

The GEN2PHEN Project

Genotype-to-phenotype databases: a holistic solution (www.gen2phen.org)

Response to the House of Lords Science and Technology Committee Call for Evidence: Genomic Medicine

Policy Framework

No evidence submitted

Research and Scientific Development

What is the state of the science? What new developments are there? What is the rate of change?

Inter-individual genetic variation strongly influences normal characteristics, disease processes, and drug responses – collectively referred to as ‘phenotypes’.

Operationally, work in this field can be divided into two parts;

a) genetic changes that work in isolation to ‘cause’ phenotypes which are generally rare (hereafter referred to as ‘Mendelian’ variants)

b) genetic changes that work in complex combinations, and in partnership with environment and lifestyle, to ‘modify risk’ of phenotypes that are generally common (hereafter referred to as ‘susceptibility’ variants)

Mendelian variants:

Effective procedures for the identification of Mendelian variants are well established and widely deployed, such that most of the genes underlying this class of genetic disorders (e.g., cystic fibrosis, Duchene muscular dystrophy, and rarer forms of cancer and common disease) have now been identified. The next challenge is to bring this information clinical utility (diagnostics, prognostics, and drug development). Progress towards this goal is, however, being held up by the lack of effective interdisciplinary partnerships between drug companies, groups who manage (database) data in the research domain, and healthcare providers. Additionally, enhanced diagnostics infrastructure would be required. Given wise strategic decisions on these fronts, then Mendelian genetics could quickly be taken ‘to the bedside’ to benefit affected individuals and their genetically-related family members.

Susceptibility variants:

Impressive procedures for the identification of susceptibility variants have become available in the last few years, and they have now been deployed to study the genetic basis of many common disorders (e.g., Crohns disease, Alzheimers disease, adult-onset diabetes, and various forms of cancer). Unfortunately, these efforts have

revealed but a small number of genes that vary naturally in the population in ways that only weakly influence disease risk. The state of the art now is such that;

a) Only a small fraction of total disease risk and incidence is explained by these discovered genes (merely a few percentage), and this is often less than the influence of environmental and lifestyle factors we already know about

b) For further human DNA studies alone to produce a substantially more complete picture of the genetic basis of common disease, there will need to be substantial method improvements, use of far larger sets of clinical materials (biobanks), allocation of significantly greater funding, and far better statistical/computational support. The main problem being faced is that the genetic basis of common disease is more complex than expected, and the law of diminishing returns starts to apply for these large-scale studies. Current knowledge and trends could even be argued to suggest that present strategies (regardless of how much improved or extensively applied) may never be able to decipher much more than ~50% of the genetic basis of common disease.

c) The clinical utility of the genetic findings achieved to date would, at best, be described as minimal.

How effective is the policy and investment framework in supporting research in this area?

Large and mainstream projects are thoroughly evaluated and well supported, and it is important to continue funding such projects. But it is increasingly difficult to get backing for more basic (non-translational), original/unconventional, and blue-sky research. Industry has picked up some of the slack here in the domain of method development, but academic research and development is now becoming quite distorted towards investigations that are argued to be directly clinically or disease relevant, and large-scale in nature. The obvious risk is that new ideas and the minds that can generate them are not encouraged.

Data Use and Interpretation

Is genomic information published, annotated and presented in a useful way? Should there be a common, public database? If so, who should fund, and have responsibility for, such an initiative?

Major deficiencies are apparent in this area, and it was the recognition of this fact that prompted the European Commission to fund the GEN2PHEN project. The challenges are complex, large, and global – but solvable given suitable funding and strategic decision making. Early experience has shown that effective solutions will need to be designed and built by involving many enthusiastic and differentially skilled teams - i.e., by the coordinated evolution of an optimal bottom-up databasing infrastructure, rather than by trying to impose a top-down solution that would be cheaper but which would not match with the communities complex needs. To practically engage researchers, to share and reward the large workload, and to

achieve optimal integration of the massive datasets that are now being generated, the solution must be based upon an inter-connected and closely federated set of tools and databases operated from many countries and institutions. It would not be effective, practical, or even acceptable (in some research or legal contexts) to place all the data in one monolithic data center. Some centralized component(s) will, however, most definitely be needed for the system to operate – e.g., to coordinate initiatives, to help develop and approve standards, to provide leadership on ethical and legal matters, and to enable top-level holistic access to core resources. From a European perspective, the UK-located European Bioinformatics Institute would be a natural home for many of these centralized activities, connecting them to dispersed databases and other resources across the community.

Given the nature and the size of the problem, funding should be provided by many sources, with international coordination efforts being funded and managed trans-nationally. The informatics and infrastructure development work should ideally be closely allied to that of researchers who generate/analyse data, and funding structures could help to ensure that such connections are enforced. GEN2PHEN is a step in the right direction, but far greater amounts of money need to be allocated to the problem to achieve an effective solution.

Who should provide the framework for optimal evaluation of data and translational opportunities? What policy and funding mechanisms are in place for recognising and utilising potential opportunities?

This is a complex, diverse, and multi-disciplinary challenge. The funding mechanisms employed to advance this area must take this into account, for which they will need to be equally complex, diverse, and multi-component in nature. Currently, most emphasis (political and funding) is being placed upon developing the needed resources and funding the primary data generation. Translational opportunities are yet to be placed in central focus, and real progress in the near term will be possible only for Mendelian disorders (see comments above).

Is other medical information recorded in a suitable format to allow optimal interpretation of genomic data? How should genomic data be brought together with other health information?

Medical information is not at all suitable structured, processed, or made available in a way that makes it useful for integration with genomic data. Furthermore, there seem to be few incentives for changing this in the healthcare world, and a poor general understanding of the issues. Many people argue correctly that the absence of a standardized way to represent phenotype information is a key bottleneck in this equation, but equally (if not more so) no-one yet has much idea of how we might handle the data that connects genetic and phenotype information (i.e., how to represent or meaningfully interpret the totality of evidence that relates genes to disease). Until these two issues are resolved, it will be very difficult to make much use of any genetic information in routine medical practice – where consistent and

predictive patterns have to be the basis of all clinical decision making.

What are the implications of the generation and storage of genome data on personal data security and privacy, and on its potential use or abuse in employment and insurance? How should these be addressed?

This subject is immensely complicated, and perhaps needs a complete examination in its own right. Two key points that, nevertheless, should perhaps be made apparent are

a) Only rare Mendelian variants are sufficiently high-impact to be usefully predictive of anything in an individual. Hence, the proliferation of companies offering personal genetic profiling (which focus primarily on common susceptibility variants) from which they supply health/disease guidance must surely be a cause for concern.

b) Genetic profiles are absolutely unique for everyone (except identical twins), and the tested DNA variants are shared to various degrees with all related family members. Therefore, it is illogical to argue that genetic information can ever be fully anonymised or made secure. In and of itself it precisely identifies individuals, and likewise all their relatives.

Translation

No evidence submitted

Biomarkers and Epidemiology

No evidence submitted