

Beyond genotype to phenotype: why the phenotype of an individual cannot always be predicted from their genome sequence and the environment that they experience

Alejandro Burga¹ and Ben Lehner^{1,2}

¹ Genetic Systems, EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG) and UPF, Barcelona, Spain

² Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Keywords

genetic predictions; genetics of individuals; genotype-to-phenotype; noise; parental effects; phenotypic variance

Correspondence

B. Lehner, Genetic Systems, EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG) and UPF, Dr. Aiguader 88, 08003 Barcelona, Spain
Fax: +34 93 316 00 99
Tel: +34 93 316 01 00
E-mail: ben.lehner@crg.es

(Received 14 May 2012, revised 7 August 2012, accepted 24 August 2012)

doi:10.1111/j.1742-4658.2012.08810.x

One promise of personalized medicine is that it will be possible to make useful predictions about the phenotypes of individuals from their complete genome sequences (e.g. concerning their susceptibility to disease). However, to what extent is knowledge about an individual's genotype, together with information about the environment that they have experienced, sufficient to predict phenotypic variation? In the present review, we argue that, although the 'typical' phenotypic outcome of an individual's genome can be predicted, it is much more difficult to predict the actual outcome for a particular individual. We highlight three reasons for this. First, the outcome of mutations can be influenced by random (stochastic) processes. Second, genetic variation present in one generation can influence phenotypic traits in the next generation, even if individuals do not inherit this variation. Third, the environment experienced by one generation can influence phenotypic variation in the next generation. These contributions to phenotypic variation have long been appreciated by quantitative geneticists, although they have only recently been studied at the molecular level. Taken together, they mean that, in many cases, the genotypes of individuals and the environment that they experience may not be sufficient to determine their phenotypes. A more comprehensive genotype-to-phenotype model will be required to make accurate predictions about the biology of individuals.

Introduction

What is the biological basis of individuality? To put it simply: what makes each of us different from each other? The phenotype of each individual is usually considered as an interaction between two variables: the genes each individual carries and the environment that they experience. In this review, the word 'environment' refers to general external factors that are (or could) be shared by groups of individuals, such as diet, exposure to pathogens and lifestyle. However, are the genes and the environment of an individual sufficient to determine their phenotype? We review the evidence that

suggests this is not always the case. In the present review, we focus on three areas. First, recent studies have provided insights into how stochastic molecular variation can be an important influence on phenotypic variation. Second, the outcome of an inherited mutation can be influenced by the genotype of the previous generation. Third, in multiple species, it is now clear that the environment of one generation can influence the phenotype of subsequent generations. Research in these three areas is enriching our conception of what determines an individual, revealing that the genes and

Abbreviations

MZ, monozygotic; RNAi, RNA interference.

the environment of an individual are not necessarily sufficient to determine their phenotype (Fig. 1). We speculate that this may also be true in our own species, and that it could be important for understanding and predicting disease susceptibility.

Mapping phenotype to genotype plus the environment

The distinction between genotype and phenotype was an important conceptual advance made by Danish researcher Wilhelm Johannsen in 1903 [1]. When studying the inheritance of traits such as seed size in beans, he noted that, even though the beans were derived from highly inbred (isogenic) lines, the size of the seeds was variable and followed a normal distribution. To explain this observation, he differentiated between the genotype and phenotype of an individual seed and attributed the differences in seed size to variability in the environment in which the plants developed. This joint contribution of genes and environment has been the reigning paradigm that has been used to explain phenotypic variation among individuals. This suggests that, if we know the genome sequence of an individual and understand how the environment that they experi-

ence modifies the action of these genes, we should be able to predict all of their phenotypic traits. But how much can we say about any individual human simply based on knowledge of their genome sequence? A recent study went some way towards addressing this question in the model organism budding yeast, showing that, at least for some phenotypic traits, reasonable predictions could be made about how they vary relative to a reference individual from individual genome sequences [2]. The prevalence of genetic interactions or epistasis in genomes [3,4] also poses a challenge for these predictions because the effect of a particular variant could depend on the genetic background. However, such studies only assess the ‘typical’ outcome of a particular genome sequence, and not the actual phenotype of each individual with that genome.

There is no doubt that both genetic and environmental variation have a major role in determining phenotypic variation [5]. What is less clear in many instances is their relative contribution. In 1920, Sewall Wright, one of the founders of quantitative genetics, addressed this problem by studying the relative importance of inheritance and environment in the piebald pattern of guinea pigs [6]. Wright could estimate the contribution of genes (heritability) to the total phenotypic variance from parental–offspring correlations and, in addition, he could estimate the contribution of the environment common to litter mates before birth (tangible environment). In an outbred control population, 42% of the phenotypic variation in piebald pattern could be attributed to genes but, unexpectedly, the contribution of the tangible environment was very small. Wright hypothesized that the remaining 58% must be a result of ‘irregularities during development due to the intangible sort of causes to which the word chance is applied’ [6].

Seventy years later, Gärtner reached similar conclusions by studying highly inbred lines of mice that showed extensive phenotypic variation in diverse traits even when grown in controlled environments and attributed this variation to a ‘third component’ in addition to genes and environment [7]. In quantitative genetics, the term ‘environment’ was initially used by Fisher to denote arbitrary external causes independent of heredity [8]. Thus, traditionally, the total phenotypic variance of a population (V_p) is partitioned into a genetic (V_g) and an environmental (V_e) component, where $V_p = V_g + V_e$. The environmental variance term, V_e , includes both tangible and intangible (stochastic) sources of variation [9] (Fig. 1).

Progress made in the last decade has increased our knowledge about the sources and consequences of this previously intangible variation. This, together with the

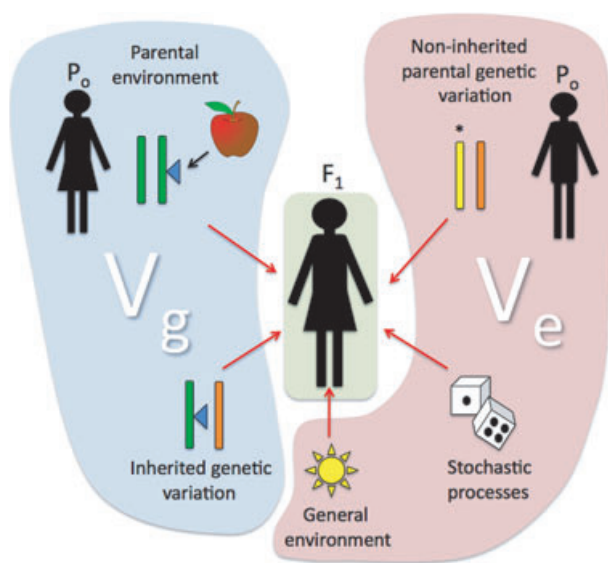


Fig. 1. Factors influencing the phenotype of an individual. In addition to the genome and environment experienced by individuals, recent advances have provided molecular insights into the influence of non-inherited parental genetic variants, parental environment and stochastic molecular variation (noise). These factors can be classified according to whether they would contribute to the traditional genetic (V_g) (light blue) or environmental (V_e) (light red) components of phenotypic variance (V_p) used in the population genetics literature.

fact that an important component of the phenotypic variance is not genetic, nor a result of general environmental factors, suggests that the measurement of stochastic variables is likely to increase our predictive power for individuals [10] and also suggests that stochastic processes should be considered as a third variable, independent from the environmental variance term [11]. Moreover, when aiming to predict the particular phenotype of individuals, it becomes a necessity to gain molecular insights into the various mechanisms in which V_g and V_e are further partitioned (Fig. 1).

Most human diseases also have a significant genetic component (i.e. heritability) [5], yet many observations indicate that genes and the tangible environment are insufficient to predict the phenotypes of individual people. Monozygotic (MZ) twins often differ substantially in their susceptibility to disease, despite sharing the same genome sequence and normally similar environments [12]. For example, if a woman develops breast cancer, the probability that her monozygotic twin sister will also develop this condition is only approximately 0.2 (concordance rate of 20%) [13]. However, if a man suffers from schizophrenia, his twin brother will suffer the same condition in approximately 50% of cases [14] (i.e. the genetic equivalent of tossing a coin). Although twin studies suggest the presence of a third factor besides genes and environment, a low concordance rate in disease liability in MZ twins cannot be taken as evidence for a lack of heritability [15]. MZ twins may show some somatic genetic variation [16,17], although this is unlikely to explain phenotypic-discordance in most cases [18,19]. Importantly, for these and many additional diseases [12], concordance rates set a limit to the best predictions we can ever hope to achieve from knowledge of an individual's genome alone. Aside from

somatic mutations arising in particular tissues [20], what are the causes of this variation in the outcome of mutations? Is the general environment responsible for these differences? No single pair of individuals has experienced exactly the same environment. However, several studies performed with twins who grew up in the same or different families have tried to estimate the contribution of the 'shared environment'. Shared environment makes reference to the environmental factors that two siblings growing in the same family have in common. The unexpected result of these studies is that the proportion of the variance explained for most traits and diseases is quite small. In other words, genetics aside, two siblings growing in the same house are not much more similar than any two children taken from the same population [21,22]. In model organisms, the failure of a simple genotype-to-phenotype model is even more striking because both the genotype and the environment can be controlled. For example, in both *Caenorhabditis elegans* [23] and mice [7,24,25], isogenic individuals often show substantial phenotypic variation in a highly-controlled homogenous environment, particularly when they carry detrimental mutations [26] (Fig. 2). What is the nature of this third variable?

Tossing a coin: how inter-individual variation in gene expression can influence the outcome of mutations

It has long been suspected [27,28] that isogenic individuals vary stochastically at the molecular level because of the low copy numbers of many important biological molecules involved in gene expression and the genome itself. Only recently, however, it has become possible to visualize and to quantify this inter-individual variation

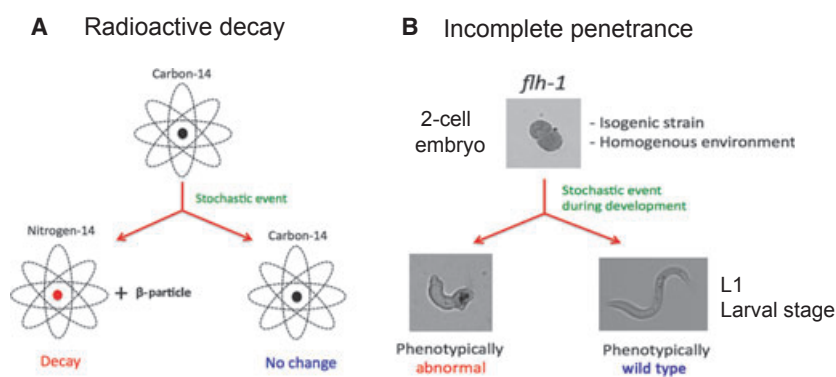


Fig. 2. Stochastic influences on the outcome of mutations. Radioactive decay is a stochastic process at the atomic level but deterministic in a population (A). Stochasticity and determinism are not mutually exclusive. Variation in the outcome of an inherited mutation in *C. elegans* (B). The mutation *flh-1(bc374)* inactivates a transcription factor required for embryonic development. However, even in isogenic strains and in a homogenous environment, only a subset of individuals is detrimentally affected by the mutation.

using fluorescent reporter constructs and single molecule detection techniques [29,30]. Inter-individual variation in gene expression can be substantial, in both bacteria and higher eukaryotes [31–33] and can have important biological consequences, such as in cell signalling and development [34,35]. It can also be a substantial cause of variation with respect to the phenotypic consequences of inherited mutations, as revealed in several recent studies [10,36–38]. In this review, we focus on variability in gene expression, which could also be influenced by variation in DNA methylation [39] or errors in transcription or translation [40]. However, promiscuous molecular interactions or catalysis [41,42], protein aggregation [43], variable cell-to-cell contacts and variation in mechanical force generation could all have a stochastic nature that influences phenotypic variation.

The first study to link variability in gene expression to the phenotypic outcome of a mutation (incomplete penetrance) was performed by Elowitz *et al.* [44]. The bacteria *Bacillus subtilis* can develop into a dormant spore when environmental conditions do not guarantee proper cell growth [45]. *SpoIIR*, together with other genes, activates this differentiation programme. Genetic modifications affecting the rate and/or time of onset of *spoIIR* expression were shown to affect the sporulation process of only a subset of the population. Quantifying the induction of a transcriptional reporter for *SpoIIR* by time-lapse microscopy revealed that variation in the rate and delay of expression of this gene was correlated with the phenotypic outcome of a mutation in *SpoIIR* in each individual. However, variation in the induction of the mutated gene only partially explained variation in the outcome of the mutation amongst individuals. This suggests that other unidentified factors, such as variation in the expression of other genes, are also important: reduced activity of the *spoIIR* gene exposed phenotypic variation that was normally buffered by the fully active gene [44].

The downstream consequences of an incompletely penetrant mutation have also been studied in *C. elegans* [36]. As a model, Van Oudenaarden and colleagues [36] used the development of the worm intestine, for which the gene regulatory network is well described [46]. The gene *skn-1* codes for a transcription factor that is maternally provided to the zygote and is required for embryonic specification of the EMS blastomere [47]. Embryos from mothers carrying mutations in *skn-1* undergo a developmental arrest (100% lethality) and the endoderm is absent in approximately 70% of embryos [47]. The SKN-1 protein initiates a highly redundant regulatory cascade by activating the expression of pairs of transcription factors, including *med-1/*

med-2 and *end-1/end-3*, which later leads to the activation of *elt-2*, the master regulator of gene expression in the intestine. Using a fluorescence *in situ* hybridization technique that can detect single mRNA molecules, the levels of several downstream targets of *skn-1* were quantified in individual embryos. *skn-1* mutants showed an increased variability among individuals in the expression of the target gene *end-1* and a lower mean level compared to wild-type worms. Animals with *end-1* expression below a certain threshold failed to activate *elt-2*, producing an approximate binary response in the activation of this gene that presumably affects intestinal development. In a similar study, mutations in the transcription factor *alr-1* were shown to increase variation in the expression of the target gene *mec-3* in *C. elegans* [37].

These studies [36,37] demonstrate how mutations in transcription factors can increase the variability in the expression of target genes and also how this variability can propagate in gene regulatory cascades. However, although single molecule detection by fluorescence *in situ* hybridization allows an elegant quantitative description of the system, the use of fixed samples limits the study of gene expression dynamics and makes it difficult to establish a causal relationship between molecular variation and the actual phenotypic variation that occurs in each individual. The use of fluorescent reporter proteins such as green fluorescent protein allows the dynamics of gene expression to be quantified *in vivo* and compared among individuals. Importantly, the use of live imaging to quantify gene expression also makes it possible to evaluate how well particular differences in gene expression predict later phenotypic variation.

Genetically identical *C. elegans* individuals growing in an homogeneous environment show a high degree of variation in their lifespan, similar to the high variability present in human populations [48]. Also, isogenic lines of rats growing under laboratory controlled conditions show a large range of variability in their life spans (ranging from 60 to 140 weeks) [49]. Rea *et al.* [50] noted that, after applying a mild heat stress, which extends lifespan, the induction of a transcriptional fluorescent reporter for the heat shock *hsp-16.2* gene was highly variable among individual worms. Furthermore, they showed that worms inducing higher levels of the reporter were longer lived compared to their low-expression counterparts [50]. However, given that this heat stress increases the mean lifespan of the population, it is not clear whether the predicted lifespan differences are the result of intrinsic differences in the nonstressed individuals or differences in the heat shock response.

In another example of the use of reporters, Pincus *et al.* [51] developed an ingenious way of measuring the growth rate and reporter fluorescence from individual worms during most of their lifespan. They found that variability in the expression levels of three microRNA reporters through mid-adulthood (*mir-71*, *mir-246* and *mir-239*) was a predictor of lifespan [51]. Given that these microRNAs act upstream in the insulin pathway, variability in their levels may be one causal determinant of lifespan. Also, there is evidence that a variable pathogenicity response among individuals, as reported by a *sod-3* transcriptional reporter, could play a role in determining the lifespan of *C. elegans* [52]. In all of these studies, however, reporter gene expression levels are measured quite late in life, meaning that they could, in reality, be reporting on variation in the life-history exposure to environmental stimuli.

We have also made use of fluorescent reporters to understand the causes of incomplete penetrance in *C. elegans* [10]. *tbx-8* and *tbx-9* are two partially redundant genes that originated from a gene duplication event and are required for the morphogenesis of the worm [53,54]. A double knockout of these genes is 100% embryonic lethal, although the deletion of each gene alone results in a subset of embryos with abnormal phenotypes, despite the fact that all of them are genetic clones developing in the same environment. We found that *tbx-9* null mutants induced higher levels *tbx-8* compared to non-mutant animals, indicating the existence of a compensatory feedback circuit [55,56]. Moreover, the expression of this reporter correlated with the phenotypic outcome of the mutation: those embryos with higher levels of induction of the redundant gene were more likely to develop into phenotypically normal animals.

We also found that variation in the expression of chaperones such as *daf-21* (homologue of mammalian HSP90) early in development predicted *tbx-9* mutation outcome. The outcome of many mutations is, directly or indirectly, dependent on chaperone activity [38,57–59], and so stochastic variation in chaperone activity among individuals may represent a somewhat general influence on mutation outcome [10,38]. In the case of the *tbx-9* mutation, variation in the two buffering mechanisms (partially redundant gene duplicate and molecular chaperone expression) is independent. Thus, simultaneously quantifying the levels of both systems increased the accuracy of predictions. For any particular mutation, variation in many genetic interaction partners [3] could contribute to incomplete penetrance. By quantifying this variation, it should be possible to increase the accuracy

of phenotypic predictions in individuals carrying a mutation.

Not your genes, but your parents'

Beyond stochastic molecular variation, there are other reasons why the genotype and environment of an individual can be insufficient to explain their phenotype. For example, some mutations do not affect the phenotype of the individuals who carry them but rather that of their offspring. Famous examples of this phenomenon are mutations in genes that code for maternal factors controlling development. The early development of a wide range of organisms depends on the activity of genes contributed maternally as mRNA or protein [60]. For example, in *Drosophila melanogaster*, genes such as *bicoid* and *hunchback* are contributed maternally and deposited as mRNA during oogenesis [61]. Consequently, a female fly homozygous mutant for *bicoid* will produce offspring with a 'mutant' phenotype even if the offspring themselves are only heterozygous for the mutation. Thus, in this relatively simple scenario, it is clear that the phenotype of the offspring also depends on the genotype of the mother.

Interestingly, the parental genotype can also influence disease susceptibility in offspring. Xing *et al.* [62] studied this phenomenon in *Drosophila* using a hyperactive JAK kinase mutant that develops blood tumours. A screen for genetic modifiers of this mutation (*hop^{Tum-1}*) revealed that many of these modifiers (such as Krüppel) act epigenetically and that their effects persisted in the offspring for at least two generations even in the absence of the modifier mutation [62] (Fig. 3). If this phenomenon also applies to humans, then the genetic variation in your parents' genome that you did not inherit could increase your risk of developing disease (e.g. cancer).

Another elegant example of an individual's phenotype being influenced by the genetics of their parents is provided in *C. elegans*. Genetic incompatibility occurs when individuals with different genotypes produce nonviable or infertile offspring. This occurs in crosses between two different strains of *C. elegans*; when the two strains are mated, a subset of the hybrid progeny die. Mapping the loci underlying this incompatibility identified two important natural variants. For one of these, it is the zygotic genotype that is important but, for the other, it is the parental genotype that matters: embryos homozygous for a mutation in the gene *zeel-1* only arrest if the parent's sperm carries the incompatible allele in the gene *peel-1* [63]. Thus, the outcome of a mutation in *zeel-1* depends on the parental genotype at a second locus.

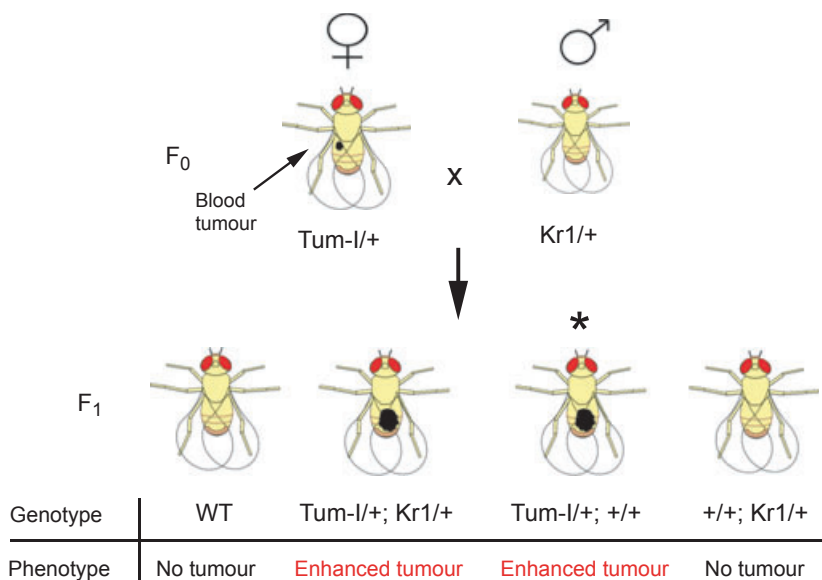


Fig. 3. Parental genotype influences on an offspring's phenotype. Paternal non-inherited genotype influences blood tumour susceptibility in the offspring of flies [62]. Mutations in Krüppel (Kr1) can enhance tumour formation in offspring carrying a *Hop^{Tum-I}* mutation even when the offspring do not inherit the Kr1 mutation (*).

Blame your parents' lifestyle: transgenerational 'epigenetic' influences on the outcome of inherited mutations

Mutations in particular genes are relatively infrequent events. However, changes in the environment can be much more frequent. How can biological systems cope with this? In response to environmental change, a single genome can produce different phenotypes providing adaptation to the new conditions. This property is known as phenotypic plasticity [64]. If a new environmental condition is also likely to be experienced by the next generation, then it could prove adaptive to pass on information about the environment to the next generation and so elicit an appropriate phenotypic response. An old and controversial idea, such 'Lamarckian' influences on phenotypic variation are again beginning to receive much attention.

Phenotypic changes that occur in response to an environmental change in the previous generation have now been reported in several species [65–68]. One example is the flowering time in the monocarpic herb *Campanulastrum americanum*, where the maternal light environment influences the choice between annual and biennial flowering time in the next generation [69]. This 'transgenerational' plasticity could provide a fitness advantage, predicting long-term environmental variation [69].

In plants, there is no early separation of soma and germline, and gametes can be derived from somatic tissue late in development. It is therefore much easier to envisage how the environment experienced by one generation could influence the phenotype of the next generation. However, there are now also increasing numbers of likely examples of epigenetic (non-DNA encoded) inheritance described in animals [67–70]. For example, the waterflea *Daphnia cucullata* responds to the presence of predators by changing its morphology and developing a larger helmet. Agrawal *et al.* [70] showed that the progeny of mothers who grew in the presence of predators were better protected (had larger helmets) than those whose mothers grew in a control environment. This adaptive phenotypic plasticity effect could be detected for two generations.

In nematodes, RNA interference (RNAi)-triggered gene silencing can be inherited for many generations when selecting for phenotypically silenced progenitors in the absence of the original dsRNA trigger [71–73]. A recent study by Hobert and colleagues [86] provided a potential explanation for the existence of such a mechanism. It was found that the anti-viral RNAi-response of *C. elegans* [74] can also be transmitted to its progeny, and also that the antiviral RNA agent could be transmitted through sperm [75]. Perhaps this RNAi-based mechanism is also responsible for a reported case of inheritance of a behavioural imprinting in *C. elegans* [76]. In addition, mutations in genes

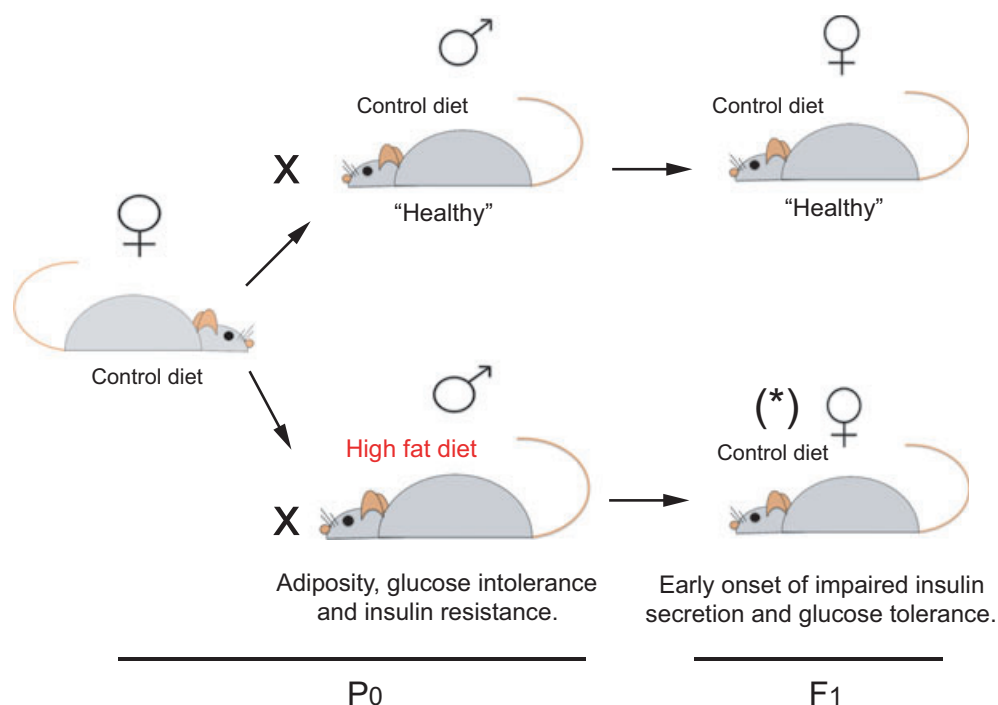


Fig. 4. Parental environment influences on the phenotype of offspring. Ng *et al.* [85] recently reported that female rats whose fathers were fed a high-fat diet (*) show altered glucose-insulin homeostasis compared to a control diet group.

coding for chromatin modifiers have been shown to have transgenerational effects on lifespan in *C. elegans*, possibly by altering the normal resetting of chromatin marks that takes place in the germline [77,78].

Transgenerational effects have also been described in mammals. Examples include the epigenetic inheritance of coat colour in mice [79–81], parental imprinting of gene expression [82], maternal behaviours influencing gene expression in offspring [83] and inherited RNA-induced ‘paramutations’ [84]. Two recent studies have also demonstrated the potential for the transgenerational inheritance of environmental information in mammals. Ng *et al.* [85] showed that changes in paternal diet can influence the metabolic status of offspring in rats. Male rats fed a high-fat or normal diet were crossed with females on a control diet. Markedly, the female offspring of fathers fed a high-fat diet had impaired glucose-insulin homeostasis (Fig. 4). Expression profiling using microarrays showed that the β -cells of affected offspring had mild changes in the expression of hundreds of genes. For at least one gene, this change in expression was correlated with changes in the DNA methylation state of its promoter region [85]. In a similar study, Rando and colleagues showed that a paternal low-protein diet in mice caused an up-regulation of proliferation and lipid biosynthesis gene expression in the livers of the next generation. These changes in hepatic

gene expression were also modest and associated with differences in promoter methylation [86]. However, as in rats, the actual mechanism of transgenerational inheritance remains unknown.

In humans too, there is some epidemiological evidence for transgenerational environmental influences on phenotypic variation [87,88]. How can environmental information be propagated through the germline to the next generation? Possibilities include patterns of histone modifications; even human sperm retain modified histones in many promoter regions [89,90] and these are transmitted to the zygote [91]. In addition, DNA methylation, noncoding RNAs, proteins and metabolite levels could all transmit epigenetic information across generations. Understanding how this information is transmitted, as well as its consequences for phenotypic variation and evolutionary theory, are major questions for future research.

Conclusions

With advances in next generation technologies, there is great excitement about the use of individual genome sequencing for the prediction of disease risk and in the development of personalized medicine. But what can we really expect from this? We have highlighted in the present review that studies in monozygotic twins and

model organisms show that precise phenotypic predictions from personal genomes are unlikely. The genome of an individual is an important factor, although clearly not sufficient to explain variation in most phenotypic traits. A role for the environment in influencing phenotypic variation is widely appreciated, although we have emphasized three additional important influences on phenotypic variation that, although long recognized by quantitative geneticists [9], have only recently begun to be studied at the molecular level. First, biological processes are inherently stochastic and this random variation can be an important influence on the outcome of a particular mutation. Second, genetic variation carried by one generation can be an important influence on phenotypic variation in the next generation, especially for developmental processes. Third, there is now good evidence that the environment experienced by one generation can influence the phenotype of the next generation. Maternal and paternal genetic effects, as well as nongenetic inheritance, clearly demonstrate that parental influence on offspring goes beyond the inherited genes. New layers of detail will undoubtedly need to be added in forthcoming years. For example, humans can be considered as 'superorganisms' with an internal ecosystem of diverse symbiotic microbiota and parasites [92], which is highly variable among individuals [93], and the implications of this in health and disease are only beginning to be studied.

When it comes to human genetics, it is the individual that matters. A person does not want to only know the typical outcome of a mutation that she carries, she wants to know whether she will actually develop a disease or not. To be able to make predictions about the phenotypes of individuals, it is clear from both historical and recent work in model organisms that knowledge of genome sequencing will be insufficient in many cases. Rather, we need to consider how genetic, environmental and stochastic variation, together with transgenerational effects, combine to determine the phenotypes of individuals.

Acknowledgements

Our research is funded by grants from the European Research Council, Ministerio de Ciencia e Innovación Plan Nacional BFU2008-00365 and BFU2011-26206, Agència de Gestió d'Ajuts Universitaris i de Recerca, ERASysBio+, the European Molecular Biology Organization Young Investigator Program, EU Framework 7 project 277899 4DCellFate, the EMBL/CRG Systems Biology Program and by a Formación de Personal Investigador–Ministerio de Ciencia e Innovación

fellowship to A.B. We thank three anonymous reviewers for their helpful suggestions.

References

- Johannsen W (1909) Elemente der exakten Erblichkeitslehre. Gustav Fischer, Jena.
- Jelier R, Semple JI, Garcia-Verdugo R & Lehner B (2011) Predicting phenotypic variation in yeast from individual genome sequences. *Nat Genet* **43**, 1270–1274.
- Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S *et al.* (2010) The genetic landscape of a cell. *Science* **327**, 425–431.
- Gerke J, Lorenz K & Cohen B (2009) Genetic interactions between transcription factors cause natural variation in yeast. *Science* **323**, 498–501.
- Visscher PM, Hill WG & Wray NR (2008) Heritability in the genomics era – concepts and misconceptions. *Nat Rev Genet* **9**, 255–266.
- Wright S (1920) The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proc Natl Acad Sci USA* **6**, 320–332.
- Gärtner K (1990) A third component causing random variability beside environment and genotype. A reason for the limited success of a 30 year long effort to standardize laboratory animals? *Lab Anim* **24**, 71–77.
- Fisher RA (1918) The correlation between relatives on the supposition of Mendelian Inheritance. *Trans R Soc Edinb* **52**, 399–433.
- Lynch M & Walsh B (1998) Genetics and Analysis of Quantitative Traits. Sinauer Associates, Inc., Sunderland, MA.
- Burga A, Casanueva MO & Lehner B (2011) Predicting mutation outcome from early stochastic variation in genetic interaction partners. *Nature* **480**, 250–253.
- Finch CE & Kirkwood TB (2000) Chance, Development and Aging. Oxford University Press, New York, NY.
- Wong AH, Gottesman II & Petronis A (2005) Phenotypic differences in genetically identical organisms: the epigenetic perspective. *Hum Mol Genet* **14**(Spec No 1), R11–R18.
- Hamilton AS & Mack TM (2003) Puberty and genetic susceptibility to breast cancer in a case-control study in twins. *N Engl J Med* **348**, 2313–2322.
- Cardno AG & Gottesman II (2000) Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *Am J Med Genet* **97**, 12–17.
- Smith C (1970) Heritability of liability and concordance in monozygous twins. *Ann Hum Genet* **34**, 85–91.
- Bruder CE, Piotrowski A, Gijsbers AA, Andersson R, Erickson S, Diaz de Stahl T, Menzel U, Sandgren J, von Tell D, Poplawski A *et al.* (2008) Phenotypically

- concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am J Hum Genet* **82**, 763–771.
- 17 Kaminsky ZA, Tang T, Wang SC, Ptak C, Oh GH, Wong AH, Feldcamp LA, Virtanen C, Halfvarson J, Tysk C *et al.* (2009) DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet* **41**, 240–245.
 - 18 Lasa A, Ramon y Cajal T, Llort G, Suela J, Cigudosa JC, Cornet M, Alonso C, Barnadas A & Baiget M (2010) Copy number variations are not modifiers of phenotypic expression in a pair of identical twins carrying a BRCA1 mutation. *Breast Cancer Res Treat* **123**, 901–905.
 - 19 Baranzini SE, Mudge J, van Velkinburgh JC, Khankhanian P, Khrebtukova I, Miller NA, Zhang L, Farmer AD, Bell CJ, Kim RW *et al.* (2010) Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* **464**, 1351–1356.
 - 20 Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan PM, Rizzu P, Smith S, Fell M *et al.* (2011) Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **479**, 534–537.
 - 21 Plomin R & Daniels D (2011) Why are children in the same family so different from one another? *Int J Epidemiol* **40**, 563–582.
 - 22 Smith GD (2011) Epidemiology, epigenetics and the ‘Gloomy Prospect’: embracing randomness in population health research and practice. *Int J Epidemiol* **40**, 537–562.
 - 23 Horvitz HR & Sulston JE (1980) Isolation and genetic characterization of cell-lineage mutants of the nematode *Caenorhabditis elegans*. *Genetics* **96**, 435–454.
 - 24 Medeiros LA, Dennis LM, Gill ME, Houbaviy H, Markoulaki S, Fu D, White AC, Kirak O, Sharp PA, Page DC *et al.* (2011) Mir-290-295 deficiency in mice results in partially penetrant embryonic lethality and germ cell defects. *Proc Natl Acad Sci USA* **108**, 14163–14168.
 - 25 Pereira R, Halford K, Sokolov BP, Khillan JS & Prockop DJ (1994) Phenotypic variability and incomplete penetrance of spontaneous fractures in an inbred strain of transgenic mice expressing a mutated collagen gene (COL1A1). *J Clin Invest* **93**, 1765–1769.
 - 26 Waddington CH (1957) *The Strategy of the Genes; A Discussion of Some Aspects of Theoretical Biology*. Allen & Unwin, London.
 - 27 Spudich JL & Koshland DE Jr (1976) Non-genetic individuality: chance in the single cell. *Nature* **262**, 467–471.
 - 28 Schrödinger E (1945) *What is Life?: The Physical Aspect of the Living Cell*. University Press, Cambridge.
 - 29 Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A & Tyagi S (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* **5**, 877–879.
 - 30 Kaern M, Elston TC, Blake WJ & Collins JJ (2005) Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* **6**, 451–464.
 - 31 Elowitz MB, Levine AJ, Siggia ED & Swain PS (2002) Stochastic gene expression in a single cell. *Science* **297**, 1183–1186.
 - 32 Raser JM & O’Shea EK (2004) Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814.
 - 33 Bar-Even A, Paulsson J, Maheshri N, Carmi M, O’Shea E, Pilpel Y & Barkai N (2006) Noise in protein expression scales with natural protein abundance. *Nat Genet* **38**, 636–643.
 - 34 Wernet MF, Mazzoni EO, Celik A, Duncan DM, Duncan I & Desplan C (2006) Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature* **440**, 174–180.
 - 35 Chang HH, Hemberg M, Barahona M, Ingber DE & Huang S (2008) Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544–547.
 - 36 Raj A, Rifkin SA, Andersen E & van Oudenaarden A (2010) Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913–918.
 - 37 Topalidou I, van Oudenaarden A & Chalfie M (2011) *Caenorhabditis elegans* aristaless/Arx gene *alr-1* restricts variable gene expression. *Proc Natl Acad Sci USA* **108**, 4063–4068.
 - 38 Casanueva MO, Burga A & Lehner B (2012) Fitness trade-offs and environmentally induced mutation buffering in isogenic *C. elegans*. *Science* **335**, 82–85.
 - 39 Becker C, Hagmann J, Muller J, Koenig D, Stegle O, Borgwardt K & Weigel D (2011) Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* **480**, 245–249.
 - 40 Meyerovich M, Mamou G & Ben-Yehuda S (2010) Visualizing high error levels during gene expression in living bacterial cells. *Proc Natl Acad Sci USA* **107**, 11543–11548.
 - 41 Aharoni A, Gaidukov L, Khersonsky O, Mc QGS, Roodveldt C & Tawfik DS (2005) The ‘evolability’ of promiscuous protein functions. *Nat Genet* **37**, 73–76.
 - 42 Vavouri T, Semple JI, Garcia-Verdugo R & Lehner B (2009) Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* **138**, 198–208.
 - 43 Hortschansky P, Schroeckh V, Christopeit T, Zandomenighi G & Fandrich M (2005) The aggregation kinetics of Alzheimer’s beta-amyloid peptide is controlled by stochastic nucleation. *Protein Sci* **14**, 1753–1759.
 - 44 Eldar A, Chary VK, Xenopoulos P, Fontes ME, Loson OC, Dworkin J, Piggot PJ & Elowitz MB (2009) Partial

- penetrance facilitates developmental evolution in bacteria. *Nature* **460**, 510–514.
- 45 Hilbert DW & Piggot PJ (2004) Compartmentalization of gene expression during *Bacillus subtilis* spore formation. *Microbiol Mol Biol Rev* **68**, 234–262.
 - 46 Maduro MF (2009) Structure and evolution of the *C. elegans* embryonic endomesoderm network. *Biochim Biophys Acta* **1789**, 250–260.
 - 47 Bowerman B, Eaton BA & Priess JR (1992) *skn-1*, a maternally expressed gene required to specify the fate of ventral blastomeres in the early *C. elegans* embryo. *Cell* **68**, 1061–1075.
 - 48 Kirkwood TB, Feder M, Finch CE, Franceschi C, Globerson A, Klingenberg CP, LaMarco K, Omholt S & Westendorp RG (2005) What accounts for the wide variation in life span of genetically identical organisms reared in a constant environment? *Mech Ageing Dev* **126**, 439–443.
 - 49 Soffritti M, Belpoggi F, Degli Esposti D, Lambertini L, Tibaldi E & Rigano A (2006) First experimental demonstration of the multipotential carcinogenic effects of aspartame administered in the feed to Sprague–Dawley rats. *Environ Health Perspect* **114**, 379–385.
 - 50 Rea SL, Wu D, Cypser JR, Vaupel JW & Johnson TE (2005) A stress-sensitive reporter predicts longevity in isogenic populations of *Caenorhabditis elegans*. *Nat Genet* **37**, 894–898.
 - 51 Pincus Z, Smith-Vikos T & Slack FJ (2011) MicroRNA predictors of longevity in *Caenorhabditis elegans*. *PLoS Genet* **7**, e1002306.
 - 52 Sanchez-Blanco A & Kim SK (2011) Variable pathogenicity determines individual lifespan in *Caenorhabditis elegans*. *PLoS Genet* **7**, e1002047.
 - 53 Pocock R, Ahringer J, Mitsch M, Maxwell S & Woollard A (2004) A regulatory network of T-box genes and the even-skipped homologue *vab-7* controls patterning and morphogenesis in *C. elegans*. *Development* **131**, 2373–2385.
 - 54 Andachi Y (2004) *Caenorhabditis elegans* T-box genes *tbx-9* and *tbx-8* are required for formation of hypodermis and body-wall muscle in embryogenesis. *Genes Cells* **9**, 331–344.
 - 55 Kafri R, Bar-Even A & Pilpel Y (2005) Transcription control reprogramming in genetic backup circuits. *Nat Genet* **37**, 295–299.
 - 56 DeLuna A, Springer M, Kirschner MW & Kishony R (2010) Need-based up-regulation of protein levels in response to deletion of their duplicate genes. *PLoS Biol* **8**, e1000347.
 - 57 Rutherford SL & Lindquist S (1998) Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342.
 - 58 Tokuriki N & Tawfik DS (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459**, 668–673.
 - 59 Queitsch C, Sangster TA & Lindquist S (2002) Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624.
 - 60 Schier AF (2007) The maternal-zygotic transition: death and birth of RNAs. *Science* **316**, 406–407.
 - 61 Berleth T, Burri M, Thoma G, Bopp D, Richstein S, Frigerio G, Noll M & Nusslein-Volhard C (1988) The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo. *EMBO J* **7**, 1749–1756.
 - 62 Xing Y, Shi S, Le L, Lee CA, Silver-Morse L & Li WX (2007) Evidence for transgenerational transmission of epigenetic tumor susceptibility in *Drosophila*. *PLoS Genet* **3**, 1598–1606.
 - 63 Seidel HS, Rockman MV & Kruglyak L (2008) Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. *Science* **319**, 589–594.
 - 64 Pigliucci M, Murren CJ & Schlichting CD (2006) Phenotypic plasticity and evolution by genetic assimilation. *J Exp Biol* **209**, 2362–2367.
 - 65 Galloway LF (2005) Maternal effects provide phenotypic adaptation to local environmental conditions. *New Phytol* **166**, 93–99.
 - 66 Youngson NA & Whitelaw E (2008) Transgenerational epigenetic effects. *Annu Rev Genomics Hum Genet* **9**, 233–257.
 - 67 Jablonka E & Raz G (2009) Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Q Rev Biol* **84**, 131–176.
 - 68 Mousseau TA & Fox CWE (1998) *Maternal Effects as Adaptations*. Oxford University Press, New York, NY.
 - 69 Galloway LF & Etterson JR (2007) Transgenerational plasticity is adaptive in the wild. *Science* **318**, 1134–1136.
 - 70 Agrawal AA, Laforsch C & Tollrian R (1999) Transgenerational induction of defences in animals and plants. *Nature* **401**, 60–63.
 - 71 Grishok A, Tabara H & Mello CC (2000) Genetic requirements for inheritance of RNAi in *C. elegans*. *Science* **287**, 2494–2497.
 - 72 Alcazar RM, Lin R & Fire AZ (2008) Transmission dynamics of heritable silencing induced by double-stranded RNA in *Caenorhabditis elegans*. *Genetics* **180**, 1275–1288.
 - 73 Grishok A (2005) RNAi mechanisms in *Caenorhabditis elegans*. *FEBS Lett* **579**, 5932–5939.
 - 74 Lu R, Maduro M, Li F, Li HW, Broitman-Maduro G, Li WX & Ding SW (2005) Animal virus replication and RNAi-mediated antiviral silencing in *Caenorhabditis elegans*. *Nature* **436**, 1040–1043.
 - 75 Rechavi O, Minevich G & Hobert O (2011) Transgenerational inheritance of an acquired small RNA-based antiviral response in *C. elegans*. *Cell* **147**, 1248–1256.

- 76 Remy JJ (2010) Stable inheritance of an acquired behavior in *Caenorhabditis elegans*. *Curr Biol* **20**, R877–R878.
- 77 Greer EL, Maures TJ, Ucar D, Hauswirth AG, Mancini E, Lim JP, Benayoun BA, Shi Y & Brunet A (2011) Transgenerational epigenetic inheritance of longevity in *Caenorhabditis elegans*. *Nature* **479**, 365–371.
- 78 Katz DJ, Edwards TM, Reinke V & Kelly WG (2009) A *C. elegans* LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. *Cell* **137**, 308–320.
- 79 Morgan HD, Sutherland HG, Martin DI & Whitelaw E (1999) Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* **23**, 314–318.
- 80 Rakyan VK, Chong S, Champ ME, Cuthbert PC, Morgan HD, Luu KV & Whitelaw E (2003) Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. *Proc Natl Acad Sci USA* **100**, 2538–2543.
- 81 Wolff GL, Kodell RL, Moore SR & Cooney CA (1998) Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *FASEB J* **12**, 949–957.
- 82 Reik W & Walter J (2001) Genomic imprinting: parental influence on the genome. *Nat Rev Genet* **2**, 21–32.
- 83 Francis D, Diorio J, Liu D & Meaney MJ (1999) Nongenomic transmission across generations of maternal behavior and stress responses in the rat. *Science* **286**, 1155–1158.
- 84 Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I & Cuzin F (2006) RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**, 469–474.
- 85 Ng SF, Lin RC, Laybutt DR, Barres R, Owens JA & Morris MJ (2010) Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. *Nature* **467**, 963–966.
- 86 Carone BR, Fauquier L, Habib N, Shea JM, Hart CE, Li R, Bock C, Li C, Gu H, Zamore PD *et al.* (2010) Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* **143**, 1084–1096.
- 87 Painter RC, Osmond C, Gluckman P, Hanson M, Phillips DI & Roseboom TJ (2008) Transgenerational effects of prenatal exposure to the Dutch famine on neonatal adiposity and health in later life. *BJOG* **115**, 1243–1249.
- 88 Gluckman PD, Hanson MA & Beedle AS (2007) Non-genomic transgenerational inheritance of disease risk. *BioEssays* **29**, 145–154.
- 89 Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT & Cairns BR (2009) Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**, 473–478.
- 90 Arpanahi A, Brinkworth M, Iles D, Krawetz SA, Paradowska A, Platts AE, Saida M, Steger K, Tedder P & Miller D (2009) Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. *Genome Res* **19**, 1338–1349.
- 91 van der Heijden GW, Ramos L, Baart EB, van den Berg IM, Derijck AA, van der Vlag J, Martini E & de Boer P (2008) Sperm-derived histones contribute to zygotic chromatin in humans. *BMC Dev Biol* **8**, 34.
- 92 Nicholson JK, Holmes E, Lindon JC & Wilson ID (2004) The challenges of modeling mammalian biocomplexity. *Nat Biotechnol* **22**, 1268–1274.
- 93 HMPC (2012) Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214.