (2009) Inferring Genetic Ancestry: Opportunities, Challenges and Implications. Report of the ASHG Task Force on Genetic Ancestry.
23. Nelson, D. A National Analysis of Minorities in Science and Engineering Faculties at Research Universities, 2007. http://chem.ou.edu/~djn/diversity/Faculty_Tables_FY07/07Report.pdf.
24. Cliff, Jimmy (1969). Many Rivers to Cross. On *Jimmy Cliff* (Santa Monica, California: A&M Records) [vinyl record album].