Available online at www.sciencedirect.com
ScienceDirect
Stud. Hist. Phil. Biol. & Biomed. Sci. 38 (2007) 238–254
Studies in History and Philosophy of Biological and Biomedical Sciences
www.elsevier.com/locate/shpsc

Gene expression and the concept of the phenotype

Ohad Nachtomy [a], Ayelet Shavit [b], Zohar Yakhini [c]

[a] *Department of Philosophy, Bar-Ilan University, Ramat-Gan 52900, Israel*
[b] *Department of Philosophy, University of California, Davis, 1 Shields Avenue, Davis CA 95616, USA*
[c] *Agilent Technologies Israel Ltd. Azorim Business Park, 94 EM Hamoshavot Road, 49527 Petach Tikva, Israel*

Received 8 August 2005; received in revised form 14 April 2006

Abstract

While the definition of the 'genotype' has undergone dramatic changes in the transition from classical to molecular genetics, the definition of the 'phenotype' has remained for a long time within the classical framework. In addition, while the notion of the genotype has received significant attention from philosophers of biology, the notion of the phenotype has not. Recent developments in the technology of measuring gene-expression levels have made it possible to conceive of phenotypic traits in terms of levels of gene expression. We demonstrate that not only has this become possible but it has also become an actual practice. This suggests a significant change in our conception of the phenotype: as in the case of the 'genotype', phenotypes can now be conceived in quantitative and measurable terms on a comprehensive molecular level. We discuss in what sense gene expression profiles can be regarded as phenotypic traits and whether these traits are better described as a novel concept of phenotype or as an extension of the classical concept. We argue for an extension of the classical concept and call for an examination of the type of extension involved.

Keywords: Phenotype; Genotype; Gene expression; Concept extension

1. Introduction

> 'To give a new concept' can only mean to introduce a new employment of the concept, a new practice. (Wittgenstein, 1978, p. 432)

E-mail addresses: ohadnachtomy@mac.com (O. Nachtomy), ashavit@ucdavis.edu (A. Shavit), zohar_yakhini@agilent.com (Z. Yakhini).

1369-8486/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsc.2006.12.014

Our central aim in this article is to present a current extension in the way the concept of the phenotype is being applied, namely its becoming more comprehensively quantitative and grounded in complex molecular level properties. This development is driven by the increasing power of molecular measurement technologies. We observe that practitioners in biology and medical genetics currently extend the classical notion of the phenotype to include measurable levels of gene expression. To better understand the current developments in the concept of the phenotype, we note the changes that took place in the concept of the genotype within the context of the transition from Mendelian to molecular genetics. This transition has brought about a radical transformation in our conception of genes: from functional and informational unit to structural molecular one. While heredity units were already presupposed in Mendel (1865) and an explicit distinction between soma and germ cells was made by Weismann (1889 [1885]), genotypes were first defined by Johannsen (1909) as abstract accounting or calculating units. Such units were postulated by Morgan (1917) and his colleagues to be 'physical genes lined up on the chromosomes in the fruit fly's nucleus' (Harman, 2006), while others, especially in Germany, considered them to be holistic and/or vitalistic entities located both within and outside the nucleolus (Harwood, 1993, pp. 49–52; Harrington, 1996, pp. 49–51).

Abstract or physical, mechanical or vitalistic, however divergent conceptions of genes were employed in the classical period of genetics, (that is, until the early nineteen fifties), genes were defined according to the phenotypic trait they were responsible for. In the molecular period, the genome was identified with the DNA molecule and genes came to be defined as segments on the DNA molecule. The genome thus came to be seen as a molecular entity defined by its sequence of nucleotide base pairs.[1] Yet, in practice, the definition of gene as a segment on the DNA molecule is not purely structural but also functional. A 'gene' is typically defined as the segment coding information for the production of a polypeptide chain of a functional protein (see Stotz & Griffiths, 2004). In fact, our current understanding of cellular regulatory mechanisms implies that an accurate definition of genes requires an accommodation of a plurality of DNA segments as well as disjunctive functions. For example, Berg & Singer (1992) define a gene as 'a combination of DNA segments that together constitute an expressible unit—that is, a unit whose expression leads to the formation of either a functional RNA or a polypeptide' (ibid., p. 135).

While the concept of the gene is not defined in current molecular genetics in purely structural and chemical terms (that is, as entirely independent of its function and activity), and even if this concept is quite ambiguous (Moss, 2002), it is beyond doubt that the molecular conception of the genotype has proven to be immensely fruitful. Moreover, Falk (1986) has convincingly argued that the functional/structural ambiguity has itself been very fruitful and influential. The molecular approach, with its definition of the genome as the complete sequence of DNA and genes as segments on the DNA sequence, is likewise the current definition in the minds of working scientists (see Stotz

[1] As Lewontin nicely points out 'a complete description of the DNA sequence is identical with a complete specification of the genotype... the developments of techniques of observing the phenotype have been revolutionary for genetic analysis, precisely because they solve the problem of inferring genotype from phenotype by eliminating development. All genotypes, irrespective of their influence on development, can be unambiguously discriminated at the molecular level [of the phenotype]' (Lewontin, 1992, p. 143).

& Griffiths, 2004, for an empirical study that supports this conclusion). We note that a genotype of an individual is a measurable quantitative property, as individuals differ in the actual content of specific loci in their genomes. The currently prevalent methods of genotyping and DNA sequencing enable us to map and compare the full genomes, or parts thereof, of such organisms as yeast, rice, nematodes and humans. Even if this point has been widely overstated in recent years, it is, no doubt, a remarkable feat of molecular genetics as well as a remarkable confirmation of the intuitions and insights of its founders.

At the same time, the great success in deciphering and mapping the genome has revealed the enormous complexity of the transition from the genotypic level to the phenotypic one. Since, in most cases, the causal relation between genes and phenotypic traits is a many–many relation, the transition from genes to phenotypic traits in an organism is far more complex than the mere mapping of the genome. As Lewontin (1992, p. 140) notes, 'the forward mapping of genotypic description into phenotypic description is not possible except in special cases'. In other words, the rapid advances in our understanding of the genome have made it clear that this is merely a first step in understanding the complex processes of development and of cell function and regulation. One might say that the clarity we have gained about the chemical structure of the genotype has shown how vague and general remains our notion of the phenotype—a notion that is still widely defined in classical terms. The widely accepted notion of phenotype is 'the observable traits of the organism', still remarkably similar to Johannsen's original definition: 'The phenotype of an individual is thus the sum total of all his expressed characters'(Johannsen, 1909, p. 163). Johannsen considered 'the individual's worth as an ancestor [as] essentially determined by the ''type'' to which it belongs and not by its purely personal condition' (Johannsen, 1905, p. 82 [Dunn, 1991, p. 91]).[2] For this reason, in his classical 1909 definition of the 'phenotype', he describes an appearance-type (*Erscheinungstypus*).[3]

Being rich and precise, Johannsen's original definition still commends our attention. It is clear that the notion Johannsen had in mind is thoroughly statistical, defined in relations to variations between individuals. We should also stress that this concept of the phenotype is, generally speaking, non-controversial. While so much in biology has changed, the definition of the phenotype is still current. To see this, let us consider the currently best available definition of the phenotype. In his article 'Genotype/phenotype distinction' in *The Stanford encyclopaedia of philosophy*, Lewontin continues Johannsen's line: 'The ''*phenotype*'' of an organism is the class to which that organism belongs as determined by the description of the physical and behavioural characteristics of the organism, for example its size and shape, its metabolic activities and its pattern of movement' (Lewontin, 2004,

[2] '''Phenotype''. . . namely what can be observed as typical' (Johannsen, 1905, p. 123).

[3] Johannsen clearly states that '''type''. . . is merely an appearance of superficial nature, which may mislead; only through further investigations will it be decided whether a single or several biological types are at hand. Therefore one could properly denote the statistical emerging type as appearance-type, or briefly and clearly, as ''phenotype''. Such phenotype are, as a matter of fact measurable realities: namely what can be observed as typical; i.e., with relation to variations, the centers about which the variants group. By the word phenotype merely the necessary reservation is made that from appearance alone no far reaching conclusion may be made' (Johanssen, 1909, p. 123). We deeply thank Rafi Falk for translating Johannsen's text from German, and for calling our attention to many other important references and critical remarks.

p. 1). Lewontin goes on to distinguish between the 'phenotype', which is a descriptor, and the material object it describes, the 'phenome'. He writes:

> It is essential to distinguish the descriptors of the organism, its genotype and phenotype, from the material objects that are being described. The genotype is the descriptor of the genome which is the set of physical DNA molecules inherited from the organism's parents. The phenotype is the descriptor of the phenome, the manifest physical properties of the organism, its physiology, morphology and behaviour. (Ibid.)

We shall return to Lewontin's insightful definition in the discussion section. At present, we would like to highlight the point that after a hundred years of genetics our present (2004) definition of the phenotype is very close to Johannsen's original definition of the phenotype. Lest there be any misunderstanding, let us state clearly that the current definition of the phenotype has an interesting history. We cannot unfold this history here, only to point out some of its highlights. For example, although 'behaviour' is not a physical trait, Johannsen regarded it as part of the phenotype. Thus, it was only natural for most biologists since the 1950s to suppose that the phenotype does not stop at skin surface. Similarly, although 'shape and size' were initially perceived as traits reserved for the organism, tissue or cell levels, most found it easy to think about the shape of a protein—for example the protein of the sickle cell anaemia—or the size of a mature mRNA molecule as a phenotypic trait of an organism.[4] Thus, it is clear that, although the notion of the phenotype has gone through some modifications and developments its definition remains very general and—most important for our purposes—essentially classical.

This leads to the first point we would like to emphasize. While the definition of the 'genotype' has undergone dramatic transformation from an abstract to a molecular notion, the definition of the 'phenotype' has remained within the classical framework. The second point is that there has been a wide gap in the attention of the philosophical community regarding these notions: while considerable attention has been dedicated to the concept of the genotype, much less work has been dedicated to the concept of the phenotype. A quick survey of the literature will indicate that the number of books and articles written about the 'genome', the 'genotype', 'genes' and the 'gene' is on an entirely different scale than the number of works written about the 'phenotype'. In fact, it is not easy to find works that focus directly on the concept of the phenotype.[5] Accordingly, it is not surprising that a prominent philosopher of biology, Evelyn Fox-Keller, has labelled the twentieth century 'The century of the gene' (Keller, 2000). In addition, the prevalent use of many derivative terms of the word genotype, such as 'gene', 'genome', 'genotyping', 'genomics' in the biological literature, in contrast to the very few derivative terms from 'phenotype' nicely reflects the difference in the attention devoted to the genotypic and the phenotypic levels.[6] Considering the fact that techniques in molecular biology have rendered genomes amenable to quantitative analysis, this is hardly surprising.

[4] We would like to thank Yaron Ramati for this point.

[5] For an exceptional direct reference to Phenotype, see Rollow (1995). Another exception to this point should be made to Dawkins' notion of the 'extended phenotype', which is irrelevant in this context. Even Lewontin's insightful discussion is in the context of an Encyclopedia or Key Words article.

[6] Those who doubt the importance of language to scientific practice and knowledge are referred to the Preface and the first chapter of Keller (1995) for an eloquent argument regarding the importance of the language of science, and Chapters 4–6 in Keller (2002), for a forceful illustration of this argument with regard to 'gene' metaphors.

In recent years, more and more attention is being directed in biological and medical research not only to gene sequencing but also to mapping and measuring their activity. A central technology that enables this is the mRNA measurement in cell populations—a measurement assay that has come to be called gene expression profiling. To make the significance of this approach clear, one has only to recall a basic fact underlying our understanding of an organism's development: while all the somatic cells of an organism are (nominally) genetically identical, they differ in almost every other respect. A bone cell and a pancreatic cell of the same organism share the same genetic content, that is, an identical DNA sequence. It is clear, therefore, that the differences between them are reflected in the expression (or, more precisely, the expression levels) of their genes. In other words, much of the difference in the cells' function is due to variation in the quantity of activation (or inhibition) of their genotypes by epigenetic and other regulation mechanisms. Functional and structural differences of cells are thus partly explained in terms of the levels of expression of certain genes in comparison to others.

It is for this reason that the study of gene expression is rapidly advancing. The great strides in this field are enhanced by new techniques which make it possible to measure the expression profiles of thousands of genes in specific cells or tissues. In effect, and here we come to our central point, these current methods for measuring and profiling the expression of genes allow us, if not force us, to rethink our classical definition of the phenotype. In fact, it would be more accurate to say that scientists are already using the new methods in gene expression analysis to extend the concept of the phenotype. In practice, a number of scientists have already taken the profile of expressed genes as phenotypic traits and therefore as an important part of the phenome. They do so on the basis of gene expression measurements not only at the level of the cell but also at the level of tissues, and at the level of entire organisms.

It is arguable that this approach and these new techniques of measurement are transforming our concept of the phenotype in a similar way to that the concept of the genotype was transformed since the 1950s, namely that it, too, is being 'made molecular' (Waters, 1994). If possible, such a 'moleculalrization' of the phenotype would have obvious similar advantages: unlike the classical notion, it is empirically well-defined, directly quantifiable, and comprehensively measurable. The 'degree of expression of each gene becomes a quantitative trait' (Darvasi, 2003, p. 269). Notice that the gene's degree of expression may be seen as a phenotypic measurable unit. As we shall indicate below, practitioners in this field are already presupposing such a concept of the phenotype.

At the same time, we would like to emphasize that the practitioners who lead this new employment of the phenotype are using and advancing the technology without considering its conceptual implications. In other words, as one might expect, the practitioners' approach is practical rather than critical. For this reason, we believe that some reflection is called for at this juncture. As Earnest Mayr clearly stated: 'Too often in the past the biologists have ignored the analyses of philosophers, and the philosophers have ignored the discoveries of the biologists' (Mayr, 1988, p. vii). In this article, we seek to examine the impact of the new technologies from a conceptual perspective.

In Section 2, we present the recently developed techniques of measuring gene expression profiles by means of microarrays. In Section 3, we demonstrate that, in practice, scientists already conceive of gene-expression profiles as phenotypic traits. In Section 4, we discuss whether these changes are better described as a novel concept of the phenotype or as an extension of its classical definition.

2. The technology

2.1. What is an expression profile of a cell or a system?

With only a few exceptions, every cell of the body of an individual organism contains a full set of chromosomes and the same set of genes with identical nucleotide base sequences. The exact content of the DNA that resides in cells constitutes the genome of the individual organism. Genes encode molecular instructions that are carried out through a mechanism that involves their expression, typically in the form of mature mRNA molecules (messenger RNA), and then, as the protein molecules which execute cellular functions. In different cells of the same individual organism various genes are expressed in different levels, and it is the profile of this expression activity that confers unique properties to each cell type or to cells in different conditions. Such conditions may include developmental stages, exposure to stimulants or drugs, and disease status, among others. Gene expression is the term used to describe the transcription of a sequence of nucleic acids contained within the DNA molecule into a mature mRNA molecule. The ensemble of all amounts of mRNA produced by a cell, indexed by their genes (or splice variants) of origin, is the cell's expression profile. Mathematically, this is represented as a vector or an indexed set of numbers. For example, the expression profile of a population of *S. cerevisiae* (a type of yeast) cell expression profiles would be represented as a vector of about 6000 numbers, each standing for a single gene; when comparing a human normal lung cell population to a lung tumour cell population expression profiles are represented by a vector of ~30000 numbers, each corresponding to the expression level of some single specific gene. Gene expression is a highly complex and tightly regulated process that allows a cell to respond dynamically to environmental stimuli and to its own changing needs. The expression profiles of all cells in an organism at a given time t can be thought of as the expression profile of the organism at t. While the genome of an organism is relatively invariant, the expression profiles of cells, of cell populations, and of organisms are highly variable, changing over time as a function of developmental, environmental and other conditions.

2.2. Expression profiling

The view of expression profiles as phenotypic traits is based on technologies that can support their measurement and relate them to other phenomena. Currently, the most widely used and accepted technology for measuring gene expression uses microarray based hybridization assays. Expression profiling is widely used in studying a wide variety of biological phenomena. These studies include human disease classification and pathogenesis, organism and tissue development processes and comparisons between organisms. In Bhattacharjee et al. (2001) the authors study the differential expression of lung cancer as compared to normal human lungs. Spellman et al. (1988) measured the expression levels of all S. cervisiae genes in different stages of the organism's cell cycle. In Stuart, Segal, Koller, & Kim (2003), as well as in Bergmann, Ihmels, & Barkai (2004) expression pathways and networks of different organisms are systematically compared. In this section, we shall describe the principles of microarray expression profiling.

Microarray-based hybridization assays work by exploiting an important natural property of mRNA molecules: the ability to bind specifically, in a chemical reaction called hybridization, to matched DNA sequences. A microarray is a membrane or a glass slide

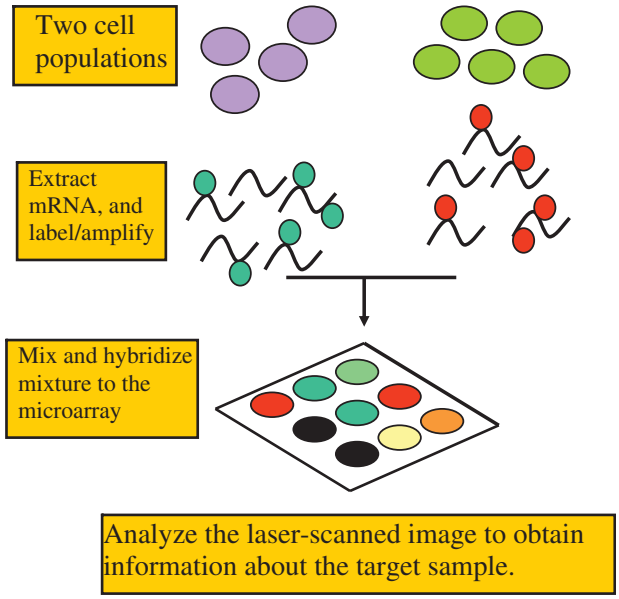

Fig. 1. Expression profiling by microarrays. The expression of genes in two cell populations (e.g. one derived from a tumor sample and one from normal cells in the same individual) are compared in this assay. Extra-nuclear mRNA is extracted from both populations, separately. The mRNA in each extract is labeled with a different fluorescence chemical dye, represented in the schematic by green and red (e.g. green for the normal cells and red for the tumor cells). The labeled samples are mixed and hybridized to a surface carrying oligonucleotide or cDNA probes in a pre-defined pattern. This surface is the microarray. Each spot on the microarray contains many copies (all nominally identical) of molecules that were designed to probe (and enable the measurement of) a specific mRNA transcript. The spots are called features. Features are depicted as circles in the bottom of the schematic description. Laser scanning determines the levels of red and green fluorescence in each feature. For features where red is dominant we infer that the gene for which this probe is designed has higher expression in the red labeled sample (e.g. tumor) than in the green sample (e.g. normal). Intermediate colors (such as yellow) represent more equal expression levels. Black spots represent no hybridization at either channel. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

on which chemical moieties can be placed in a pre-defined pattern. In each spatially defined and addressable location thereof, a microarray contains many copies of DNA molecules (called probes) that would specifically bind to an mRNA molecule of interest. When mRNA extracted from a sample is brought in contact with the microarray, the amount of mRNA bound to each site on the array indicates the expression level of that mRNA molecule. Each array has thousands of defined locations, or sites, and thus enables the measurement of thousands of mRNA variants, or the expression levels of thousands of different genes. In this way, a microarray enables the measurement of the expression profile of a sample.[7] (See Fig. 1.) Data that result from microarray-based expression profiling studies are often graphically represented as figures similar to Fig. 2 (e.g. Hedenfalk et al., 2001).

Microarrays represent a significant technology improvement because they enable the simultaneous measurement of the expression levels of thousands and tens of thousands of genes (compared to single genes or dozens which could be handled by earlier technologies such as qRT-PCR) and because they facilitate the study of small samples. The latter is an important feature, enabling the study of clinical samples, which are often limited in size. Microarrays may be used to profile single samples or to compare the profiles of several

[7] More details of microarray-based expression profiling are described in the National Center for Biotechnology Information (2004) website.

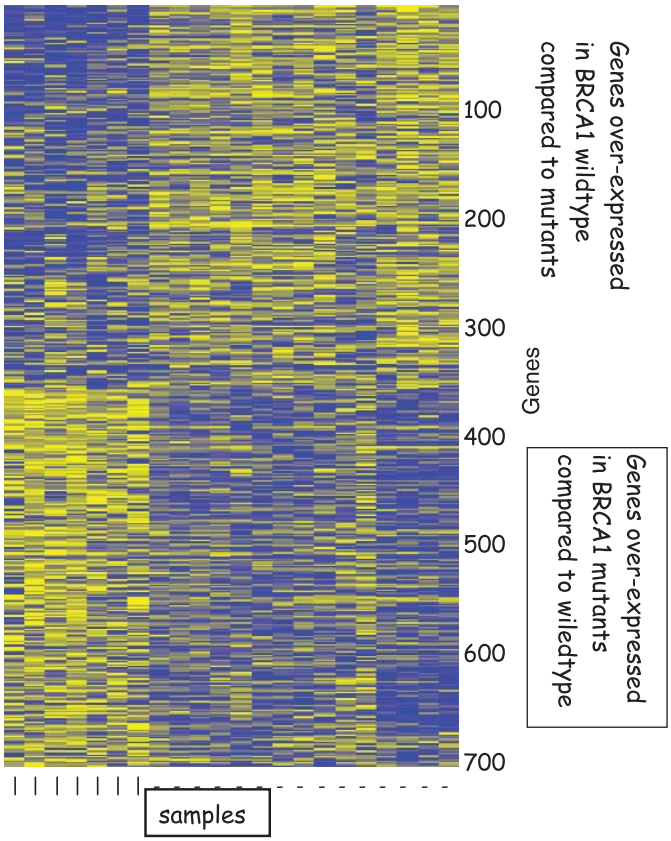

Fig. 2. Differential expression of BRCA1 mutant breast cancer tumor samples and BRCA1 wildtype breast cancer tumor samples. The results of a microarray based expression profiling study are represented as a matrix as follows: the rows represent genes; the columns represent samples (in this case each tumor sample is taken from an individual patient); yellow intensity represents high expression levels and blue intensity represents low expression levels. In this figure the BRCA1 mutant samples are in the 7 left hand columns and the BRCA1 wildtype samples are on the right. Genes are partitioned according to wildtype vs mutant differential expression. Note that the sample indicated by the arrow at the bottom of the matrix has an expression profile of a mutant even though it is BRCA1 wildtype. For this patient an aberrant methylation pattern of the BRCA1 promotor region was measured – representing a different silencing mechanism that results in an expression profile that is much like the one possessed by cells with defective BRCA1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

different cell types or tissue samples, such as in healthy and diseased tissue. In so doing, this technology identifies small subpopulations that homogenously respond to a drug, and advances the vision of an individualized approach to medicine. Microarray technology is rapidly developing. New applications of microarrays, such as profiling the structure of cancer genomes are under development (e.g. Barrett et al., 2004; Bignell et al., 2004; Harbison et al., 2004), and promise more insight into cellular processes, gained by exploring thousands of parameters and mining the results to find the significant ones. The high throughput nature of microarray-based expression profiling allows practitioners to treat expression profile signatures as well as the expression levels of single genes as molecular representations of phenotypes—a central aspect of this article that we shall now demonstrate.

3. Examples

In this section, we demonstrate the practice of several current research groups to measure expression profiles and treat them as measurements of phenotypes.

3.1. Cheung & Spielman (2002): 'The genetics of variation in gene expression'

In a review article in Nature Genetics of 'recent studies that have used micoroarrays to obtain data on gene expression phenotype, genotypes, or both', Cheung & Spielman (2002) write:

> The genetic basis of variation in gene expression lends itself to investigation by microarrays. For genetic analysis, we view the expression level of a gene as a quantitative or 'complex' trait, analogous to an individual's height or cholesterol level, and, therefore, as an inherited phenotype. Several genetic analyses of 'gene expression phenotypes' have been carried out in experimental organisms, and initial steps have been taken toward similar studies in humans.... (Ibid., p. 502)

Cheung & Spielman proceed to note that, in the studies under review, (references 1–8 in their article) 'the expression level of a highly variable gene in an individual is considered as the ''phenotype''...'. Let us take note of the fact that the authors indicate with inverted commas that their use of the word phenotype in this context is not in its ordinary sense, certainly not the classical one. Interestingly, in the sequel they abandon the use of inverted commas and are more explicit about identifying gene expression as a specific phenotype. They remain careful, though, to explicitly state this new conceptual linkage between expression levels and phenotypes. As they write: 'When the expression levels of genes are defined as phenotypes genetic analysis can be done to map, identify and characterize the genetic determinants that are responsible' (ibid., p. 523). By the end of the article they reiterate the new linkage between expression profiles and the concept of the phenotype and relate it to microarrays: 'The development of microarrays has made it possible to expand the phenotype to include another form of variation: genome-wide gene expression levels' (ibid., p. 524). 'Ultimately, it will be necessary to extend the studies to include variation in proteins. At present, however, microarrays have made it possible to analyze phenotype at the transcript level' (ibid.). These citations speak for themselves and require little comment. One can almost sense how the notion of the phenotype is being revised or, more precisely, expanded as they write, that is, as they shift slightly from a classical notion of phenotype to one defined in terms of gene expression, and as they move from reserved caution to hopeful forecasts. We take this to be good evidence of the current change in the articulation of the notion of the phenotype. At the same time, Cheung & Spielman also very nicely mark the current state of the art: we can now measure variations in RNA transcripts but not in proteins.

3.2. Darvasi: 'Gene expression meets genetics'

In a review of the comprehensive and innovative study by Schadt et al. (2003) entitled 'Gene expression meets genetics' and published in *Nature*, Darvasi (2003) writes:

> The expression level of each gene is ... treated as a quantitative trait. Quantitative traits are determined by more than one gene and show a graded variation across a population, such that the variation can only be measured quantitatively. Height and weight are typical examples, but it is perfectly reasonable to consider gene-expression levels as a quantitative trait, too. (Ibid., pp. 269–270)

An additional degree of sophistication can be introduced to such analyses by including a particular trait—a disease, for instance. Gene-expression data might help to define such a trait more accurately, generating genetically more homogeneous groups of individuals that have that characteristic. Schadt et al. (2003) looked at obesity, specifically the mass of fat pads, in mice. They first divided mice into groups with high versus low fat-pad mass, and then analyzed gene expression data. They subsequently subdivided the high-mass class into two subgroups. Genetic analysis of these subgroups then allowed the authors to identify QTLs (chromosomal regions that are significantly correlated with a quantitative trait, fat-pad mass in this case) that affect one subgroup but not the other (ibid., p. 270).

In this example, the concept of gene expression profile is already put to work, as a more accurate way to categorize and treat the phenotype of obesity. In other words, Schadt et al. assume that the trait of obesity, traditionally located by visual means or by measurements of fat mass, is better located by pinpointing the specific region on the chromosome that is transcribed into mRNA. What seemed a single disease—obesity—turns out to be two different diseases, which should perhaps be medicated differently.

3.3. Morley et al. and Cox, 'An expression of interest'

Commenting on a *Nature* review article by Morley et al. (2004), Nancy J. Cox (2004) distinguishes quite explicitly between the classical phenotypic traits such as height and eye colour and the baseline level of gene expression. She writes:

> In searching for the genetic basis of a measurable trait, or phenotype—height or eye colour, for example—geneticists start by identifying people with variation in that trait. In the simplest situations, the variation is traced to a single gene, and particular variations in the sequence of that gene are shown to determine the observed phenotypic variation. A less tangible, but no less significant, trait is the baseline level of gene expression. (Ibid., p. 733)

Cox goes on to explain the rationale, as well as the employment of treating the baseline level of gene expression, as a phenotypic trait:

> Much of our progress so far in understanding the basis of rare genetic diseases has come through identifying changes in gene sequence that change the nature of the encoded protein, so that it is insufficiently or inappropriately functional. But changes in the amount of protein produced might also affect the health of an organism. This amount is determined in part by gene expression—or how much messenger RNA (mRNA) is transcribed from the relevant DNA sequence—and the relative abundances of mRNA for many thousands of genes can now be routinely assessed for any accessible tissue. (Ibid., p. 733)

Cox here explains the approach taken by Morley et al., stated explicitly in their Discussion section:

> Our study combined microarray expression data with publicly available SNP genotype data, and applied genome-wide mapping techniques to identify the chromosomal regions linked to the gene expression phenotypes. Level of gene expression

> is thus a trait like many others, and is amenable to genetic analysis. (Morley et al., 2004, p. 746)

The most significant point for our present purposes is that (once again, but more explicitly) gene expression level is treated as a phenotypic trait, 'like many others'. This point is underlined by Morley et al.'s approach in seeking the causes of these phenotypic traits at the DNA level. As Cox explains:

> [G]ene expression is generally regulated by DNA regions outside the parts of genes that actually encode proteins, and there is much that is not yet understood about this process. In particular, little is known about how variation in DNA sequences might affect the variability in baseline levels of gene expression among individuals—the topic of Morley and colleagues' investigations. (Cox, 2004, p. 733)

Morley and colleagues seek the genetic determinants of these phenotypic traits by means of genetic analysis. In thus relating the novel molecular notion of the phenotype (levels of mRNA) with the molecular notion of the genotype (DNA sequences), their approach strongly supports this emerging notion of phenotype.

Cox highlights this point:

> The investigators characterize their results as indicating that the human gene-expression phenotype is 'a trait like many others... amenable to genetic analysis'. To veterans of linkage-mapping studies on complex human phenotypes, this may seem to be an understatement on a par with Watson and Crick's 'it has not escaped our notice...'. Indeed, these gene-expression phenotypes seem to be particularly amenable to genetic analysis, which bodes well for the future studies that will build on these results. (Ibid., p. 734)

This comment indicates the great importance which leading scientists ascribe to this approach and the methods that permeate it—all of which strongly attest to the fact that the emerging notion of the phenotype is already well at work as well as firmly entrenched in the minds and the practices of working scientists.

3.4. Tsalenko et al.: 'Analysis of SNP-expression association matrices'

To substantiate Cox's point a bit further, let us draw on a yet unpublished article. In the introduction to the article, entitled 'Analysis of SNP-expression association matrices', the authors summarize the emerging approach to the phenotype (as well as their own), as follows:

> The development of high throughput techniques for expression profiling and genotyping enables the study of the genetic determinants of expression variation, both in humans and in other organisms. Expression levels are taken as quantitative phenotypes, of independent interest, as well as determinants or indications of end-point clinical phenotypes. Much of our understanding of the genetic base of disease comes from identifying polymorphisms that affect protein structure or integrity. We know, however, that protein abundance and expression levels of mRNA are also responsible for disease processes. It is therefore important to explore the genetic base of variation in gene expression, regarded as a quantitative phenotype. (Tsalenko et al., 2005, p. 1)

In this section, we have tried to illustrate the current expansion of the concept of the phenotype. This process is still under way, and by following these noteworthy examples, one may see how the concept of gene-expression-as-a-phenotype is gradually being formed. It is interesting to observe that a conceptual shift that has been explicitly noticed and discussed within the community of genomics in 2002 is only mentioned *en passant* in 2005. We predict that, in the near future, this transformation will not even require mentioning and will undergo what Latour (1987, pp. 3–4) has called 'black boxing'.

4. Discussion

The main point we would like to address in concluding this article is whether the changes we have described and exemplified in Section 3 are better seen as a novel concept of phenotype or as an extension of the classical concept of the phenotype. As we noted in the first section, the widely accepted definition of the phenotype is still the classical 'sum total of [an individual's] expressed characters' (Johannsen, 1909, p. 163). Lewontin's recent work on the genotype/phenotype distinction is an important exception to the lack of attention the notion of the phenotype has received from philosophers of biology. It is clear that Lewontin's concept of phenotype is much more sophisticated—both biologically and philosophically—than Johannsen's. Yet, it is remarkable that the most sophisticated definition of the phenotype currently available is still comparable to Johannsen's original definition.[8] In contrast to Johannsen's emphasis on the type rather than the individual, Lewontin notes the impracticality of the type–token distinction in this case. Lewontin argues that an organism's development involves many genetic mutations and micro environments, so that its genotype and phenotype turn out to be classes with only a single member. Most important to our purposes, Lewontin argues that 'In practice genotypic and phenotypic descriptions are not total but partial, restricted to some subset of the characteristics of the organism that is regarded as relevant for a particular explanatory or experimental purpose' (ibid., p. 4). We shall return to this insightful observation.

While the phenotypic traits we have considered, such as mRNA profiles of cell populations of an organism in various conditions, were clearly not among the observable traits of the organism in the classical sense of 'observable' (that is, seen by a naked eye or under an optic microscope, in contrast to the non-observable and unexpressed genotype), they have certainly become observable and measurable by virtue of the new methods we have outlined in Section 2. In addition, since these new traits require complex manipulations in order to be 'seen', they may be at variance with the classical sense of phenotypic traits. The definition of the phenotype as 'all the observable features of an organism' is, however, broad and inclusive, so that it can be taken to also include mRNA levels.

As previously mentioned, while the definition of the phenotype remained more or less the same, the traits included under the concept of phenotype have certainly changed. It is clear that the tendency to moleculalrize the phenotype did not begin with the development of high throughput expression profiling technologies. Since its early days, molecular biology emphasized the importance of the structure of gene products in bringing about macroscopic phenotypic changes (e.g. CF and sickle cell anaemia), so that not only molecular structure was

[8] This point is made evident by the bibliography to Lewontin's article.

important but also the degree of molecular differences.[9] As an example, consider Habby and Lewontin's 1966 measurements of electrophoresis gel migration of proteins from different natural populations of *Drosophila pseudoobscura*.[10] In this sense, the molecular structure and quantity of proteins has been considered a part of the phenotype for a long time. In addition, the current shift to the molecular level of the concept of the phenotype, as we discuss in this article, is part of a wider development within medical genetics that redefines and reclassifies disease conditions by quantifying biochemical properties that are likely to represent the many genes involved in the disease instead of its visible clinical-syndrome of the organism as a whole (Belmaker 2004; Gottesman & Gould, 2003).[11]

Given this context, let us consider the significance of concept shift represented by mRNA levels being seen as phenotypic traits. Morley et al. (2004, p. 746) write that, 'Level of gene expression is... a trait like many others', thus stressing its similarity rather than difference from classical traits. So, what is new about such traits, beyond the excitement in the rhetoric we have recorded? We just noted that a molecular approach to the phenotype is not new. Nor is a quantitative and statistical approach. A quantitative approach already appears in Johannsen, and during the past two decades biologists have been measuring the presence and levels of various biochemical moieties, including mRNA.

Despite the above, the current changes in the way we think about of the phenotype are highly significant, rather than merely incremental, in two respects: (1) The current technology allows for a *comprehensive* quantitative and molecular snapshot of the totality of mRNA activity in a certain cell-population under a certain condition. The comprehensiveness here means that we measure not several mRNA transcripts but tens of thousands of them. (2) This comprehensiveness together with the flexibility and relative simplicity of expression profiling makes it possible to construe phenotypes as classes of different developmental stages of a single organism and (alternatively) classes of different cells or tissues of a single organism.[12] Such a considerable change in scale may bear conceptual implications. It constitutes a significant advance in the attempt to observe classes of different cells or tissues of a single organism, track their changes during the organism's development and compare them to other organisms in its population or meta-population. Comprehensive snapshots of gene activity are thus new and it is certainly new that they are treated as phenotypic traits. Simply put, the phenome is now taken to include traits that could not be observed and measured previously.

To illustrate point 2 above, let us note some features of gene expression profiling. While none of these features is entirely new, taken together they indicate the change in the classical notion of the phenotype as driven by current tendency for more refined classification in medical genetics and the emergence of profiling technologies. (1) Expression profiling can potentially facilitate protocols to define the existence, development or retreat of a disease, and by so doing might change the working context of diagnosing and treating disease. (2) It permits the quantitative study of developmental processes. The same is true for technologies that will enable a comprehensive understanding of the next stage in the causal pathway of protein production. (3) Comparing organisms on the basis of molecular

[9] We would like to thank Yaron Ramati for this point.
[10] We would like to thank Rafi Falk for this example.
[11] We thank an anonymous referee for this point.
[12] We thank an anonymous referee for this formulation.

and cellular networks is enabled by the existence of comprehensive functional data (Stuart, Segal, Koller, & Kim 2003; Bergmann, Ihmels, & Barkai, 2004). (4) What is denoted by 'phenotype'—a vector of numbers representing quantities of mRNA molecules—can only be an open ended trait with very fine grained degrees of activity which does not apply to discrete traits such as having a certain blood type.

Yet we should emphasize that, unlike in the case of the genome, a full description of the phenome is practically unattainable. As Lewontin (2004) noted, it is clear that an exhaustive description of the inherently dynamic phenotype is not feasible. We take this to illustrate an important difference between the genotype and the phenotype that remains intact. While the phenotype is intrinsically variable, the genotype is relatively stable. We should not lose sight of the fact that, while genetic variation can be now described by means of different combinations of a few simple molecular elements (base pairs), the phenotypic level, on the other hand, encompasses enormous multiplicity, complexity and diversity of phenomena, which operate at very different levels of organization. A molecular notion of phenotype, defined in terms of gene expression profiles, however useful in some contexts, may be too narrowly defined in many others. For example, when we describe the migration patterns of certain birds, it is not clear that a gene expression profile is the most useful mode of description. In such contexts, the gap between the molecular level and the macro, phenomenal level may be too wide to attempt a description in terms of gene expression profiles alone. Such an attempt to reduce the notion of the phenotype oversimplifies the biological complexity. We would like to note that the attempt to capture all the aspects of phenome (and thus all phenotypic traits) in terms of mRNA transcripts is misguided. In addition to its indisputable dependence on the varying conditions of the physical environment and on the organism's stage of development (time index), the notion of a phenotype is intrinsically context sensitive in other respects as well. One aspect of this context-sensitivity derives from the practical need to focus on a phenotypic trait among infinitely many traits. The focus on a phenotypic trait is normally confined to a particular interest. We may be looking at a disease, such as cancer, or we may be looking at obesity or at the exceptional beauty of an orchid or at the symmetrical pattern of a butterfly or at some complex relation between them. In this connection, it is useful to recall Lewontin's recent point that, 'In practice, genotypic and phenotypic descriptions are not total but partial, restricted to some subset of the characteristics of the organism that is regarded as relevant for a particular explanatory or experimental purpose' (Lewontin, 2004, p. 4). In practice, we are unable to investigate a whole phenotype. We must focus on a certain aspect within a certain context. Thus we fully agree with Lewontin's point that, 'The problem of what parts of the genome and phenome are to be included in the partial genotypic and phenotypic descriptions of the organism in particular cases is one of the most problematic in biology. While it is undoubtedly true that every part of the genome is connected causally with the phenome by some pathway, it is simply impossible to consider all pathways of connection' (ibid., pp. 4–5). This implies that what we call phenotype depends on what we are looking for—it may be cancer, which applies at the level of certain cells, tissues or organs; it may be adventurism, which applies at the level of individual behaviour in comparison to others; or it may be sexual tendency which applies at the level of several organisms. And this points to yet another aspect in which phenotypes are intrinsically context sensitive. The common (and classical) presumption is that the phenotype is a phenotype of a whole organism. However, this largely simplifies the rich complexity and diversity of biological phenomena. For example, when trying to map a genotype to a

phenotype of a cell it is crucial to study the buffering (canalization) of the relevant trait across its different micro cellular environments, while the macro environment of the organism as a whole is not always relevant. In fact, the new technique to determine the number, location and magnitude of QTLs is already put to use for tracking the molecular mechanisms of canalization directly (Flatt, 2005). Since biological entities have a complex structure, and are certainly not limited to multi-cellular organisms, and since organisms themselves entail different levels of organization, a phenotype may vary according to the particular trait under consideration. In other words, the question of what 'phenotype' refers to (cell, tissue, organism, species, developmental system etc.) remains entirely underdetermined without a given context.

This point becomes acutely clear in light of the examples of phenotypic traits we have considered in Section 3. Some of these are phenotypes of cells or even of the activity of some of their genes, rather than phenotypes of the whole organism. In this sense, it certainly seems that the new technology is opening the way for thinking about genetic phenomena differently. Classically, phenotypes have been construed as classes of organisms. As we have noted above, new array technologies makes it possible to construe phenotypes as classes of different developmental stages of a single organism as well as classes of different cells or tissues of a single organism. The implications of this recent development merit attention and careful reflection from the philosophical community.[13]

Having made this point, let us observe that there are also rare cases where a specific phenotypic context is well defined and in a one-to-one correspondence with a certain genotype. In such a case, we may benefit from treating 'gene expression profile' as the most relevant phenotype. An example of such a case is the hereditary breast cancer involving a dysfunctional BRCA1 gene. A patient may show the symptoms of a dysfunctional BRCA1 while possessing a non-mutant genotype (see Fig. 2 above). In the case studied by Hedenfalk et al. (2001), the expression of BRCA1 is inhibited due to aberrant methylation of the gene's promoter. In effect, this situation implies that this disease, while hereditary, is best understood (diagnosed and/or treated) at the level of the mRNA rather than at the level of DNA. Therefore, in this context, the most pertinent phenotype is the amount of mRNA molecules.

To conclude, we think that the contextual and partial character of the phenotype highlighted in this discussion supports a pluralistic, rather than a reductive notion of a phenotype. A pluralistic notion of the phenotype accommodates different traits of various biological entities and operates at different levels of complexity which the new technology enables us to articulate, observe, and measure. This conclusion confirms Lewontin's account (2004) which argues that applying the operative concept of partial phenotype requires making a decision about 'what set of phenotypes and genotypes are to be regarded as indistinguishable and so are to be included in the definitions of the partial genotypic and phenotypic classes'. This implies that, as methodologies change, the partitioning of objects into phenotypes will change as well. In this respect, expression profiles are best seen as new phenotypic traits that extend the classical concept of the phenotype. This is in line with Wittgenstein's insightful remark that '"To give a new concept" can only mean to introduce a new employment of the concept, a new practice'.

[13] We owe this formulation to a referee of this journal.

Acknowledgements

We would like to thank Rafael Falk, Emily Grosholz, Oren Harman, Yaron Ramati, and two referees of this journal for very helpful comments and critical reading of early versions of this article. Finally, we would like to thank the Israeli Science Foundation, grant number 354/06 and The Mobility 6 program, Marie Curie (OIF), no. 040436, for supporting this research.

References

Barrett, M. T., Scheffer A., Ben-Dor, A., Sampas, N., Lipson, D., Kincaid, R., Tsang, P., Curry, B., Baird, K., Meltzer, P. S., Yakhini, Z., Bruhn, L., & Laderman S. (2004). Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proceedings of the National Academy of Science*, *101*, 17765–17770.

Belmaker, R. H. (2004). Medical progress: Bipolar disorder. *The New England Journal of Medicine*, *351*, 476–486.

Berg, P., & Singer, M. (1992). *Dealing with genes*. Mill Valley, CA: University of Science Books.

Bergmann, S., Ihmels, J., & Barkai, N. (2004). Similarities and differences in genome-wide expression data of six organisms. *PLoS Biology*, *2*, 85–92.

Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., Mark, E. J., Lander, E. S., Wong, W., Johnson, B. E., Golub, T. R., Sugarbaker, D. J., & Meyerson, M. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proceedings in the National Academy of Science USA*, *98*(24), 13790–13795.

Bignell, G. R., Huang, J., Greshock, J., Watt, S., Butler, A., West, S., Grigorova, M., Jones, K. W., Wei, W., Stratton M. R., Futreal, A. P., Weber, B., Shapero, M. H., & Wooster, R. (2004). High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Research*, *14*, 287–295.

Cheung, V. G., & Spielman, R. C. (2002). The genetics of variation in gene expression. *Nature Genetics*, *32*, 522–526.

Cox, N. J. (2004). Human genetics: An expression of interest. *Nature*, *430*, 733.

Darvasi, A. (2003). Genomics: Gene expression meets genetics. *Nature*, *422*, 269–290.

Dunn, L. C. (1991). *A short history of genetics*. Ames, IA: Iowa State University Press.

Falk, R. (1986). What is a gene? *Studies in History and Philosophy of Science*, *17*, 133–173.

Flatt, T. (2005). The evolutionary genetics of canalization. *The Quarterly Review of Biology*, *80*, 287–316.

Gottesman, I. I., & Gould, T. D. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *The American Journal of Psychiatry*, *160*, 636–645.

Habby, J. L., & Lewontin, R. C. (1966). A molecular approach to the hetrozygosity in natural populations. *Genetics*, *54*, 577–590.

Harbison, C. T., Gordon, B., Lee, T. I., Rinaldi, N. I., Macisaac, K. D., Danford, T. W., Hannett, N. M., Tagne J. B., Reynolds, D. B., Yoo, J., Jennings, E. G., Zeitlinger, J., Pokholok, D. K., Kellis, M., Rolfe, A. P., Takusagawa, K. T., Lander, E. S., Gifford, D. K., Fraenkel, E., & Young R. A. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature*, *431*, 99–104.

Harman, O. S. (2006). History of classical genetics. In *Encyclopedia of life sciences*. http://www.els.net. Chichester: John Wiley & Sons, Ltd.

Harrington, A. (1996). *Re-enchanted science: Holism in German culture from Wilhelm II to Hitler*. Princeton: Princeton University Press.

Harwood, J. (1993). *Styles of scientific thought*. Chicago: The University of Chicago Press.

Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer P., Gusterson, B., Esteller, M., Raffeld, M., Yakhini, Z., Ben-Dor A., Dougherty, E., Kononen, J., Bubendorf, L., Fehrle, W., Pittaluga, S., Gruvberger, S., Loman, N., Johannsson, O., Olsson, H., Wilfond, B., Sauter, G., Kallioniemi, O. P., Borg, A., & Trent, J. (2001). Gene expression profiles in hereditary breast cancer. *The New England Journal of Medicine*, *344*, 539–548.

Johannsen, W. (1905). *Arvelighedslaerens elementer*. København: Gyldendalske Boghandel, Nordiske Forlag.

Johannsen, W. (1909). *Elemente der exakten Erblichkeitslehre*. Jena: Gustav Fischer.

Keller, E. F. (1995). *Refiguring life: Metaphors of twentieth century biology*. New York: Columbia University Press.

Keller, E. F. (2000). *The century of the gene*. Cambridge, MA: Harvard University Press.

Keller, E. F. (2002). *Making sense of life*. Cambridge, MA: Harvard University Press.

Latour, B. (1987). *Science in action*. Cambridge, MA: Harvard University Press.

Lewontin, R. C. (1992). Genotype and phenotype. In E. F. Keller, & E. A. Lloyd (Eds.), *Keywords in evolutionary biology* (pp. 137–144). Cambridge, MA: Harvard University Press.

Lewontin, R. C. (2004). The genotype/phenotype distinction. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. http://plato.stanford.edu. Stanford CA: Metaphysics Research Lab.

Mayr, E. (1988). *Towards a new philosophy of biology*. Cambridge, MA: Harvard University Press.

Morgan, T. H. (1917). The theory of the gene. *The American Naturalist*, *51*, 513–544.

Morley, M., Molony, C. M., Weber, T. M., Devlin, J. L., Ewens, K. G., Spielman, R. S., & Cheung, V. G. (2004). Genetic analysis of genome-wide variation in human gene expression. *Nature*, *430*, 743–747.

Moss, L. (2002). *What genes can't do*. Cambridge, MA: MIT Press.

National Center for Biotechnology Information. (2004). Microarrays: Chipping away at the mysteries of science and medicine. In *A science primer*. http://www.ncbi.nlm.nih.gov/About/primer/microarrays.html.

Rollow, C. D. (1995). *Phenotypes, their epigenetics, ecology and evolution*. London: Chapman and Hall.

Schadt, E. E., Monks, S. A., Drake, T. A., Lusis, A. J., Che, N., Colinayo, V., Ruff, T. G., Milligan, S. B., Lamb, J. R., Cavet, G., Linsley, P. S., Mao, M., Stoughton, R. B., & Friend, S. H. (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature*, *422*, 297–302.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D., & Futcher, B. (1988). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell*, *9*, 3273–3297.

Stotz, K., & Griffiths, P. (2004). How scientists conceptualise genes: An empirical study. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *35*, 647–673.

Stuart, J. M., Segal, E., Koller, D., & Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science*, *302*, 249–255.

Tsalenko, A., Sharan R., Kristensen V., Børessen-Dale, A.-L., Ben-Dor, A., & Yakhini, Z. (2005). Analysis of SNP-expression association matrices. In *Proceedings of the 2005 IEEE Computational Systems Bioinformatics Conference* (pp. 135–143). Washington, DC: IEEE Computer Society.

Waters, K. C. (1994). Genes made molecular. *Philosophy of Science*, *61*, 163–185.

Weismann, A. (1889). The continuity of the germ-plasm as the foundation of a theory of heredity (E. Poulton, S. Schönland, & A. E. Shipley, Eds. & Trans.). In idem, *Essays upon heredity and kindred biological problems* (pp. 161–249). Oxford: Clarendon Press.

Wittgenstein, L. (1978). *Remarks on the foundations of mathematics* (3rd ed.) (G. H. von Wright, R. Rhees, & G. E. M. Anscombe, Eds.; G. E. M. Anscombe, Trans.). Oxford: Basil Blackwell.