bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

Clinical genetics of neurodevelopmental disorders

Gholson J. Lyon[1,2,3,*], Jason O'Rawe[1,3]

[1]Stanley Institute for Cognitive Genomics, One Bungtown Road, Cold Spring Harbor Laboratory, NY, USA, 11724
[2]Institute for Genomic Medicine, Utah Foundation for Biomedical Research, E 3300 S, Salt Lake City, Salt Lake City, UT, USA, 84106
[3]Stony Brook University, 100 Nicolls Rd, Stony Brook, NY, USA, 11794

* Corresponding author: Gholson J. Lyon
Email: GholsonJLyon@gmail.com

Other author emails:
Jason O'Rawe: jazon33y@gmail.com

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

Introduction

"our incomplete studies do not permit actual classification; but it is better to leave things by themselves rather than to force them into classes which have their foundation only on paper" – Edouard Seguin [1]

"The fundamental mistake which vitiates all work based upon Mendel's method is the neglect of ancestry, and the attempt to regard the whole effect upon offspring, produced by a particular parent, as due to the existence in the parent of particular structural characters; while the contradictory results obtained by those who have observed the offspring of parents apparently identical in certain characters show clearly enough that not only the parents themselves, but their race, that is their ancestry, must be taken into account before the result of pairing them can be predicted" – Walter Frank Raphael Weldon [2].

There are ~6 billion nucleotides in every cell of the human body, and there are ~25-100 trillion cells in each human body. Given somatic mosaicism, epigenetic changes and environmental differences, no two human beings are the same, particularly as there are only ~7 billion people on the planet. One of the next great challenges for studying human genetics will be to acknowledge and embrace complexity [3-13]. Every human *is* unique, and the study of human disease phenotypes (and phenotypes in general) will be greatly enriched by moving from a deterministic to a more stochastic/probabilistic model [14-19]. The dichotomous distinction between 'simple' and 'complex' diseases is completely artificial, and we argue instead for a model that considers a spectrum of diseases that are variably manifesting in each person. The rapid adoption of whole genome sequencing (WGS) and the Internet-mediated networking of people promise to yield more insight into this century-old debate [2,20-25]. Comprehensive ancestry tracking and detailed family history data, when combined with WGS or at least cascade-carrier screening [26], might eventually facilitate a degree of genetic prediction for some diseases in the context of their familial and ancestral etiologies. However, it is important to remain humble, as our current state of knowledge is not yet sufficient, and in principle, any number of nucleotides in the genome, if mutated or modified in a certain way and at a certain time and place, might influence some phenotype during embryogenesis or postnatal life [9,27-44].

In this chapter, we will traverse contemporary understandings of the genetic architecture of human disease, and explore the clinical implications of the current state of our knowledge. Many molecular models have been postulated as being important in genetic disease, and, despite our incomplete knowledge of the genetic workings of many diseases, significant progress has been made over the past 50 years. Many different classes of genetic mutations have been implicated as being involved in predisposition to certain diseases, and we are continually uncovering other means by which genetics plays an important role in human disease, such as with somatic genetic mosaicism. An explosion in the development of new biomedical techniques, molecular technologies, and analytical tools has enriched our knowledge of the many molecular bases of disease, underscored by the fact that we now exist in a world where each person can be characterized on the level of their 'genome', 'transcriptome' and 'proteome'. We discuss these exciting new developments and the current applications of these technologies, their limitations, their implications for prenatal diagnosis and implantation genetics, as well as future prospects.

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

Clinical classifications and the genetic architecture of disease

"Those who have given any attention to congenital mental lesions, must have been frequently puzzled how to arrange, in any satisfactory way, the different classes of this defect which may have come under their observation. Nor will the difficulty be lessened by an appeal to what has been written on the subject. The systems of classification are generally so vague and artificial, that, not only do they assist but feebly, in any mental arrangement of the phenomena represented, but they completely fail in exerting any practical influence on the subject." – John Langdon Down [45]

As most clinicians know from experience, it is quite difficult to characterize the range of human experience in the two-dimensional world of the printed page, as we are attempting to do here. In addition, classifications can sometimes lead people to try to force round pegs into square holes, and so we are reluctant to further promulgate these classifications. Such classifications include terms such as: 'Mendelian', 'complex disease', 'penetrance', 'expressivity', 'oligogenic', and 'polygenic'. For example, some have used the word 'Mendelian' to refer to a disease that appears to somehow be 'caused' by mutations in a single gene. As such, cystic fibrosis, Huntington's disease, and Fragile X are all diseases that some people refer to as being 'caused' by mutations occurring in single genes. However, the expression of the phenotype within these diseases is extremely variable, depending in part on the exact mutations in each gene, and it is not at all clear that any mutation really and truly 'causes' any phenotype, at least not according to thoughtful definitions of causation that we are aware of [46,47]. For example, some children with certain mutations in *CFTR* may only have pancreatitis as a manifestation of cystic fibrosis, without any lung involvement [48,49], and there is evidence that mutations in other genes in the genomes can have a modifying effect on the phenotype [50,51]. In the case of Huntington's, there is extreme variability in the expression of the phenotype, both in time, period and scope of illness, and all of this is certainly modified substantially by the number of trinucleotide repeats [52], genetic background [53] and environmental influences [54]. Even in the case of whole chromosome disorders, such as Down Syndrome, there is ample evidence of substantial phenotypic expression differences, modified again by genetic background [55,56], somatic mosaicism [57] and environmental influences [58,59], including synaptic and brain plasticity [19,60-63]. The same is true for genomic deletion and duplication syndromes, such as velocardiofacial syndrome and other deletions [64-69]. And, of course, there is constant interaction of the environment with a person, both prenatally and postnatally. As just one example, cretinism is related to a lack of iodine in the mother's diet, and there is incredibly variable expression of this illness based in part on the amount of iodine deficiency and how this interacts with fetal development [70].

The words 'penetrance' and 'expressivity' can be defined as:

- *Penetrance*: The number of individuals in a population carrying a disease predisposing allele that are also categorically defined as being affected by the associated disease.
- *Expressivity*: The extremeness, or number of symptoms, in the presentation of a disease in the context of individuals who have the associated disease predisposing allele."

Unfortunately, these two separate terms have led to a great deal of confusion in the field, and this sort of categorical thinking tends to miss complexity. Some use the word 'penetrance' when they really mean 'expressivity' of disease in any one person. As such, perhaps we should get rid of the two terms altogether and just discuss the expression of each trait in the context of a phenotypic spectrum, which is of course what led Walter Frank Raphael Weldon to establish the field of biometry [15,71,72]. Another way to express this point is to say that we have yet to characterize the full breadth of expression for virtually any mutation in humans, as we have not systematically sequenced or karyotyped any genetic alteration in thousands to millions of randomly selected people from a whole range of ethnic classes, i.e. clans [73,74]. There is an ongoing clash of world-views, with some

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

wanting to believe that single mutations predominately drive outcome while others are explicitly acknowledging the importance of substantial phenotypic modification via genetic background and/or environmental influence(s) [5,27,75-81]. Some recent population-based sequencing efforts have shown the complexity of demonstrating how much any one genetic variant contributes to disease in any one particular individual, and we disagree with overly simplistic and artificial categorizations of mutations as "causative", "pathogenic" or "non-pathogenic" [82-85].

It is very likely that there will be a continuum of disease, given that the 'effect size' of any particular mutation will obviously vary according to genetic background and environment, as demonstrated repeatedly in model organisms [75-78,86-93]. Thus, while a mutation associated with hemochromatosis or breast cancer might have high expression in one particular pedigree or clan, that same mutation may have very low expression in another pedigree, clan or group of unrelated people [94]. The reasons for variable expression can be myriad and are currently unknown in many instances; however, problems start to appear when scientists attempt to invoke a third allele as necessary and perhaps sufficient for the expression of any symptoms from within a typical disease. This disease model has been most clearly advocated for Bardet-Biedl Syndrome, in which the authors contend that some subjects have zero disease symptoms while possessing two autosomal recessive mutations in a known 'disease gene"; the authors also show that some affected people have a mutation in another gene, i.e. a third allele, which they speculate is necessary and perhaps sufficient for expression of any symptoms of the disease [95-97]. However, this model has been challenged by others [98-102], and at least one group maintains that all people that they have studied with two autosomal recessive mutations have some manifestations of disease but with variable expression, i.e. one person might only have retinitis pigmentosa whereas another person might have the full-blown symptoms of Bardet-Biedl syndrome [98]. One wonders whether the debate about triallelism might really just be a semantic one due to problems with the phenotyping of 'unaffected' people, particularly if these people were not evaluated longitudinally. Detailed online longitudinal characterizations of all such reportedly 'unaffected' people could aid in documenting, with some degree of certainty, that these people did indeed have zero symptoms of Bardet-Biedl syndrome, as that would then be further proof that mutations are not deterministic in any way at all. Said another way, this would be demonstration of the enormous variability in expression for mutations that do contribute more to a phenotype in some people with their own genetic backgrounds and environmental differences, and this observation ought to have dramatic implications for any ideas concerning prenatal diagnosis and 'prediction' of any genotype/phenotype relationship (discussed more below).

Surprisingly, a precise definition of the term 'oligogenic' is not apparent or consistent in the world literature. Some people have invoked the term 'oligogenic' to mean an interaction between mutations in two genes to somehow collectively 'cause' a disease, such as with this above case of triallelism in Bardet-Biedl syndrome [103]. These authors define oligogenic inheritance as occurring "when specific alleles at more than one locus affect a genetic trait by causing and/or modifying the severity and range of a phenotype" [103]. Another case in point involves the 22q11.2 locus, also known as velocardiofacial syndrome. This deletion does not involve only a single gene, but rather ~X number of genes, depending on the exact size of the deletion interval. The phenotypic manifestations can be incredibly heterogeneous, illustrated by the fact that some ~30% develop psychotic symptoms and get labeled as 'schizophrenic' [104]. Of course, heuristic diagnoses for schizophrenia are usually made based on certain semantic criteria, so it is likely that subthreshold symptoms are not counted (or perhaps not even detected). But, at least one has the advantage of knowing which people possess the deletion, allowing one to perform detailed phenotyping to determine whether subthreshold symptoms were missed within a family, and this has indeed been done in the case of a well-known translocation involving *DISC1* [105,106]. Unfortunately, genome-wide studies are not yet performed routinely for people with 'idiopathic schizophrenia', so it has been difficult to identify and group many people by genotype(s). As we discuss below, we believe that the routine clinical use of exome and eventually whole genome sequencing might finally enable this to occur, assuming that aggregation of genotype and phenotype data is allowed on a massive scale.

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

The definition of 'polygenic' literally means "many genes", including the combined effects of dozens (or perhaps even hundreds) of different mutations in different genes on a particular phenotype, although it is sometimes not very clear whether these multiple genes are meant to be spread across individuals or within individuals. We tend to favor the definition involving multiple mutations within the same individual somehow contributing toward phenotypic development. Height has historically been characterized as being a polygenic phenotype, with GWAS studies implicating the possible involvement of hundreds of loci [107,108]. Height is an easily measured phenotype and is generally described as being distributed continuously within human populations, modified of course by gender and ancestral backgrounds. If one looks at height in males or females of a certain ethnic background and from the same geographic locale, one can typically draw a semi-Gaussian (normal) function, but with tails that deviate from what is expected, encompassing rare cases of dwarfism and gigantism. We tend to also think that a single vertical measurement does not capture the true phenotypic variability involving height, as this measurement does not adequately capture the variability that exists in the many determinants of height (i.e., bone dimensions, age, environment, etc.). So, for a trait that seems conceptually simple to measure, there exists difficulty in uncovering its genetic component(s) due, in part, to uncharacterized uncertainty (variability) introduced at the phenotypic measurement level. If we now consider psychometrically defined traits, a large amount of further uncertainty is introduced at the phenotypic measurement level, as we are still unable to accurately characterize even a single measurement for most psychiatric disorders. These difficulties are underscored by the fact that psychiatric definitions are ephemeral and can change in a dramatic fashion over the course of even a few years. It seems premature to argue that schizophrenia, for example, is Gaussian in nature [109]. We would argue that we simply do not know enough about the phenotypic expression of the many different diseases that this amorphous concept of 'schizophrenia' encompasses to be able to make any conclusions regarding its genetic inheritance on a population or individual level [110]. Until there is substantial evidence to support another viewpoint, it is therefore important to treat each family as a special case. One must study people within families to determine whether some people in families have illness due to mutations with variable expression, modified by genetic background and environmental influences.

There have been numerous reviews concerning the ongoing debate for common and rare variants, with arguments made for various 'camps' of thought, including the common disease-common variant (CDCV) model, the infinitesimal model, the rare allele model and the broad sense heritability model [111]. Frankly, these models are simply semantic and reductionistic arguments that do not reflect the complexity of the human condition, and we are not sure that arguing for and against various models is useful, given that these models are basically straw men artificially constructed to be knocked down. This is very similar to the psychiatric literature in which several people decided, about 100 years ago, to introduce various names (or models) for certain diseases, such as the words 'schizophrenia' [112] and 'manic-depressive illness or bipolar' [113]. It is quite apparent to most clinicians that the phenotypic heterogeneity of these illnesses is so tremendous as to render these names basically moot and not particularly useful. This is akin to 50 years ago when people simply stated that someone had 'cancer'. Now, it is not useful to say only that someone has cancer, as there are literally hundreds of molecular etiologies for cancer, divided up not only by organ expression but also by specific pathways in the cell [114]. We anticipate that in 50 years, these terms 'schizophrenia' and 'bipolar' will be replaced by much more precise molecularly defined terms, as is occurring now in the cancer field [115,116]. Locus heterogeneity will likely play an important role in most diseases, but particularly in psychiatric disease, given the extensive phenotypic heterogeneity. Some of this complexity has been documented in reports of individual people [117-129], and a review by one of us of the literature related to schizophrenia [130] rendered the distinct impression that we really hardly know anything about the mechanistic basis of these many illnesses that we currently lump together as 'schizophrenia'. This is primarily due to overly broad descriptions and categorizations of these illnesses into these artificially named syndromes, despite the obvious heterogeneous and inconsistent nature of these categorizations. Remarkably, bipolar and schizophrenia have been artificially 'split' into different syndromes [131,132], in spite of the existence of a well document-

ed literature demonstrating overlap in at least some families with symptoms from both 'syndromes' [133].

Oddly enough, some diseases such as Fragile X, Rett Syndrome and other now molecularly defined disorders are sometimes removed from the 'nonsyndromic idiopathic autism' camp, leaving the remaining disorders still eligible for a semantic debate about which 'genetic model' they fit into [134]. One wonders if the same thing has occurred for velocardiofacial syndrome, with its relevance to schizophrenia, given the overwhelming evidence that the single 22q11.2 deletion event predisposes its carriers to some version of 'schizophrenia' with some exhibiting anywhere between 20 and 30% of the symptoms currently being defined as consistent with 'schizophrenia'[104]. All of these disorders were at one point labeled as 'idiopathic' until molecular lesions associated with them were identified. It has been known by at least some researchers and clinicians for quite some time that there are likely many minor physical anomalies in people labeled as 'nonsyndromic' [135,136], all of which is further proof of the substantial phenotypic expression differences of all disorders. Therefore, the dichotomous use of the words 'syndromic' and 'nonsyndromic' is completely artificial and does not reflect the reality or complexity of the situation in any one person.

A recent paper using exome sequencing to study hypertension pedigrees made the following statements: "These findings demonstrate the utility of exome sequencing in disease gene identification despite the combined complexities of locus heterogeneity, mixed models of transmission and frequent de novo mutation. Gene identification was complicated by the combined effects of locus heterogeneity, two modes of transmission at one locus, and few informative meioses. Many so far unsolved Mendelian traits may have similar complexities. Use of control exomes as comparators for analysis of mutation burden may be broadly applicable to discovery of such loci "[137]. This paper illustrates exactly what we are discussing above, in terms of the possible heterogeneity of many illnesses on many levels, making it impossible to predict (or even need) any particular model that may or may not fit the disease. It is far better to allow the data to speak for themselves.

De novo mutations, germline mosaicism and other complexities

Although this concept of somatic mosaicism has been in the literature for many years [138-142], it is really only recently that more people are beginning to realize that it might be much more extensive in humans than previously thought [25,143-158]. In fact, hardly anything is truly known regarding the extent of somatic mosaicism in humans and its effect on phenotype in even well studied diseases. For example, little is known regarding pathogenesis of the phenotype in people with trisomy 21 mosaicism and Down syndrome, although there is likely variation in phenotype associated with the percentage of trisomic cells and their tissue-specificity [159-161]. A more recent study looked at this issue of somatic mosaicism in Timothy syndrome type 1 (TS-1), which is a rare disorder that affects multiple organ systems and has a high incidence of sudden death due to profound QT prolongation and resultant ventricular arrhythmias. All previously described cases of TS-1 are associated with a missense mutation in exon 8A (p.G406R) of the L-type calcium channel gene (Ca(v)1.2, *CACNA1C*). Most cases reported in the literature represent highly affected people who present early in life with severe cardiac and neurological manifestations, but these authors found somatic mosaicism in people with TS-1 with less severe manifestations than the typical person with TS-1 [162]. There are therefore likely large ascertainment biases, given that people with subtler phenotypes are likely not coming to anyone's attention. The implications of these findings with somatic mosaicism are that one cannot currently predict phenotype from genotype, particularly in the absence of any comprehensive characterization of which tissues are mutated in any one person. Also, putative 'de novo' mutations can instead represent cases of parental mosaicism (including in the germline), which could be revealed by careful genotyping of parental tissues other than peripheral blood lymphocytes. In fact, we are increasingly becoming aware of many instances of germline mosaicism, in which a mutation is not present or is present only at a very low level in the blood sample from a parent, but clearly

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

must be in their germline, as they have two or more children with the same mutation that must therefore have originated through the parent's germline [163-183]. Clearly, we are truly ignorant concerning the extent of diversity brought about by somatic mosaicism, and it is therefore far too simplistic to assume that a single blood draw truly represents the entire genome of a human being, with anywhere from 25-100 trillion cells in their body divided up among multiple organs and other tissue systems. Of course, even the words "whole genome sequencing" are misleading, as there might very well be millions to trillions of similar (but not the exact same) genomes in each person's body .

Rare and compensatory mutations

There is an increasingly rich literature regarding rare mutations with seemingly large phenotypic effects [184-187]. An example of this is Liam Hoekstra, known as the world's strongest toddler when he was age 3, and who has an extremely rare mutation in the gene encoding myostation, leading to myostatin-related muscle hypertrophy with increased muscle mass and reduced body fat [188]. However, the effects of these mutations have mainly been reported in the context of particular genetic backgrounds, and so our knowledge of the expression of these mutations in the context of any number of genetic backgrounds is lacking. It is likely that there can be, and are, many genomic elements that act in concert to influence these traits in a phenotypic spectrum. Of course, compensatory mutations can be explored in the context of other organisms [189-191], but human migration and breeding is certainly not something that can be experimentally manipulated!

There are many disabling psychiatric syndromes, which have been lumped under certain artificial categories, such as schizophrenia, Tourette Syndrome (TS), obsessive compulsive disorder (OCD), and attention deficit hyperactivity disorder (ADHD). A very good way forward is to study these syndromes in large families living in the same geographic region, so as to control for ancestry differences, minimize environmental influences, and focus on specific genotypes in these families. It is possible that a low number of genetic mutations will be shared in a relatively small combination (on the order of 1-3 such variants) among affected relatives within some pedigrees, and that these variants will not be present in the same combination in unaffected relatives or in other families with very little to no neuropsychiatric disorders [25,110,192-196]. An alternative is that some affected people in these families have these illnesses due to additive and/or epistatic interactions among dozens to hundreds of loci within each person [197-199]. The currently classified syndromes of schizophrenia, obsessive compulsive disorder (OCD), attention deficit hyperactivity disorder (ADHD), autism and other mental illnesses are quite heterogeneous within and between families, and these symptoms have also been observed in known single locus disorders such as Fragile X and 22q11.2 velocardiofacial syndrome [110,193].

Some of these syndromes are referred to as 'complex' diseases simply because the presentation is so incredibly heterogeneous that is it very likely that there will be multiple different genetic and environmental explanations. One possible genetic explanation is that some symptoms of severe mental illness may emerge in a particular family due to a genetic constellation including dozens to hundreds of loci acting in each person either additively or via epistasis (and possibly modified by environment; G X E), which some refer to as the 'polygenic' model [197-200], as previously discussed. If true, for predictive efforts in any particular family, the solution will ultimately require whole genome sequencing to tease out the numerous mutations involved. On the other hand, some discuss this concept of "many rare variants of large effect", which they refer to as the 'oligogenic' model of inheritance [201,202], as previously discussed. Some families have deleterious copy number variants [193,203-206], and de novo single nucleotide mutations have recently been implicated as important for spontaneous 'singleton' cases in at least some families [207-212]. There could also be a set of families with single, pair or triplet interactions among 1-3 gene mutations of high expression that can largely, on their own, contribute to a set of symptoms currently overlapping with named syndromes, such as 'autism' and 'schizophrenia' [213]. As there is no way of really distinguishing between these two artificially created models in any one

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

particular family, it is reasonable (with current costs) to perform microarray genotyping and whole genome sequencing as a comprehensive way to ascertain most of the relevant genetic variance in any particular family.

It is becoming generally accepted that at least 5% of the 'autisms' appear to be associated with various large copy number variants [214]. So, it is likely that some additional portion of the 'autisms' will be influenced by other types of mutations, with some evidence pointing to a role for 'de novo' mutations in singleton, uninherited cases of autism [208-211,215] and other evidence suggesting that there might be multiple genetic and environmental influences in each person [197].

Current ability / approaches

There has been an explosive growth in exome and whole genome sequencing (WGS) [25], led in part by dramatic cost reductions. The same is true for genotyping microarrays, which are becoming increasingly denser with various markers while maintaining a relatively stable cost [216]. With rapid advancements in sequencing technologies [217] and improved haplotype-phasing[218,219], high-throughput sequencing (HTS) data on the genomes of a diverse number of species are being generated at an unprecedented rate. The development of bioinformatics tools for handling these data has been somewhat lagged in response, creating a gap between the massive data being generated, and the ability to fully exploit their biological content. Many short read alignment software tools are now available, along with several single nucleotide variants (SNVs) and copy number variant (CNVs) calling algorithms [25]. However, there is a paucity of methods that can simultaneously handle a large number of genetic variants and annotate their functional impacts (particularly for a human genome, which typically hosts >3 million variants), despite the fact that this is an important task in many sequencing applications. Functional interpretation of genetic variants therefore becomes one of the major obstacles to connect sequencing data with biomedical researchers who are willing to embrace the sequencing technology.

In the medical world, WGS has since led to the discovery of the genetic basis of Miller Syndrome [220] and in another instance, it was used to investigate the genetic basis of Charcot-Marie-Tooth neuropathy [221], alongside a discussion of the 'return of results' [222]. In 2011, the diagnosis of a pair of twins with dopa (3,4-dihydroxyphenylalanine) responsive dystonia (DRD; OMIM #128230) and the discovery that they carried compound heterozygous mutations in the SPR gene encoding sepiapterin reductase led to supplementation of l-dopa therapy with 5-hydroxytryptophan, a serotonin precursor, resulting in clinical improvements in both twins [223].

Despite current technological limitations, mutations are continually being identified in research settings [220,224-228]. However, the human genomics community has recognized a number of distinct challenges, including with phenotyping, sample collection, sequencing strategies, bioinformatics analysis, biological validation of variant function, clinical interpretation and validity of variant data, and delivery of genomic information to various constituents [25,229]. In particular, there is a need for large pedigree sample collection, high-quality sequencing data acquisition, rigorous generation of variant calls, and comprehensive functional annotation of variants [25,230,231]. Empirical estimates seem to suggest that exome sequencing can identify a putative disease associated variant in only about 10-50% of the cases for which it is applied [25], and the genetic architecture of most neuropsychiatric illness is still largely undefined and controversial [110,192,197,198]. The sequencing of entire genomes in large families will create a dataset that can be analysed and re-analysed for years to come as new biology and new methods emerge. The cost of a whole genome will likely decrease much more rapidly in relation to the cost of exome sequencing, given the relatively fixed labor and reagent costs for capturing the exons in the genome. Also, there is emerging evidence that exon capture and sequencing only achieves high depth of sequencing coverage in about 90% of the exons, whereas WGS does not involve a capture step and thus obtains better coverage on >95% of all exons in the genome. Of course, even the definition of the exome is a moving target, as the research community is constantly annotating and finding new exons not

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

previously discovered [232,233], and therefore WGS is a much more comprehensive way to assess coding and non-coding regions of the genome.

It is obvious that in both research and clinical settings WGS can dramatically impact clinical care, and it is now a matter of economics and feasibility in terms of WGS being adopted widely in a clinical setting [25,231]. There are, however, still many challenges in showing how any one mutation can contribute toward a clear phenotype, particularly in the context of genetic background and possible environmental influences [234]. Bioinformatics confounders, such as poor data quality [235], sequence inaccuracy, and variation introduced by different methodological approaches [236] can further complicate biological and genetic inferences. Furthermore, one cannot exclude polygenic and epistatic modes of inheritance [96,237-242]. To address these issues, future work will need to focus on evaluating next generation sequencing data coming from multiple sequencing and informatics platforms, and involving multiple other family members. By using a combination of data from many family members and from different sequencing technologies evaluated by a number of bioinformatics pipelines, we can maximize accuracy and thus the biological inference stemming from these data.

Prenatal diagnosis, preimplantation genetic diagnosis/screening

"Before a new function can arise, it may be essential for a lineage to evolve a potentiating genetic background that allows the actualizing mutation to occur or the new function to be expressed. Finally, novel functions often emerge in rudimentary forms that must be refined to exploit the ecological opportunities. This three-step process—in which potentiation makes a trait possible, actualization makes the trait manifest, and refinement makes it effective—is likely typical of many new functions." – Richard Lenski [92]

A great clinical geneticist, John Opitz, has observed the following: "More fetuses die prenatally than are born alive. Many die because of genetic conditions, malformations, and syndromes. Most are not autopsied, and in such cases appropriate genetic counseling is not provided or possible. In such 'cases' (fetuses, infants) a huge amount of genetic pathology is yet to be discovered (our last frontier!)" [243].

In this regard, some have suggested a canalization model, which describes phenotypes as being robust to small perturbations, seemingly stuck within "phenotypic canals". Phenotypes may 'slosh' against the sides of the canal during development, but with little effect on the final outcome of development [244-246]. In such a model, it is only perturbations with a magnitude exceeding a certain threshold that can direct the developmental path out of the canal (see Figure 1 for an illustrative model of canalization). Accordingly, phenotypes are robust up to a limit, with little robustness beyond this limit. This pattern may increase rates of evolution in fluctuating environments, as phenotypes are more likely to be perturbed with increased frequency and magnitude, thus leading to

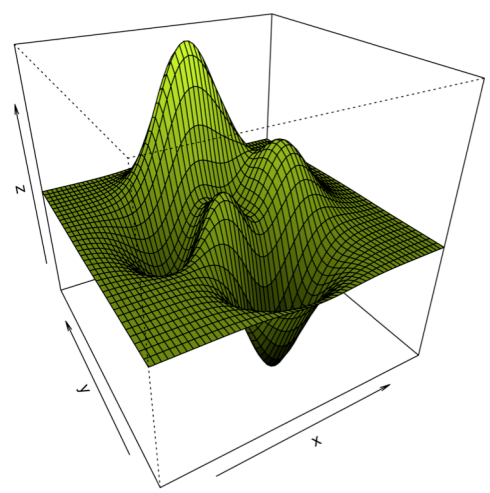

Figure 1. A conceptual model of canalization. The *y* plane represents a phenotypic spectrum, the *x* plane represents the canalized progression of development through time, and the *z* plane represents environmental fluctuations. As any particular phenotype progresses through development, it can encounter environmental fluctuations that either repel (a local maximum) or attract (a local minimum) its developmental path. Either force, if strong enough, can cause a shift in the developmental path, fundamentally altering the end resulting phenotype.

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

more rapid delineations and differentiations of canalized phenotypes.

One could argue that the birth of a child in one particular famliy with a clear phenotype, such as cystic fibrosis, along with previously identified associated mutations, dramatically increases the 'prior probability' that a future child with these same mutations being born in that same family would have a similar 'canalized' phenotype. It is really only in that particular situation in which one could make a somewhat informed prediction of genotype going down one particular phenotypic "canal". And yet, a study in Australia from 2000-2004 showed that of the 82 children born with cystic fibrosis (CF) in Victoria, Australia, 5 (6%) were from families with a known history of CF. The authors found that "even when a family history is known, most relatives do not undertake carrier testing. In an audit of cascade carrier testing after a diagnosis of CF through newborn screening, only 11.8% of eligible (non-parent) (82/716) relatives were tested" [247]. These same researchers also showed that in a clinical setting, the diagnosis of a baby with CF by newborn screening "does not lead to carrier testing for the majority of the baby's non-parent relatives" [26]. This is incredibly unfortunate, given that predictions of any reliability ought to include the prior probability of someone being born in that 'ancestry group' with the mutations and phenotype of interest.

Despite the above facts, non-invasive sequencing of fetal genomes is an area of intense interest in genomic medicine, and a cynical person might argue that the rush to implement this technology is driven mainly by financial interests. Current techniques are based on the observation that a small proportion of the cell-free DNA in a pregnant woman's blood is derived from the fetus, so that aneuploidy or genomic sequence of a fetus may be inferred by sequencing of maternal plasma DNA and algorithmic decoupling of maternal and fetal DNA variants. A few companies are already marketing non-invasive prenatal screening (NIPS) tests for non-invasive detection of trisomy 21 associated with Down's syndrome [248]. One can reasonably argue that detecting Down's syndrome is a conceptually and practically much simpler task than detecting individual variants within the fetal genome to assess mutations associated with disorders such as cystic fibrosis and hearing loss. However, with sufficiently high sequence depth, it is technically feasible to detect single nucleotide alterations in a fetal genome, as shown in several recent papers [249-252]. But, to allow accurate detection of individual variants, very high sequencing depth is required (potentially hundreds-fold higher than sequencing germline genomes); therefore, it is likely that targeted exon capture and sequencing might dominate the market until sufficiently high depth whole-genome sequencing becomes an economically feasible alternative. Given these technological developments, it is likely that some form of fetal genome testing will be available in the next few years. Others have noted that we might be reaching a point in the near-term future where it may be feasible to incorporate genetic, genomic and transcriptomic data to develop new approaches to fetal treatment [253,254]. One concern is that greed and financial conflicts of interest could lead to indiscriminate marketing and use of NIPS as diagnostic tests, rather than simply as screening, and that this technology will be implemented without any regard for genetic background or environmental differences, alongside a complete misunderstanding of this concept of extreme variability in phenotypic expression.

Implications for acceptance, prognosis and treatment

"When a complex system starts to dysfunction, it is generally best to fix it early. The alternative often means delaying until the system has degenerated into a disorganized, chaotic mess — at which point it may be beyond repair. Unfortunately, the general approach to cancer has ignored such common sense. The vast majority of cancer research is devoted to finding cures, rather than finding new ways to prevent disease" – Michael Sporn [114].

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

Prevention of illness through environmental modification has been, and likely always will be, the major driver for global health [114,116]. With this in mind, the sequencing of whole genomes on a large scale promises to enable the discovery and prediction of disease in some people. The ability to sequence an infant at birth and to be able to predict a higher probability of certain phenotypes, such as developmental delay, would allow for educational and behavioral interventions to influence the phenotype, thus altering the trajectory of that phenotype [255-260]. One recent study of chromosomal microarray (CMA) testing found that "among 1792 patients with developmental delay (DD), intellectual disability (ID), multiple congenital anomalies (MCA), and/or autism spectrum disorders (ASD), 13.1% had clinically relevant results, either abnormal (n = 131; 7.3%) or variants of possible significance (VPS; n = 104; 5.8%). Abnormal variants generated a higher rate of recommendation for clinical action (54%) compared with VPS (34%; Fisher exact test, P = 0.01)" [261]. The authors concluded that "CMA results influenced medical management in a majority of patients with abnormal variants and a substantial proportion of those with VPS" thus supporting the use of CMA in this population [261]. We agree that the identification of certain CNVs and other mutations can suggest a range of phenotypes that might occur in any one individual with that mutation or mutations.

However, there are some major barriers to the widespread implementation of genomic medicine in the clinic. These include:

1) Lack of public education
2) Lack of physician knowledge about genetics
3) Apathy on the part of the populace in terms of preventive efforts
4) Refusal of insurance companies and governments to pay for genetic testing
5) Focus in our society on treatment, not on early diagnosis and prevention
6) Privacy concerns
7) Limits of our current knowledge

The emphasis should be on diagnosis and prevention, not just on treatment. During the medical training of one of the authors (GJL), two episodes helped to illustrate this. The first involved a 15-year old girl with Type I diabetes, who was hospitalized dozens of times with diabetic ketoacidosis. Literally hundreds of thousands of dollars were spent to repeatedly save her life, but very little time or money was spent on therapy or education to teach her about taking her insulin and ensuring that she did. Unfortunately, in America at least, this is due to a relative lack of reimbursement for such activities, whereas saving someone already in diabetic ketoacidosis is quite lucrative to everyone involved. A second episode involved a 14-year old boy, who had been hospitalized well over 10 times with acute pancreatitis over a ten year period, with very little thought concerning why he had recurring pancreatitis. Finally, someone obtained a genetics consult, and they recommended cystic fibrosis (CF) genetic screening, which had never been ordered before due to a prior 'negative' sweat test. It turns out that this boy had two rare mutations in *CFTR*, undiagnosed till then, which had been contributing to recurrent pancreatitis. He had never had any lung manifestations, and he had never had a positive sweat test for CF, mainly due to the fact that these mutations appeared to only be exerting effects in his pancreas, not in his skin or lungs. After this diagnosis, this person benefited from pancreatic enzyme supplementation, along with therapy and education. Once again, the reason it took so long to diagnose this person is because the incentive structure in many developed nations is not on early diagnosis and prevention, but rather on treatment of people only once they become severely ill [262,263]. This is illustrated by the fact that there are only about ~1000 medical geneticists in America and ~3000 genetic counselors, for a population of ~315 million, which makes it basically impossible for these limited number of professionals to implement genomic medicine in any meaningful way [264]. The numbers of such health care professionals are even smaller in developing regions of the world, thus making it currently very difficult to provide widespread genetic counseling [73,265,266]. Stepping into this void are direct-to-consumer for-profit genetic testing companies, and this is certainly one disruptive way of trying to help people manage their genetic results online [267,268], although financial motives and lack of transparency can create problems [269].

Privacy concerns have added to the difficulties of implementing genomics-guided medicine. Genetic data have the potential of being informative across a wide variety of human traits and health conditions, and some worry about the potential

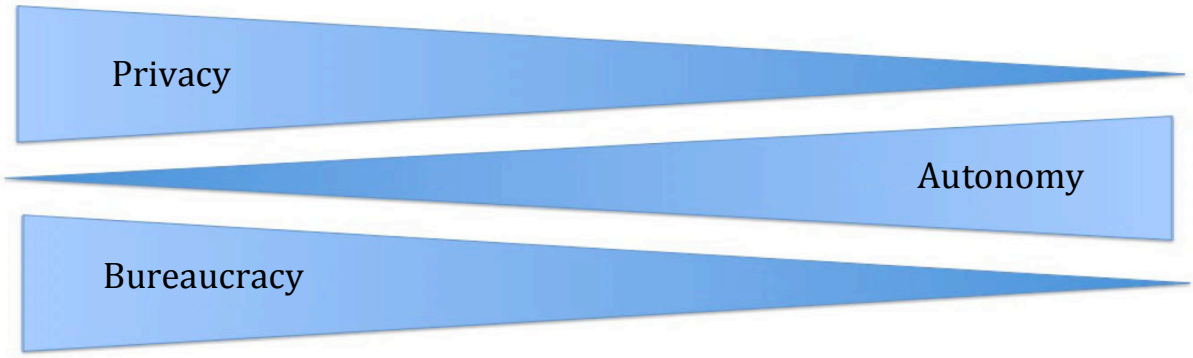

Figure 2. An illustration of the tradeoff between privacy and autonomy when receiving results from genetic testing. Models that guarantee an increased level of privacy are generally accompanied by a great deal of bureaucratic and paternalistic decision-making on the part of medical and advisory institutions (left). Models that propose and advocate for increased autonomy when receiving genetic test results come with the risk of reduced privacy (right). A whole genome sequence from a single person could, in principle, inform many aspects of his/her health care as well as allow for the prospect of future health predictions. This leads to speculations on how insurance agencies and health care providers could/would use this information. One can envision a 'sinister scenario' where people are rejected from hospitals and denied insurance based on putative genetic aberrations that may associate with costly, long term, care. Others worry about the potential implications of results found by genome scale testing, and would rather not know about risks pertaining to untreatable illnesses. Recent movements push for the democratization as well as large-scale adoption of this type of testing for every person, which could help to prove that we are all truly genetically unique and all carry any number of mutations and/or large genetic aberrations that may or may not be associated with disease. In reality, current technologies are far from the realm of genotype to phenotype predictions, and so genetic discrimination could only create illusory economic gains for any institution for the foreseeable future.

misuse of these data by insurance agencies as well as by health care providers [270]. Genetic testing has historically been focused on targeting and examining a small number of known genetic aberrations [271]; however, since the advent of high-throughput sequencing technologies, the landscape is starting to change. With the emergence of tests that can target and examine all coding regions of the genome, or even the genome in its entirety [272], testing can now be performed on a more global and exploratory scale. Some people worry about returning the results of such a test, whose findings can have questionable clinical significance, and in response have advocated for selectively restricting the returnable medical content. Others have proposed complicated anonymization techniques that could allow for a safe return of research results to participants whose genome is suspect to contain 'clinically actionable' information. One such proposition involves the cryptographic transformation of genomic data in which only by the coalescence of keys held by many different intermediate parties would the identity of the participant be revealed, and only in cases where all parties agree that there is indeed the presence of clinically actionable information [273]. These types of recommendations take a more paternalistic approach in returning test results to people, and generally involve a deciding body of people that can range in size from a single medical practitioner to a committee of experts. In contrast, there is a growing movement among the populace to learn more about their own 'personalized' health and health care. There has also been a renewed push for the unfiltered sharing and networking of health related data, which has been facilitated and hastened by the explosion of digitally mediated social networking over the past decade, as well as by private institutions such as 23andMe [274] and the Personal Genomes Project [275] that aim to popularize and democratize genetic testing. Clearly, between these contrasting approaches, there is a tradeoff between the privacy and personal safety one can expect to retain by either freely acquiring and sharing the full breadth of one's genetic testing data, or by allowing deciding bodies to choose what information you will receive.

Public databases containing human sequence data have grown in magnitude and in number, and relatively comprehensive sequencing data have already been generated and published on thousands of people [276,277]. Similar privacy concerns have since been expressed about the degree of medical and personal privacy that these and other research participants can expect [278], given that each person is genetically unique. As a demonstration of current vulnerabilities, researchers have shown that the identities of participants can be discovered using these publicly available data [279]. Although these data have been instrumental in furthering our understanding of human genetics, medicine, and biological processes in general, some advocate for caution when sharing and publishing human genetic sequence information [280].

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

As the cost and difficulty of sequencing continually decreases, a wealth of data are becoming available to researchers, privately funded institutions and individual consumers. More people are willing to share a larger portion of their personal life in the public arena, and we fully expect that, given the popularization of 'personalized' genomic health related data, more people will want to share these data and offer their own DNA sequence for others to explore. There is a trade off between the risks inherent in sharing vast quantities of health data, and maintaining personal privacy in the burgeoning age of personalized medicine and genomics. As the technology and science mature, our power to interpret and use these health data for practical and preventative measures will certainly improve. Conventions for privacy and autonomy will likely be driven by popular demand, and could vary from person to person, as all people differ in their desire for privacy and autonomy (see Figure 2 for a conceptual model of this tradeoff).

In addition, within the current paradigm of genetic determinism, which stretches back to the time of William Bateson [281,282], some people would have us believe that variants can and should be binned into different classes based on clinical utility and validity [283-285], without any obvious regard to genetic background or environmental differences. Environment and ancestry matter [2,3,281,282], and yet some clinical geneticists trained in the current paradigm of genetic determinism clearly do not wish to acknowledge this. Categorical thinking misses complexity. In fact, one medical academy in America recently released guidelines in which they recommended the "return of secondary findings" for only 57 genes, without any real guidance for the rest of the genome or environmental influences [286]. This is therefore a very conservative set of recommendations, given that there are approximately 20,000 protein-coding genes in the human genome, along with the thousands of other identified, important noncoding elements of the genome [9,30-44]! As stated above, but worth repeating, there are 6 billion nucleotides of DNA in every cell of the human body, and there are 25-100 trillion cells in each human body. Given genetic modifiers, somatic mosaicism, epigenetic changes, and environmental differences, no two human beings are the same, and therefore the expression of any mutation will be different in each person. At best, phenotypes will follow canalized pathways in direct relatives, such as mother and child, so the analysis of mutations over several generations in the same families is a worthwhile effort. But, how we will ever get to a world of millions of whole genomes shared and analyzed for numerous additive, epistatic interactions and gene by environment interactions, so that we can make any reliable predictions for any one human being, if we are only recommending 'return of results' from ~57 genes? We need to sequence and collate online the raw exome and genome data and phenotypic information from thousands and then millions of people, so that we can actually begin to really understand the expression patterns of any mutation in the human genome in particular families. In medicine, people tend to create illusions of certainty, when in fact everything is probabilistic [14]. Some humans like to be told things in a 'yes/no' manner, but there always exists a degree of unresolvable uncertainty.

Conclusions

"A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it." -- Max Planck

With the advent of exome and whole genome sequencing, we need to focus again on families over several generations, so as to attempt to minimize genetic differences, locus heterogeneity and environmental influences. Forging strong ties with families will also enable access to other tissues to continue to study newly discovered loci with many emerging technologies. Some might consider it to be 'social activism' to advocate for a more comprehensive collection and collation of human pedigrees, whole genome sequencing data and phenotypic information. But, in the words of one

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

author: "Scientists, whether we like it or not, are members of society, and we are prone to the ideas and beliefs of the times in which we live [287]." We currently live within a paradigm of genetic determinism, but we should not be forever condemned to this simplistic mode of thinking. One can imagine or hope that in the not too distant future, each person will be able to keep track of detailed longitudinal phenotyping data on themselves online, and they will be able to link this to records of their relatives, both living and deceased. One can also hope that we are approaching a time where sufficient information is available within many large families for calculating highly accurate probabilistic outcomes [14-18], at which point we might be able to more effectively alter the trajectory for many diseases. One can see this beginning already to occur in certain geographically isolated clans, such as in Iceland [184,185], so there is some optimism that this can indeed occur on a global level, including in the currently less developed regions of the world [266].

Acknowledgements

We thank the editor, Kevin Mitchell, for detailed comments and suggestions regarding an earlier draft of this manuscript. Others who have made very helpful suggestions include: Anne Buchanan, George Church, Nathaniel Comfort, Jesse Gillis, Nathaniel Pearson, William Provine, Gregory Radick, Michael Stone, and Kai Wang. We also acknowledge members of the Lyon laboratory, particularly Han Fang and Max Doerfel, for their helpful suggestions as well. We would also like to thank Gail Sherman at the CSHL library for her efforts to procure some of the older literature on our behalf.

References

1 Seguin, E. *Idiocy : and its treatment by the physiological method*. (William Wood & Co., 1866).
2 Weldon, W. F. R. Mendel's laws of alternative inheritance in peas. *Biometrika* **1**, 228-254 (1902).
3 Radick, G. Physics in the Galtonian sciences of heredity. *Stud Hist Philos Biol Biomed Sci* **42**, 129-138, doi:10.1016/j.shpsc.2010.11.019 (2011).
4 Bearn, A. G. *Archibald Garrod and the individuality of Man*. (Clarendon Press; Oxford University Press, 1993).
5 Comfort, N. C. *The science of human perfection : how genes became the heart of American medicine*. (Yale University Press, 2012).
6 Tennessen, J. A., Bigham, A. W., O'Connor, T. D., Fu, W., Kenny, E. E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., Kang, H. M., Jordan, D., Leal, S. M., Gabriel, S., Rieder, M. J., Abecasis, G., Altshuler, D., Nickerson, D. A., Boerwinkle, E., Sunyaev, S., Bustamante, C. D., Bamshad, M. J. & Akey, J. M. Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. *Science*, doi:10.1126/science.1219240 (2012).
7 Grillo, E., Lo Rizzo, C., Bianciardi, L., Bizzarri, V., Baldassarri, M., Spiga, O., Furini, S., De Felice, C., Signorini, C., Leoncini, S., Pecorelli, A., Ciccoli, L., Mencarelli, M. A., Hayek, J., Meloni, I., Ariani, F., Mari, F. & Renieri, A. Revealing the complexity of a monogenic disease: rett syndrome exome sequencing. *PLoS One* **8**, e56599, doi:10.1371/journal.pone.0056599 (2013).
8 Misteli, T. The cell biology of genomes: bringing the double helix to life. *Cell* **152**, 1209-1212, doi:10.1016/j.cell.2013.02.048 (2013).
9 Sabin, L. R., Delas, M. J. & Hannon, G. J. Dogma derailed: the many influences of RNA on the genome. *Mol Cell* **49**, 783-794, doi:10.1016/j.molcel.2013.02.010 (2013).
10 Scriver, C. R. The PAH gene, phenylketonuria, and a paradigm shift. *Hum Mutat* **28**, 831-845, doi:10.1002/humu.20526 (2007).
11 Allchin, D. The dilemma of dominance. *Biol Philos* **20**, 427-451, doi:Doi 10.1007/S10539-005-2561-Z (2005).
12 Terwilliger, J. D. & Weiss, K. M. Confounding, ascertainment bias, and the blind quest for a genetic 'fountain of youth'. *Ann Med* **35**, 532-544 (2003).
13 Weiss, K. M. & Terwilliger, J. D. How many diseases does it take to map a gene with SNPs? *Nat Genet* **26**, 151-157, doi:10.1038/79866 (2000).
14 Gigerenzer, G. *Calculated risks : how to know when numbers deceive you*. (Simon & Schuster, 2002).

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

15 Sokal, R. R. *Biometry : the principles and practice of statistics in biological research*. [Extensively rev.] 4th edn, (W.H. Freeman, 2012).

16 Gigerenzer, G. & Galesic, M. Why do single event probabilities confuse patients? *Bmj* **344**, e245, doi:10.1136/bmj.e245 (2012).

17 Gigerenzer, G., Wegwarth, O. & Feufel, M. Misleading communication of risk. *Bmj* **341**, c4830, doi:10.1136/bmj.c4830 (2010).

18 Kurz-Milcke, E., Gigerenzer, G. & Martignon, L. Transparency in risk communication: graphical and analog tools. *Ann N Y Acad Sci* **1128**, 18-28, doi:10.1196/annals.1399.004 (2008).

19 Freund, J., Brandmaier, A. M., Lewejohann, L., Kirste, I., Kritzler, M., Kruger, A., Sachser, N., Lindenberger, U. & Kempermann, G. Emergence of individuality in genetically identical mice. *Science* **340**, 756-759, doi:10.1126/science.1235294 (2013).

20 Bateson, W. & Mendel, G. *Mendel's principles of heredity a defense*. (University Press, 1902).

21 Olby, R. The dimensions of scientific controversy: the biometric--Mendelian debate. *Br J Hist Sci* **22**, 299-320 (1989).

22 Provine, W. B. *The origins of theoretical population genetics*. 2nd edn, (University of Chicago Press, 2001).

23 Nielsen, M. A. *Reinventing discovery : the new era of networked science*. (Princeton University Press, 2012).

24 Lyon, G. J. & Segal, J. P. Practical, ethical and regulatory considerations for the evolving medical and research genomics landscape. *Applied & Translational Genomics*, doi: http://dx.doi.org/10.1016/j.atg.2013.02.001, (2013).

25 Lyon, G. J. & Wang, K. Identifying disease mutations in genomic medicine settings: current challenges and how to accelerate progress. *Genome Med* **4**, 58, doi:10.1186/gm359 (2012).

26 McClaren, B. J., Metcalfe, S. A., Aitken, M., Massie, R. J., Ukoumunne, O. C. & Amor, D. J. Uptake of carrier testing in families after cystic fibrosis diagnosis through newborn screening. *Eur J Hum Genet* **18**, 1084-1089, doi:10.1038/ejhg.2010.78 (2010).

27 Keller, E. F. *The mirage of a space between nature and nurture*. (Duke University Press, 2010).

28 Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A. & Bejerano, G. Enhancers: five essential questions. *Nat Rev Genet* **14**, 288-295, doi:10.1038/nrg3458 (2013).

29 Dickel, D. E., Visel, A. & Pennacchio, L. A. Functional anatomy of distant-acting mammalian enhancers. *Philos Trans R Soc Lond B Biol Sci* **368**, 20120359, doi:10.1098/rstb.2012.0359 (2013).

30 Cartault, F., Munier, P., Benko, E., Desguerre, I., Hanein, S., Boddaert, N., Bandiera, S., Vellayoudom, J., Krejbich-Trotot, P., Bintner, M., Hoarau, J. J., Girard, M., Genin, E., de Lonlay, P., Fourmaintraux, A., Naville, M., Rodriguez, D., Feingold, J., Renouil, M., Munnich, A., Westhof, E., Fahling, M., Lyonnet, S. & Henrion-Caude, A. Mutation in a primate-conserved retrotransposon reveals a noncoding RNA as a mediator of infantile encephalopathy. *Proc Natl Acad Sci U S A* **109**, 4980-4985, doi:10.1073/pnas.1111596109 (2012).

31 Salzman, J., Gawad, C., Wang, P. L., Lacayo, N. & Brown, P. O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* **7**, e30733, doi:10.1371/journal.pone.0030733 (2012).

32 Batista, P. J. & Chang, H. Y. Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152**, 1298-1307, doi:10.1016/j.cell.2013.02.012 (2013).

33 Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K. & Kjems, J. Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384-388, doi:10.1038/nature11993 (2013).

34 Kapusta, A., Kronenberg, Z., Lynch, V. J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M. & Feschotte, C. Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. *PLoS Genet* **9**, e1003470, doi:10.1371/journal.pgen.1003470 (2013).

35 Khoddami, V. & Cairns, B. R. Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nat Biotechnol*, doi:10.1038/nbt.2566 (2013).

36 Ledford, H. Circular RNAs throw genetics for a loop. *Nature* **494**, 415, doi:10.1038/494415a (2013).

37 Maxmen, A. RNA: The genome's rising stars. *Nature* **496**, 127-129 (2013).

38 Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S. D., Gregersen, L. H., Munschauer, M., Loewer, A., Ziebold, U., Landthaler, M., Kocks, C., le Noble, F. & Rajewsky, N. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333-338, doi:10.1038/nature11928 (2013).

39 Mercer, T. R. & Mattick, J. S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat Struct Mol Biol* **20**, 300-307, doi:10.1038/nsmb.2480 (2013).

40 Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J. O. & Lai, E. C. Widespread and extensive lengthening of 3' UTRs in the mammalian brain. *Genome Res* **23**, 812-825, doi:10.1101/gr.146886.112 (2013).

41 Moreau, M. P., Bruse, S. E., Jornsten, R., Liu, Y. & Brzustowicz, L. M. Chronological Changes in MicroRNA Expression in the Developing Human Brain. *PLoS One* **8**, e60480, doi:10.1371/journal.pone.0060480 (2013).

42 Ning, S., Wang, P., Ye, J., Li, X., Li, R., Zhao, Z., Huo, X., Wang, L. & Li, F. A global map for dissecting phenotypic variants in human lincRNAs. *Eur J Hum Genet*, doi:10.1038/ejhg.2013.7 (2013).

43 Perrat, P. N., DasGupta, S., Wang, J., Theurkauf, W., Weng, Z., Rosbash, M. & Waddell, S. Transposition-driven genomic heterogeneity in the Drosophila brain. *Science* **340**, 91-95, doi:10.1126/science.1231965 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

44 Wilusz, J. E. & Sharp, P. A. Molecular biology. A circuitous route to noncoding RNA. *Science* **340**, 440-441, doi:10.1126/science.1238522 (2013).

45 Down, J. L. Observations on an ethnic classification of idiots. 1866. *Ment Retard* **33**, 54-56 (1995).

46 Hume, D. & Selby-Bigge, L. A. *A treatise of human nature*. (Clarendon press, 1896).

47 Fins, J. J. Deep brain stimulation, deontology and duty: the moral obligation of non-abandonment at the neural interface. *J Neural Eng* **6**, 050201, doi:10.1088/1741-2552/6/5/050201 (2009).

48 Corleto, V. D., Gambardella, S., Gullotta, F., D'Apice, M. R., Piciucchi, M., Galli, E., Lucidi, V., Novelli, G. & Delle Fave, G. New PRSS1 and common CFTR mutations in a child with acute recurrent pancreatitis, could be considered an "Hereditary" form of pancreatitis ? *BMC Gastroenterol* **10**, 119, doi:10.1186/1471-230X-10-119 (2010).

49 Derikx, M. H. & Drenth, J. P. Genetic factors in chronic pancreatitis; implications for diagnosis, management and prognosis. *Best Pract Res Clin Gastroenterol* **24**, 251-270, doi:10.1016/j.bpg.2010.02.001 (2010).

50 Rosendahl, J., Landt, O., Bernadova, J., Kovacs, P., Teich, N., Bodeker, H., Keim, V., Ruffert, C., Mossner, J., Kage, A., Stumvoll, M., Groneberg, D., Kruger, R., Luck, W., Treiber, M., Becker, M. & Witt, H. CFTR, SPINK1, CTRC and PRSS1 variants in chronic pancreatitis: is the role of mutated CFTR overestimated? *Gut* **62**, 582-592, doi:10.1136/gutjnl-2011-300645 (2013).

51 Emond, M. J., Louie, T., Emerson, J., Zhao, W., Mathias, R. A., Knowles, M. R., Wright, F. A., Rieder, M. J., Tabor, H. K., Nickerson, D. A., Barnes, K. C., Gibson, R. L. & Bamshad, M. J. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. *Nat Genet* **44**, 886-889, doi:10.1038/ng.2344 (2012).

52 Orr, H. T. & Zoghbi, H. Y. Trinucleotide repeat disorders. *Annu Rev Neurosci* **30**, 575-621, doi:10.1146/annurev.neuro.29.051605.113042 (2007).

53 Tome, S., Manley, K., Simard, J. P., Clark, G. W., Slean, M. M., Swami, M., Shelbourne, P. F., Tillier, E. R., Monckton, D. G., Messer, A. & Pearson, C. E. MSH3 polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice. *PLoS Genet* **9**, e1003280, doi:10.1371/journal.pgen.1003280 (2013).

54 Ciancarelli, I., Tozzi Ciancarelli, M. G. & Carolei, A. Effectiveness of intensive neurorehabilitation in patients with Huntington's disease. *Eur J Phys Rehabil Med* (2013).

55 Li, H., Cherry, S., Klinedinst, D., DeLeon, V., Redig, J., Reshey, B., Chin, M. T., Sherman, S. L., Maslen, C. L. & Reeves, R. H. Genetic modifiers predisposing to congenital heart disease in the sensitized Down syndrome population. *Circ Cardiovasc Genet* **5**, 301-308, doi:10.1161/CIRCGENETICS.111.960872 (2012).

56 Ackerman, C., Locke, A. E., Feingold, E., Reshey, B., Espana, K., Thusberg, J., Mooney, S., Bean, L. J., Dooley, K. J., Cua, C. L., Reeves, R. H., Sherman, S. L. & Maslen, C. L. An excess of deleterious variants in VEGF-A pathway genes in Down-syndrome-associated atrioventricular septal defects. *American journal of human genetics* **91**, 646-659, doi:10.1016/j.ajhg.2012.08.017 (2012).

57 Papavassiliou, P., York, T. P., Gursoy, N., Hill, G., Nicely, L. V., Sundaram, U., McClain, A., Aggen, S. H., Eaves, L., Riley, B. & Jackson-Cook, C. The phenotype of persons having mosaicism for trisomy 21/Down syndrome reflects the percentage of trisomic cells present in different tissues. *Am J Med Genet A* **149A**, 573-583, doi:10.1002/ajmg.a.32729 (2009).

58 Solomon, A. *Far from the tree : parents, children and the search for identity*. 1st Scribner hardcover edn, (Scribner, 2012).

59 Dodd, K. J. & Shields, N. A systematic review of the outcomes of cardiovascular exercise programs for people with Down syndrome. *Arch Phys Med Rehabil* **86**, 2051-2058, doi:10.1016/j.apmr.2005.06.003 (2005).

60 Wang, L., Kloc, M., Gu, Y., Ge, S. & Maffei, A. Layer-specific experience-dependent rewiring of thalamocortical circuits. *J Neurosci* **33**, 4181-4191, doi:10.1523/jneurosci.4423-12.2013 (2013).

61 Maffei, A., Bucher, D. & Fontanini, A. Homeostatic plasticity in the nervous system. *Neural Plast* **2012**, 913472, doi:10.1155/2012/913472 (2012).

62 Maffei, A. Enriching the environment to disinhibit the brain and improve cognition. *Front Cell Neurosci* **6**, 53, doi:10.3389/fncel.2012.00053 (2012).

63 Maffei, A. & Turrigiano, G. The age of plasticity: developmental regulation of synaptic plasticity in neocortical microcircuits. *Prog Brain Res* **169**, 211-223, doi:10.1016/s0079-6123(07)00012-x (2008).

64 Guris, D. L., Duester, G., Papaioannou, V. E. & Imamoto, A. Dose-dependent interaction of Tbx1 and Crkl and locally aberrant RA signaling in a model of del22q11 syndrome. *Dev Cell* **10**, 81-92, doi:10.1016/j.devcel.2005.12.002 (2006).

65 Iascone, M. R., Vittorini, S., Sacchelli, M., Spadoni, I., Simi, P. & Giusti, S. Molecular characterization of 22q11 deletion in a three-generation family with maternal transmission. *Am J Med Genet* **108**, 319-321 (2002).

66 Liao, J., Kochilas, L., Nowotschin, S., Arnold, J. S., Aggarwal, V. S., Epstein, J. A., Brown, M. C., Adams, J. & Morrow, B. E. Full spectrum of malformations in velo-cardio-facial syndrome/DiGeorge syndrome mouse models by altering Tbx1 dosage. *Hum Mol Genet* **13**, 1577-1585, doi:10.1093/hmg/ddh176 (2004).

67 McDonald-McGinn, D. M., Fahiminiya, S., Revil, T., Nowakowska, B. A., Suhl, J., Bailey, A., Mlynarski, E., Lynch, D. R., Yan, A. C., Bilaniuk, L. T., Sullivan, K. E., Warren, S. T., Emanuel, B. S., Vermeesch, J. R., Zackai, E. H. & Jerome-Majewska, L. A. Hemizygous mutations in SNAP29 unmask autosomal recessive conditions and contribute to

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

atypical findings in patients with 22q11.2DS. *J Med Genet* **50**, 80-90, doi:10.1136/jmedgenet-2012-101320 (2013).

68 Stalmans, I., Lambrechts, D., De Smet, F., Jansen, S., Wang, J., Maity, S., Kneer, P., von der Ohe, M., Swillen, A., Maes, C., Gewillig, M., Molin, D. G., Hellings, P., Boetel, T., Haardt, M., Compernolle, V., Dewerchin, M., Plaisance, S., Vlietinck, R., Emanuel, B., Gittenberger-de Groot, A. C., Scambler, P., Morrow, B., Driscol, D. A., Moons, L., Esguerra, C. V., Carmeliet, G., Behn-Krappa, A., Devriendt, K., Collen, D., Conway, S. J. & Carmeliet, P. VEGF: a modifier of the del22q11 (DiGeorge) syndrome? *Nat Med* **9**, 173-182, doi:10.1038/nm819 (2003).

69 Moreno-De-Luca, A., Myers, S. M., Challman, T. D., Moreno-De-Luca, D., Evans, D. W. & Ledbetter, D. H. Developmental brain dysfunction: revival and expansion of old concepts based on new genetic evidence. *Lancet Neurol* **12**, 406-414, doi:10.1016/S1474-4422(13)70011-5 (2013).

70 Zimmermann, M. B. The effects of iodine deficiency in pregnancy and infancy. *Paediatr Perinat Epidemiol* **26 Suppl 1**, 108-117, doi:10.1111/j.1365-3016.2012.01275.x (2012).

71 Bescond, M. & Rahmani, Z. Dual-specificity tyrosine-phosphorylated and regulated kinase 1A (DYRK1A) interacts with the phytanoyl-CoA alpha-hydroxylase associated protein 1 (PAHX-AP1), a brain specific protein. *Int J Biochem Cell Biol* **37**, 775-783, doi:10.1016/j.biocel.2004.12.006 (2005).

72 Jamieson, A. & Radick, G. in *THE PHILOSOPHY OF BIOLOGY: A COMPANION FOR EDUCATORS* (ed K. Kampourakis) 577-595 (Springer, 2013).

73 Bittles, A. H. & Black, M. L. Evolution in health and medicine Sackler colloquium: Consanguinity, human evolution, and complex diseases. *Proc Natl Acad Sci U S A* **107 Suppl 1**, 1779-1786, doi:10.1073/pnas.0906079106 (2010).

74 Lupski, J. R., Belmont, J. W., Boerwinkle, E. & Gibbs, R. A. Clan genomics and the complex architecture of human disease. *Cell* **147**, 32-43, doi:10.1016/j.cell.2011.09.008 (2011).

75 Dolinoy, D. C., Weidman, J. R., Waterland, R. A. & Jirtle, R. L. Maternal genistein alters coat color and protects Avy mouse offspring from obesity by modifying the fetal epigenome. *Environ Health Perspect* **114**, 567-572 (2006).

76 Bernal, A. J. & Jirtle, R. L. Epigenomic disruption: the effects of early developmental exposures. *Birth Defects Res A Clin Mol Teratol* **88**, 938-944, doi:10.1002/bdra.20685 (2010).

77 Weinhouse, C., Anderson, O. S., Jones, T. R., Kim, J., Liberman, S. A., Nahar, M. S., Rozek, L. S., Jirtle, R. L. & Dolinoy, D. C. An expression microarray approach for the identification of metastable epialleles in the mouse genome. *Epigenetics* **6**, 1105-1113, doi:10.4161/epi.6.9.17103 (2011).

78 Casanueva, M. O., Burga, A. & Lehner, B. Fitness trade-offs and environmentally induced mutation buffering in isogenic C. elegans. *Science* **335**, 82-85, doi:10.1126/science.1213491 (2012).

79 Burga, A., Casanueva, M. O. & Lehner, B. Predicting mutation outcome from early stochastic variation in genetic interaction partners. *Nature* **480**, 250-253, doi:10.1038/nature10665 (2011).

80 Comfort, N. C. *The tangled field : Barbara McClintock's search for the patterns of genetic control*. (Harvard University Press, 2001).

81 Beaudet, A. L. The utility of chromosomal microarray analysis in developmental and behavioral pediatrics. *Child Dev* **84**, 121-132, doi:10.1111/cdev.12050 (2013).

82 Refsgaard, L., Holst, A. G., Sadjadieh, G., Haunso, S., Nielsen, J. B. & Olesen, M. S. High prevalence of genetic variants previously associated with LQT syndrome in new exome data. *Eur J Hum Genet*, doi:10.1038/ejhg.2012.23 (2012).

83 Andreasen, C., Refsgaard, L., Nielsen, J. B., Sajadieh, A., Winkel, B. G., Tfelt-Hansen, J., Haunso, S., Holst, A. G., Svendsen, J. H. & Olesen, M. S. Mutations in Genes Encoding Cardiac Ion Channels Previously Associated With Sudden Infant Death Syndrome (SIDS) Are Present With High Frequency in New Exome Data. *Can J Cardiol*, doi:10.1016/j.cjca.2012.12.002 (2013).

84 Risgaard, B., Jabbari, R., Refsgaard, L., Holst, A., Haunso, S., Sadjadieh, A., Winkel, B., Olesen, M. & Tfelt-Hansen, J. High prevalence of genetic variants previously associated with Brugada syndrome in new exome data. *Clin Genet*, doi:10.1111/cge.12126 (2013).

85 Andreasen, C., Nielsen, J. B., Refsgaard, L., Holst, A. G., Christensen, A. H., Andreasen, L., Sajadieh, A., Haunso, S., Svendsen, J. H. & Olesen, M. S. New population-based exome data are questioning the pathogenicity of previously cardiomyopathy-associated genetic variants. *Eur J Hum Genet* **21**, 918-928, doi:10.1038/ejhg.2012.283 (2013).

86 Holmes, F. L. & Summers, W. C. *Reconceiving the gene : Seymour Benzer's adventures in phage genetics*. (Yale University Press, 2006).

87 Greenspan, R. J. Biological indeterminacy. *Sci Eng Ethics* **18**, 447-452, doi:10.1007/s11948-012-9379-2 (2012).

88 Greenspan, R. J. Selection, gene interaction, and flexible gene networks. *Cold Spring Harb Symp Quant Biol* **74**, 131-138, doi:10.1101/sqb.2009.74.029 (2009).

89 Greenspan, R. J. Seymour Benzer (1921-2007). *Curr Biol* **18**, R106-110 (2008).

90 Kendler, K. S. & Greenspan, R. J. The nature of genetic influences on behavior: lessons from "simpler" organisms. *The American journal of psychiatry* **163**, 1683-1694, doi:10.1176/appi.ajp.163.10.1683 (2006).

91 van Swinderen, B. & Greenspan, R. J. Flexibility in a gene network affecting a simple behavior in Drosophila melanogaster. *Genetics* **169**, 2151-2163, doi:10.1534/genetics.104.032631 (2005).

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

92 Blount, Z. D., Barrick, J. E., Davidson, C. J. & Lenski, R. E. Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature* **489**, 513-518, doi:10.1038/nature11514 (2012).

93 Meyer, J. R., Dobias, D. T., Weitz, J. S., Barrick, J. E., Quick, R. T. & Lenski, R. E. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* **335**, 428-432, doi:10.1126/science.1214449 (2012).

94 Kohane, I. S., Hsing, M. & Kong, S. W. Taxonomizing, sizing, and overcoming the incidentalome. *Genet Med*, doi:10.1038/gim.2011.68 (2012).

95 Eichers, E. R., Lewis, R. A., Katsanis, N. & Lupski, J. R. Triallelic inheritance: a bridge between Mendelian and multifactorial traits. *Ann Med* **36**, 262-272 (2004).

96 Katsanis, N., Ansley, S. J., Badano, J. L., Eichers, E. R., Lewis, R. A., Hoskins, B. E., Scambler, P. J., Davidson, W. S., Beales, P. L. & Lupski, J. R. Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. *Science* **293**, 2256-2259, doi:10.1126/science.1063525 (2001).

97 Katsanis, N., Eichers, E. R., Ansley, S. J., Lewis, R. A., Kayserili, H., Hoskins, B. E., Scambler, P. J., Beales, P. L. & Lupski, J. R. BBS4 is a minor contributor to Bardet-Biedl syndrome and may also participate in triallelic inheritance. *American journal of human genetics* **71**, 22-29, doi:10.1086/341031 (2002).

98 Abu-Safieh, L., Al-Anazi, S., Al-Abdi, L., Hashem, M., Alkuraya, H., Alamr, M., Sirelkhatim, M. O., Al-Hassnan, Z., Alkuraya, B., Mohamed, J. Y., Al-Salem, A., Alrashed, M., Faqeih, E., Softah, A., Al-Hashem, A., Wali, S., Rahbeeni, Z., Alsayed, M., Khan, A. O., Al-Gazali, L., Taschner, P. E., Al-Hazzaa, S. & Alkuraya, F. S. In search of triallelism in Bardet-Biedl syndrome. *Eur J Hum Genet*, doi:10.1038/ejhg.2011.205 (2012).

99 Smaoui, N., Chaabouni, M., Sergeev, Y. V., Kallel, H., Li, S., Mahfoudh, N., Maazoul, F., Kammoun, H., Gandoura, N., Bouaziz, A., Nouiri, E., M'Rad, R., Chaabouni, H. & Hejtmancik, J. F. Screening of the eight BBS genes in Tunisian families: no evidence of triallelism. *Invest Ophthalmol Vis Sci* **47**, 3487-3495, doi:10.1167/iovs.05-1334 (2006).

100 Laurier, V., Stoetzel, C., Muller, J., Thibault, C., Corbani, S., Jalkh, N., Salem, N., Chouery, E., Poch, O., Licaire, S., Danse, J. M., Amati-Bonneau, P., Bonneau, D., Megarbane, A., Mandel, J. L. & Dollfus, H. Pitfalls of homozygosity mapping: an extended consanguineous Bardet-Biedl syndrome family with two mutant genes (BBS2, BBS10), three mutations, but no triallelism. *Eur J Hum Genet* **14**, 1195-1203, doi:10.1038/sj.ejhg.5201688 (2006).

101 Nakane, T. & Biesecker, L. G. No evidence for triallelic inheritance of MKKS/BBS loci in Amish Mckusick-Kaufman syndrome. *Am J Med Genet A* **138**, 32-34, doi:10.1002/ajmg.a.30593 (2005).

102 Mykytyn, K., Nishimura, D. Y., Searby, C. C., Beck, G., Bugge, K., Haines, H. L., Cornier, A. S., Cox, G. F., Fulton, A. B., Carmi, R., Iannaccone, A., Jacobson, S. G., Weleber, R. G., Wright, A. F., Riise, R., Hennekam, R. C., Luleci, G., Berker-Karauzum, S., Biesecker, L. G., Stone, E. M. & Sheffield, V. C. Evaluation of complex inheritance involving the most common Bardet-Biedl syndrome locus (BBS1). *American journal of human genetics* **72**, 429-437 (2003).

103 Beales, P. L., Badano, J. L., Ross, A. J., Ansley, S. J., Hoskins, B. E., Kirsten, B., Mein, C. A., Froguel, P., Scambler, P. J., Lewis, R. A., Lupski, J. R. & Katsanis, N. Genetic interaction of BBS1 mutations with alleles at other BBS loci can result in non-Mendelian Bardet-Biedl syndrome. *American journal of human genetics* **72**, 1187-1199, doi:10.1086/375178 (2003).

104 Philip, N. & Bassett, A. Cognitive, behavioural and psychiatric phenotype in 22q11.2 deletion syndrome. *Behav Genet* **41**, 403-412, doi:10.1007/s10519-011-9468-z (2011).

105 Hamshere, M. L., Bennett, P., Williams, N., Segurado, R., Cardno, A., Norton, N., Lambert, D., Williams, H., Kirov, G., Corvin, A., Holmans, P., Jones, L., Jones, I., Gill, M., O'Donovan, M. C., Owen, M. J. & Craddock, N. Genomewide linkage scan in schizoaffective disorder: significant evidence for linkage at 1q42 close to DISC1, and suggestive evidence at 22q11 and 19p13. *Arch Gen Psychiatry* **62**, 1081-1088, doi:10.1001/archpsyc.62.10.1081 (2005).

106 Blackwood, D. H., Fordyce, A., Walker, M. T., St Clair, D. M., Porteous, D. J. & Muir, W. J. Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. *American journal of human genetics* **69**, 428-433 (2001).

107 Berndt, S. I., Gustafsson, S., Magi, R., Ganna, A., Wheeler, E., Feitosa, M. F., Justice, A. E., Monda, K. L., Croteau-Chonka, D. C., Day, F. R., Esko, T., Fall, T., Ferreira, T., Gentilini, D., Jackson, A. U., Luan, J., Randall, J. C., Vedantam, S., Willer, C. J., Winkler, T. W., Wood, A. R., Workalemahu, T., Hu, Y. J., Lee, S. H., Liang, L., Lin, D. Y., Min, J. L., Neale, B. M., Thorleifsson, G., Yang, J., Albrecht, E., Amin, N., Bragg-Gresham, J. L., Cadby, G., den Heijer, M., Eklund, N., Fischer, K., Goel, A., Hottenga, J. J., Huffman, J. E., Jarick, I., Johansson, A., Johnson, T., Kanoni, S., Kleber, M. E., Konig, I. R., Kristiansson, K., Kutalik, Z., Lamina, C., Lecoeur, C., Li, G., Mangino, M., McArdle, W. L., Medina-Gomez, C., Muller-Nurasyid, M., Ngwa, J. S., Nolte, I. M., Paternoster, L., Pechlivanis, S., Perola, M., Peters, M. J., Preuss, M., Rose, L. M., Shi, J., Shungin, D., Smith, A. V., Strawbridge, R. J., Surakka, I., Teumer, A., Trip, M. D., Tyrer, J., Van Vliet-Ostaptchouk, J. V., Vandenput, L., Waite, L. L., Zhao, J. H., Absher, D., Asselbergs, F. W., Atalay, M., Attwood, A. P., Balmforth, A. J., Basart, H., Beilby, J., Bonnycastle, L. L., Brambilla, P., Bruinenberg, M., Campbell, H., Chasman, D. I., Chines, P. S., Collins, F. S., Connell, J. M., Cookson, W. O., de Faire, U., de Vegt, F., Dei, M., Dimitriou, M., Edkins, S., Estrada, K., Evans, D. M., Farrall, M., Ferrario, M. M., Ferrieres, J., Franke, L., Frau, F., Gejman, P. V., Grallert, H., Gronberg, H., Gudnason, V., Hall, A. S., Hall, P., Hartikainen, A. L., Hayward, C., Heard-Costa, N. L., Heath, A. C., Hebebrand, J., Homuth, G., Hu, F. B., Hunt, S. E., Hypponen, E., Iribarren, C., Jacobs, K. B., Jansson, J. O., Jula, A., Kahonen, M., Kathiresan, S., Kee, F., Khaw, K. T., Kivimaki, M., Koenig, W., Kraja, A. T., Kumari, M., Kuulasmaa, K., Kuusisto, J., Laitinen, J. H., Lakka, T. A., Langenberg, C., Launer, L. J., Lind, L., Lindstrom, J., Liu, J., Liuzzi, A., Lokki, M. L., Lorentzon, M., Madden, P. A., Magnusson, P. K., Manunta, P., Marek, D., Marz, W., Mateo Leach, I., McKnight,

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

B., Medland, S. E., Mihailov, E., Milani, L., Montgomery, G. W., Mooser, V., Muhleisen, T. W., Munroe, P. B., Musk, A. W., Narisu, N., Navis, G., Nicholson, G., Nohr, E. A., Ong, K. K., Oostra, B. A., Palmer, C. N., Palotie, A., Peden, J. F., Pedersen, N., Peters, A., Polasek, O., Pouta, A., Pramstaller, P. P., Prokopenko, I., Putter, C., Radhakrishnan, A., Raitakari, O., Rendon, A., Rivadeneira, F., Rudan, I., Saaristo, T. E., Sambrook, J. G., Sanders, A. R., Sanna, S., Saramies, J., Schipf, S., Schreiber, S., Schunkert, H., Shin, S. Y., Signorini, S., Sinisalo, J., Skrobek, B., Soranzo, N., Stancakova, A., Stark, K., Stephens, J. C., Stirrups, K., Stolk, R. P., Stumvoll, M., Swift, A. J., Theodoraki, E. V., Thorand, B., Tregouet, D. A., Tremoli, E., Van der Klauw, M. M., van Meurs, J. B., Vermeulen, S. H., Viikari, J., Virtamo, J., Vitart, V., Waeber, G., Wang, Z., Widen, E., Wild, S. H., Willemsen, G., Winkelmann, B. R., Witteman, J. C., Wolffenbuttel, B. H., Wong, A., Wright, A. F., Zillikens, M. C., Amouyel, P., Boehm, B. O., Boerwinkle, E., Boomsma, D. I., Caulfield, M. J., Chanock, S. J., Cupples, L. A., Cusi, D., Dedoussis, G. V., Erdmann, J., Eriksson, J. G., Franks, P. W., Froguel, P., Gieger, C., Gyllensten, U., Hamsten, A., Harris, T. B., Hengstenberg, C., Hicks, A. A., Hingorani, A., Hinney, A., Hofman, A., Hovingh, K. G., Hveem, K., Illig, T., Jarvelin, M. R., Jockel, K. H., Keinanen-Kiukaanniemi, S. M., Kiemeney, L. A., Kuh, D., Laakso, M., Lehtimaki, T., Levinson, D. F., Martin, N. G., Metspalu, A., Morris, A. D., Nieminen, M. S., Njolstad, I., Ohlsson, C., Oldehinkel, A. J., Ouwehand, W. H., Palmer, L. J., Penninx, B., Power, C., Province, M. A., Psaty, B. M., Qi, L., Rauramaa, R., Ridker, P. M., Ripatti, S., Salomaa, V., Samani, N. J., Snieder, H., Sorensen, T. I., Spector, T. D., Stefansson, K., Tonjes, A., Tuomilehto, J., Uitterlinden, A. G., Uusitupa, M., van der Harst, P., Vollenweider, P., Wallaschofski, H., Wareham, N. J., Watkins, H., Wichmann, H. E., Wilson, J. F., Abecasis, G. R., Assimes, T. L., Barroso, I., Boehnke, M., Borecki, I. B., Deloukas, P., Fox, C. S., Frayling, T., Groop, L. C., Haritunian, T., Heid, I. M., Hunter, D., Kaplan, R. C., Karpe, F., Moffatt, M. F., Mohlke, K. L., O'Connell, J. R., Pawitan, Y., Schadt, E. E., Schlessinger, D., Steinthorsdottir, V., Strachan, D. P., Thorsteinsdottir, U., van Duijn, C. M., Visscher, P. M., Di Blasio, A. M., Hirschhorn, J. N., Lindgren, C. M., Morris, A. P., Meyre, D., Scherag, A., McCarthy, M. I., Speliotes, E. K., North, K. E., Loos, R. J. & Ingelsson, E. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet* **45**, 501-512, doi:10.1038/ng.2606 (2013).

108 Visscher, P. M., McEvoy, B. & Yang, J. From Galton to GWAS: quantitative genetics of human height. *Genet Res (Camb)* **92**, 371-379, doi:10.1017/S0016672310000571 (2010).

109 Visscher, P. M., Goddard, M. E., Derks, E. M. & Wray, N. R. Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. *Mol Psychiatry* **17**, 474-485, doi:10.1038/mp.2011.65 (2012).

110 Mitchell, K. J. What is complex about complex disorders? *Genome Biol* **13**, 237, doi:10.1186/gb-2012-13-1-237 (2012).

111 Gibson, G. Rare and common variants: twenty arguments. *Nat Rev Genet* **13**, 135-145, doi:10.1038/nrg3118 (2011).

112 Bleuler, E. *Dementia praecox; or, The group of schizophrenias*. (International Universities Press, 1958).

113 Kraepelin, E. *Manic-depressive insanity and paranoia*. (Livingstone, 1921).

114 Sporn, M. B. Perspective: The big C - for Chemoprevention. *Nature* **471**, S10-11, doi:10.1038/471S10a (2011).

115 Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr. & Kinzler, K. W. Cancer genome landscapes. *Science* **339**, 1546-1558, doi:10.1126/science.1235122 (2013).

116 Mukherjee, S. *The emperor of all maladies : a biography of cancer*. 1st Scribner hardcover edn, (Scribner, 2010).

117 Lyon, G. J. & Coffey, B. J. Complex tics and complex management in a case of severe Tourette's disorder (TD) in an adolescent. *J Child Adolesc Psychopharmacol* **19**, 469-474, doi:10.1089/cap.2009.19402 (2009).

118 Lyon, G. J., Coffey, B. & Silva, R. Postraumatic stress disorder and reactive attachment disorder: outcome in an adolescent. *J Child Adolesc Psychopharmacol* **18**, 641-646, doi:10.1089/cap.2008.1863 (2008).

119 Lyon, G. J. Possible varenicline-induced paranoia and irritability in a patient with major depressive disorder, borderline personality disorder, and methamphetamine abuse in remission. *J Clin Psychopharmacol* **28**, 720-721, doi:10.1097/JCP.0b013e31818db354 (2008).

120 Sacks, O. W. *An anthropologist on Mars : seven paradoxical tales*. 1st edn, (Alfrd A. Knopf, 1995).

121 Sacks, O. W. *The man who mistook his wife for a hat and other clinical tales*. 1st Touchstone edn, (Simon & Schuster, 1998).

122 Luria, A. R. *The man with a shattered world; the history of a brain wound*. (Basic Books, 1972).

123 Luria, A. R. *The mind of a mnemonist : a little book about a vast memory*. (H. Regnery, 1976).

124 Van Horn, J. D., Irimia, A., Torgerson, C. M., Chambers, M. C., Kikinis, R. & Toga, A. W. Mapping connectivity damage in the case of Phineas Gage. *PLoS One* **7**, e37454, doi:10.1371/journal.pone.0037454 (2012).

125 Ratiu, P., Talos, I. F., Haker, S., Lieberman, D. & Everett, P. The tale of Phineas Gage, digitally remastered. *J Neurotrauma* **21**, 637-643, doi:10.1089/089771504774129964 (2004).

126 Eichenbaum, H. What H.M. taught us. *J Cogn Neurosci* **25**, 14-21, doi:10.1162/jocn_a_00285 (2013).

127 Worthey, E. A., Mayer, A. N., Syverson, G. D., Helbling, D., Bonacci, B. B., Decker, B., Serpe, J. M., Dasu, T., Tschannen, M. R., Veith, R. L., Basehore, M. J., Broeckel, U., Tomita-Mitchell, A., Arca, M. J., Casper, J. T., Margolis, D. A., Bick, D. P., Hessner, M. J., Routes, J. M., Verbsky, J. W., Jacob, H. J. & Dimmock, D. P. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet Med* **13**, 255-262, doi:10.1097/GIM.0b013e3182088158 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

128 Penrose, L. S. *The biology of mental defect*. [3d rev. and reset edn, (Sidgwick and Jackson, 1963).

129 Ward, O. C. *John Langdon Down, 1828-1896 : a caring pioneer*. (Royal Society of Medicine Press, 1998).

130 Lyon, G. J., Abi-Dargham, A., Moore, H., Lieberman, J. A., Javitch, J. A. & Sulzer, D. Presynaptic regulation of dopamine transmission in schizophrenia. *Schizophr Bull* **37**, 108-117, doi:10.1093/schbul/sbp010 (2011).

131 Craddock, N. & Owen, M. J. The Kraepelinian dichotomy - going, going... but still not gone. *Br J Psychiatry* **196**, 92-95, doi:10.1192/bjp.bp.109.073429 (2010).

132 Williams, H. J., Craddock, N., Russo, G., Hamshere, M. L., Moskvina, V., Dwyer, S., Smith, R. L., Green, E., Grozeva, D., Holmans, P., Owen, M. J. & O'Donovan, M. C. Most genome-wide significant susceptibility loci for schizophrenia and bipolar disorder reported to date cross-traditional diagnostic boundaries. *Hum Mol Genet* **20**, 387-391, doi:10.1093/hmg/ddq471 (2011).

133 Lichtenstein, P., Yip, B. H., Bjork, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F. & Hultman, C. M. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**, 234-239, doi:10.1016/S0140-6736(09)60072-6 (2009).

134 Reiss, A. L. Childhood developmental disorders: an academic and clinical convergence point for psychiatry, neurology, psychology and pediatrics. *J Child Psychol Psychiatry* **50**, 87-98, doi:10.1111/j.1469-7610.2008.02046.x (2009).

135 Aldridge, K., George, I. D., Cole, K. K., Austin, J. R., Takahashi, T. N., Duan, Y. & Miles, J. H. Facial phenotypes in subgroups of prepubertal boys with autism spectrum disorders are correlated with clinical phenotypes. *Mol Autism* **2**, 15, doi:10.1186/2040-2392-2-15 (2011).

136 Miles, J. H. Autism spectrum disorders--a genetics review. *Genet Med* **13**, 278-294, doi:10.1097/GIM.0b013e3181ff67ba (2011).

137 Boyden, L. M., Choi, M., Choate, K. A., Nelson-Williams, C. J., Farhi, A., Toka, H. R., Tikhonova, I. R., Bjornson, R., Mane, S. M., Colussi, G., Lebel, M., Gordon, R. D., Semmekrot, B. A., Poujol, A., Valimaki, M. J., De Ferrari, M. E., Sanjad, S. A., Gutkin, M., Karet, F. E., Tucci, J. R., Stockigt, J. R., Keppler-Noreuil, K. M., Porter, C. C., Anand, S. K., Whiteford, M. L., Davis, I. D., Dewar, S. B., Bettinelli, A., Fadrowski, J. J., Belsha, C. W., Hunley, T. E., Nelson, R. D., Trachtman, H., Cole, T. R., Pinsk, M., Bockenhauer, D., Shenoy, M., Vaidyanathan, P., Foreman, J. W., Rasoulpour, M., Thameem, F., Al-Shahrouri, H. Z., Radhakrishnan, J., Gharavi, A. G., Goilav, B. & Lifton, R. P. Mutations in kelch-like 3 and cullin 3 cause hypertension and electrolyte abnormalities. *Nature* **482**, 98-102, doi:10.1038/nature10814 (2012).

138 Vig, B. K. Somatic mosaicism in plants with special reference to somatic crossing over. *Environ Health Perspect* **27**, 27-36 (1978).

139 Bakker, E., Veenema, H., Den Dunnen, J. T., van Broeckhoven, C., Grootscholten, P. M., Bonten, E. J., van Ommen, G. J. & Pearson, P. L. Germinal mosaicism increases the recurrence risk for 'new' Duchenne muscular dystrophy mutations. *J Med Genet* **26**, 553-559 (1989).

140 Hall, J. G. Review and hypotheses: somatic mosaicism: observations related to clinical genetics. *American journal of human genetics* **43**, 355-363 (1988).

141 Sastry, G. R., Cooper, H. B., Jr. & Brink, R. A. Paramutation and somatic mosaicism in maize. *Genetics* **52**, 407-424 (1965).

142 Hollander, W. F. Sectorial mosaics in the domestic pigeion: 25 more years. *J Hered* **66**, 177-202 (1975).

143 Biesecker, L. G. & Spinner, N. B. A genomic view of mosaicism and human disease. *Nat Rev Genet* **14**, 307-320, doi:10.1038/nrg3424 (2013).

144 Jongmans, M. C., Verwiel, E. T., Heijdra, Y., Vulliamy, T., Kamping, E. J., Hehir-Kwa, J. Y., Bongers, E. M., Pfundt, R., van Emst, L., van Leeuwen, F. N., van Gassen, K. L., Geurts van Kessel, A., Dokal, I., Hoogerbrugge, N., Ligtenberg, M. J. & Kuiper, R. P. Revertant somatic mosaicism by mitotic recombination in dyskeratosis congenita. *American journal of human genetics* **90**, 426-433, doi:10.1016/j.ajhg.2012.01.004 (2012).

145 Steinbusch, C., van Roozendaal, K., Tserpelis, D., Smeets, E., Kranenburg-de Koning, T., de Waal, K., Zweier, C., Rauch, A., Hennekam, R., Blok, M. & Schrander-Stumpel, C. Somatic mosaicism in a mother of two children with Pitt-Hopkins syndrome. *Clin Genet*, doi:10.1111/j.1399-0004.2012.01857.x (2012).

146 Yamada, M., Okura, Y., Suzuki, Y., Fukumura, S., Miyazaki, T., Ikeda, H., Takezaki, S., Kawamura, N., Kobayashi, I. & Ariga, T. Somatic mosaicism in two unrelated patients with X-linked chronic granulomatous disease characterized by the presence of a small population of normal cells. *Gene* **497**, 110-115, doi:10.1016/j.gene.2012.01.019 (2012).

147 Lindhurst, M. J., Sapp, J. C., Teer, J. K., Johnston, J. J., Finn, E. M., Peters, K., Turner, J., Cannons, J. L., Bick, D., Blakemore, L., Blumhorst, C., Brockmann, K., Calder, P., Cherman, N., Deardorff, M. A., Everman, D. B., Golas, G., Greenstein, R. M., Kato, B. M., Keppler-Noreuil, K. M., Kuznetsov, S. A., Miyamoto, R. T., Newman, K., Ng, D., O'Brien, K., Rothenberg, S., Schwartzentruber, D. J., Singhal, V., Tirabosco, R., Upton, J., Wientroub, S., Zackai, E. H., Hoag, K., Whitewood-Neal, T., Robey, P. G., Schwartzberg, P. L., Darling, T. N., Tosi, L. L., Mullikin, J. C. & Biesecker, L. G. A mosaic activating mutation in AKT1 associated with the Proteus syndrome. *N Engl J Med* **365**, 611-619, doi:10.1056/NEJMoa1104017 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

148 Choate, K. A., Lu, Y., Zhou, J., Choi, M., Elias, P. M., Farhi, A., Nelson-Williams, C., Crumrine, D., Williams, M. L., Nopper, A. J., Bree, A., Milstone, L. M. & Lifton, R. P. Mitotic recombination in patients with ichthyosis causes reversion of dominant mutations in KRT10. *Science* **330**, 94-97, doi:10.1126/science.1192280 (2010).

149 Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Marchetto, M. C., Muotri, A. R., Mu, Y., Carson, C. T., Macia, A., Moran, J. V. & Gage, F. H. Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proc Natl Acad Sci U S A* **108**, 20382-20387, doi:10.1073/pnas.1100273108 (2011).

150 Margari, L., Lamanna, A. L., Buttiglione, M., Craig, F., Petruzzelli, M. G. & Terenzio, V. Long-term follow-up of neurological manifestations in a boy with incontinentia pigmenti. *Eur J Pediatr*, doi:10.1007/s00431-013-2021-8 (2013).

151 Huisman, S. A., Redeker, E. J., Maas, S. M., Mannens, M. M. & Hennekam, R. C. High rate of mosaicism in individuals with Cornelia de Lange syndrome. *J Med Genet* **50**, 339-344, doi:10.1136/jmedgenet-2012-101477 (2013).

152 Lindhurst, M. J., Parker, V. E., Payne, F., Sapp, J. C., Rudge, S., Harris, J., Witkowski, A. M., Zhang, Q., Groeneveld, M. P., Scott, C. E., Daly, A., Huson, S. M., Tosi, L. L., Cunningham, M. L., Darling, T. N., Geer, J., Gucev, Z., Sutton, V. R., Tziotzios, C., Dixon, A. K., Helliwell, T., O'Rahilly, S., Savage, D. B., Wakelam, M. J., Barroso, I., Biesecker, L. G. & Semple, R. K. Mosaic overgrowth with fibroadipose hyperplasia is caused by somatic activating mutations in PIK3CA. *Nat Genet* **44**, 928-933, doi:10.1038/ng.2332 (2012).

153 Tanaka, T., Takahashi, K., Yamane, M., Tomida, S., Nakamura, S., Oshima, K., Niwa, A., Nishikomori, R., Kambe, N., Hara, H., Mitsuyama, M., Morone, N., Heuser, J. E., Yamamoto, T., Watanabe, A., Sato-Otsubo, A., Ogawa, S., Asaka, I., Heike, T., Yamanaka, S., Nakahata, T. & Saito, M. K. Induced pluripotent stem cells from CINCA syndrome patients as a model for dissecting somatic mosaicism and drug discovery. *Blood* **120**, 1299-1308, doi:10.1182/blood-2012-03-417881 (2012).

154 Kurek, K. C., Luks, V. L., Ayturk, U. M., Alomari, A. I., Fishman, S. J., Spencer, S. A., Mulliken, J. B., Bowen, M. E., Yamamoto, G. L., Kozakewich, H. P. & Warman, M. L. Somatic mosaic activating mutations in PIK3CA cause CLOVES syndrome. *American journal of human genetics* **90**, 1108-1115, doi:10.1016/j.ajhg.2012.05.006 (2012).

155 Macosko, E. Z. & McCarroll, S. A. Exploring the variation within. *Nat Genet* **44**, 614-616, doi:10.1038/ng.2311 (2012).

156 Weiss, K. M. Cryptic causation of human disease: reading between the (germ) lines. *Trends Genet* **21**, 82-88, doi:10.1016/j.tig.2004.12.008 (2005).

157 Baugher, J. D., Baugher, B. D., Shirley, M. D. & Pevsner, J. Sensitive and specific detection of mosaic chromosomal abnormalities using the Parent-of-Origin-based Detection (POD) method. *BMC Genomics* **14**, 367, doi:10.1186/1471-2164-14-367 (2013).

158 Shirley, M. D., Tang, H., Gallione, C. J., Baugher, J. D., Frelin, L. P., Cohen, B., North, P. E., Marchuk, D. A., Comi, A. M. & Pevsner, J. Sturge-Weber syndrome and port-wine stains caused by somatic mutation in GNAQ. *N Engl J Med* **368**, 1971-1979, doi:10.1056/NEJMoa1213507 (2013).

159 Hulten, M. A., Jonasson, J., Nordgren, A. & Iwarsson, E. Germinal and Somatic Trisomy 21 Mosaicism: How Common is it, What are the Implications for Individual Carriers and How Does it Come About? *Curr Genomics* **11**, 409-419, doi:10.2174/138920210793176056 (2010).

160 Kovaleva, N. V. Germ-line transmission of trisomy 21: Data from 80 families suggest an implication of grandmaternal age and a high frequency of female-specific trisomy rescue. *Mol Cytogenet* **3**, 7, doi:10.1186/1755-8166-3-7 (2010).

161 Iourov, I. Y., Vorsanova, S. G. & Yurov, Y. B. Chromosomal mosaicism goes global. *Mol Cytogenet* **1**, 26, doi:10.1186/1755-8166-1-26 (2008).

162 Etheridge, S. P., Bowles, N. E., Arrington, C. B., Pilcher, T., Rope, A., Wilde, A. A., Alders, M., Saarel, E. V., Tavernier, R., Timothy, K. W. & Tristani-Firouzi, M. Somatic mosaicism contributes to phenotypic variation in Timothy syndrome. *Am J Med Genet A* **155A**, 2578-2583, doi:10.1002/ajmg.a.34223 (2011).

163 Aldred, M. A., Bagshaw, R. J., Macdermot, K., Casson, D., Murch, S. H., Walker-Smith, J. A. & Trembath, R. C. Germline mosaicism for a GNAS1 mutation and Albright hereditary osteodystrophy. *J Med Genet* **37**, E35 (2000).

164 Barbosa, R. H., Vargas, F. R., Aguiar, F. C., Ferman, S., Lucena, E., Bonvicino, C. R. & Seuanez, H. N. Hereditary retinoblastoma transmitted by maternal germline mosaicism. *Pediatr Blood Cancer* **51**, 598-602, doi:10.1002/pbc.21687 (2008).

165 Chaturvedi, L. S., Mittal, R. D., Srivastava, S., Mukherjee, M. & Mittal, B. Analysis of dinucleotide repeat loci of dystrophin gene for carrier detection, germline mosaicism and de novo mutations in Duchenne muscular dystrophy. *Clin Genet* **58**, 234-236 (2000).

166 Evans, J. C., Archer, H. L., Whatley, S. D. & Clarke, A. Germline mosaicism for a MECP2 mutation in a man with two Rett daughters. *Clin Genet* **70**, 336-338, doi:10.1111/j.1399-0004.2006.00691.x (2006).

167 Frank, J. & Happle, R. Cutaneous mosaicism: right before our eyes. *J Clin Invest* **117**, 1216-1219, doi:10.1172/jci32111 (2007).

168 Hosoki, K., Takano, K., Sudo, A., Tanaka, S. & Saitoh, S. Germline mosaicism of a novel UBE3A mutation in Angelman syndrome. *Am J Med Genet A* **138A**, 187-189, doi:10.1002/ajmg.a.30926 (2005).

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

169 Jongmans, M. C., Hoefsloot, L. H., van der Donk, K. P., Admiraal, R. J., Magee, A., van de Laar, I., Hendriks, Y., Verheij, J. B., Walpole, I., Brunner, H. G. & van Ravenswaaij, C. M. Familial CHARGE syndrome and the CHD7 gene: a recurrent missense mutation, intrafamilial recurrence and variability. *Am J Med Genet A* **146A**, 43-50, doi:10.1002/ajmg.a.31921 (2008).

170 Mari, F., Caselli, R., Russo, S., Cogliati, F., Ariani, F., Longo, I., Bruttini, M., Meloni, I., Pescucci, C., Schurfeld, K., Toti, P., Tassini, M., Larizza, L., Hayek, G., Zappella, M. & Renieri, A. Germline mosaicism in Rett syndrome identified by prenatal diagnosis. *Clin Genet* **67**, 258-260, doi:10.1111/j.1399-0004.2005.00397.x (2005).

171 Meyer, K. J., Axelsen, M. S., Sheffield, V. C., Patil, S. R. & Wassink, T. H. Germline mosaic transmission of a novel duplication of PXDN and MYT1L to two male half-siblings with autism. *Psychiatr Genet* **22**, 137-140, doi:10.1097/YPG.0b013e32834dc3f5 (2012).

172 Parodi, S., Bachetti, T., Lantieri, F., Di Duca, M., Santamaria, G., Ottonello, G., Matera, I., Ravazzolo, R. & Ceccherini, I. Parental origin and somatic mosaicism of PHOX2B mutations in Congenital Central Hypoventilation Syndrome. *Hum Mutat* **29**, 206, doi:10.1002/humu.9516 (2008).

173 Pauli, S., Pieper, L., Haberle, J., Grzmil, P., Burfeind, P., Steckel, M., Lenz, U. & Michelmann, H. W. Proven germline mosaicism in a father of two children with CHARGE syndrome. *Clin Genet* **75**, 473-479 (2009).

174 Rand, C. M., Yu, M., Jennings, L. J., Panesar, K., Berry-Kravis, E. M., Zhou, L. & Weese-Mayer, D. E. Germline mosaicism of PHOX2B mutation accounts for familial recurrence of congenital central hypoventilation syndrome (CCHS). *Am J Med Genet A* **158A**, 2297-2301, doi:10.1002/ajmg.a.35499 (2012).

175 Sato, N., Ohyama, K., Fukami, M., Okada, M. & Ogata, T. Kallmann syndrome: somatic and germline mutations of the fibroblast growth factor receptor 1 gene in a mother and the son. *J Clin Endocrinol Metab* **91**, 1415-1418, doi:10.1210/jc.2005-2266 (2006).

176 Sbidian, E., Feldmann, D., Bengoa, J., Fraitag, S., Abadie, V., de Prost, Y., Bodemer, C. & Hadj-Rabia, S. Germline mosaicism in keratitis-ichthyosis-deafness syndrome: pre-natal diagnosis in a familial lethal form. *Clin Genet* **77**, 587-592, doi:10.1111/j.1399-0004.2009.01339.x (2010).

177 Shanske, A. L., Goodrich, J. T., Ala-Kokko, L., Baker, S., Frederick, B. & Levy, B. Germline mosacism in Shprintzen-Goldberg syndrome. *Am J Med Genet A* **158A**, 1574-1578, doi:10.1002/ajmg.a.35388 (2012).

178 Slavin, T. P., Lazebnik, N., Clark, D. M., Vengoechea, J., Cohen, L., Kaur, M., Konczal, L., Crowe, C. A., Corteville, J. E., Nowaczyk, M. J., Byrne, J. L., Jackson, L. G. & Krantz, I. D. Germline mosaicism in Cornelia de Lange syndrome. *Am J Med Genet A* **158A**, 1481-1485, doi:10.1002/ajmg.a.35381 (2012).

179 Sol-Church, K., Stabley, D. L., Demmer, L. A., Agbulos, A., Lin, A. E., Smoot, L., Nicholson, L. & Gripp, K. W. Male-to-male transmission of Costello syndrome: G12S HRAS germline mutation inherited from a father with somatic mosaicism. *Am J Med Genet A* **149A**, 315-321, doi:10.1002/ajmg.a.32639 (2009).

180 Tajir, M., Fergelot, P., Lancelot, G., Elalaoui, S. C., Arveiler, B., Lacombe, D. & Sefiani, A. Germline mosaicism in Rubinstein-Taybi syndrome. *Gene* **518**, 476-478, doi:10.1016/j.gene.2012.12.105 (2013).

181 Trevisson, E., Forzan, M., Salviati, L. & Clementi, M. Neurofibromatosis type 1 in two siblings due to maternal germline mosaicism. *Clin Genet*, doi:10.1111/cge.12177 (2013).

182 Venancio, M., Santos, M., Pereira, S. A., Maciel, P. & Saraiva, J. M. An explanation for another familial case of Rett syndrome: maternal germline mosaicism. *Eur J Hum Genet* **15**, 902-904, doi:10.1038/sj.ejhg.5201835 (2007).

183 Wuyts, W., Biervliet, M., Reyniers, E., D'Apice, M. R., Novelli, G. & Storm, K. Somatic and gonadal mosaicism in Hutchinson-Gilford progeria. *Am J Med Genet A* **135**, 66-68, doi:10.1002/ajmg.a.30663 (2005).

184 Jonsson, T., Atwal, J. K., Steinberg, S., Snaedal, J., Jonsson, P. V., Bjornsson, S., Stefansson, H., Sulem, P., Gudbjartsson, D., Maloney, J., Hoyte, K., Gustafson, A., Liu, Y., Lu, Y., Bhangale, T., Graham, R. R., Huttenlocher, J., Bjornsdottir, G., Andreassen, O. A., Jonsson, E. G., Palotie, A., Behrens, T. W., Magnusson, O. T., Kong, A., Thorsteinsdottir, U., Watts, R. J. & Stefansson, K. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488**, 96-99, doi:10.1038/nature11283 (2012).

185 Styrkarsdottir, U., Thorleifsson, G., Sulem, P., Gudbjartsson, D. F., Sigurdsson, A., Jonasdottir, A., Oddsson, A., Helgason, A., Magnusson, O. T., Walters, G. B., Frigge, M. L., Helgadottir, H. T., Johannsdottir, H., Bergsteinsdottir, K., Ogmundsdottir, M. H., Center, J. R., Nguyen, T. V., Eisman, J. A., Christiansen, C., Steingrimsson, E., Jonasson, J. G., Tryggvadottir, L., Eyjolfsson, G. I., Theodors, A., Jonsson, T., Ingvarsson, T., Olafsson, I., Rafnar, T., Kong, A., Sigurdsson, G., Masson, G., Thorsteinsdottir, U. & Stefansson, K. Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. *Nature*, doi:10.1038/nature12124 (2013).

186 Williams, M. S. Myostatin mutation associated with gross muscle hypertrophy in a child. *N Engl J Med* **351**, 1030-1031; author reply 1030-1031 (2004).

187 Boyden, L. M., Mao, J., Belsky, J., Mitzner, L., Farhi, A., Mitnick, M. A., Wu, D., Insogna, K. & Lifton, R. P. High bone density due to a mutation in LDL-receptor-related protein 5. *N Engl J Med* **346**, 1513-1521, doi:10.1056/NEJMoa013444 (2002).

188 *Liam Hoekstra Strongest Boy in the World*, <http://videos.disabled-world.com/video/159/liam-hoekstra-strongest-boy-in> (2013).

189 Fu, M., Zhang, X., Lai, X., Wu, X., Feng, F., Peng, J., Zhong, H., Zhang, Y., Wang, Y., Zhou, Q., Wang, S., Chen, L., He, Z., Gao, Y., Ma, X., He, R. & Liu, Q. Generation of Sequence Variants via Accelerated Molecular Evolution Methods. *Recent Pat DNA Gene Seq* (2013).

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

190 Leconte, A. M., Dickinson, B. C., Yang, D. D., Chen, I. A., Allen, B. & Liu, D. R. A population-based experimental model for protein evolution: effects of mutation rate and selection stringency on evolutionary outcomes. *Biochemistry* **52**, 1490-1499, doi:10.1021/bi3016185 (2013).

191 Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499-503, doi:10.1038/nature09929 (2011).

192 Mitchell, K. J. & Porteous, D. J. Rethinking the genetic architecture of schizophrenia. *Psychol Med* **41**, 19-32, doi:10.1017/S003329171000070X (2011).

193 Girirajan, S., Rosenfeld, J. A., Coe, B. P., Parikh, S., Friedman, N., Goldstein, A., Filipink, R. A., McConnell, J. S., Angle, B., Meschino, W. S., Nezarati, M. M., Asamoah, A., Jackson, K. E., Gowans, G. C., Martin, J. A., Carmany, E. P., Stockton, D. W., Schnur, R. E., Penney, L. S., Martin, D. M., Raskin, S., Leppig, K., Thiese, H., Smith, R., Aberg, E., Niyazov, D. M., Escobar, L. F., El-Khechen, D., Johnson, K. D., Lebel, R. R., Siefkas, K., Ball, S., Shur, N., McGuire, M., Brasington, C. K., Spence, J. E., Martin, L. S., Clericuzio, C., Ballif, B. C., Shaffer, L. G. & Eichler, E. E. Phenotypic Heterogeneity of Genomic Disorders and Rare Copy-Number Variants. *N Engl J Med*, doi:10.1056/NEJMoa1200395 (2012).

194 Shi, L., Zhang, X., Golhar, R., Otieno, F. G., He, M., Hou, C., Kim, C., Keating, B., Lyon, G. J., Wang, K. & Hakonarson, H. Whole-genome sequencing in an autism multiplex family. *Mol Autism* **4**, 8, doi:10.1186/2040-2392-4-8 (2013).

195 Fullston, T., Gabb, B., Callen, D., Ullmann, R., Woollatt, E., Bain, S., Ropers, H. H., Cooper, M., Chandler, D., Carter, K., Jablensky, A., Kalaydjieva, L. & Gecz, J. Inherited balanced translocation t(9;17)(q33.2;q25.3) concomitant with a 16p13.1 duplication in a patient with schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* **156**, 204-214, doi:10.1002/ajmg.b.31157 (2011).

196 Crepel, A., Breckpot, J., Fryns, J. P., De la Marche, W., Steyaert, J., Devriendt, K. & Peeters, H. DISC1 duplication in two brothers with autism and mild mental retardation. *Clin Genet* **77**, 389-394, doi:10.1111/j.1399-0004.2009.01318.x (2010).

197 Klei, L., Sanders, S. J., Murtha, M. T., Hus, V., Lowe, J. K., Willsey, A. J., Moreno-De-Luca, D., Yu, T. W., Fombonne, E., Geschwind, D., Grice, D. E., Ledbetter, D. H., Lord, C., Mane, S. M., Lese Martin, C., Martin, D. M., Morrow, E. M., Walsh, C. A., Melhem, N. M., Chaste, P., Sutcliffe, J. S., State, M. W., Cook, E. H., Jr., Roeder, K. & Devlin, B. Common genetic variants, acting additively, are a major source of risk for autism. *Mol Autism* **3**, 9, doi:10.1186/2040-2392-3-9 (2012).

198 Visscher, P. M., Goddard, M. E., Derks, E. M. & Wray, N. R. Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. *Mol Psychiatry*, doi:10.1038/mp.2011.65 (2011).

199 Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* **109**, 1193-1198, doi:10.1073/pnas.1119675109 (2012).

200 Anney, R., Klei, L., Pinto, D., Almeida, J., Bacchelli, E., Baird, G., Bolshakova, N., Bolte, S., Bolton, P. F., Bourgeron, T., Brennan, S., Brian, J., Casey, J., Conroy, J., Correia, C., Corsello, C., Crawford, E. L., de Jonge, M., Delorme, R., Duketis, E., Duque, F., Estes, A., Farrar, P., Fernandez, B. A., Folstein, S. E., Fombonne, E., Gilbert, J., Gillberg, C., Glessner, J. T., Green, A., Green, J., Guter, S. J., Heron, E. A., Holt, R., Howe, J. L., Hughes, G., Hus, V., Igliozzi, R., Jacob, S., Kenny, G. P., Kim, C., Kolevzon, A., Kustanovich, V., Lajonchere, C. M., Lamb, J. A., Law-Smith, M., Leboyer, M., Le Couteur, A., Leventhal, B. L., Liu, X. Q., Lombard, F., Lord, C., Lotspeich, L., Lund, S. C., Magalhaes, T. R., Mantoulan, C., McDougle, C. J., Melhem, N. M., Merikangas, A., Minshew, N. J., Mirza, G. K., Munson, J., Noakes, C., Nygren, G., Papanikolaou, K., Pagnamenta, A. T., Parrini, B., Paton, T., Pickles, A., Posey, D. J., Poustka, F., Ragoussis, J., Regan, R., Roberts, W., Roeder, K., Roge, B., Rutter, M. L., Schlitt, S., Shah, N., Sheffield, V. C., Soorya, L., Sousa, I., Stoppioni, V., Sykes, N., Tancredi, R., Thompson, A. P., Thomson, S., Tryfon, A., Tsiantis, J., Van Engeland, H., Vincent, J. B., Volkmar, F., Vorstman, J., Wallace, S., Wing, K., Wittemeyer, K., Wood, S., Zurawiecki, D., Zwaigenbaum, L., Bailey, A. J., Battaglia, A., Cantor, R. M., Coon, H., Cuccaro, M. L., Dawson, G., Ennis, S., Freitag, C. M., Geschwind, D. H., Haines, J. L., Klauck, S. M., McMahon, W. M., Maestrini, E., Miller, J., Monaco, A. P., Nelson, S. F., Nurnberger, J. I., Jr., Oliveira, G., Parr, J. R., Pericak-Vance, M. A., Piven, J., Schellenberg, G. D., Scherer, S. W., Vicente, A. M., Wassink, T. H., Wijsman, E. M., Betancur, C., Buxbaum, J. D., Cook, E. H., Gallagher, L., Gill, M., Hallmayer, J., Paterson, A. D., Sutcliffe, J. S., Szatmari, P., Vieland, V. J., Hakonarson, H. & Devlin, B. Individual common variants exert weak effects on the risk for autism spectrum disorderspi. *Hum Mol Genet* **21**, 4781-4792, doi:10.1093/hmg/dds301 (2012).

201 Schaaf, C. P., Sabo, A., Sakai, Y., Crosby, J., Muzny, D., Hawes, A., Lewis, L., Akbar, H., Varghese, R., Boerwinkle, E., Gibbs, R. A. & Zoghbi, H. Y. Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. *Hum Mol Genet* **20**, 3366-3375, doi:10.1093/hmg/ddr243 (2011).

202 Gagnon, F., Roslin, N. M. & Lemire, M. Successful identification of rare variants using oligogenic segregation analysis as a prioritizing tool for whole-exome sequencing studies. *BMC Proc* **5 Suppl 9**, S11, doi:10.1186/1753-6561-5-S9-S11 (2011).

203 Malhotra, D. & Sebat, J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* **148**, 1223-1241, doi:10.1016/j.cell.2012.02.039 (2012).

204 Gai, X., Xie, H. M., Perin, J. C., Takahashi, N., Murphy, K., Wenocur, A. S., D'Arcy, M., O'Hara, R. J., Goldmuntz, E., Grice, D. E., Shaikh, T. H., Hakonarson, H., Buxbaum, J. D., Elia, J. & White, P. S. Rare structural variation of synapse and neurotransmission genes in autism. *Mol Psychiatry* **17**, 402-411, doi:10.1038/mp.2011.10 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

205 Shaikh, T. H., Haldeman-Englert, C., Geiger, E. A., Ponting, C. P. & Webber, C. Genes and biological processes commonly disrupted in rare and heterogeneous developmental delay syndromes. *Hum Mol Genet* **20**, 880-893, doi:10.1093/hmg/ddq527 (2011).

206 Elia, J., Gai, X., Xie, H. M., Perin, J. C., Geiger, E., Glessner, J. T., D'Arcy, M., deBerardinis, R., Frackelton, E., Kim, C., Lantieri, F., Muganga, B. M., Wang, L., Takeda, T., Rappaport, E. F., Grant, S. F., Berrettini, W., Devoto, M., Shaikh, T. H., Hakonarson, H. & White, P. S. Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. *Mol Psychiatry* **15**, 637-646, doi:10.1038/mp.2009.57 (2010).

207 Novarino, G., El-Fishawy, P., Kayserili, H., Meguid, N. A., Scott, E. M., Schroth, J., Silhavy, J. L., Kara, M., Khalil, R. O., Ben-Omran, T., Ercan-Sencicek, A. G., Hashish, A. F., Sanders, S. J., Gupta, A. R., Hashem, H. S., Matern, D., Gabriel, S., Sweetman, L., Rahimi, Y., Harris, R. A., State, M. W. & Gleeson, J. G. Mutations in BCKD-kinase Lead to a Potentially Treatable Form of Autism with Epilepsy. *Science*, doi:10.1126/science.1224631 (2012).

208 O'Roak, B. J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B. P., Levy, R., Ko, A., Lee, C., Smith, J. D., Turner, E. H., Stanaway, I. B., Vernot, B., Malig, M., Baker, C., Reilly, B., Akey, J. M., Borenstein, E., Rieder, M. J., Nickerson, D. A., Bernier, R., Shendure, J. & Eichler, E. E. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246-250, doi:10.1038/nature10989 (2012).

209 Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y. H., Narzisi, G., Leotta, A., Kendall, J., Grabowska, E., Ma, B., Marks, S., Rodgers, L., Stepansky, A., Troge, J., Andrews, P., Bekritsky, M., Pradhan, K., Ghiban, E., Kramer, M., Parla, J., Demeter, R., Fulton, L. L., Fulton, R. S., Magrini, V. J., Ye, K., Darnell, J. C., Darnell, R. B., Mardis, E. R., Wilson, R. K., Schatz, M. C., McCombie, W. R. & Wigler, M. De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285-299, doi:10.1016/j.neuron.2012.04.009 (2012).

210 Neale, B. M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K. E., Sabo, A., Lin, C. F., Stevens, C., Wang, L. S., Makarov, V., Polak, P., Yoon, S., Maguire, J., Crawford, E. L., Campbell, N. G., Geller, E. T., Valladares, O., Schafer, C., Liu, H., Zhao, T., Cai, G., Lihm, J., Dannenfelser, R., Jabado, O., Peralta, Z., Nagaswamy, U., Muzny, D., Reid, J. G., Newsham, I., Wu, Y., Lewis, L., Han, Y., Voight, B. F., Lim, E., Rossin, E., Kirby, A., Flannick, J., Fromer, M., Shakir, K., Fennell, T., Garimella, K., Banks, E., Poplin, R., Gabriel, S., DePristo, M., Wimbish, J. R., Boone, B. E., Levy, S. E., Betancur, C., Sunyaev, S., Boerwinkle, E., Buxbaum, J. D., Cook, E. H., Jr., Devlin, B., Gibbs, R. A., Roeder, K., Schellenberg, G. D., Sutcliffe, J. S. & Daly, M. J. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242-245, doi:10.1038/nature11011 (2012).

211 Sanders, S. J., Murtha, M. T., Gupta, A. R., Murdoch, J. D., Raubeson, M. J., Willsey, A. J., Ercan-Sencicek, A. G., DiLullo, N. M., Parikshak, N. N., Stein, J. L., Walker, M. F., Ober, G. T., Teran, N. A., Song, Y., El-Fishawy, P., Murtha, R. C., Choi, M., Overton, J. D., Bjornson, R. D., Carriero, N. J., Meyer, K. A., Bilguvar, K., Mane, S. M., Sestan, N., Lifton, R. P., Gunel, M., Roeder, K., Geschwind, D. H., Devlin, B. & State, M. W. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237-241, doi:10.1038/nature10945 (2012).

212 Xu, B., Ionita-Laza, I., Roos, J. L., Boone, B., Woodrick, S., Sun, Y., Levy, S., Gogos, J. A. & Karayiorgou, M. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat Genet*, doi:10.1038/ng.2446 (2012).

213 Girirajan, S., Rosenfeld, J. A., Cooper, G. M., Antonacci, F., Siswara, P., Itsara, A., Vives, L., Walsh, T., McCarthy, S. E., Baker, C., Mefford, H. C., Kidd, J. M., Browning, S. R., Browning, B. L., Dickel, D. E., Levy, D. L., Ballif, B. C., Platky, K., Farber, D. M., Gowans, G. C., Wetherbee, J. J., Asamoah, A., Weaver, D. D., Mark, P. R., Dickerson, J., Garg, B. P., Ellingwood, S. A., Smith, R., Banks, V. C., Smith, W., McDonald, M. T., Hoo, J. J., French, B. N., Hudson, C., Johnson, J. P., Ozmore, J. R., Moeschler, J. B., Surti, U., Escobar, L. F., El-Khechen, D., Gorski, J. L., Kussmann, J., Salbert, B., Lacassie, Y., Biser, A., McDonald-McGinn, D. M., Zackai, E. H., Deardorff, M. A., Shaikh, T. H., Haan, E., Friend, K. L., Fichera, M., Romano, C., Gecz, J., DeLisi, L. E., Sebat, J., King, M. C., Shaffer, L. G. & Eichler, E. E. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* **42**, 203-209, doi:10.1038/ng.534 (2010).

214 Sanders, S. J., Ercan-Sencicek, A. G., Hus, V., Luo, R., Murtha, M. T., Moreno-De-Luca, D., Chu, S. H., Moreau, M. P., Gupta, A. R., Thomson, S. A., Mason, C. E., Bilguvar, K., Celestino-Soper, P. B., Choi, M., Crawford, E. L., Davis, L., Wright, N. R., Dhodapkar, R. M., DiCola, M., DiLullo, N. M., Fernandez, T. V., Fielding-Singh, V., Fishman, D. O., Frahm, S., Garagaloyan, R., Goh, G. S., Kammela, S., Klei, L., Lowe, J. K., Lund, S. C., McGrew, A. D., Meyer, K. A., Moffat, W. J., Murdoch, J. D., O'Roak, B. J., Ober, G. T., Pottenger, R. S., Raubeson, M. J., Song, Y., Wang, Q., Yaspan, B. L., Yu, T. W., Yurkiewicz, I. R., Beaudet, A. L., Cantor, R. M., Curland, M., Grice, D. E., Gunel, M., Lifton, R. P., Mane, S. M., Martin, D. M., Shaw, C. A., Sheldon, M., Tischfield, J. A., Walsh, C. A., Morrow, E. M., Ledbetter, D. H., Fombonne, E., Lord, C., Martin, C. L., Brooks, A. I., Sutcliffe, J. S., Cook, E. H., Jr., Geschwind, D., Roeder, K., Devlin, B. & State, M. W. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**, 863-885, doi:10.1016/j.neuron.2011.05.002 (2011).

215 O'Roak, B. J., Vives, L., Fu, W., Egertson, J. D., Stanaway, I. B., Phelps, I. G., Carvill, G., Kumar, A., Lee, C., Ankenman, K., Munson, J., Hiatt, J. B., Turner, E. H., Levy, R., O'Day, D. R., Krumm, N., Coe, B. P., Martin, B. K., Borenstein, E., Nickerson, D. A., Mefford, H. C., Doherty, D., Akey, J. M., Bernier, R., Eichler, E. E. & Shendure, J. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619-1622, doi:10.1126/science.1227764 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

216 lllumina. *Genotyping Microarray*, <http://www.illumina.com/applications/genotyping.ilmn> (2013).

217 Schneider, G. F. & Dekker, C. DNA sequencing with nanopores. *Nat Biotechnol* **30**, 326-328, doi:10.1038/nbt.2181 (2012).

218 Williams, A. L., Patterson, N., Glessner, J., Hakonarson, H. & Reich, D. Phasing of many thousands of genotyped samples. *American journal of human genetics* **91**, 238-251, doi:10.1016/j.ajhg.2012.06.013 (2012).

219 Peters, B. A., Kermani, B. G., Sparks, A. B., Alferov, O., Hong, P., Alexeev, A., Jiang, Y., Dahl, F., Tang, Y. T., Haas, J., Robasky, K., Zaranek, A. W., Lee, J. H., Ball, M. P., Peterson, J. E., Perazich, H., Yeung, G., Liu, J., Chen, L., Kennemer, M. I., Pothuraju, K., Konvicka, K., Tsoupko-Sitnikov, M., Pant, K. P., Ebert, J. C., Nilsen, G. B., Baccash, J., Halpern, A. L., Church, G. M. & Drmanac, R. Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells. *Nature* **487**, 190-195, doi:10.1038/nature11236 (2012).

220 Roach, J. C., Glusman, G., Smit, A. F., Huff, C. D., Hubley, R., Shannon, P. T., Rowen, L., Pant, K. P., Goodman, N., Bamshad, M., Shendure, J., Drmanac, R., Jorde, L. B., Hood, L. & Galas, D. J. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328**, 636-639, doi:10.1126/science.1186802 (2010).

221 Lupski, J. R., Reid, J. G., Gonzaga-Jauregui, C., Rio Deiros, D., Chen, D. C., Nazareth, L., Bainbridge, M., Dinh, H., Jing, C., Wheeler, D. A., McGuire, A. L., Zhang, F., Stankiewicz, P., Halperin, J. J., Yang, C., Gehman, C., Guo, D., Irikat, R. K., Tom, W., Fantin, N. J., Muzny, D. M. & Gibbs, R. A. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* **362**, 1181-1191, doi:10.1056/NEJMoa0908094 (2010).

222 McGuire, A. L. & Lupski, J. R. Personal genome research : what should the participant be told? *Trends Genet* **26**, 199-201, doi:10.1016/j.tig.2009.12.007 (2010).

223 Bainbridge, M. N., Wiszniewski, W., Murdock, D. R., Friedman, J., Gonzaga-Jauregui, C., Newsham, I., Reid, J. G., Fink, J. K., Morgan, M. B., Gingras, M. C., Muzny, D. M., Hoang, L. D., Yousaf, S., Lupski, J. R. & Gibbs, R. A. Whole-genome sequencing for optimized patient management. *Sci Transl Med* **3**, 87re83, doi:10.1126/scitranslmed.3002243 (2011).

224 Ng, S. B., Bigham, A. W., Buckingham, K. J., Hannibal, M. C., McMillin, M. J., Gildersleeve, H. I., Beck, A. E., Tabor, H. K., Cooper, G. M., Mefford, H. C., Lee, C., Turner, E. H., Smith, J. D., Rieder, M. J., Yoshiura, K., Matsumoto, N., Ohta, T., Niikawa, N., Nickerson, D. A., Bamshad, M. J. & Shendure, J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* **42**, 790-793, doi:ng.646 [pii] 10.1038/ng.646 (2010).

225 Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., Huff, C. D., Shannon, P. T., Jabs, E. W., Nickerson, D. A., Shendure, J. & Bamshad, M. J. Exome sequencing identifies the cause of a mendelian disorder. *Nature genetics* **42**, 30-35, doi:ng.499 [pii] 10.1038/ng.499 (2010).

226 Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A. & Shendure, J. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* **12**, 745-755, doi:10.1038/nrg3031 (2011).

227 Hedges, D. J., Burges, D., Powell, E., Almonte, C., Huang, J., Young, S., Boese, B., Schmidt, M., Pericak-Vance, M. A., Martin, E., Zhang, X., Harkins, T. T. & Zuchner, S. Exome sequencing of a multigenerational human pedigree. *PLoS One* **4**, e8232, doi:10.1371/journal.pone.0008232 (2009).

228 Lyon, G. J. Personal account of the discovery of a new disease using next-generation sequencing. Interview by Natalie Harrison. *Pharmacogenomics* **12**, 1519-1523, doi:10.2217/pgs.11.117 (2011).

229 Katsanis, S. H. & Katsanis, N. Molecular genetic testing and the future of clinical genomics. *Nat Rev Genet* **14**, 415-426, doi:10.1038/nrg3493 (2013).

230 Lyon, G. J. Personalized medicine: Bring clinical standards to human-genetics research. *Nature* **482**, 300-301, doi:10.1038/482300a (2012).

231 Lyon, G. J. *Guest post: Time to bring human genome sequencing into the clinic*, <http://www.genomesunzipped.org/2012/02/guest-post-time-to-bring-human-genome-sequencing-into-the-clinic.php> (2012).

232 Zumbo, P. & Mason, C. E. in *Genome Analysis: Current Procedures and Applications.* (Horizon Press. , 2014).

233 Aceview. *The AceView genes*, <http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/> (2013).

234 Moreno-De-Luca, A., Myers, S. M., Challman, T. D., Moreno-De-Luca, D., Evans, D. W. & Ledbetter, D. H. Developmental brain dysfunction: revival and expansion of old concepts based on new genetic evidence. *The Lancet Neurology* **12**, 406-414 (2013).

235 Nielsen, R., Paul, J. S., Albrechtsen, A. & Song, Y. S. Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet* **12**, 443-451, doi:10.1038/nrg2986 (2011).

236 O'Rawe, J., Guangqing, S., Wang, W., Hu, J., Bodily, P., Tian, L., Hakonarson, H., Johnson, E., Wei, Z., Jiang, T., Wang, K., Lyon, G. & Wu, Y. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Medicine* **5**, 28 (2013).

237 Kajiwara, K., Berson, E. & Dryja, T. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science* **264**, 1604-1608, doi:10.1126/science.8202715 (1994).

238 El-Hattab, A. W., Zhang, F., Maxim, R., Christensen, K. M., Ward, J. C., Hines-Dowell, S., Scaglia, F., Lupski, J. R. & Cheung, S. W. Deletion and duplication of 15q24: Molecular mechanisms and potential modification by additional copy number variants. *Genet Med* **12**, 573-586 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

239 Davis, E. E., Zhang, Q., Liu, Q., Diplas, B. H., Davey, L. M., Hartley, J., Stoetzel, C., Szymanska, K., Ramaswami, G., Logan, C. V., Muzny, D. M., Young, A. C., Wheeler, D. A., Cruz, P., Morgan, M., Lewis, L. R., Cherukuri, P., Maskeri, B., Hansen, N. F., Mullikin, J. C., Blakesley, R. W., Bouffard, G. G., Gyapay, G., Rieger, S., Tonshoff, B., Kern, I., Soliman, N. A., Neuhaus, T. J., Swoboda, K. J., Kayserili, H., Gallagher, T. E., Lewis, R. A., Bergmann, C., Otto, E. A., Saunier, S., Scambler, P. J., Beales, P. L., Gleeson, J. G., Maher, E. R., Attie-Bitach, T., Dollfus, H., Johnson, C. A., Green, E. D., Gibbs, R. A., Hildebrandt, F., Pierce, E. A. & Katsanis, N. TTC21B contributes both causal and modifying alleles across the ciliopathy spectrum. *Nat Genet*, doi:ng.756 [pii] 10.1038/ng.756 (2011).

240 Lemmers, R. J. L. F., Tawil, R., Petek, L. M., Balog, J., Block, G. J., Santen, G. W. E., Amell, A. M., van der Vliet, P. J., Almomani, R., Straasheijm, K. R., Krom, Y. D., Klooster, R., Sun, Y., den Dunnen, J. T., Helmer, Q., Donlin-Smith, C. M., Padberg, G. W., van Engelen, B. G. M., de Greef, J. C., Aartsma-Rus, A. M., Frants, R. R., de Visser, M., Desnuelle, C., Sacconi, S., Filippova, G. N., Bakker, B., Bamshad, M. J., Tapscott, S. J., Miller, D. G. & van der Maarel, S. M. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat Genet* **44**, 1370-1374, doi:http://www.nature.com/ng/journal/v44/n12/abs/ng.2454.html - supplementary-information (2012).

241 Lai, J., Li, R., Xu, X., Jin, W., Xu, M., Zhao, H., Xiang, Z., Song, W., Ying, K., Zhang, M., Jiao, Y., Ni, P., Zhang, J., Li, D., Guo, X., Ye, K., Jian, M., Wang, B., Zheng, H., Liang, H., Zhang, X., Wang, S., Chen, S., Li, J., Fu, Y., Springer, N. M., Yang, H., Wang, J., Dai, J., Schnable, P. S. & Wang, J. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat Genet* **42**, 1027-1030, doi:http://www.nature.com/ng/journal/v42/n11/abs/ng.684.html - supplementary-information (2010).

242 Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T. L. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **494**, 234-237, doi:10.1038/nature11867 (2013).

243 Opitz, J. M. 2011 William Allan Award: development and evolution. *American journal of human genetics* **90**, 392-404, doi:10.1016/j.ajhg.2011.12.025 (2012).

244 Waddington, C. <http://en.wikipedia.org/wiki/Conrad_Hal_Waddington> (

245 Waddington, C. H. Canalization of development and genetic assimilation of acquired characters. *Nature* **183**, 1654-1655 (1959).

246 Waddington, C. H. The epigenotype. 1942. *International journal of epidemiology* **41**, 10-13, doi:10.1093/ije/dyr184 (2012).

247 McClaren, B. J., Metcalfe, S. A., Amor, D. J., Aitken, M. & Massie, J. A case for cystic fibrosis carrier testing in the general population. *Med J Aust* **194**, 208-209 (2011).

248 ACMG. *ACMG Releases Policy Statement on Noninvasive Prenatal Screening (NIPS)*, <http://www.acmg.net/docs/NIPS_Release.pdf> (2013).

249 Fan, H. C., Gu, W., Wang, J., Blumenfeld, Y. J., El-Sayed, Y. Y. & Quake, S. R. Non-invasive prenatal measurement of the fetal genome. *Nature* **487**, 320-324, doi:10.1038/nature11251 (2012).

250 Kitzman, J. O., Snyder, M. W., Ventura, M., Lewis, A. P., Qiu, R., Simmons, L. E., Gammill, H. S., Rubens, C. E., Santillan, D. A., Murray, J. C., Tabor, H. K., Bamshad, M. J., Eichler, E. E. & Shendure, J. Noninvasive whole-genome sequencing of a human fetus. *Sci Transl Med* **4**, 137ra176, doi:10.1126/scitranslmed.3004323 (2012).

251 Papageorgiou, E. A. & Patsalis, P. C. Maternal plasma sequencing: a powerful tool towards fetal whole genome recovery. *BMC Med* **11**, 56, doi:10.1186/1741-7015-11-56 (2013).

252 Cheng, P., M, D. C., Chen, C. F., Hoffmann, R. F., Armitage, R. & Deldin, P. J. Sleep-disordered breathing in major depressive disorder. *J Sleep Res*, doi:10.1111/jsr.12029 (2013).

253 Bianchi, D. W. From prenatal genomic diagnosis to fetal personalized medicine: progress and challenges. *Nat Med* **18**, 1041-1051, doi:10.1038/nm.2829 (2012).

254 Guedj, F. & Bianchi, D. W. Noninvasive prenatal testing creates an opportunity for antenatal treatment of down syndrome. *Prenat Diagn*, doi:10.1002/pd.4134 (2013).

255 Bates, B. R., Graham, D., Striley, K., Patterson, S., Arora, A. & Hamel-Lambert, J. Examining Antecedents of Caregivers' Access to Early Childhood Developmental Screening: Implications for Campaigns Promoting Use of Services in Appalachian Ohio. *Health Promot Pract*, doi:10.1177/1524839913479955 (2013).

256 McIntyre, L. L. Parent training for young children with developmental disabilities: randomized controlled trial. *Am J Ment Retard* **113**, 356-368, doi:10.1352/2008.113:356-368 (2008).

257 Rickards, A. L., Walstab, J. E., Wright-Rossi, R. A., Simpson, J. & Reddihough, D. S. A randomized, controlled trial of a home-based intervention program for children with autism and developmental delay. *J Dev Behav Pediatr* **28**, 308-316, doi:10.1097/DBP.0b013e318032792e (2007).

258 Rickards, A. L., Walstab, J. E., Wright-Rossi, R. A., Simpson, J. & Reddihough, D. S. One-year follow-up of the outcome of a randomized controlled trial of a home-based intervention programme for children with autism and developmental delay and their families. *Child Care Health Dev* **35**, 593-602, doi:10.1111/j.1365-2214.2009.00953.x (2009).

259 Salem, Y., Gropack, S. J., Coffin, D. & Godwin, E. M. Effectiveness of a low-cost virtual reality system for children with developmental delay: a preliminary randomised single-blind controlled trial. *Physiotherapy* **98**, 189-195, doi:10.1016/j.physio.2012.06.003 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

260 Velleman, S. L. & Mervis, C. B. Children with 7q11.23 Duplication Syndrome: Speech, Language, Cognitive, and Behavioral Characteristics and their Implications for Intervention. *Perspect Lang Learn Educ* **18**, 108-116, doi:10.1044/lle18.3.108 (2011).

261 Coulter, M. E., Miller, D. T., Harris, D. J., Hawley, P., Picker, J., Roberts, A. E., Sobeih, M. M. & Irons, M. Chromosomal microarray testing influences medical management. *Genet Med* **13**, 770-776, doi:10.1097/GIM.0b013e31821dd54a (2011).

262 Brawley, O. W. & Goldberg, P. *How we do harm : a doctor breaks ranks about being sick in America*. 1st edn, (St. Martin's Press, 2012).

263 Makary, M. *Unaccountable : what hospitals won't tell you and how transparency can revolutionize health care*. 1st U.S. edn, (Bloomsbury Press, 2012).

264 Brandt, D. S., Shinkunas, L., Hillis, S. L., Daack-Hirsch, S. E., Driessnack, M., Downing, N. R., Liu, M. F., Shah, L. L., Williams, J. K. & Simon, C. M. A Closer Look at the Recommended Criteria for Disclosing Genetic Results: Perspectives of Medical Genetic Specialists, Genomic Researchers, and Institutional Review Board Chairs. *J Genet Couns*, doi:10.1007/s10897-013-9583-5 (2013).

265 Hamamy, H. Consanguineous marriages : Preconception consultation in primary health care settings. *J Community Genet* **3**, 185-192, doi:10.1007/s12687-011-0072-y (2012).

266 Bittles, A. H. Genetics and global healthcare. *J R Coll Physicians Edinb* **43**, 7-10, doi:10.4997/jrcpe.2013.102 (2013).

267 Francke, U., Dijamco, C., Kiefer, A. K., Eriksson, N., Moiseff, B., Tung, J. Y. & Mountain, J. L. Dealing with the unexpected: consumer responses to direct-access BRCA mutation testing. *PeerJ* **1**, e8, doi:10.7717/peerj.8 (2013).

268 Chua, E. W. & Kennedy, M. A. Current State and Future Prospects of Direct-to-Consumer Pharmacogenetics. *Front Pharmacol* **3**, 152, doi:10.3389/fphar.2012.00152 (2012).

269 Sterckx, S., Cockbain, J., Howard, H., Huys, I. & Borry, P. "Trust is not something you can reclaim easily": patenting in the field of direct-to-consumer genetic testing. *Genet Med* **15**, 382-387, doi:10.1038/gim.2012.143 (2013).

270 Allain, D., Friedman, S. & Senter, L. Consumer awareness and attitudes about insurance discrimination post enactment of the Genetic Information Nondiscrimination Act. *Familial Cancer* **11**, 637-644, doi:10.1007/s10689-012-9564-0 (2012).

271 Bakker, E. Is the DNA Sequence the Gold Standard in Genetic Testing? Quality of Molecular Genetic Tests Assessed. *Clinical Chemistry* **52**, 557-558, doi:10.1373/clinchem.2005.066068 (2006).

272 *Individual Genome Sequencing (IGS) Test*, <http://www.illumina.com/clinical/illumina_clinical_laboratory.ilmn> (2013).

273 Hunter, L. E., Hopfer, C., Terry, S. F. & Coors, M. E. Reporting Actionable Research Results: Shared Secrets Can Save Lives. *Sci Transl Med* **4**, 143cm148, doi:10.1126/scitranslmed.3003958 (2012).

274 *23andMe*, <https://http://www.23andme.com/> (2013).

275 *Personal Genome Project*, <http://www.personalgenomes.org/about.html> (2013).

276 Fu, W., O/'Connor, T. D., Jun, G., Kang, H. M., Abecasis, G., Leal, S. M., Gabriel, S., Altshuler, D., Shendure, J., Nickerson, D. A., Bamshad, M. J., Project, N. E. S. & Akey, J. M. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216-220, doi:http://www.nature.com/nature/journal/v493/n7431/abs/nature11690.html - supplementary-information (2013).

277 An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65, doi:http://www.nature.com/nature/journal/v491/n7422/abs/nature11632.html - supplementary-information (2012).

278 McGuire, A. L. & Gibbs, R. A. No Longer De-Identified. *Science* **312**, 370-371, doi:10.1126/science.1125339 (2006).

279 Gymrek, M., McGuire, A. L., Golan, D., Halperin, E. & Erlich, Y. Identifying Personal Genomes by Surname Inference. *Science* **339**, 321-324, doi:10.1126/science.1229566 (2013).

280 Lowrance, W. W. & Collins, F. S. Identifiability in Genomic Research. *Science* **317**, 600-602, doi:10.1126/science.1147699 (2007).

281 Radick, G. in *Philosophy, Biology and Life* 21-47 (Cambridge University Press, 2005).

282 Radick, G. *Scientific Inheritance - an Inaugural Lecture from Gregory Radick* <http://www.youtube.com/watch?v=D3nyB2lqmRo> (2013).

283 Goddard, K. A., Whitlock, E. P., Berg, J. S., Williams, M. S., Webber, E. M., Webster, J. A., Lin, J. S., Schrader, K. A., Campos-Outcalt, D., Offit, K., Feigelson, H. S. & Hollombe, C. Description and pilot results from a novel method for evaluating return of incidental findings from next-generation sequencing technologies. *Genet Med*, doi:10.1038/gim.2013.37 (2013).

284 Berg, J. S., Adams, M., Nassar, N., Bizon, C., Lee, K., Schmitt, C. P., Wilhelmsen, K. C. & Evans, J. P. An informatics approach to analyzing the incidentalome. *Genet Med* **15**, 36-44, doi:10.1038/gim.2012.112 (2013).

285 Green, R. C., Berg, J. S., Berry, G. T., Biesecker, L. G., Dimmock, D. P., Evans, J. P., Grody, W. W., Hegde, M. R., Kalia, S., Korf, B. R., Krantz, I., McGuire, A. L., Miller, D. T., Murray, M. F., Nussbaum, R. L., Plon, S. E., Rehm, H. L. & Jacob,

bioRxiv preprint doi: https://doi.org/10.1101/000687; this version posted November 18, 2013.

H. J. Exploring concordance and discordance for return of incidental findings from clinical sequencing. *Genet Med* **14**, 405-410, doi:10.1038/gim.2012.21 (2012).

ACMG. (2013).

Mole. How we know II: bad dreams *Journal of Cell Science* **119**, 197-198 doi:doi:10.1242/jcs.02763 (2006).