Mutagenesis, 2015, 30, 719–722
doi:10.1093/mutage/gev068
Commentary
Advance Access publication 14 September 2015

Commentary

Exposomics: mathematics meets biology

Paolo Vineis*

MRC/PHE Centre for Environment and Health, School of Public Health, Imperial College London, Room 511 (5th floor), St Mary's Campus, Norfolk Place, London W2 1PG, UK

*To whom correspondence should be addressed. Tel: +44 207 594 3372; Fax: +44 207 594 3196; Email: p.vineis@imperial.ac.uk

Received 11 February 2015; Revised 20 March 2015; Accepted 24 March 2015.

Abstract

Although 'exposome' research has started to appear, and the concept is fascinating, we still have little proof-of-principle. This issue of *Mutagenesis* reports a few examples of exposome research, showing that the approach is providing the first results. In this Commentary, I develop the example of epigenome-wide methylation studies related to smoking as a success story, that fits well with previous research in humans and *in vitro* on mechanisms of carcinogenesis, and also with conceptual models such as Cairns' model based on asymmetric division of stem cells. The field of exposomics merges different disciplines, notably biology and mathematics, but also the evolutionary theory, and can possibly lead to interesting breakthroughs in the next years.

Exposome and exposomics have become fashionable words, and the Special Topic papers in the current issue of *Mutagenesis* show that relevant research has started to appear. Although the meanings of these words have been repeatedly explained previously, it is not redundant to restate them. The exposome concept refers to the totality of exposures from a variety of sources including, but not limited to, chemical agents, biological agents, radiation and psycho-social components from conception onwards, over a complete lifetime, and offers a conceptual leap in studying the role of the environment in human disease (1,2). The term omics (hence 'exposomics') refers to the quantitative measurement of global sets of molecules in bio-samples using high-throughput techniques, in combination with advanced biostatistics and bioinformatics tools (3,4). Exposomics, in addition to being a specific EU-funded multicentre project (http://www.exposomicsproject.eu/) has also become synonymous of omics-based exposome.

The model of tobacco smoke

Perhaps the most intriguing finding so far has come from the study of the effects of tobacco smoke through epigenome-wide methylation studies. Methylation of certain CpG sites—in particular in the *AHRR* and *F2RL3* genes—has been associated with smoking in an extremely strong way (5,6). This is the main example I will use in this Commentary to discuss some of the challenges of the field.

Long-term exposure to carcinogens such as those contained in tobacco smoke is able to induce molecular changes that involve DNA damage and epigenetic changes. The effects of tobacco smoking may be reversible: the risk of lung cancer drops from >16% (cumulative risk to the age of 80) to 5% when comparing individuals who smoked without interruption from the age of 20 with individuals who stopped at the age of 50 (with a cumulative risk of 1% for never smokers). Reversibility of cancer risk after cessation of exposure suggests that the mechanisms involved are likely to include epigenetic events because mutations or chromosome damage are fixed after cell replication.

What happens in ex-smokers has been described in an effective way by John Cairns:

> At any moment, the frequency of any class of cancer presumably reflects the number of cells that have undergone all but the final step in carcinogenesis, multiplied by the rate of whatever happens to be the final step. If this final step required DNA damage, the death rate from lung cancer would quickly drop when the mutagenic stimulus (smoking) was removed. In fact (…) at that point the death rate shows no decline and for the next 10–20 years actually stays roughly at the level it had reached in the year before the smoker stopped smoking (…). This suggests that the final step (…) is an event of a different kind from the steps that lead to the accumulation of mutant cells because, unlike the earlier steps, it does not have to be stimulated by the toxicity and mutagenicity of tobacco (…). It must, however, be a rare event because the incidence of lung cancer in ex-smokers, even in those who smoked for >40 years, is <1% a year.

Cairns continues by suggesting that what happens is an epigenetic event at the stem cell level:

> Normally, each time a stem cell divides, its two daughters have different fates: one has to replace its parent and stay within the confines of the stem cell niche, while the other is programmed to differentiate and multiply and eventually all its descendants die and are discarded. The final step could be the rare failure of this dichotomy, when a fully mutant cancerous stem cell produces a daughter cell that is freed from some nongenetic imperative to differentiate and die (an event analogous to Prehn's 'symmetrical division') (7).

I show below how some confirmation of this model has come from recent epidemiological and epigenetic ('exposome') research.

Mathematical models of carcinogenesis

Studying the timing of carcinogenic events has provided important hints on the putative mechanisms of cancer onset, i.e., of the processes that lead to the first malignant cell (M-cell). We have previously compared two modelling approaches to lung carcinogenesis: an evolutionary model based on the Wright–Fisher (WF) process and the MVK model of Moolgavkar, Venzon and Knudson, also known as the two-stage clonal expansion model. The main result of our study on lung cancer was that with both models (WF and MVK) we obtained similar values for the median waiting time to the first M-cell and that there was a dominating effect of clonal cell selection in carcinogenesis. Both the WF and MVK models predicted waiting times that are consistent across several previously reported data sets. Thus, while both mutation and selection seem necessary for carcinogenesis, the model comparison suggests that the formation of the first M-cell is driven mainly by clonal expansion of selectively advantageous cells, whereas the effect of the mutation rate appears smaller (8). We hypothesised that carcinogenesis is a Darwinian process involving stem cells that acquire selective advantage due to epigenetic (i.e. reversible) changes (9).

This work was preceded by work in cells in culture where we simulated a Darwinian environment by treating cells with high doses of toxic and carcinogenic chemicals. We investigated the effects of 4-aminobiphenyl (4-ABP) and *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine (MNNG) on genetically stable colorectal (HCT116) and bladder (RT112) cancer cells. Cells were treated with carcinogens to generate resistant clones that were then subjected to genetic analysis to assess whether they displayed either chromosome instability (CIN) or microsatellite instability (MIN). We found that 50–60% of cells treated with 4-ABP developed CIN but none developed MIN. In contrast, all MNNG-treated clones (12/12) developed MIN but none developed CIN as shown by the microsatellite assay. CIN or MIN remained stable after several cycles of cell replication. By providing a mechanistic link between exposure to a tobacco constituent and the development of CIN, our results contributed to a better understanding of the origins of genetic instability (10). Although not investigated in our experiment, it is now hypothesised that CIN arises mainly as a consequence of epigenetic phenomena. Therefore, mathematical models and experimental work on tobacco-related cancers (lung and bladder) suggested that the action of tobacco smoke can be mediated by epigenetic mechanisms.

Epigenome-wide methylation study

To investigate the dynamics of methylation in smoking, we ran a set of epigenome-wide analyses in a population sample of 1000 subjects (6). The evolution of the number of significant associations was broken down into two main phases. First, we observed a decrease in the number of associations for values of time since cessation lower than 19 years. The corresponding sites potentially represented loci whose smoking-induced methylation alterations revert back to levels that are typical of the unexposed after several years following smoking cessation. For a second group of probes, methylation levels reverted towards those of never smokers but never fully reached values typical of never smokers, suggesting the existence of sites whose methylation remains fixed even >40 years after last active exposure to tobacco smoke. The stability of these changes is not compatible with white blood cell half-life (replication rate), i.e. it suggests that haematopoietic stem cells of the bone marrow must be involved (an alternative—but unlikely—explanation is that toxic substances remain in the bone marrow for decades).

In our interpretation, these findings give rise to several important hypotheses: (i) haematopoietic stem cells are involved, since the persistence of altered gene methylation goes far beyond the half-life of mature white blood cells; (ii) methylation changes in key genes (such as *AHRR*) are likely to confer selective advantage to cells and lead to clonal cell selection. In fact, we have to bear in mind that what we call 'hypomethylation' is a change in the proportion of cells that are unmethylated at a certain CpG site compared to those that are methylated. Methylation at the single cell-single CpG site level is either present or not. Therefore, stem cells preserve a 'memory' of past exposures in the form of a greater proportion of unmethylated CpG sites versus methylated CpG sites. We speculate that exposure to toxic agents selects a clone of cells that are unmethylated in a CpG involved in the activation of a pathway reactive to environmental insults. This imprinting remains in the memory of stem cells and is comparable to immunological memory. The studies we conducted in smokers were in white blood cells, i.e. 'memory' involved haematopoietic stem cells. However, we also investigated the lung tissue of smokers and non-smokers (5). Methylation levels in the *AHRR* gene probes were significantly decreased ($P < 0.001$) and expression increased ($P = 0.0047$) in the lung tissue of current smokers compared with non-smokers. This was further validated in a mouse model of smoke exposure with similar results.

Meet-in-the-middle and the concept of emergence

The greatest challenge in epidemiology is to prove causality in the absence of experiment, i.e. to connect exposure to an agent to the onset of a disease via intermediate steps. The simplest approach is to start by determining the associations between exposures, intermediate markers and disease. Building on this, we have proposed an approach that is known as 'meet-in-the-middle' (MITM) (11) and includes by construction a multilayer causal framework. Its implementation aims to address the challenge of identifying causal relationships that link exposures and disease outcomes. This approach is based on a combination, within a population study, of (i) prospective search for intermediate biomarkers that are elevated in subjects who eventually go on to develop disease and (ii) retrospective search for links of such biomarkers to past environmental exposures.

However, the MITM concept is a very rudimentary approach to causality, corresponding to what Wesley Salmon called 'propagation of a mark' as a way to distinguish between genuine causality and

pure statistical association. A step further is the search for meaningful biochemical/molecular pathways. However, this step also tends to be elusive. It is in fact unlikely that we can identify a single pathway leading, e.g., from smoking to lung cancer. Several pathways are more likely to be in action and sophisticated mathematical approaches such as compartmental models and manifold embedding curves can help in unravelling them.

But the real challenge is that of emergence of a new property, as described by the literature on complexity, i.e. how exogenous molecules (e.g. pollutants) modify endogenous molecules (DNA, RNA, proteins) and how the latter make a difference in cells; then how cells with such changes emerge as having selective advantage or influencing surrounding or distant cells, up to the level of the organism and the population. The challenge is to understand which modifications (qualitative and quantitative) at each layer (molecules, cells, tissues, organisms and populations) are critical to cause a modification at the upper layer, i.e. the emergence of new properties. From this point of view, the onset of cancer is very similar to 'speciation', the appearance of a new species, and depends on changes in organic molecules followed by cell selection, the latter being influenced by the surrounding environment (the difference lies obviously in the fact that a new species is a well-organised organism made of a number of differentiated cells).

> In philosophy, systems theory, science, and art, emergence is conceived as a process whereby larger entities, patterns, and regularities arise through interactions among smaller or simpler entities that themselves do not exhibit such properties. Emergence is central in theories of integrative levels and of complex systems. For instance, the phenomenon life as studied in biology is commonly perceived as an emergent property of interacting molecules (…) (from Wikipedia: http://en.wikipedia.org/wiki/Emergence).

The formation of snowflakes is a good example of the emergence of new properties (shape and even beauty) from simple molecular interactions (Figure 1).

The wealth of research that is emerging on omics and the exposome, as testified by the Special Topic in the current issue of *Mutagenesis*, is likely to proceed step-by-step through the identification of good MITM candidates, to the elucidation of pathways, to the development of cross-omic validation of signals (e.g. the concurrent use of epigenomics, metabolomics and transcriptomics to reinforce causal claims on pathways from exposure to diseases). The papers in this issue cover different aspects of exposomics, referring to the exposome *in utero* and life-course influences (12); to nutrigenomics and transcriptomics (13) and metabolomics (14); to new statistical approaches to metabolomics (15) and to a new type of exposome, the systematic analysis of mutational spectra in relation to environmental exposures (16). What they have in common are a few features that are characteristic of the exposome philosophy: (i) explaining disease requires a life-course approach starting from conception onward; (ii) single molecular changes or even simple pathways will not suffice if we want to understand the connections between environmental exposures and disease and (iii) an evolutionary paradigm underlies the exposome paradigm, e.g. if we consider the evolution of the cancer phenotype through the history of mutational and epigenetic patterns.

Conclusions

One of the new concepts introduced in medicine in the last 50 years is that of 'action at a distance', e.g. through the 'early life theory of

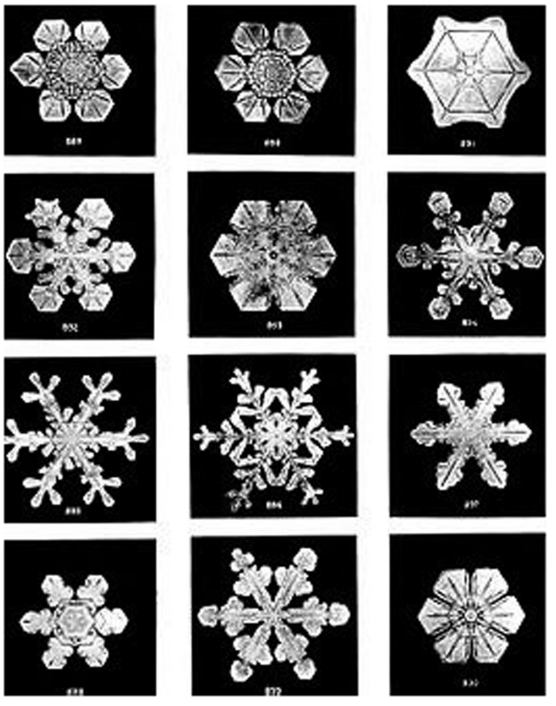

Figure 1. Snowflakes (from Wikipedia: http://en.wikipedia.org/wiki/Emergence).

disease' associated with the name of David Barker (17,18). In fact, there is strong evidence that (i) the risk of disease is influenced by early exposures, including *in utero*; (ii) influential life stages include critical periods (during which changes in exposure have long-term effects on disease risks or related, intermediate markers) and sensitive periods (during which an exposure has stronger effect on development and, hence, disease risk than at other time) (19). While action at a distance is commonly accepted in physics, it is rather new in medical causation. To embed this concept in biology, we need to hypothesise that early acquired changes are transferred to multiple generations of cells. A plausible mechanism comes from our observations in smokers because epigenetic changes are in fact transmitted through generations of cells. However, since methylation is an all or none event at the single cell-single CpG site level, we have to invoke selective advantage of unmethylated stem cells like in the example of *AHRR*. A similar mechanism may be at work in the famous example of the children who were *in utero* during the Dutch Hunger in the winter of 1944, whose *IGF2* gene probes were still hypomethylated 60 years later (20).

Based on these early studies, we may predict that great contributions will come from exposomics to the unravelling of the causes and mechanisms of common diseases, but the challenges are still enormous, both conceptually and practically. A challenge with high priority is the development of tools for the integration of existing and emerging knowledge on the evolution in time of intermediate phenotypes. The 'hallmarks of cancer' concept needs to be extended to diseases other than malignant tumours. Also, it is currently a static concept, i.e. it does not incorporate time dependency of intermediate phenotypes. Most bioinformatics databases and analytical tools that are currently available summarise information on the links between molecular features (and their networks and pathways), on the one hand, and disease phenotypes on the other hand. However, this integration does not yet incorporate time dependence. The time-dependent evolution between successive intermediate phenotypes during the

pathogenesis of most diseases does not seem to be adequately covered by existing bioinformatics databases, and this methodological development is probably one of the key aspects of future exposomics research.

Funding

This work was supported by the grant FP7 of the European Commission 'Enhanced exposure assessment and omic profiling for high priority environmental exposures in Europe' (no. 308610).

Acknowledgements

I am grateful to Gianluca Campanella, Vincent Cunliffe and an anonymous reviewer for fruitful comments.
Conflict of interest statement: None declared.

References

1. Rappaport, S. M. and Smith, M. T. (2010) Epidemiology. Environment and disease risks. *Science*, 330, 460–461.
2. Wild, C. P. (2012) The exposome: from concept to utility. *Int. J. Epidemiol*., 41, 24–32.
3. Vineis, P., Khan, A. E., Vlaanderen, J. and Vermeulen, R. (2009) The impact of new research technologies on our understanding of environmental causes of disease: the concept of clinical vulnerability. *Environ. Health*, 8, 54.
4. Vineis, P., van Veldhoven, K., Chadeau-Hyam, M. and Athersuch, T. J. (2013) Advancing the application of omics-based biomarkers in environmental epidemiology. *Environ. Mol. Mutagen*., 54, 461–467.
5. Shenker, N. S., Ueland, P. M., Polidoro, S., van Veldhoven, K., Ricceri, F., Brown, R., Flanagan, J. M. and Vineis, P. (2013) DNA methylation as a long-term biomarker of exposure to tobacco smoke. *Epidemiology*, 24, 712–716.
6. Guida, F., Sandanger, T., Polidoro, S., *et al.* (2015) Genome-wide methylation and gene expression profiling identifies reversible and irreversible changes after smoking cessation. *Hum. Mol. Genet*, 24, 2349–2359.
7. Cairns, J. (2006) Cancer and the immortal strand hypothesis. *Genetics*, 174, 1069–1072.
8. Schöllnberger, H., Beerenwinkel, N., Hoogenveen, R. and Vineis, P. (2010) Cell selection as driving force in lung and colon carcinogenesis. *Cancer Res*., 70, 6797–6803.
9. Vineis, P., Schatzkin, A. and Potter, J. D. (2010) Models of carcinogenesis: an overview. *Carcinogenesis*, 31, 1703–1709.
10. Saletta, F., Matullo, G., Manuguerra, M., Arena, S., Bardelli, A. and Vineis, P. (2007) Exposure to the tobacco smoke constituent 4-aminobiphenyl induces chromosomal instability in human cancer cells. *Cancer Res*., 67, 7088–7094.
11. Chadeau-Hyam, M., Athersuch, T. J., Keun, H. C., *et al.* (2011) Meeting-in-the-middle using metabolic profiling—a strategy for the identification of intermediate biomarkers in cohort studies. *Biomarkers*, 16, 83–88.
12. Ghantous, A., Hernandez-Vargas, H., Byrnes, G., Dwyer, T. and Herceg, Z. (2015) Characterizing the epigenome as a key component of the fetal exposome in evaluating in utero exposures and childhood cancer risk. *Mutagenesis*, 30, 733–742.
13. van Breda, S. G. J., Wilms, L. C., Gaj, S., Jennen, D. G. J., Briedé, J. J., Kleinjans, J. C. S. and de Kok, T. M. C. M. (2015) The exposome concept in a human nutrigenomics study: evaluating the impact of exposure to a complex mixture of phytochemicals using transcriptomics signatures. *Mutagenesis*, 30, 723–731.
14. Athersuch, T. and Keun, H. (2015) Metabolic profiling in human exposome studies. *Mutagenesis*, 30, 755–762.
15. Assi, N., Fages, A., Vineis, P., Chadeau-Hyam, M., *et al.* (2015) A statistical framework to model the meet-in-the-middle principle using metabolomic data: application to hepatocellular carcinoma in the EPIC study. *Mutagenesis*, 30, 743–753.
16. Nik-Zainal, S., Kucab, J. E., Morganella, S., *et al.* (2015) The genome as a record of environmental exposure. *Mutagenesis*, 30, 763–770.
17. Hales, C. N. and Barker, D. J. (2001) The thrifty phenotype hypothesis. *Br. Med. Bull*., 60, 5–20.
18. Cunliff, V. (2015) Experience-sensitive epigenetic mechanisms, developmental plasticity and the biological embedding of chronic disease risk. *Wiley Interdiscip. Rev. Syst. Biol. Med*., 7, 53–71.
19. Ben-Shlomo, Y. and Kuh, D. (2002) A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int. J. Epidemiol*., 31, 285–293.
20. Tobi, E. W., Lumey, L. H., Talens, R. P., Kremer, D., Putter, H., Stein, A. D., Slagboom, P. E. and Heijmans, B. T. (2009) DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum. Mol. Genet*., 18, 4046–4053.