



A new conceptual framework for investigating complex genetic disease

Shobbir Hussain*

Department of Biology and Biochemistry, University of Bath, Bath, UK

Some common diseases are known to have an inherited component, however, their population- and familial-incidence patterns do not conform to any known monogenic Mendelian pattern of inheritance and instead they are currently much better explained if an underlying polygenic architecture is posited. Studies that have attempted to identify the causative genetic factors have been designed on this polygenic framework, but so far the yield has been largely unsatisfactory. Based on accumulating recent observations concerning the roles of somatic mosaicism in disease, in this article a second framework which posits a single gene-two hit model which can be modulated by a mutator/anti-mutator genetic background is suggested. I discuss whether such a model can be considered a viable alternative based on current knowledge, its advantages over the current polygenic framework, and describe practical routes via which the new framework can be investigated.

OPEN ACCESS

Edited by:

Noor Ahmad Shaik,
King Abdulaziz University,
Saudi Arabia

Reviewed by:

Maria Paola Lombardi,
University of Amsterdam, Netherlands
Suresh Kumar Chitta,
Sri Krishnadevaraya University, India

*Correspondence:

Shobbir Hussain
s.hussain@bath.ac.uk

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Genetics

Received: 02 July 2015

Accepted: 21 October 2015

Published: 04 November 2015

Citation:

Hussain S (2015) A new conceptual framework for investigating complex genetic disease. *Front. Genet.* 6:327. doi: 10.3389/fgene.2015.00327

Keywords: somatic mosaicism, mutator alleles, anti-mutator alleles, nanopore sequencing, rare variants, RVAS, genome wide association studies, rare variant association studies

INTRODUCTION

Some diseases within the human population are described as being ‘complex’ or ‘multifactorial’ in that they have an underlying inherited genetic etiology which contributes to disease risk along with environmental factors. These occur relatively common within the population and examples include congenital disorders such as cleft lip and palate, pyloric stenosis, clubfoot, and neural tube defects; those that can manifest at some point in childhood/early adulthood such as autism and schizophrenia; and those that can appear later in life such as Alzheimer’s disease. The proportion of genetic and environmental contributions to these types of diseases varies, but the proportion of variance of each of these particular diseases observed within a population which is due to variance in genetic factors is sometimes referred to as the ‘heritability’ and can be calculated via multiple methods. The simplest method, to obtain an estimate, is by observing the difference in concordance for the disease between monozygotic and dizygotic twins. The aforementioned diseases as well as several others are estimated to have a high heritability. The current consensus is that this genetic contribution is polygenic, such that multiple genetic loci interact in some fashion in order for the disease to manifest and such a model was put forth primarily based on the observed patterns of population- and familial-incidences of such diseases. In the last decade or so, various population-based large scale efforts to unmask the precise genetic factors responsible for disease manifestation have been based on this model, and have included investigations probing genetic variations that occur commonly, and more recently those that occur more rarely within the population. Overall, the output from such investigations have been disappointing in terms of providing little in the

way of identifying and characterizing causal variants that actually play a significant role in disease manifestation. Here, I propose and describe an alternative theoretical model and consider whether the new model would be consistent with the population and manifestation characteristics of the diseases as well as current knowledge in genetics. I also discuss the merits and potential flaws of such a framework in comparison to current models, and define routes by which the framework could be tested and utilized.

POLYGENIC THEORY, DISEASE-THRESHOLD, AND THE CARTER EFFECT

Fisher (1918), arguably one of the most important thinkers in the history of the biological sciences, published a work which he entitled 'The correlation between relatives on the supposition of Mendelian inheritance'. It was a landmark moment in our understanding of the genetics of inheritance. With the use of several mathematical applications, which many geneticists would struggle to grasp the inner-workings of, he settled an ongoing and at times heated debate between two major schools of thought regarding fundamental aspects of inheritance. On one side were the 'Mendelians' who championed the ideas of Mendel that single entities, now known as genes, were sufficient to determine phenotypic traits; and on the other side the statistically inclined 'biometricians' who argued that the majority of phenotypic traits displayed a continuous variation within the population and therefore at odds with this Mendelian idea of inheritance which rather predicted a dichotomous appearance of characters. Fisher's (1918) mathematical intuition and training enabled him to introduce novel statistical models to demonstrate that 'cumulative Mendelian factors seem to fit the facts very accurately' as he put it; meaning that several genetic factors contribute in an additive manner to determine the value of an observed continuous trait. This came to be known as 'polygenic theory' and the work further laid the foundations of 'quantitative genetics' which concerns the study of genotype-phenotype correlations within populations.

The statistical models, Fisher (1918) introduced in his seminal paper and in further works (Fisher, 1922) are still widely used today in various fields of science. Though Fisher (1918) developed polygenic theory with the particular intention to explain continuous traits within populations, shortly following his death, his theory was extended in an attempt to explain the presence of dichotomous characters (Falconer, 1965), i.e., those characters previously acknowledged to be consistent with a Mendelian single-gene model. There was good motivation behind making such an extension: it was observed that many common diseases of a dichotomous nature that displayed some degree of familial clustering, displayed a rather low recurrence incidence amongst relatives and thus did not conform to any known pattern of Mendelian inheritance. Already inherent to Fisher's (1918) polygenic model was that an individual is much less likely to develop the disease if he/she needs to inherit several alleles in order for the disease to manifest compared to

if he/she needed to inherit just one or two (as in Mendelian inheritance), thus potentially explaining the reduced familial recurrence risks. However, to reconcile the observations of the dichotomous nature of diseases with Fisher's (1918) model, the idea of 'threshold' was included where it is postulated that many genes do in fact contribute to disease manifestation in an additive fashion, but that only when a particular threshold biological value is breached, does the disease manifest. Thus the 'polygenic threshold model' as it became known neatly ticked all the required boxes, and in one form or another, has remained a mainstay as an explanation for inherited diseases which do not conform to any classic Mendelian pattern whilst displaying a dichotomous manifestation.

Indeed, this very concept of disease-threshold was incorporated into an explanation for another more specific observation regarding the population characteristics of some of these diseases: that a few displayed sex dimorphism where one particular sex is more commonly affected than the other (Carter and Evans, 1969). For example, in some disorders males are more commonly affected, and it is proposed that females must possess a higher threshold for disease expression: therefore a higher number of disease susceptibility alleles need to be inherited for the disorder to manifest in females. As a consequence, this would also mean that when females are affected, it would suggest a higher 'genetic load' of disease susceptibility alleles within the family and thus a higher recurrence risk for her relatives. Thus in a sex dimorphic polygenic disease, when the usually less commonly affected sex presents with the disease within a family, there is predicted to be a higher recurrence risk for relatives. This phenomenon became known as the 'Carter effect'.

It may be of some note at this point that Fisher (1918) appeared quite happy to let the Mendelians 'have' the dichotomous traits, as is quite apparent from the fact that he made little effort to explain them using his polygenic theory, though he had plenty of scope, and no doubt, the intellectual vision to do so. Further, it is imperative to point out here that though the later proposed disease-threshold model can sometimes be useful in helping to estimate the recurrence risk of a disease once the population- and familial-incidences are known, as well as in making heritability predictions, there is currently no direct evidence to support the model. Indeed, it should at least be of a slight concern that the Carter effect is widely cited as evidence for the threshold model, as conversely there is no evidence to support the basis of the Carter effect, nor even a proposed biological mechanism via which one particular sex would have a higher/lower threshold for the expression of diseases.

FROM COMMON VARIANTS TO RARE VARIANTS

Rarer inherited diseases tend to display a monogenic Mendelian pattern of inheritance within families, and toward the end of the last century, linkage and mapping studies had been very successful in identifying the genetic causes of many of these. However, similar types of studies proved insufficient

for identifying the causes of more common inherited diseases suspected to have an underlying polygenic determination. Indeed this difficulty was attributed to the existence of several causative genes, each only contributing a small-effect on disease manifestation, making them difficult to pinpoint using standard genetics approaches. Human genome sequencing projects earlier on this century, however, revealed that some genetic variations occurred fairly commonly within the population, usually as single nucleotide polymorphisms (SNPs). It was hypothesized that these common genetic variants were likely to be responsible for polygenic disease manifestation, i.e., phenotypic variance would be dependent on the contribution of several genetic variants each of which would make only a small contribution toward the manifestation of the disease; and thus these genetic variants needed to occur fairly commonly within the population to account for the prevalence of the disorders. This became known as the ‘common disease-common variant’ hypothesis, and it seemed somewhat fitting that technological advances being made at the time meant that this could be tested using physical molecular genetics experiments in approaches termed genome wide association studies (GWAS). The highs and lows of GWAS studies have been discussed and debated extensively elsewhere and such a discussion is not necessary here. But to cut a long story short, the current consensus is that these common genetic variants at most account for very little of the causative genetic contribution to these diseases.

Focus in the field has more recently turned its attention toward rare genetic variants to search for this otherwise ‘missing heritability’ (Manolio et al., 2009). In this case it is hypothesized that low frequency genetic variants may have larger-effect sizes, thus with fewer alleles needing to contribute in a polygenic manner toward disease manifestation. Investigations to identify these causative rare genetic variants in patients can now be performed with next-generation-sequencing technologies and are sometimes referred to as Rare-Variant-Association-Studies (RVAS). Investigations have shown that rare genetic variants are highly abundant in human populations and indeed that these variants are more likely to have larger deleterious biological effects (Nelson et al., 2012; Tennessen et al., 2012) as predicted by the ‘rare-variant-common disease’ hypothesis. Indeed, it has been suggested that an extensive ‘polygenic burden’ of rare variants within individuals might contribute to the prevalence of the diseases within human populations (Purcell et al., 2014). Currently, the main consensus is that the case/control population cohort sizes would need to be substantially large for the actual causative variants to be identified, thus making efforts to identify disease-causing genes very challenging (Derkach et al., 2014; Lee et al., 2014; Robinson et al., 2014; Sham and Purcell, 2014; Zhang, 2015). There are additional concerns over the reliability of *in silico* methods that are being used to predict whether rare-variants are likely to be deleterious (Henn et al., 2015). Further, the RVAS studies that have been performed so far indicate that most rare variants in fact only have modest-to-small effect sizes on phenotypic variation (Auer and Lettre, 2015). However, it is still relatively early days for such RVAS studies, and as GWAS studies did until a few years ago, they currently dominate the complex genetic diseases research landscape.

BOX 1 | Fundamental components of the new model.

- (1) A generally high, but varying, degree of locus heterogeneity exists for these diseases.
- (2) Biallelic inactivation of a single gene is both necessary and sufficient for disease manifestation.
- (3) The first mutation is inherited from a carrier parent. A random second hit occurs somatically; most likely at some rapid growth phase during embryonic/fetal development, but may also occur during childhood.
- (4) Genetic variants which are inherited from both parents alter the frequency of an individual's mutation rate, and these are a component of the genetic background. Several such mutator/anti-mutator alleles contribute in an additive fashion to modulate the probability of the second hit occurring.
- (5) When sex dimorphism is observed, it can be explained by X-linked inheritance within some families, though with deviations from the expected classic Mendelian pattern, rather than the Carter effect.
- (6) The framework is more likely to apply to diseases with a dichotomous manifestation and to diseases with a suspected developmental component, but may not be limited to diseases of this character.
- (7) Environmental factors might contribute to disease risk, but in a lot of cases, they play a much lesser role than predicted by the current multifactorial model.

AN ALTERNATIVE THEORETICAL FRAMEWORK

When quizzed by an astute undergraduate student, ‘why such a big fuss about rare variants now?’ it is somewhat embarrassing to have no much better answer than ‘because we were mostly wrong about common variants’. We could attempt to lessen the blow by explaining that we are technologically in a much better position today to investigate rare variants in the laboratory, but in truth the student has a point: since both hypotheses were proposed, little has changed regarding our overall understanding of these variants and nothing has changed regarding the theoretical frameworks upon which both of the hypotheses were synthesized. Thus without the hindsight that the common variant hypothesis is deficient, there is little that we have learnt to suggest that the polygenic rare variant hypothesis is any more scientifically sound (Gibson, 2012; Robinson et al., 2014; Auer and Lettre, 2015; Henn et al., 2015). This makes it very acceptable, even prudent, to seriously consider alternative theoretical frameworks, even though they may be against major consensus held for decades. Here, I attempt to piece together a consistent framework which is based on empirical and experimental observations of colleagues in the field, past and present. The fundamental aspects of the conceptual framework are outlined in **Box 1**; and the rationale, merits, and potential shortcomings, are discussed in the sections below.

CONSIDERING THE GENERAL PATTERNS OF OBSERVED FAMILIAL- AND POPULATION-INCIDENCES

A Single-Gene Two-Hit Model

Much of the inspiration for writing this article came from considering observations and ideas of a study entitled ‘Genetics, Chance, and Morphogenesis’ by Kurnit et al. (1987) published

almost 30 years ago. The study investigated the roles of natural random probabilistic events that occur during early development, such as endocardial cell migration, and their influence over biological process by performing computer simulations. They observed a high degree of probabilistic variability in the outcomes of their simulations, which was distributed continuously: and in fact, the output appeared to resemble the normal distribution curves predicted by polygenic theory. They further demonstrated that this also held true if they included a single gene into their probabilistic model, and propose that multifactorial diseases may thus in fact be caused by inheritance of one or two copies of a single defective gene but also by chance events that occur during normal embryonic development as an important contributing factor. A limitation acknowledged by Kurnit et al. (1987) and extended on later by colleagues (Hook, 1988), was that the model didn't satisfactorily account for steep falling-off of recurrence risk with diminishing relatedness as well as it should (addressed in 'Mutator/Anti-mutator Genetic Background section' below). Further, whereas the occurrence of sporadic birth defects may be attributed to some random very bad luck event during development, whether such events occur frequently enough to interact with genetic factors to significantly contribute to relatively common diseases within the population is perhaps much less likely. Here, I suggest a variation to the model such that only one defective copy of a single gene is inherited; and that rather than an adverse random biological processes during development, the chance element comes from inactivation of the second copy of the gene via random mutation during early development or growth. This two-hit model would explain the reduced recurrence-risk within affected families compared to classic Mendelian disorders.

Such a two-hit model is of course at the foundation of our understanding of how mutations in tumor suppressor genes contribute to some forms of cancer (Knudson, 1971). Cancerous cells and tissue represent special examples where accumulating mutations in cell cycle regulators and DNA repair factors accelerate the rates of mutation such that likelihood of the second-hit is increased, and where a new mutations can confer a growth advantage to the cell, thus expanding the target for the next cancer-promoting mutation to occur. The primary reason why such a model is not normally extended to non-cancerous disease is that such events of successive growth-promoting mutations are unlikely in healthy tissue. However, the model presented here posits that only a single mutation needs to occur for disease manifestation, and that this most likely occurs during a rapid growth phase such as in embryonic or fetal development, where a rapid rate of cell division is already naturally inherent to most cells. It is also important to note here that most gene disruptions will not significantly alter growth rates, whether it's a reduction or an increase, or trigger apoptosis, as is sometimes erroneously assumed. Depending on how early these mutations occur in development they can either effect an entire organ or physiological system, or specific circuits within an organ – but enough to manifest as disease. Highly complex organs, such as the brain, probably require a smaller amount of tissue to be affected for the disease to manifest. Such considerations would also explain why these diseases would not display a classic Mendelian

pattern, i.e., when both parents are carriers: if both mutations are inherited this would result in all cells in the offspring being affected, immediately from the single-cell zygote stage, and thus a different observed phenotype. A rather striking example of this comes from our knowledge of mutations in the BRCA2 gene: when one mutation is inherited, there is a predisposition to breast cancer; whereas when two mutations are inherited it causes a congenital syndrome, Fanconi's anemia, where patients instead display high incidences of leukemia, bone marrow failure, and other wide and varying phenotypes (Howlett et al., 2002).

What is a crucial consideration, however, is the natural mutation rate within healthy tissues. This is still an area of active enquiry and estimates vary (Callaway, 2015). But what is clear is that estimates of the somatic mutation rate are consistently considerably much higher than that in the germline (Garcia et al., 2007; Lynch et al., 2008; Lynch, 2010a,b). The appearance of mutations within our genomes include errors incorporated during replication as well those inflicted by mutagens, but an additional very important source is the commonly occurring spontaneous deamination of methylated cytosines of CG dinucleotides to thymine. CG dinucleotides have generally reduced in frequency in genomes over history due to these spontaneous deamination events, but they remain more frequently in exons of protein coding genes where their presence to encode the correct sequence of amino acids is more crucial. Consequently, when deaminations at these sites do occur, they are very often identified as hotspots of causative point mutations in Mendelian and sporadic genetic disease. No truly reliable estimate of somatic mutation rate, however, exists, and this is largely due to our inability over the years to perform a wide survey to reliably determine the spectrum of sequence changes within the genomes of multiple tissue regions, although it is noteworthy that important initiatory studies in this regard using modern sequencing and computational approaches find very extensive somatic genetic variation when diverse tissues are sampled (Frank, 2010; O'Huallachain et al., 2012; Behjati et al., 2014); and more common roles for somatic mosaicism in disease manifestation has been suggested (De, 2011; Cohen et al., 2015). Indeed, our observations of non-cancerous human diseases caused by somatic mutations has been steadily increasing in recent years as our ability to detect them improves; and this includes mutations that may accumulate during the lifetime of an individual in age-related illnesses (Forsberg et al., 2012, 2013; Bonnefond et al., 2013; Aghili et al., 2014), and also new mutations that occur in early development that may contribute to the formation of sporadic forms of non-cancerous disease (Lindhurst et al., 2011; Poduri et al., 2012, 2013; Li et al., 2013; Dal et al., 2014).

A Mutator/Anti-mutator Genetic Background

Mutation rates vary between species and they can also be influenced by variant alleles of genes encoding DNA replication and repair machinery within populations (Fijalkowska and Schaaper, 1995; Wibley et al., 2003; Baer, 2007; Senejani et al., 2012). Further, the observation that germline and somatic

mutation incidences are often widely different within eukaryotic organisms indicates the existence of factors, likely genetically encoded, that can exert control over the rate of mutation (Lynch, 2010b). In addition, there are likely several yet other diverse genetically encoded factors that can influence mutation rate by regulating the rates/efficiency of mutagen clearance within cells. Importantly, other classes of genetically encoded factors that can influence the mutation rate have also been described; and these are currently known to range from microRNAs, to factors which regulate dNTP pools, amongst others (Valeri et al., 2010a,b; Tili et al., 2011; Ahluwalia and Schaaper, 2013; Alderton, 2013; Scanlon and Glazer, 2015; Sohl et al., 2015). Thus we likely have an architecture where genetic variants of a very broad range of cellular factors capable of significantly influencing the mutation rate exist, thus potentially resulting in significant variance of the mutation rate amongst individuals within a population.

With regards to the two-hit model presented in the current framework, once the first mutation has been inherited, since the chance second hit is both necessary and sufficient for disease manifestation, any factors that influence the probability of the second hit occurring, will play a significant role in determining disease manifestation. We know that classic Mendelian disease can sometimes be influenced by the genetic background which can modulate the penetrance. Here, it is proposed that the genetic background also otherwise contains an additional very important component that contributes to disease manifestation in this particular framework: a number of mutator and anti-mutator alleles; and that these variants can be inherited from both the carrier-parent and the non-carrier parent. Such a proposal would help explain the steeper falling-off of recurrence risk with decreasing relatedness observed for these diseases compared to the proportionate falling-off observed in classic Mendelian disorders. This is perhaps the most important aspect of this entire framework, and may be indirectly supported by observations. Mutation rates are known to be elevated in multicellular species relative to unicellular organisms, and in a fascinating theoretical undertaking, Lynch (2008) hypothesizes that multicellularity increases the probability of fitness reduction associated with somatic damage, but paradoxically in addition actually encourages the accumulation of mutator alleles that cause such damage, thereby magnifying the vulnerability to somatic mutations and disease. He also notes that when mutator alleles rise to moderate frequencies, considerable heterogeneity in the mutation rate is expected among individuals. The general idea of genetic variants commonly influencing mutation rates across a population is supported by a large-scale sequencing study which revealed that mutation rates vary substantially between families, and the authors thus suggest that mutation rates may be influenced by an individual's genetic background (Conrad et al., 2011). The notion that genetic variants in the genetic background which are capable of significantly influencing the penetrance of disease was also previously suggested (Demogines et al., 2008). Indeed, investigations have shown that patients with schizophrenia caused by new mutations, displayed generally increased mutation rates compared to control individuals (Girard et al., 2011; Xu et al., 2012).

Locus Heterogeneity

The pleiotropic effect refers to the situation where a single locus affects more than one phenotypic trait, and plays established roles in the manifestation of genetic disease (Stearns, 2010). In contrast, locus heterogeneity describes the situation where a particular phenotypic trait or disorder can result from mutations at any one of several distinct loci; this particular concept likely plays key roles within the proposed framework, as it could potentially help explain the more frequent occurrence of complex genetic diseases within the population. A proposal of locus heterogeneity would also have equally important implications for other observations regarding the manifestation characteristics of some of these diseases. For example, it is sometimes observed that some families affected by a particular multifactorial disease have a higher incidence risk than others, most often amongst siblings. In fact if a couple have two children with a particular disease, the recurrence risk for the next child steeply increases. Within these families, it is possible that the particular causative gene is an example of a mutation hotspot locus thus significantly increasing the risk to offspring. Additionally, the parents could harbor a high number of mutator alleles/low number of anti-mutator alleles, thus further increasing the risk. Another observation in a few examples of multifactorial disease, such as cleft palate and cleft lip and neural tube defects, is that if a child is severely affected with the disease, then the recurrence-risk for next child also increases. This is a little more difficult to fully explain with the current framework, however, the proposed locus heterogeneity likely also plays a key role: some genes will encode factors which when disrupted cause different aspects of a biological process or physiological system to fail, thus though they manifest as the same observable disease, they do so with varying degrees of severity.

CONSIDERING CHARACTERISTICS OF DISEASE MANIFESTATION AND SEX DIMORPHISM

Many multifactorial diseases display a dichotomous manifestation. This is most apparent in the congenital diseases such as cleft palate and cleft lip, pyloric stenosis, clubfoot, and neural tube defects, but also in others such as autism, schizophrenia, and bipolar disorder. These and others are much better explained by the current single-gene model than by the polygenic model, as the former does not require invoking the idea of disease-thresholds, the actual biological mechanisms for which are ill-defined.

As described earlier, the Carter effect is posited as an explanation for observations of sex dimorphism in the polygenic threshold model, for which no actual proposed biological mechanism exists. Within the current framework I suggest that the examples of sex dimorphic multifactorial diseases might instead be explained by X-linked inheritance, which only occurs in some families due to the high degree of locus heterogeneity. For diseases which display a rather high level of sex dimorphism

such as in pyloric stenosis, clubfoot, and autism, more of the causative genes or even just one or two which are mutation hotspots could be located on the X chromosome; whereas those that display a more modest degree of dimorphism contain fewer, perhaps even just one causative X-chromosome gene. When X-linked inheritance does occur within families, somatic mosaicism complicates the pattern of X-linked inheritance expected from the classic Mendelian prediction much more frequently than it normally would do. For example, since random X-inactivation occurs at the ~20 cell stage of the developing female embryo, and since somatic mosaicism is inherent in determining disease occurrence under this framework (i.e., only a smaller amount of tissue needs to be affected), then it is much more likely that random X-inactivation could lead to manifestation of the disease, as inactivation of the normal X-chromosome would only need to occur in a much more limited amount of tissue. Thus in such situations, daughters of mother-carriers would be more frequently affected than expected, thus masking the pattern of X-linked inheritance that we expect from the usual Mendelian prediction where males are mainly affected after inheriting the faulty gene from a carrier mother.

Indeed, the most common cases of sex dimorphism in multifactorial disease are where the number of affected males significantly outnumber the number of affected females within the population. In such cases, if the Carter effect does occur, the most readily detectable observation would be that relatives of affected females should be at higher risk than relatives of affected males. A Danish study which included a cohort of ~two million children, ~3,400 of which had surgery for pyloric stenosis, found no such correlation (Krogh et al., 2010). In a second further large-scale study, it was reported that the recurrence rates of autism spectrum disorders (ASDs) were the same in families with affected female individuals as with affected male individuals (Ozonoff et al., 2011). Indeed, the authors of both of these large-scale studies commented that the observed data do not support the polygenic threshold model. An interesting observation, however, was that maternal half-siblings are at a significantly increased risk compared to paternal half-siblings in a large Danish study reporting on ASD incidence (Grønberg et al., 2013). The authors propose, that factors associated with pregnancy and the maternal intrauterine environment contribute to the risk of developing ASDs; but, the observations may instead be explained if females are more often carriers of a faulty gene(s), i.e., one that is X-linked. As is apparent, only a very limited amount of information can be gleaned from such studies, and the complications to predicted patterns of inheritance added by the current framework exacerbates these limitations.

OTHER IMPORTANT CONSIDERATIONS

Genetic Interactions

In Fisher's (1918) framework, genetic factors contribute to a continuous trait, such as height, in an additive fashion. It's a revered theory amongst geneticists and describes continuous traits, in an elegant, easy to envisage manner. However, when used in this fashion in attempts to account for common diseases

within populations, shortcomings become apparent (Nelson et al., 2013); and multiplicative models have been suggested where gene-gene interactions likely play a significant role (Risch, 1990; Wray and Goddard, 2010). Higher order genetic interactions (HGIs) probably do exist in nature (Taylor and Ehrenreich, 2015), though the roles for these in disease manifestation have not previously been described, and we have knowledge of only very few examples of them in biological systems. It has been suggested that genetic studies fail to detect HGIs, because of a lack of statistical power (Cordell, 2009; Taylor and Ehrenreich, 2015), but even so, it may indeed be that the roles in disease manifestation are actually rare. This is since when considering multiplicative genetic interactions from a biological perspective, we are referring to specific genetically encoded allelic products, rather than any form of the genetically encoded product, that are able to influence each other in a biologically meaningful way. Although attractive as a concept in genetics, from a cell biology point of view, this is very difficult to imagine as occurring with any meaningful frequency, as would be expected in the polygenic model. Genetic background, however, is a well-recognized biological feature in influencing phenotypic variation (Nadeau, 2001; Dowell et al., 2010; Taylor and Ehrenreich, 2015). In the current framework though mutator/anti-mutator variants are introduced into the genetic background, they most likely act in an additive fashion, as described by Fisher (1918), to influence an individual's general mutation rate.

Recessiveness vs. Dominance

As recently discussed by Henn et al., 2015, dominance is perhaps the most important quantity that has not been estimated from genome-wide data characterizing the proportion of deleterious mutations within genomes, though it's fair to stipulate that the proportion of deleterious mutations that are recessive or dominant remains an open question. However, we know particularly from examples of classic Mendelian disease but even some cases of sporadic genetic disease, that a recessive mode is the most commonly observed route to disease manifestation. Thus for a particular locus, examples where half the amount of 'wild-type' genetic product is insufficient, or where a variant product is dominant over a 'wild-type' product, are in fact rare. Though this is what the polygenic disease model generally assumes, and it assumes that this occurs with a high degree of frequency, particularly in the rare-variant hypothesis where new genetic variants are almost exclusively found in heterozygous form. The two-hit model presented in the current framework instead posits a generally recessive mode of inheritance where both copies of a gene need to be inactivated.

Contributions from the Environment

Multifactorial disease are always thought of as having an environmental component that also significantly contributes to disease risk. Sometimes it is much more obvious what the environmental components may be; they include diet, exercise, and smoking habits for conditions such as hypertension, diabetes, obesity, and cancer for example. In other cases the factors are much more difficult to ascertain, and this is particularly the case for multifactorial congenital disorders. It has been shown

for example that the concordance rates for monozygotic twins in clubfoot, cleft lip, and pyloric stenosis is fairly modest, ranging from 17 to 46% (Christensen and Fogh-Anderson, 1993; Engell et al., 2006; Krogh et al., 2010). Pyloric stenosis is not truly congenital, in that it becomes apparent in the first 2–8 weeks of life; and based on previous suggestions, Krogh et al. (2010) attribute the 54% discordance amongst monozygotic twins from their large-scale study to factors such as feeding practice, maternal smoking, infant sleeping position, and postnatal use of macrolides, as possible contributing environmental factors. For clubfoot and cleft lip and palate, environmental factors that have been suggested are maternal smoking and alcohol intake. Although some of these environmental factors are sometimes responsible for the incidence of sporadic birth defects, their contributions to the aforementioned multifactorial diseases, and others, may have been overestimated. Instead, discordance between monozygotic twins may be much more readily explained by the current two-hit framework, where the second post-zygotic hit is required for disease manifestation.

TESTING THE MODEL AND DEMONSTRATING GENOTYPE–PHENOTYPE CAUSALITY

Currently, much focus in identifying associative genetic factors in complex genetic disease is on RVAS studies, and it has been suggested that such population-based studies need to be of a substantially large sample size in order to be able to detect the causal rare variants (Lee et al., 2014; Robinson et al., 2014; Sham and Purcell, 2014). The appropriateness of this line of investigation can be questioned (Wagenmakers et al., 2014). It has also been suggested that smaller family-based studies would be ineffective in disease-gene identification (Robinson et al., 2014), though this observation is based on a framework where an underlying polygenic architecture exists. Although in the framework proposed here, the disease-causing alleles are likely to be rare variants, family-based studies rather than large-scale population based studies would be better suited for disease-gene identification. The framework posits a single causative gene, however, segregation-analysis would not be entirely appropriate, as unaffected siblings may also harbor the inherited mutation. Instead blood/cheek swabs from parents and affected individuals should be obtained for sequencing, but also critical here is the identification of the second hit in affected tissues from the affected individuals; indeed testing of unaffected vs. affected tissues was previously used to identify causative mutations in sporadic forms of Proteus syndrome (Lindhurst et al., 2011). It may not always be possible to obtain affected tissue from living patients, but wherever it is feasible to obtain such material in a stress-free manner, they should be biopsied and sequenced. In some situations identifying the precise regions of tissues that are affected may be less-obvious, in which case tissue of the appropriate developmental lineage may instead be analyzed. In other cases, a mixture of normal/affected tissue might be unavoidable, however, next-generation-sequencing approaches and recent developments in computational methods should be

able to identify and distinguish the somatic mutations (Roth et al., 2012; Schmitt et al., 2012; Kim et al., 2013; Josephidou et al., 2015). A key consideration is that it would need to be determined that the second hit has occurred on the opposite chromosome and not in the same copy of the gene that harbors the first hit. This may now be possible with Nanopore sequencing methods which allows the direct DNA sequencing of ultra-long reads, which would allow haplotypes to be readily determined. Although not yet commercially available, Oxford Nanopore Technologies (ONT), have made available their first sequencing device (the minION) as part of an early-access program for researchers. Early reports have indeed confirmed the ability of the device to obtain sequencing reads of tens of kilobases in length (Laver et al., 2015). Once biallelic hits in a gene from the affected tissue have been identified, only one of which would be more likely to be present in unaffected tissue, and they are both predicted to significantly disrupt gene function, they should be considered candidate genes. However, demonstration of causality needs to be a central consideration, and should be assessed by biological means. Animal models may be useful in such regards, but care must be taken to recapitulate as closely as possible the physiological manifestation characteristics of the disease, i.e., the second hit needs to be induced in the correct developmental lineage and in a relevant point in time, which may not always be readily identifiable. However, cheaper and quicker genome editing methods, such as the Crispr-Cas9 system, mean that such investigation can be more routinely performed, and such biological tools would also further be useful in defining and characterizing routes to therapy.

Depending on the degree of locus heterogeneity, several disease genes may exist for a particular disease, one of which will be disrupted in each family. Each of these genes will first need to be identified as described above, but thereafter, blood/cheek swab mutation screening of these candidate genes could be performed in all families for future diagnostics applications. Research costs are often a limiting factor in medical genetics, but it is important that all of the causative disease genes for a particular disease are comprehensively identified, if at least to provide a more satisfactory level of genetic counseling than is currently available for all families affected. In this regard, yet another very alluring feature of ONT devices is their low cost, since although the sequencing cost per genome is rapidly improving for the currently commercially available technologies (van Dijk et al., 2014; Watson, 2014), these generally operate via a limited number of core sequencing facilities. However, the ONT devices themselves will be much more routinely affordable and we may thus envisage smaller family-based studies which all or most medical genetics laboratories around the world are able to perform; this should speed up comprehensive disease gene-identification, providing much better defined routes to therapy in a family-tailored manner. Roughly 85% of disease causing mutations identified in classic Mendelian disease occurs in the protein-coding portions of the genome, and it is probably likely we would see a similar percentage under the current framework. Therefore, where cost is an especially limiting factor, studies might focus on whole-exome-sequencing approaches, until whole-genome-sequencing becomes more routinely affordable.

Finally, since an advantage of this framework is that it explains diseases that have a more dichotomous manifestation better than the polygenic model does, these particular diseases may be more suitable for initial investigations.

CONCLUDING REMARKS

The primary, perhaps only reason we should harbor hopes that such a framework is correct, is that it would make it considerably easier to define and develop routes to therapy, as this could be specifically targeted against the single causative

monogenic factor within families. Such hopes of course have little or nothing to do with scientific reality and it is very possible that the framework presented here is incorrect. The jury is still very much out for the polygenic rare-variant hypothesis, and the previous overwhelming consensus regarding the polygenic common-variant hypothesis was, for most intents and purposes, wrong. I thus feel rather apprehensive to conclude with a comment regarding how likely it is for this proposed framework to hold true to fact. However, the framework appears broadly consistent with recorded observations past and present, and if correct, the potential benefits are huge; testing of the model is clearly incumbent.

REFERENCES

- Aghili, L., Foo, J., DeGregori, J., and De, S. (2014). Patterns of somatically acquired amplifications and deletions in apparently normal tissues of ovarian cancer patients. *Cell Rep.* 7, 1310–1319. doi: 10.1016/j.celrep.2014.03.071
- Ahluwalia, D., and Schaaper, R. M. (2013). Hypermutability and error catastrophe due to defects in ribonucleotide reductase. *Proc. Natl. Acad. Sci. U.S.A.* 110, 18596–18601. doi: 10.1073/pnas.1310849110
- Alderton, G. K. (2013). Genomics: mutator catalogues. *Nat. Rev. Cancer* 13:681. doi: 10.1038/nrc3608 PMID:NOPMIm
- Auer, P. L., and Lettre, G. (2015). Rare variant association studies: considerations, challenges and opportunities. *Genome Med.* 7:16. doi: 10.1186/s13073-015-0138-2
- Baer, C. F. (2007). Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nat. Rev. Genet.* 8, 619–631. doi: 10.1038/nrg2158
- Behjati, S., Huch, M., van Boxtel, R., Karthaus, W., Wedge, D. C., Tamuri, A. U., et al. (2014). Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* 513, 422–425. doi: 10.1038/nature13448
- Bonnefond, A., Skrobek, B., Lobbens, S., Eury, E., Thuillier, D., Cauchi, S., et al. (2013). Association between large detectable clonal mosaicism and type 2 diabetes with vascular complications. *Nat. Genet.* 45, 1040–1043. doi: 10.1038/ng.2700
- Callaway, E. (2015). DNA mutation clock proves tough to set. *Nature* 519, 139–140. doi: 10.1038/519139a
- Carter, C. O., and Evans, K. A. (1969). Inheritance of congenital pyloric stenosis. *J. Med. Genet.* 6, 233–254. doi: 10.1136/jmg.6.3.233
- Christensen, K., and Fogh-Anderson, P. (1993). Cleft lip (\pm cleft palate) in Danish twins, 1970–1990. *A. J. Med. Genet.* 47, 910–916. doi: 10.1002/ajmg.1320470620
- Cohen, A. S., Wilson, S. L., Trinh, J., and Ye, X. C. (2015). Detecting Somatic mosaicism: considerations and clinical implications. *Clin. Genet.* 87, 554–562. doi: 10.1111/cge.12502
- Conrad, D. F., Keebler, J. E., DePristo, M. A., Lindsay, S. J., Zhang, Y., Casals, F., et al. (2011). Variation in genome-wide mutation rates within and between human families. *Nat. Genet.* 43, 712–715. doi: 10.1038/ng.862
- Cordell, H. J. (2009). Detecting gene-gene interactions that underlie human diseases. *Nat. Rev. Genet.* 10, 392–404. doi: 10.1038/nrg2579
- Dal, G. M., Ergüner, B., Sağroğlu, M. S., Yüksel, B., Onat, O. E., Alkan, C., et al. (2014). Early postzygotic mutations contribute to de novo variation in a healthy monozygotic twin pair. *J. Med. Genet.* 51, 455–459. doi: 10.1136/jmedgenet-2013-102197
- De, S. (2011). Somatic mosaicism in healthy human tissues. *Trends Genet.* 27, 217–223. doi: 10.1016/j.tig.2011.03.002
- Demogines, A., Wong, A., Aquadro, C., and Alani, E. (2008). Incompatibilities involving yeast mismatch repair genes: a role for genetic modifiers and implications for disease penetrance and variation in genomic mutation rates. *PLoS Genet.* 6:e1000103. doi: 10.1371/journal.pgen.1000103
- Derkach, A., Lawless, J. F., and Sun, L. (2014). Pooled association tests for rare genetic variants: a review and some new results. *Stat. Sci.* 29, 302–321. doi: 10.1214/13-STS456
- Dowell, R. D., Ryan, O., Jansen, A., Cheung, D., Agarwala, S., Danford, T., et al. (2010). Genotype to phenotype: a common problem. *Science* 328, 469. doi: 10.1126/science.1189015
- Engell, V., Damborg, F., Andersen, M., Kyvik, K. O., and Thomsen, K. (2006). Clubfoot: a twin study. *J. Bone Joint Surg.* 88-B, 374–376. doi: 10.1302/0301-620X.88B3.16685
- Falconer, D. S. (1965). The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann. Hum. Genet.* 29, 51–76. doi: 10.1111/j.1469-1809.1965.tb00500.x
- Fijalkowska, I. J., and Schaaper, R. M. (1995). Effects of *Escherichia coli* dnaE Antimutator Alleles in a Proofreading-Deficient mutD5 Strain. *J. Bacteriol.* 177, 5979–5986.
- Fisher, R. A. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Trans. R. Soc. Edinburgh* 52, 399–433. doi: 10.1017/S0080456800012163
- Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. Lond. Ser. A* 222, 309–368. doi: 10.1098/rsta.1922.0009
- Forsberg, L. A., Absher, D., and Dumanski, J. P. (2013). Non-heritable genetics of human disease: spotlight on post-zygotic genetic variation acquired during lifetime. *J. Med. Genet.* 50, 1–10. doi: 10.1136/jmedgenet-2012-101322
- Forsberg, L. A., Rasi, C., Razzaghian, H. R., Pakalapati, G., Waite, L., Thilbeault, K. S., et al. (2012). Age-related somatic structural changes in the nuclear genome of human blood cells. *Am. J. Hum. Genet.* 90, 217–228. doi: 10.1016/j.ajhg.2011.12.009
- Frank, S. A. (2010). Evolution in health and medicine Sackler colloquium: somatic evolutionary genomics: mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. *Proc. Natl. Acad. Sci. U.S.A.* 107(Suppl. 1), 1725–1730. doi: 10.1073/pnas.0909343106
- Garcia, A. M., Derventzi, A., Busuttill, R., Calder, R. B., Perez, E. Jr., Chadwell, L., et al. (2007). A model system for analyzing somatic mutations in *Drosophila melanogaster*. *Nat. Methods* 4, 401–403.
- Gibson, G. (2012). Rare and common variants: twenty arguments. *Nat. Rev. Genet.* 13, 135–145. doi: 10.1038/nrg3118
- Girard, S. L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Jouan, L., et al. (2011). Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat. Genet.* 43, 860–863. doi: 10.1038/ng.886
- Grønborg, T. K., Schendel, D. E., and Parner, E. T. (2013). Recurrence of autism spectrum disorders in full- and half-siblings and trends over time: a population-based cohort study. *JAMA Pediatr.* 167, 947–953. doi: 10.1001/jamapediatrics.2013.2259
- Henn, B. M., Botigué, L. R., Bustamante, C. D., Clark, A. G., and Gravel, S. (2015). Estimating the mutation load in human genomes. *Nat. Rev. Genet.* 16, 333–343. doi: 10.1038/nrg3931
- Hook, E. B. (1988). Polygenes versus one gene plus chance: what are the real differences? *Am. J. Hum. Genet.* 43, 783.
- Howlett, N. G., Taniguchi, T., Olson, S., Cox, B., Waisfisz, Q., De Die-Smulders, C., et al. (2002). Biallelic inactivation of Brca2 in Fanconi anemia. *Science* 297, 606–609. doi: 10.1126/science.1073834
- Josephidou, M., Lynch, A. G., and Tavaré, S. (2015). multiSNV: a probabilistic approach for improving detection of somatic point mutations from multiple related tumour samples. *Nucleic Acids Res.* 43:e61. doi: 10.1093/nar/gkv135

- Kim, S., Jeong, K., Bhutani, K., Lee, J., Patel, A., Scott, E., et al. (2013). Virmid: accurate detection of somatic mutations with sample impurity inference. *Genome Biol.* 14:R90. doi: 10.1186/gb-2013-14-8-r90
- Knudson, A. G. Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. U.S.A.* 68, 820–823. doi: 10.1073/pnas.68.4.820
- Krogh, C., Fischer, T. K., Skotte, L., Biggar, R. J., Øyen, N., Skytthe, A., et al. (2010). Familial aggregation and heritability of pyloric stenosis. *JAMA* 303, 2393–2399. doi: 10.1001/jama.2010.784
- Kurnit, D. M., Layton, W. M., and Matthyse, S. (1987). Genetics. Chance and Morphogenesis. *Am. J. Hum. Genet.* 41, 979–995.
- Laver, T., Harrisona, J., O'Neill, P. A., Moorea, K., Farbosa, A., Paszkiewicz, K., et al. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol. Detect. Quantif.* 3, 1–8. doi: 10.1016/j.bdq.2015.02.001
- Lee, S., Abecasis, G. R., Boehnke, M., and Lin, X. (2014). Rare-variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet.* 95, 5–23. doi: 10.1016/j.ajhg.2014.06.009
- Li, R., Montpetit, A., Rousseau, M., Wu, S. Y., Greenwood, C. M., Spector, T. D., et al. (2013). Somatic point mutations occurring early in development: a monozygotic twin study. *J. Med. Genet.* 51, 28–34. doi: 10.1136/jmedgenet-2013-101712
- Lindhurst, M. J., Sapp, J. C., Teer, J. K., Johnston, J. J., Finn, E. M., Peters, K., et al. (2011). A mosaic activating mutation in *AKT1* associated with the Proteus syndrome. *N. Engl. J. Med.* 365, 611–619. doi: 10.1056/NEJMoa1104017
- Lynch, M. (2008). The cellular, developmental and population-genetic determinants of mutation-rate evolution. *Genetics* 180, 933–943. doi: 10.1534/genetics.108.090456
- Lynch, M. (2010a). Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. U.S.A.* 107, 961–968. doi: 10.1073/pnas.0912629107
- Lynch, M. (2010b). Evolution of the mutation rate. *Trends Genet.* 26, 345–352. doi: 10.1016/j.tig.2010.05.003
- Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C. R., Dopman, E. B., et al. (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 105, 9272–9277. doi: 10.1073/pnas.0803466105
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorf, L. A., Hunter, D. J., et al. (2009). Finding the missing heritability of complex diseases. *Nature* 461, 747–753. doi: 10.1038/nature08494
- Nadeau, J. H. (2001). Modifier genes in mice and humans. *Nat. Rev. Genet.* 2, 165–174. doi: 10.1038/35056533
- Nelson, M. R., Wegmann, D., Ehm, M. G., Kessner, D., St. Jean, P., Verzilli, C., et al. (2012). An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* 337, 100–104. doi: 10.1126/science.1217876
- Nelson, R. M., Pettersson, M. E., and Carlborg, Ö (2013). A century after Fisher: time for a new paradigm in quantitative genetics. *Trends Genet.* 29, 669–676. doi: 10.1016/j.tig.2013.09.006
- O'Huallachain, M., Karczewski, K. J., Weissman, S. M., Urban, A. E., and Snyder, M. P. (2012). Extensive genetic variation in somatic human tissues. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18018–18023. doi: 10.1073/pnas.1213736109
- Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., et al. (2011). recurrence risk for autism spectrum disorders: a baby siblings research consortium study. *Pediatrics* 128, e1–e8. doi: 10.1542/peds2010-2825
- Poduri, A., Evrony, G. D., Cai, X., Elhosary, P. C., Beroukhim, R., Lehtinen, M. K., et al. (2012). Somatic activation of *AKT3* causes hemispheric developmental brain malformations. *Neuron* 74, 41–48. doi: 10.1016/j.neuron.2012.03.010
- Poduri, A., Evrony, G. D., Cai, X., and Walsh, C. A. (2013). Somatic mutation, genomic variation, and neurological disease. *Science* 341:1237758. doi: 10.1126/science.1237758
- Purcell, S. M., Moran, J. L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., et al. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 506, 185–190. doi: 10.1038/nature12975
- Risch, N. (1990). Linkage strategies for genetically complex traits. Multilocus models. *Am. J. Hum. Genet.* 46, 222–228.
- Robinson, M. R., Wray, N. R., and Visscher, P. M. (2014). Explaining additional genetic variation in complex traits. *Trends Genet.* 30, 124–132. doi: 10.1016/j.tig.2014.02.003
- Roth, A., Ding, J., Morin, R., Crisan, A., Ha, G., Giuliany, R., et al. (2012). JointSNVMix: a probabilistic model for accurate detection of somatic mutations in normal/tumour paired next-generation sequencing data. *Bioinformatics* 28, 907–913. doi: 10.1093/bioinformatics/bts053
- Scanlon, S. E., and Glazer, P. M. (2015). Multifaceted control of DNA repair pathways by the hypoxic tumor microenvironment. *DNA Repair* 32, 180–189. doi: 10.1016/j.dnarep.2015.04.030
- Schmitt, M. W., Kennedy, S. R., Salk, J. J., Fox, E. J., Hiatt, J. B., and Loeb, L. A. (2012). Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14508–14513. doi: 10.1073/pnas.1208715109
- Senejani, A. G., Dalal, S., Liu, Y., Nottoli, T. P., McGrath, J. M., Clairmont, C. S., et al. (2012). Y265C DNA polymerase beta knockin mice survive past birth and accumulate base excision repair intermediate substrates. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6632–6637. doi: 10.1073/pnas.1200800109
- Sham, P. C., and Purcell, S. M. (2014). Statistical power and significance testing in large-scale genetic studies. *Nat. Rev. Genet.* 15, 335–346. doi: 10.1038/nrg3706
- Sohl, C. D., Ray, S., and Sweasy, J. B. (2015). Pools and Poles: mechanism of a mutator phenotype. *Proc. Natl. Acad. Sci. U.S.A.* 112, 5864–5865. doi: 10.1073/pnas.1505169112
- Stearns, F. W. (2010). One hundred years of pleiotropy: a retrospective. *Genetics* 186, 767–773. doi: 10.1534/genetics.110.122549
- Taylor, M. B., and Ehrenreich, I. M. (2015). Higher-order genetic interactions and their contribution to complex traits. *Trends Genet.* 31, 34–40. doi: 10.1016/j.tig.2014.09.001
- Tennessen, J. A., Bigham, A. W., O'Connor, T. D., Fu, W., Kenny, E. E., Gravel, S., et al. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337, 64–69. doi: 10.1126/science.1219240
- Tili, E., Michaille, J. J., Wernicke, D., Alder, H., Costinean, S., Volinia, S., et al. (2011). Mutator activity induced by microRNA-155 (miR-155) links inflammation and cancer. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4908–4913. doi: 10.1073/pnas.1101795108
- Valeri, N., Gasparini, P., Braconi, C., Paone, A., Lovat, F., Fabbri, M., et al. (2010a). MicroRNA-21 induces resistance to 5-fluorouracil by down-regulating human DNA MutS homolog 2 (hMSH2). *Proc. Natl. Acad. Sci. U.S.A.* 107, 21098–21103. doi: 10.1073/pnas.1015541107
- Valeri, N., Gasparini, P., Fabbri, M., Braconi, C., Veronese, A., Lovat, F., et al. (2010b). Modulation of mismatch repair and genomic stability by miR-155. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6982–6987. doi: 10.1073/pnas.1002472107
- van Dijk, E. L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet.* 30, 418–426. doi: 10.1016/j.tig.2014.07.001
- Wagenmakers, E. J., Verhagen, J., Ly, A., Bakker, M., Lee, M. D., Matzke, D., et al. (2014). A power fallacy. *Behav. Res. Methods* doi: 10.3758/s13428-014-0517-4 [Epub ahead of print].
- Watson, M. (2014). Illuminating the future of DNA sequencing. *Genome Biol.* 15:108. doi: 10.1186/gb4165
- Wible, J. E., Waters, T. R., Haushalter, K., Verdine, G. L., and Pearl, L. H. (2003). Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1. *Mol. Cell* 11, 1647–1659. doi: 10.1016/S1097-2765(03)00235-1
- Wray, N. R., and Goddard, M. E. (2010). Multi-locus models of genetic risk of disease. *Genome Med.* 2:10. doi: 10.1186/gm131
- Xu, B., Ionita-Laza, I., Roos, J. L., Boone, B., Woodruff, S., Sun, Y., et al. (2012). De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat. Genet.* 44, 1365–1369. doi: 10.1038/ng.2446
- Zhang, Q. (2015). Associating rare genetic variants with human diseases. *Front. Genet.* 6:133. doi: 10.3389/fgene.2015.00133

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Hussain. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.