

 Links

**FURTHER INFORMATION Hartl lab page**  
**ENCYCLOPEDIA OF LIFE SCIENCES Genome**  
**organization/humans**

- Mirsky, A. E. & Ris, H. The DNA content of animal cells and its evolutionary significance. *J. Gen. Physiol.* **34**, 451–462 (1951).
- Thomas, C. A. The genetic organization of chromosomes. *Annu. Rev. Genet.* **5**, 237–256 (1971).
- Bonner, J. *et al.* Functional organization of the mammalian genome. *Cold Spring Harbor Symp. Quant. Biol.* **38**, 303–310 (1973).
- Davidson, E. H., Hough, B. R., Amenson, C. S. & Britten, R. J. General interspersed repetitive with nonrepetitive sequence elements in the DNA of *Xenopus*. *J. Mol. Biol.* **77**, 1–23 (1973).
- John, B. & Miklos, G. L. G. *The Eukaryotic Genome in Development and Evolution* 1–416 (Allen and Unwin, London, 1988).
- Dunham, I. *et al.* The DNA sequence of human chromosome 22. *Nature* **402**, 489–495 (1999).
- Hattori, M. *et al.* The DNA sequence of human chromosome 21. *Nature* **405**, 311–319 (2000).
- Cavaller-Smith, T. *The Evolution of Genome Size* (John Wiley, New York, 1985).
- Bliss, M. *William Osler: A Life in Medicine* (Oxford, New York, 1999).
- Cavaller-Smith, T. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the C-value paradox. *J. Cell Sci.* **34**, 247–278 (1978).
- Vinogradov, A. E. Buffering: A possible passive-homeostasis role for redundant DNA. *J. Theor. Biol.* **193**, 197–199 (1998).
- Ohno, S. in *Evolution of Genetic Systems, Brookhaven Symp. Biol.* (ed. Smith, H. H.) 366–370 (1972).
- Orgel, L. E. & Crick, F. H. C. Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607 (1980).
- Doolittle, W. F. & Sapienza, C. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601–603 (1980).
- Cavaller-Smith, T. & Beaton, M. J. The skeletal function of non-genic nuclear DNA: New evidence from ancient cell chimaeras. *Genetica* **106**, 3–13 (1999).
- Beaton, M. J. & Cavaller-Smith, T. Eukaryotic non-coding DNA is functional: evidence from the differential scaling of cryptomonad genomes. *Proc. R. Soc. Lond. B* **266**, 2053–2059 (1999).
- Bernardi, G. Isochores and the evolutionary genetics of vertebrates. *Gene* **241**, 3–17 (2000).
- Lozovskaya, E. R., Nurminsky, D. I., Petrov, D. A. & Hartl, D. L. Genome size as a mutation-selection-drift process. *Genes Genet. Syst.* **74**, 201–207 (1999).
- Laurent, A. M., Puechberty, J. & Roizès, G. Hypothesis: for the worst and for the best, L1Hs retrotransposons actively participate in the evolution of the human centromeric alphoid sequences. *Chromosome Res.* **7**, 305–317 (1999).
- Csink, A. K. & Henikoff, S. Something from nothing: The evolution and utility of satellite repeats. *Trends Genet.* **14**, 200–204 (1998).
- Russo, C. A. M., Takezaki, N. & Nei, M. Molecular phylogeny and divergence times of Drosophilid species. *Mol. Biol. Evol.* **12**, 391–404 (1995).
- Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
- Ogata, H., Fujibuchi, W. & Kanehisa, M. The size differences among mammalian introns are due to the accumulation of small deletions. *FEBS Lett.* **390**, 99–103 (1996).
- Moriyama, E. N., Petrov, D. A. & Hartl, D. L. Genome size and intron size in *Drosophila*. *Mol. Biol. Evol.* **15**, 770–773 (1997).
- Deutsch, M. & Long, M. Intron–exon structures of eukaryotic model organisms. *Nucleic Acids Res.* **27**, 3219–3228 (1999).
- Vinogradov, A. E. Intron–genome size relationship on a large evolutionary scale. *J. Mol. Evol.* **49**, 376–384 (1999).
- SanMiguel, P. & Bennetzen, J. L. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Ann. Bot.* **82**, 37–44 (1998).
- Petrov, D. A. & Hartl, D. L. High rate of DNA loss in the *D. melanogaster* and *D. virilis* species groups. *Mol. Biol. Evol.* **15**, 293–302 (1998).
- Bennetzen, J. L. & Kellogg, E. A. Do plants have a one-way ticket to genomic obesity? *Plant Cell* **9**, 1509–1514 (1997).
- San Miguel, P., Gaut, B. S., Tikhonov, A., Nakajima, Y. & Bennetzen, J. L. The paleontology of intergene retrotransposons of maize. *Nature Genet.* **20**, 43–45 (1998).
- Andrews, J. D. & Gloor, G. B. A role for the *KP* leucine zipper in regulating *P* element transposition in *Drosophila melanogaster*. *Genetics* **141**, 587–594 (1995).
- Lohe, A. R. & Hartl, D. L. Autoregulation of *mariner* transposase activity by overproduction and dominant-negative complementation. *Mol. Biol. Evol.* **13**, 549–555 (1996).
- Stellwagen, A. E. & Craig, N. L. Mobile DNA elements: controlling transposition with ATP-dependent molecular switches. *Trends Biochem. Sci.* **23**, 486–490 (1998).
- Braam, L. A. M., Goryshin, I. Y. & Reznikoff, W. S. A mechanism for Tn5 inhibition: Carboxyl-terminal dimerization. *J. Biol. Chem.* **274**, 86–92 (1999).
- Sakai, J. S., Kleckner, N., Yang, X. & Guhathakurta, A. Tn10 transpososome assembly involves a folded intermediate that must be unfolded for target capture and strand transfer. *EMBO J.* **19**, 776–785 (2000).
- Bennetzen, J. L. Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol.* **42**, 251–269 (2000).
- Ketting, R. F. & Plasterk, R. H. A. A genetic link between co-suppression and RNA interference in *C. elegans*. *Nature* **404**, 296–298 (2000).
- Chaboissier, M. C., Bucheton, A. & Finnegan, D. J. Copy number control of a transposable element, the *I* factor, a LINE-like element in *Drosophila*. *Proc. Natl Acad. Sci. USA* **95**, 11781–11785 (1998).
- Birchler, J. A., Pal-Bhadra, M. & Bhadra, U. Less from more: cosuppression of transposable elements. *Nature Genet.* **21**, 148–149 (1999).
- Jensen, S., Gassama, M. P. & Heidmann, T. Cosuppression of *I* transposon activity in *Drosophila* by *I*-containing sense and antisense transgenes. *Genetics* **153**, 1767–1774 (1999).
- Casavant, N. C. *et al.* The end of the LINE?: Lack of recent L1 activity in a group of South American rodents. *Genetics* **154**, 1809–1817 (2000).
- Graur, D., Shuali, Y. & Li, W.-H. Deletions in processes pseudogenes accumulate faster in rodents than in humans. *J. Mol. Evol.* **28**, 279–285 (1989).
- Petrov, D. A., Lozovskaya, E. R. & Hartl, D. L. High intrinsic rate of DNA loss in *Drosophila*. *Nature* **384**, 346–349 (1996).
- Petrov, D. A. & Hartl, D. L. Trash DNA is what gets thrown away: High rate of DNA loss in *Drosophila*. *Gene* **205**, 279–289 (1997).
- Wright, D. A. *et al.* Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana*. *Genetics* **142**, 569–578 (1996).
- Petrov, D. A., Sangster, T., Johnston, J. S., Hartl, D. L. & Shaw, K. L. Evidence for DNA loss as a determinant of genome size. *Science* **287**, 1060–1062 (2000).
- Robertson, H. M. The large *srh* family of chemoreceptor genes in *Caenorhabditis* nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. *Genome Res.* **10**, 192–203 (2000).
- Bensasson, D., Petrov, D. A., Zhang, D. -X., Hartl, D. L. & Hewitt, G. M. Genomic gigantism: DNA loss is slow in mountain grasshoppers. *Mol. Biol. Evol.* (in the press).
- Charlesworth, B. The changing sizes of genes. *Nature* **384**, 315–316 (1996).
- Petrov, D. A. & Hartl, D. L. Pseudogene evolution and natural selection for a compact genome. *Heredity* **91**, 221–227 (2000).
- Kimura, M. On the probability of fixation of mutant genes in a population. *Genetics* **47**, 713–719 (1962).
- Simmons, M. J. & Bucholz, L. M. Transposase titration in *Drosophila melanogaster*: A model of cytotype in the P-M system of hybrid dysgenesis. *Proc. Natl Acad. Sci. USA* **82**, 8119–8123 (1985).
- Sawyer, S. A. *et al.* Distribution and abundance of insertion sequences among natural isolates of *Escherichia coli*. *Genetics* **115**, 51–63 (1987).

## Acknowledgements

I am very grateful to my colleagues D. Petrov and E. Lozovskaya for their contributions to the work described here, and to D. Petrov for having allowed me to read an early draft of a manuscript of his own. This work was supported by a grant from the National Institutes of Health.

## TIMELINE

## Environmental health and genomics: visions and implications

*Kenneth Olden and Samuel Wilson*

The relationship between genes and the environment can be compared to a loaded gun and its trigger. A loaded gun by itself causes no harm; it is only when the trigger is pulled that the potential for harm is released. Genetic susceptibility creates an analogous situation, where the loaded gun is one or a combination of susceptibility genes (alleles) and the trigger is an environmental exposure. The key objective of the Environmental Genome Project is to identify alleles that confer susceptibility to the adverse effects of environmental agents. Here we discuss the goals of the Environmental Genome Project, its implications and, in particular, its potential effect on our ability to assess human disease risk in the future.

Scientists in biomedicine, environmental health and public health are working to understand and prevent human disease. The identification and functional characterization of susceptibility genes is critical to achieving this goal and for predicting risk from environmental exposure and response to pharmaceuticals. Many chronic diseases in humans arise from a complex array of factors, which could

include several genes, environmental conditions or exposures, the age, nutritional status or stage of development of a person, and other predisposing factors. Therefore, most chronic diseases will not be fully understood until both the genetic and environmental contributions to their aetiology are elucidated. Unfortunately, the relationship between genes and the environment is neither well

## PERSPECTIVES

understood nor extensively studied at the moment. Until recently, limited and inadequate knowledge of human genetics and human gene sequences has hampered progress in this area, and has limited scientists to relatively simplistic models — models that assume that diseases are caused by mutations in a single gene or by exposure to a single environmental agent. Interactions between several genes, or between genes and several environmental agents, have only rarely been considered as the cause of human illness. So our knowledge has many information gaps, and it has remained difficult to predict human disease risk accurately.

In a general sense, this missing information is needed to create the framework for accurately assessing human disease risk in the community and ultimately at the level of the individual. In a more specific sense, the missing information belongs to three categories: information relating to dose–response relationships for environmentally significant compounds, from well-characterized animal models tested at biologically relevant doses (that is, very low to moderate dose ranges); a comprehensive catalogue of the human gene polymorphisms that can influence human disease risk; and extensive second-generation epidemiological studies and other population-based studies that definitively link human disease to environmental exposure.

Recent developments in genomics are enabling scientists to progress towards closing these information gaps. The sequencing of the human genome will soon provide a human ‘reference sequence’. This ‘reference sequence’ will be a vital tool in many areas of molecular biology. Importantly, in disease risk assessment, it will allow us to assess the level of variation (deviation from the reference sequence) in the genes of an individual or a specific human subpopulation. During

**“During the past ten years, no subject has generated more interest, debate and activity than the effort to identify and characterize genes that predispose people to disease.”**

the past ten years, no subject has generated more interest, debate and activity than the effort to identify and characterize genes that predispose people to disease. When the sequence of the human genome is complete, this subject will be even more at the forefront of people’s thoughts because of its many ramifications for both environmental health regulatory policy and healthcare.

The aim of this article is to describe the goals and status of the Environmental Genome Project (EGP) and to discuss its potential to bring a new vision to the areas of human risk and exposure assessment, environmental health, public health and environmental policy.

Susceptibility and exposure Evidence that genetics is important in the development of disease has come from studies of familial clusters in which one or several alleles of a gene have been identified that are highly penetrant and associated with increased risk for a specific disease, for example, Huntington disease. The inheritance of such alleles in the general population is rare and probably accounts for no more than 5% of known diseases. So the contribution of monogenic disease genes to the overall incidence of disease is relatively small, although

the risk for a person with a specific disease allele is relatively high.

Many common human diseases seem to be polygenic, resulting from the complex interactions of several genes. A variant of one gene may not be detrimental, but it might become detrimental in combination with specific alleles of one or more other genes. Such susceptibility-conferring genes increase disease risk only a fewfold, but they can have a large effect on the incidence of disease in the human population because of their frequency. Susceptibility genes are not sufficient to cause disease; they modify risk in combination with other genes and with exposure to specific environmental agents.

Every organism is exposed to hazardous agents in its environment on a continual basis. As a result, organisms have evolved sophisticated pathways that can minimize the biological consequences of hazardous environmental agents. These pathways constitute the ‘environmental response machinery’ (FIG. 1). All human genes, including those that encode components of the environmental response machinery, are subject to genetic variability, which can be associated with the altered efficiency of a biological pathway. So a person’s risk for developing an illness as a result of an environmental exposure might be dependent on the efficiency of their own unique set of environmental response genes (FIG. 1). These genes, for example, might determine how a person responds to and metabolizes drugs or carcinogenic compounds after exposure.

The Environmental Genome Project The EGP was initiated in 1997 at the United States National Institute of Environmental Health Sciences to stimulate population-based and other research into the role of genetic variation in response to environmen-

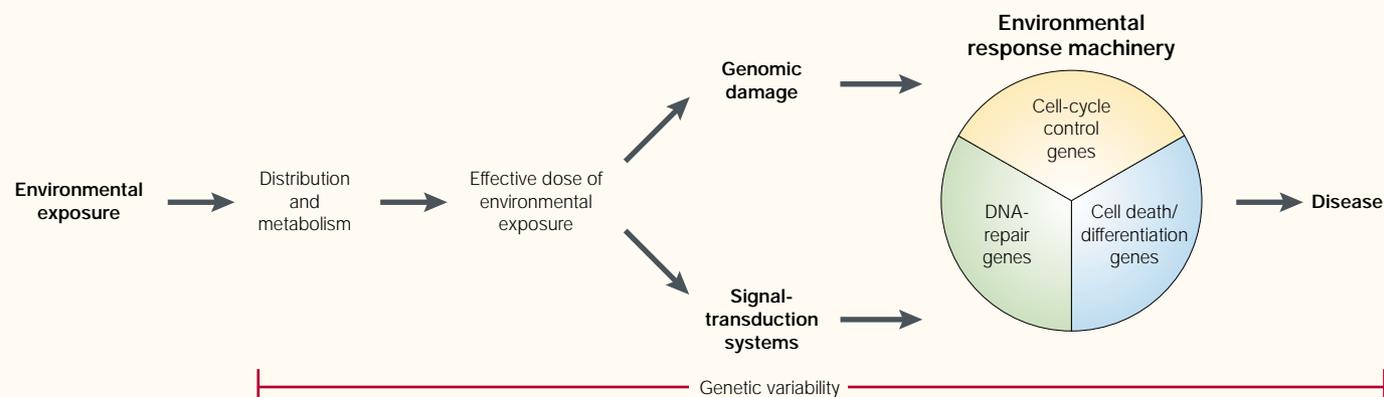


Figure 1 | **The environmental exposure–disease model.** Polymorphisms in environmental response genes can modify a person’s risk for disease. Potential roles of metabolic-activation enzymes, detoxification enzymes, signal-transduction systems and genome-metabolism systems are shown. The indicated categories of enzymes can potentially interact with one another, and genetic variation can occur in one or more enzymes in each category.

tal exposure<sup>1–4</sup>. It does this primarily through investigator-initiated grants. The EGP (BOX 1) was the first large-scale systematic study of human functional genomics of polymorphisms, but **other projects with similar goals** have also been initiated. The EGP is a 'second-generation' human genome project, because the EGP depends on the successful completion of the human reference sequence, as mentioned earlier. The goal of the EGP, which will be carried out in three phases (BOX 1), is to understand genetic susceptibility to disease as a function of response to environmental exposures.

**Phase I: Re-sequencing human genes.** The first phase of the EGP will identify polymorphism in a set of genes that will probably be important in environmentally associated diseases. Although the final human reference sequence is not yet complete, the systematic identification of sequence variations in the human genome has already begun. One motivation for starting this initiative has been the successful use of association studies to identify disease risk from environmental exposure as a function of relatively common allelic variants<sup>5,6</sup>, such as the association between cigarette smoking and bladder cancer risk as a function of allelic variants of *N*-acetyltransferases 1 and 2 (REF. 7). Methods for gene discovery and genotyping have also greatly improved, allowing faster and less expensive<sup>6</sup> large-scale population studies.

The strategy is to first catalogue and then test variants in the coding and non-coding regions of candidate susceptibility-conferring genes for functional implications, as described below. Initial plans were to survey 200 candidate genes that were selected because of their known involvement in metabolic pathways that are critical for normal cellular function (BOX 2). However, as this approach would have provided detailed information on only a few of the estimated 40,000 to 100,000 human genes, the planned survey will be expanded. The primary mechanism for selecting new genes or new categories of genes will be solicited recommendations from investigators actively working on environmental susceptibility genes in the area of toxicogenomics. It is expected, for example, that genomic-scale mRNA analysis by microarray will help us to identify many new environmental response genes. It is anticipated that variant alleles in thousands of genes might be identified within the next five years. In the later stages of the EGP, the selection of population DNA samples for variant characteriza-

#### Box 1 | Phases of the Environmental Genome Project

##### Phase I

- Develop a special sample repository representing the United States population (in collaboration with the NHGRI, NIGMS and the CDC)\*.
- Re-sequence 200 candidate genes from a special sample repository to identify variants<sup>‡</sup>.
- Develop technology to facilitate variant identification.
- Develop databases on polymorphic variation in genes with information from re-sequencing and other data sets.
- Consider the ethical, legal and social implications of this research.

##### Phase II

- Multidisciplinary functional studies of allelic variants.
- Population-based studies of exposure and allele–disease associations.
- Develop technology for Phase II studies, including toxicogenomics.
- Refine databases.
- Consider the ethical, legal and social implications.
- Initiate variant identification of more genes.

##### Phase III

- Population-based and other epidemiological studies stemming from Phases I and II.
- Develop animal models for disease susceptibility to study dose–response relationships.
- Risk assessment studies (understanding dose–response relationships in the population).
- Develop technology for Phase III studies.
- Targeted disease screening of high-risk populations.
- Consider the ethical, legal and social implications.
- Expand databases and sequence variant catalogue.

\* A repository of ~450 cell lines is maintained by the **Coriell Institute for Medical Research**. Re-sequencing genes in these samples should detect polymorphisms that occur at around 1% in the United States population. (**NHGRI** and **NIGMS**, National Human Genome Research Institute and National Institute of General Medical Sciences at the National Institutes of Health; **CDC**, Centers for Disease Control and Prevention.)

‡ Candidate genes selected from the categories listed in Box 2.

tion will be on the basis of preliminary studies regarding exposure, functional implications of a variation, and allele prevalence in specific human sub-populations.

**Phase II: Functional variants.** Polymorphisms in coding regions that alter the amino-acid sequence of proteins and their functions could explain a significant fraction of human genetic susceptibility to disease. Polymorphisms in regulatory regions can also contribute significantly to disease risk, because these gene regions control gene expression and splicing. Therefore, the EGP plans to study the functional implications of polymorphisms in both coding and regulatory regions of genes. The most abundant DNA polymorphisms are expected to be single nucleotide polymorphisms (SNPs)<sup>6</sup>. The number of SNPs and other DNA variations in the entire human population is probably large ( $10^5$  to  $10^7$ ), but the number of non-silent variations in the coding and regulatory regions of genes is expected to be much smaller, and it

should be possible to test most of them using functional assays.

The identification of new functional polymorphisms will be relatively straightforward in comparison with the immense challenge of establishing whether or not each polymorphism functions in exposure-associated disease. This is expected to be a critical and most difficult part of the EGP. Until recently, genetic linkage analysis has been used to identify genes that contribute to monogenic human diseases. However, conventional linkage analysis is not optimal for complex polygenic diseases because the effects of individual alleles are relatively weak and they may exist at high frequencies in the population, making family-based studies problematic. Therefore, the EGP will rely on several approaches for the association of combinations of polymorphisms with disease: one will be to assess the frequency with which given alleles occur in a large population of affected people, compared with the frequency of the same alleles

in a population of unaffected people. Relatively large population groups will be required for this approach ( $10^3$  to  $10^5$ ), as well as replication of the results in independent studies. Detailed environmental exposure data will be critical to such studies, because the risk associated with a particular allele may be evident only among people with a specific exposure. Associations of variant alleles with disease that have been revealed by such analysis must be interpreted with caution because of the possibility that the variant allele of interest is linked with another important, but unknown, allele<sup>8</sup>. Integration of mechanistic information, concerning how the variant alleles might influence responses to environmental toxicants, together with information from population studies, will enhance interpretations and the efficacy of study design.

Finally, the new knowledge of the human genome reference sequence offers another important opportunity for environmental health — studies of mRNA and protein expression on a genome-wide scale. Microarray assays for nucleic acid measurements and physical separation techniques for protein measurements are evolving rapidly and have already been applied to the study of response to environmental stress. Initial research in this area reveals that global genomic expression patterns are exquisitely sensitive to environmental stress and also to the cell's genetic history<sup>9</sup>. This research area, which we term 'toxicogenomics', will elucidate how the ebb and flow of genomic-scale mRNA and protein expression are involved in biological responses to causes of environmental stress. Toxicogenomics initiatives will combine information from genetics, genomic-scale mRNA and protein expression patterns, and bioinformatics to understand the role of gene–environment interactions in disease; eventually, toxicogenomics will provide us with more sensitive measurements of both biological responses to exposures and the early molecular changes of disease.

**Phase III: Animal models.** An objective of phase III of the EGP is to establish and characterize animal and cell models that are useful in understanding human susceptibility to disease. Once an allele has been implicated in a phenotype by association studies or other approaches (BOX 1), the function of the variant can be examined in more detail by generating mouse 'knockout' and 'knock-in' models. Several gene variants can be introduced into a single genetic background, in various combinations, to investigate how risk for a specific disease is altered by a par-

**“Perhaps the biggest challenge for the EGP will be... in public policy: how will those who formulate environmental and health policy accommodate the extraordinary new knowledge available to them?”**

ticular genotype and environmental exposure. Animal studies will also be important in discriminating between polymorphisms that predispose to disease without exposure and polymorphisms that predispose to disease only when a specific environmental exposure acts as the 'trigger'. In the latter case, animal models can be used to quantify the relative importance of genetic and environmental susceptibility factors.

Another current goal in the field, which is outside the EGP, is to search randomly for SNPs in coding and non-coding regions in the human genome; as it is estimated that there are many SNPs per gene, thousands of polymorphic sites could be identified, and the emerging *SNP catalogue* could be a resource to discover susceptibility genes in the human genome<sup>6</sup>. Although these SNPs may not themselves be the sequence variations that influence susceptibility, it is hoped that high-density SNP maps will provide markers that can be used in association studies to discover nearby allelic alterations that influence susceptibility.

Benefits of the EGP

As mentioned above, a goal of the EGP is to catalogue information about human genetic polymorphisms and to apply that information to understanding disease susceptibility and environmental exposure responses in population-based studies. It is clear that the EGP, combined with related advances in toxicogenomics, has the potential to have a significant effect on environmental health in the future. It is expected that the later stages of the EGP will focus on defining population subgroups with unusual susceptibility to environmental exposures. Focused studies will work towards developing preventive measures that are tailored to the needs of such 'at-risk' subgroups. However, as this goal is achieved, a new set of goals and issues is anticipated. For example, the effective use of this new

information to improve healthcare and environmental health policy will require considerable restructuring of these systems. Knowledge of susceptibility will potentially change the 'contract' between patient and physician, promoting a shift from the current emphasis on curative treatment to a greater emphasis on prevention. Prevention, diagnosis and treatment will eventually become more individualized as differences in response to environmental stress or pharmaceutical interventions can be tailored to a patient's specific genotype. The policy mechanisms by which information on susceptibility will be used to reduce risk in the human population have not been determined. But one of the biggest potential benefits of this new knowledge and understanding is that environmental health regulatory agencies will be able to develop more rational policies. At present, human genetic variation is not implicitly considered in estimating dose–response relationships, nor is it considered when setting exposure limits. Data on the prevalence and characteristics of susceptibility genes offers the potential to reduce the guesswork in risk assessment, and therefore it is likely that the ability to issue fair and appropriate regulations concerning environmental hazards will increase markedly.

In summary, the effects of the EGP could be broad, but the key areas of environmental health to which the EGP will contribute are: identification and protection of 'at-risk' subgroups; ability to understand the combination of environmental and genetic components of important human diseases (that is, asthma, diabetes, cardiovascular disease, childhood cancer, birth defects, neurological diseases and cancers of the hormonally responsive tissues, such as ovary, breast and prostate); and

#### Box 2 | Gene categories in Phase I

Categories of environmental-response genes used in Phase I of the EGP:

- Xenobiotic metabolism and detoxification
- Cell-surface receptors
- DNA repair
- Cell cycle
- Cell death
- Immune and inflammatory response
- Hormone metabolism
- Nutrition
- Oxidative metabolism and stress
- Membrane pumps and/or drug resistance
- Signal transduction

ability to understand the importance of new molecular indicators of exposure.

**Ethical, legal and social implications.** It is recognized that the new knowledge and understanding of human disease risk emerging from the EGP is accompanied by a burden of responsibility to use that knowledge fairly and wisely. Therefore, the transformation of public health policy and healthcare in the light of the goals of the EGP is burdened with significant social, ethical and legal concerns<sup>10</sup>. For example, how will we deal with the fact that we may be able to predict risk for many common diseases long before effective and acceptable medical interventions are available to treat them? What are the ethical and economic implications for people who knowingly ignore a risk about which they have been informed? Personal compliance and responsibility will take on a whole new meaning. How do we protect research participants from discrimination, stigmatization and psychological stress? How will we deal with the concern that many, on learning that they carry a predisposing genotype, will develop a fatalistic attitude and assume that they can do nothing to prevent the disease? How do we deal with the issue of informed consent given that the risks and benefits cannot be fully anticipated at the outset of a particular study? And ultimately, how can a person be assured that their personal genetic information will remain private? Policies are urgently needed to ensure the appropriate and ethical use of susceptibility data.

Addressing these ethical, legal and social implications (ELSI), and developing safeguards that appropriately protect us in the future, will represent an important challenge for the EGP. So the EGP has made ELSI issues a priority topic. The EGP is promoting research projects and broad-based discussions on these issues, and will try to ensure that the public is adequately protected. This is essential even to allow the EGP to move forwards towards completion.

Although ELSI issues are one important challenge for the EGP, the knowledge of human disease susceptibility that may soon emerge will also raise scientific and technological challenges. For example, developing rapid screening to identify predisposing polymorphisms can be difficult; human genes are often large, carry many mutations, and may vary significantly from population to population. Therefore, the use of a single test may not be feasible. As more complex models of the genetic and exposure framework of disease are discovered, mathematical modelling and data management will

have to be improved. Predictions on the basis of genetic polymorphisms will be far more complex than can be envisaged at the present time. Furthermore, improved understanding of disease susceptibility will intensify the need to identify the most beneficial and cost-effective intervention and prevention strategies.

Perhaps the biggest challenge for the EGP will be to manage its influence in the arena of public policy: how will those who formulate environmental and health policy accommodate the extraordinary new knowledge available to them? In the face of these challenges, we will need broader education and communications tools to find ways to make effective use of this knowledge and its enormous potential for biomedicine and human health.

*Kenneth Olden and Samuel Wilson are at the National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina 27709, USA. Correspondence to K.O. e-mail: olden@niehs.nih.gov*

## TIMELINE

## Engineering American society: the lesson of eugenics

*David Micklos\* and Elof Carlson†*

We stand at the threshold of a new century, with the whole human genome stretched out before us. Messages from science, the popular media, and the stock market suggest a world of seemingly limitless opportunities to improve human health and productivity. But at the turn of the last century, science and society faced a similar rush to exploit human genetics. The story of eugenics — humankind's first venture into a 'gene age' — holds a cautionary lesson for our current preoccupation with genes.

Eugenics was the effort to apply the principles of genetics and agricultural breeding towards improving the human race. The term "eugenics" — meaning well born — was coined in 1883 by Francis Galton<sup>1</sup>, a British scientist who used data from biographical dictionaries and alumni records at Oxford and Cambridge Universities to con-

 Links

FURTHER INFORMATION EGP | SNP catalogue | Center for Environmental Health Sciences | Coriell Institute for Medical Research | NHGRI | NIGMS | CDC

1. Kaiser, J. Environment institute lays plan for gene hunt. *Science* **278**, 569–570 (1997).
2. Guengerich, F. P. The environmental genome project: Functional analysis of polymorphisms. *Environ. Health Perspectives* **106**, 365–368 (1998).
3. Shalat, S. L., Hong, J. Y. & Gallo, M. The environmental genome project. *Epidemiology* **9**, 211–212 (1998).
4. Brown, P. O. & Hartwell, L. Genomics and human disease — variations on variation. *Nature Genet.* **18**, 91–93 (1998).
5. Schafer, A. J. & Hawkins, J. R. DNA variation and the future of human genetics. *Nature Biotechnol.* **16**, 33–39 (1998).
6. Wang, D. G. *et al.* Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077–1082 (1998).
7. Taylor, J. A. *et al.* The role of *N*-acetylation polymorphisms in smoking-associated bladder cancer: evidence of a gene–gene exposure three way interaction. *Cancer Res.* **58**, 3603–3610 (1998).
8. Holtzman, N. A., & Marreau, T. M. Will genetics revolutionize medicine? *N. Engl. J. Med.* **343**, 141–144 (2000).
9. Hughes, T. R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
10. Sharp, R. R. & Barrett, J. C. The environmental genome project: ethical, legal, and social implications. *Environ. Health Perspect.* **108**, 279–281 (2000).

clude that superior intelligence and abilities were traits that could be inherited<sup>2</sup>.

Most people equate eugenics with atrocities that were committed in Nazi Germany for the sake of racial purity. In this context, eugenics is easy to dismiss as purely aberrant behaviour. However, the story of eugenics in the United States is, perhaps, more important than that of Nazi Germany as a cautionary tale to take with us into our new century. Here we describe the tale of the subtle ways in which the science of genetics was, by degrees, transformed from an agricultural experiment into a popular movement to engineer American society. The fact that eugenics flourished in the land of liberty, involved numerous prominent scientists and civic leaders, and made its intellectual home at the forerunner of the now prestigious Cold Spring Harbor Laboratory shows just how far America fell from grace during this period.