How to link genomics to physiology through epigenomics

Denis Noble^{*,1} & Peter Hunter²

¹Department of Physiology, Anatomy & Genetics, University of Oxford, Oxford OX1 3PT, UK

²Auckland Bioengineering Institute, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

*Author for correspondence: denis.noble@dpag.ox.ac.uk

First draft submitted: 8 January 2020; Accepted for publication: 8 January 2020; Published online: 21 February 2020

Keywords: GWAS • physiome • CellML • epigenome • computational biology

Genome wide sequencing has been spectacularly successful in our understanding of evolutionary biology, for example, in showing widespread lateral transfer of DNA even between different species, the extensive contribution of viral DNA to genomes and the tree diagrams of human evolution. By contrast, the results for healthcare have been disappointing. Only in rare genetic diseases do we find strong associations. For most multi-factorial diseases we find low associations with individual genes. Some now favor the omnigenic hypothesis: all genes are involved one way or another in any particular function of the body [1]. Moreover, it is also widely acknowledged that statistical association does not necessarily entail biological causation. Biology is not just the chemical consequences of DNA sequences. Much more than the genome is involved in controlling the way proteins fold and become functional [2]. What is not so widely acknowledged is that the reverse is also true. Even a negligible association score does not necessarily mean lack of a biological role. It might mean that the functional biological networks are capable of working even when what may normally be a key protein that is absent. Other parts of the network take over.

The processes by which this is achieved are epigenetic, both in the original sense introduced in 1957 by Waddington [3], which was the properties of the networks themselves independently of any genome control and in the more recent sense of DNA control by various markers of DNA or histones [4].

This phenomenon was shown many years ago by demonstrating how a protein channel contributing as much as 80% functionality in cardiac rhythm can be removed with only a 10–15% change in rhythm [5]. The mechanism is that, when one protein is down-regulated or knocked out, small changes in membrane potential activate a second mechanism so that the rhythm is maintained. That discovery eventually led to a new treatment, a cardiac slowing drug, ivabradine [6]. As another example, in yeast, as many as 80% of knockouts appear to have little or no functional effect in well-fed conditions [7], yet we know that the majority of the proteins involved are functional. The reason is that physiological/biochemical networks are very effective at buffering their functionality from genomic change. They have to be. That came about through the evolution of robustness; networks employ epigenetic control mechanisms to achieve this.

The progression genomics → epigenomics → physiology therefore also works in the opposite direction: physiology → epigenomics → genomics. In these causal relationships, genomics refers to DNA sequences, epigenomics refers to gene control through DNA and histone marking and physiology refers to the networks that can cause all of this control to happen.

Physiology is therefore about causation. At the moment, genomics is focused on improving the data on association by sequencing and phenotyping ever larger cohorts [8]. We think it should also start linking up with quantitative (computational) physiology. Physiological modeling (there are hundreds of examples on the CellML website <https://models.cellml.org/cellml>) enables causation to be quantified. It also explains why the association data research usually has difficulty establishing causation with biological significance. The more we discover about particular physiological networks, the more evidence we find for robustness: the ability to adapt to genomic change even when that might be expected to be catastrophic. For example, the *CLOCK* gene can be knocked out in mice without stopping circadian rhythm [9]. To understand these kinds of results we need to connect physiology to genomics through epigenomics.

We can foresee at least two ways in which this connection between genomics & physiology might be achieved

The first might be achieved quite quickly. This would be to tackle some good examples where physiological experiments and modeling show important causal roles for particular proteins yet the GWAS results show low correlations between the genes involved and disturbances of function. The prediction would be that the models may show the multi-network properties that generate such robustness. The heart rhythm models achieve that goal already, while the phenomenon is well-established in yeast [7]. Such studies would also clarify another important issue. Does it make sense to sum the correlations in GWAS to obtain overall genetic causation? In highly robust networks, it is obvious that simply summing up individual gene association scores must be misleading. The logic is more like a set of conditionals that do not lead to simple summation of effects. Since this must be true, there may be no way – even in sequencing millions of genomes – to arrive at a quantification of all genetic contribution. For most human traits, inheritance is very low as estimated by twin studies [10]. Identical twins can diverge very rapidly after birth [11]; inheritance is therefore much more than the transmission of DNA. Transmission of epigenetic effects is also important.

The second approach is much more ambitious. Could the results from genomics be used directly to inform physiological modeling? That would require more than association studies, which tell us what correlations there are between the presence or absence of particular DNA sequences and physiological function. Physiological modeling also requires information on the extent to which each relevant gene is expressed. As an example, being able to investigate different profiles of expression levels for a number of genes involved in cardiac arrhythmia gives much better prediction of toxic effects of drugs than does any single gene [12].

To some extent, gene expression levels can be assessed from association studies combined with expression quantitative trait loci (eQTL) mapping studies [13]. It remains to be seen whether any of that kind of data could be useful in physiological modeling. It would need to become more quantitative than showing simply whether particular genes are expressed.

How ready is physiological modeling for this kind of application?

For expert handlers of CellML models, the answer is very positive. There are over 20 categories of model in the model repository, ranging from calcium dynamics through to synthetic biology, taking in a very large fraction of cell types in the body. But even with expert handlers of such models, exploitation that is immediately useful to genomic studies is weak.

Part of the reason is that there is very little guidance so far on enabling nonexpert users to access the material. This deficiency is now being put right with the recent launch of the IUPS journal, *Physiome*.

Physiome will ensure that published models will be fully tested, so that simple instructions for reproducing any of the figures in those papers will be available. Anyone, anywhere in the world and from whatever specialty should then be able to run a model to obtain any of the published results and then alter any parameters to investigate the behavior of the model under new conditions, including changes of gene expression and gene knock-outs.

Physiome models are also annotated with semantic terms so that, if appropriate, a mathematical component always has both a biological meaning and a biophysical meaning. This should make it easier to link models that contain a description of the physiological phenotype for a particular protein to the accession number for that protein in a bioinformatic database. As models of gene expression are developed, this will also facilitate the connection with a corresponding DNA and RNA sequence and the presence of epigenetic markers that may be included in the model.

One of the key aims of *Physiome* is to provide a library of CellML modules that will allow a modeler to build a new model by importing some pre-existing components from the library. For example, if a new model requires a sodium-potassium ATPase ion channel pump, a well-validated model of that ion exchanger will be available for import into OpenCOR as a CellML module. Wherever possible, these protein level models will be available in bond graph form [14] to ensure that appropriate biophysical laws are obeyed.

Some examples of protein modules from the Physiome Model Repository are:

- *Sodium/hydrogen exchanger 3* <https://models.physiomeproject.org/e/236/>
- *Thiazide-sensitive Na-Cl cotransporter* <https://models.physiomeproject.org/e/231/>
- *Sodium/glucose cotransporter 1* <https://models.physiomeproject.org/e/232/>
- *Sodium/glucose cotransporter 2* <https://models.physiomeproject.org/e/233/>

Note that in each case, as well as the CellML-encoded mathematical model, links are provided to the UniProt Knowledgebase for that protein and to the Foundational Model of Anatomy ontology (via the EMBLE-EBI Ontology Lookup Service) for information about tissue regions relevant to the expression of that protein (e.g., proximal convoluted tubule, apical plasma membrane; epithelial cell of proximal tubule; proximal straight tubule).

We very much hope that a link-up between genome wide research and Physiome modeling research can be achieved. It has the potential to bridge the genome to phenotype gap. Epigenomics forms the key to the bridging since it is through epigenetic control that the physiological causal networks control the expression of genes and can also trigger genome rearrangement [15–18]. The linking of genomics to physiology through modeling of epigenomic control processes could introduce causation into genomics and epigenomics research.

Acknowledgments

We are grateful to H Fang at the Wellcome Centre for Human Genetics at Oxford University for valuable comments on an early draft of this editorial.

Financial & competing interests disclosure

D Noble is a member of the Technical Committee of the Engineering in Medicine and Biology Society (EMBS) and President of the Physiome Society. He receives no funding or remuneration from these organizations. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Open Access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

References

- Boyle EA, Li Y, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169, 1177–1186 (2017).
- Baverstock K. Polygenic scores: are they a public health hazard? *Prog. Biophys. Mol. Biol.* 149, 4–8 (2019).
- Waddington CH. *The Strategy of the Gene*. Allen and Unwin, London, UK (1957).
- Waddington CH. Canalization of development and the inheritance of acquired characteristics. *Nature* 150, 563–565 (1942).
- Noble D. Differential and integral views of genetics in computational systems biology. *Interface Focus* 1, 7–15 (2011).
- DiFrancesco D, Camm JA. Heart rate lowering by specific and selective I(f) current inhibition with ivabradine: a new therapeutic perspective in cardiovascular disease. *Drugs* 64, 1757–1765 (2004).
- Hillenmeyer ME, Fung E, Wildenhain J *et al.* The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* 320, 362–365 (2008).
- Bycroft C, Freeman C, Petkova D *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209 (2018).
- Debruyne JP, Noton E, Lambert CM, Maywood ES, Weaver DR, Reppert SM. A clock shock: mouse *CLOCK* is not required for circadian oscillator function. *Neuron* 50(3), 465–477 (2006).
- Polderman TJC, Benjamin B, De Leeuw CA *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genetics* 47, 702–709 (2015).
- Bathgate KE, Bagley JR, Jo E *et al.* Muscle health and performance in monozygotic twins with 30 years of discordant exercise habits. *Eur. J. Appl. Physiol.* 118(10), 2097–2110 (2018).
- Mirams GR, Cui Y, Sher A *et al.* Simulation of multiple ion channel block provides improved prediction of compounds' clinical torsadogenic risk. *Cardiovasc. Res.* 91(1), 53–61 (2011).
- Wainberg M, Sinnott-Armstrong N, Mancuso N *et al.* Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* 51(4), 592–599 (2019).
- Gawthrop PJ, Cursons J, Crampin EJ. Hierarchical bond graph modeling of biochemical networks. *Proc. R. Soc. A Math. Phys. Eng. Sci.* 471(2184) (2015).
- Shapiro JA. *Evolution: A View from the 21st Century*. Pearson Education Inc, NJ, USA (2011).
- Shapiro JA. Epigenetic control of mobile DNA as an interface between experience and genome change. *Front. Genet.* 5, 87 (2014).
- Noble D. *Dance to the Tune of Life: Biological Relativity*. Cambridge University Press, Cambridge, UK (2016).

18. Noble D. Central dogma or central debate? *Physiology* 33(4), 246–249 (2018).