Digenic inheritance in medical genetics

Alejandro A Schäffer

ABSTRACT

Digenic inheritance (DI) is the simplest form of inheritance for genetically complex diseases. By contrast with the thousands of reports that mutations in single genes cause human diseases, there are only dozens of human disease phenotypes with evidence for DI in some pedigrees. The advent of high-throughput sequencing (HTS) has made it simpler to identify monogenic disease causes and could similarly simplify proving DI because one can simultaneously find mutations in two genes in the same sample. However, through 2012, I could find only one example of human DI in which HTS was used; in that example, HTS found only the second of the two genes. To explore the gap between expectation and reality, I tried to collect all examples of human DI with a narrow definition and characterise them according to the types of evidence collected, and whether there has been replication. Two strong trends are that knowledge of candidate genes and knowledge of protein–protein interactions (PPIs) has been helpful in most published examples of human DI. By contrast, the positional method of genetic linkage analysis, has been mostly unsuccessful in identifying genes underlying human DI. Based on the empirical data, I suggest that combining HTS with growing networks of established PPIs may expedite future discoveries of human DI and strengthen the evidence for them.

INTRODUCTION

Digenic inheritance (DI) has fascinated geneticists since the early 20th century. In the early decades of studies on genetics, the term ‘epistasis’ was used by some to describe some forms of digenic inheritance, but in recent decades ‘epistasis’ has been used to describe a much broader category of locus–locus interactions in polygenic diseases, including but not limited to interactions of loci identified by genome-wide association studies. This review is a synthesis of knowledge about digenic inheritance in a narrow sense, not about epistasis in a broad sense.

Defrèze–Gussenhoven suggested more than 50 years ago that there would be many human disease pedigrees showing reduced penetrance when treated in genetic analysis as monogenic, but that the inheritance could be explained more accurately by a two-locus model. The first prediction came true, but few studies of pedigrees with incomplete penetrance consider two-locus analysis, even though good methods have been developed. In this context, ‘reduced penetrance’ means that while all or almost all affected pedigree members are modelled as having the mutant genotype at the primary locus, one or more relatives with the primary mutant genotype are unaffected; genetic modelling allows for the imperfect correspondence between genotype and phenotype.

The first report of DI in a human disease was in 1994 for retinitis pigmentosa (RP). This report was convincing because it included data from multiple pedigrees, and the protein products of the two genes had a known interaction. After 1994, there was a trickle of additional DI reports until 2001, which saw prominent reports of human DI in Bardet–Biedl syndrome (BBS), deafness and other phenotypes. These discoveries stimulated a trio of influential reviews in 2002–2003. Since 2002, discoveries of human DI have been appearing at a steady pace (see Discussion), but were not reviewed systematically, except for specific diseases, such as deafness and Hirschsprung’s disease.

The three reviews and other contemporaneous papers engaged in a lively but inconclusive debate on how to define human DI. Here, I use a narrow, operational definition that inheritance is digenic when the variant genotypes at two loci explain the phenotypes of some patients and their unaffected (or more mildly affected) relatives more clearly than the genotypes at one locus alone. This includes cases where both loci determine who is affected, a substantial change in severity, or a substantial change in age of onset. The definition includes cases in which one locus is the primary locus, and by itself has variable expressivity, as well as cases where the two loci are roughly equal in importance. I generally exclude cases where the inheritance is polygenic with many more than two loci involved. I generally exclude ‘modifier loci’ that have a modest effect on the phenotype and for which the evidence is only statistical. For diseases whose aetiology involves more than two genes, formalisms, such as Bayesian networks, may be needed to describe the role of each gene and its variants in the ‘cause’ of the disease.

In the prominent example of BBS and others in the section on five examples below, a simple deterministic model that explained some pedigrees proved to be too simple for all pedigrees. When large collections of pedigrees are available, probabilistic models that assign higher probabilities to patients who have more mutations will likely fit the collection of data better. The impetus to collect hundreds of patients and fit a statistical genotype–phenotype model comes after initial observations of one or a few pedigrees that fit simpler models of multigenic inheritance. Therefore, this review focuses attention on how to find the initial digenic patients and pedigrees.

Besides the lack of recent reviews, another stimulus for this review is the hypothesis that high-throughput DNA sequencing (HTS) would be an enabling technology to accelerate discoveries of human DI. Because HTS makes it possible to sequence many genes simultaneously, disease-relevant mutations in two genes can be discovered in a single experiment.
HTS does not solve the problem of deciding which mutations are relevant to the phenotype, and doing so is more difficult when the inheritance is digenic as compared with monogenic.

There has been one recent example of HTS enabling a proof of DI. The disease is facioscapulohumeral muscular dystrophy (FSHD) type 2. The primary locus for both type 1 and type 2 FSHD is DUX4, and that had been discovered by pre-HTS methods. In many pedigrees with type 2 FSHD, the penetrance of the DUX4 variant is incomplete. Therefore, Lemmers et al.16 sought a second locus via HTS. They found that heterozygous, rare variants in the gene SMCHD1 could explain the inheritance pattern in 21/26 individuals in various pedigrees. Patients with a variant at DUX4 and a variant at SMCHD1 are mostly affected, while patients with a variant at DUX4 and wild type at SMCHD1 are mostly unaffected. SMCHD1 controls epigenetic marks affecting gene expression, so the basis for the DI is likely to be a protein–DNA interaction between SMCHD1 and DUX4, affecting the expression of DUX4.

I could not find any other studies in which HTS has facilitated a discovery of human DI, though a second example was published online after this article was submitted.17 To investigate why, I started by building a catalogue of human DI examples. Next, I analysed what study designs had been tried. Then, I considered some of the more publicised examples to see if there was anything special about the most replicated cases of human DI. The successes and some not-so-successful examples suggest three lessons that may aid future studies of human DI. In the Discussion, I use an epistemological approach to suggest how HTS and other new technologies may be used to accelerate the pace of future discoveries.

COLLECTION OF EXAMPLES OF HUMAN DIGENIC INHERITANCE

To collect examples of human digenic inheritance, I used previous reviews,10–14 Online Mendelian Inheritance in Man (OMIM),18 PubMed, PubMed Central, and Google Scholar. I used the Citation Index to find more examples and to look for replications and refutations of previous publications. The search ended in January 2013. Some items in early reviews were excluded here because: they were for model organisms; they had been subsequently refuted; the evidence for the second locus looks weak; or the second locus is a modifier locus based primarily on statistical evidence. The collection of references on BBS and other ciliopathies is incomplete, since the evidence for and against DI in those disorders has been extensively explored elsewhere.10 12 19 20

All DI examples are collected in online supplementary table S1, along with two studies at the bottom in which possible DI of a multisystem syndrome turned out to be two different diseases. The examples that are not primarily replication studies are presented succinctly in table 1. For DI examples that have been repeatedly replicated, such as CDKN2A and MC1R in melanoma susceptibility,21 only a few replicating papers are included. The inheritance at each locus can be usually described as autosomal dominant (AD), autosomal recessive (AR), or X-linked recessive (XLR). The main exception is triallelic inheritance, explained below, in the subsection on BBS. Other examples of possible triallelic inheritance can be found in online supplementary table S1 by searching the two columns titled ‘Inh.’ for the word ‘triallelic’.

Ideal evidence for DI would include identification of the two genes involved in multiple pedigrees with multiple affected individuals in at least one pedigree. The evidence is strengthened by a comparison of the phenotypes of individuals having the mutations in both genes to the phenotypes of individuals with only one gene or the other gene mutated. Ideal digenic pedigrees may be hard to find. Therefore, in a few cases, such as Long QT syndrome (LQTS), the evidence accumulates over multiple studies.22–26 Various studies suggested DI based on one or a few patients without pedigree evidence. To evaluate whether genetic linkage analysis (GLA) is useful to find DI, I included studies in which strong evidence of two loci was found by linkage analysis, without requiring that the two genes have been found.

The data about each study include: the loci and genes, whether there was pedigree evidence, whether the study was replicated later, had internal replication only, or mostly replicated a prior study (see online supplementary table S1). Two additional useful pieces of information are: whether the loci are genetically linked, and whether there is any functional relationship between the two genes or their protein products. The functional relationships could be: protein–protein interaction (PPI), protein–DNA interaction, or being on a shared pathway without a known direct interaction. The importance of whether the loci are linked, and whether the genes have any interaction, is considered in the section entitled: ‘Three lessons for future studies of human digenic inheritance’.

COMMONLY USED STUDY DESIGNS

Two experimental study designs predominate among the reported cases of human digenic inheritance. These are illustrated abstractly in figure 1. I consider one alternative design in this section and another alternative design in the Discussion.

The majority of examples in table 1 and almost all examples in which both genes have been identified are based on a candidate gene (CG) design, proceeding as follows:

1. Identify a small set of genes \( G = \{g_1, g_2, \ldots\} \) that are mutated, or might be mutated, in monogenic forms of some disease, D, that has locus heterogeneity.
2. Use Sanger sequencing to sequence at least two of the genes in \( G \) in a set of patients with disease D and perhaps in their relatives.
3. Identify patients with mutations in two genes.

The evidence from the CG study design is more impressive when relatives having only one of the two genes mutated are unaffected or have a different phenotype than the patients with two genes mutated. Additional experiments to identify how the two genes/proteins interact or reproducing the DI in an animal model27 strengthen the evidence.

The CG study design (figure 1A) has been successful, but a limitation is that the genes to sequence must be selected in advance. DI in which mutations in each of the pair of genes lead to no phenotype is nearly impossible to detect by the CG study design.

A second study design, which avoids the need to preselect genes, is based on GLA. The GLA design (figure 1B) proceeds roughly as follows:

1. Identify one or more pedigrees with cases of a disease and preferably with evidence of reduced penetrance (eg, likely dominant inheritance in which some obligate mutation carriers are unaffected).
2. Genotype markers across the genome.
3. Analyse for linkage either one locus at a time or using two-locus linkage analysis.
4. (Ideal but rarely completed) By sequencing, find mutations in one gene in each of the linkage regions.

I could not find a single example where the sequencing step (4) was completed successfully to identify both genes on different chromosomes. Rotor syndrome is one recent successful example of GLA in which the two genes are tightly linked, so linkage analysis was done as if the disease is monogenic.28 Also, when the first gene is known, then linkage analysis to find the second locus can
<table>
<thead>
<tr>
<th>Disease</th>
<th>Gene/locus1</th>
<th>Gene/locus2</th>
<th>Replication</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long QT syndrome (LQTS)</td>
<td>KCNH2/7q</td>
<td>KCNQ1/11p</td>
<td>Yes</td>
<td>22</td>
</tr>
<tr>
<td>LQTS</td>
<td>various LQ genes</td>
<td>various LQ genes</td>
<td>Yes</td>
<td>23–26</td>
</tr>
<tr>
<td>Deafness</td>
<td>GJB2/13q</td>
<td>GJB6/13q</td>
<td>DNA level</td>
<td>9</td>
</tr>
<tr>
<td>Deafness suppression</td>
<td>TECTA/11q</td>
<td>1p</td>
<td>No</td>
<td>30–31</td>
</tr>
<tr>
<td>Pendred syndrome/deafness</td>
<td>SLC26A4/7q</td>
<td>FOX1/15q</td>
<td>No</td>
<td>67</td>
</tr>
<tr>
<td>Deafness suppression</td>
<td>1q</td>
<td>4q</td>
<td>No</td>
<td>72</td>
</tr>
<tr>
<td>Deafness</td>
<td>CDH23/10q</td>
<td>ATP2B2/3p</td>
<td>Partial</td>
<td>49</td>
</tr>
<tr>
<td>Usher syndrome</td>
<td>MYO7A/11q</td>
<td>3q</td>
<td>No</td>
<td>88</td>
</tr>
<tr>
<td>Usher syndrome</td>
<td>CDH23/10q</td>
<td>PCDH15/10q</td>
<td>Yes</td>
<td>27</td>
</tr>
<tr>
<td>Bartter’s syndrome (antenatal w/deafness)</td>
<td>CLCNKA/1p</td>
<td>CLCNKB/1p</td>
<td>Yes</td>
<td>90</td>
</tr>
<tr>
<td>Bardet–Biedl syndrome</td>
<td>BBS2/16q</td>
<td>various BBS loci</td>
<td>Yes</td>
<td>8</td>
</tr>
<tr>
<td>Bardet–Biedl syndrome</td>
<td>BBS4/15q</td>
<td>various BBS genes</td>
<td>Yes</td>
<td>61</td>
</tr>
<tr>
<td>Bardet–Biedl syndrome</td>
<td>BBS5/11q</td>
<td>various BBS genes</td>
<td>Yes</td>
<td>62</td>
</tr>
<tr>
<td>Joubert syndrome (+ciliopathies)</td>
<td>CEP4/17q</td>
<td>various genes</td>
<td>No</td>
<td>65</td>
</tr>
<tr>
<td>Leber’s congenital amaurosis (+ciliopathies)</td>
<td>CEP290/12q</td>
<td>MKKS/BBS6/20p</td>
<td>Internal</td>
<td>66</td>
</tr>
<tr>
<td>Short-rib polydactyly (ciliopathy)</td>
<td>NEK1/1q</td>
<td>DYNC2H1/11q</td>
<td>No</td>
<td>67</td>
</tr>
<tr>
<td>Nephronic syndrome</td>
<td>NPHS1/19q</td>
<td>NPHS2/1q</td>
<td>Yes</td>
<td>68</td>
</tr>
<tr>
<td>Hypogonadotropic hypogonadism</td>
<td>PROKR2/20q</td>
<td>KAL1/3p</td>
<td>Yes</td>
<td>69</td>
</tr>
<tr>
<td>Hypogonadotropic hypogonadism</td>
<td>FGFR1/8p</td>
<td>NSMF/9q</td>
<td>Yes</td>
<td>70</td>
</tr>
<tr>
<td>Hypogonadotropic hypogonadism</td>
<td>FGFR-related</td>
<td>FGFR-related</td>
<td>Internal</td>
<td>71</td>
</tr>
<tr>
<td>Hypogonadotropic hypo. (syndromic)</td>
<td>RNF216/7p</td>
<td>OTUD4/1q</td>
<td>No</td>
<td>17</td>
</tr>
<tr>
<td>Hirschsprung disease</td>
<td>RET/10q</td>
<td>EDNIRB/13q</td>
<td>Yes</td>
<td>91</td>
</tr>
<tr>
<td>Parkinson’s disease</td>
<td>PARK7/1p</td>
<td>PINK1/1p</td>
<td>No</td>
<td>92</td>
</tr>
<tr>
<td>Retinitis pigmentosa</td>
<td>PRPH2/6p</td>
<td>ROM1/11q</td>
<td>Internal</td>
<td>7</td>
</tr>
<tr>
<td>Glaucoma (earlier onset)</td>
<td>MYOC/1q</td>
<td>CYP1B1/10p</td>
<td>Yes</td>
<td>93</td>
</tr>
<tr>
<td>Waardenburg syndrome/albinism</td>
<td>MITF/3p</td>
<td>TYR/11q</td>
<td>Partial</td>
<td>94</td>
</tr>
<tr>
<td>Oculocutaneous albinism (OCA)</td>
<td>TYR/11q</td>
<td>OCA2/15q</td>
<td>Yes</td>
<td>95</td>
</tr>
<tr>
<td>Junctional epidermolysis bullosa</td>
<td>COL17A1/10q</td>
<td>LAMB3/1q</td>
<td>No</td>
<td>96</td>
</tr>
<tr>
<td>Disfibrinogenemia (slow clotting)</td>
<td>FGA/4q</td>
<td>FGFI/4q</td>
<td>No</td>
<td>97</td>
</tr>
<tr>
<td>Polycystic kidney disease</td>
<td>PKD1/16p</td>
<td>PKD2/1q</td>
<td>No</td>
<td>98</td>
</tr>
<tr>
<td>Holoprosencephaly</td>
<td>SHH/7q</td>
<td>TGF/118p</td>
<td>Internal</td>
<td>11</td>
</tr>
<tr>
<td>Familial hypercholest. suppression</td>
<td>LDLR/19p</td>
<td>13q</td>
<td>No</td>
<td>73</td>
</tr>
<tr>
<td>Cystinuria</td>
<td>SLCA/12p</td>
<td>SLCA9/19q</td>
<td>Yes</td>
<td>99</td>
</tr>
<tr>
<td>Hyperinsulininemia</td>
<td>PPARG/3p</td>
<td>PPP1R3A/7q</td>
<td>No</td>
<td>100</td>
</tr>
<tr>
<td>Hypercholanemia</td>
<td>TP2/9q</td>
<td>BAA1/7q</td>
<td>No</td>
<td>79</td>
</tr>
<tr>
<td>Pheochromocytoma</td>
<td>TMEM127/2q</td>
<td>16p</td>
<td>No</td>
<td>32–33</td>
</tr>
<tr>
<td>Familial exudative vitreoretinopathy</td>
<td>FZD4/11q</td>
<td>F5/1q</td>
<td>No</td>
<td>101</td>
</tr>
<tr>
<td>Factor VII thromboembolism</td>
<td>LMNAN1q</td>
<td>EMD/1q</td>
<td>Yes</td>
<td>106</td>
</tr>
<tr>
<td>Porphyria (acute)</td>
<td>various genes</td>
<td>HFE/6p</td>
<td>Pathway level</td>
<td>107</td>
</tr>
<tr>
<td>Porphyria (acute)</td>
<td>UROD/1p</td>
<td>HMB5/11q</td>
<td>Pathway level</td>
<td>108</td>
</tr>
<tr>
<td>Porphyria (acute)</td>
<td>CPX/3q</td>
<td>ALAD/9q</td>
<td>Pathway level</td>
<td>109</td>
</tr>
<tr>
<td>Epilepsy w/febrile seizures</td>
<td>1q</td>
<td>18q</td>
<td>No</td>
<td>110</td>
</tr>
<tr>
<td>Hemachromatosis</td>
<td>HFE/6p</td>
<td>HAMP/19q</td>
<td>No</td>
<td>111</td>
</tr>
<tr>
<td>Progressive external ophthalmoplegia</td>
<td>C10orf2/10q</td>
<td>POLG/15q</td>
<td>No</td>
<td>112</td>
</tr>
<tr>
<td>Epilepsy w/febrile seizures</td>
<td>SCN1A/2q</td>
<td>SCN2A/2q</td>
<td>No</td>
<td>113</td>
</tr>
<tr>
<td>Photosensitivity in epilepsy</td>
<td>7q</td>
<td>16p</td>
<td>No</td>
<td>114</td>
</tr>
<tr>
<td>Split-hand/foot malformation</td>
<td>1q</td>
<td>6q</td>
<td>No</td>
<td>115</td>
</tr>
<tr>
<td>Iminoglycinuria</td>
<td>SLC36A2/5q</td>
<td>SLC6A20/3p</td>
<td>Internal</td>
<td>116</td>
</tr>
<tr>
<td>Hereditary motor neuropathy</td>
<td>1p</td>
<td>8q</td>
<td>No</td>
<td>117</td>
</tr>
<tr>
<td>Limb-girdle muscular dystrophy</td>
<td>SGCB/1q</td>
<td>SGCD/5q</td>
<td>No</td>
<td>118</td>
</tr>
<tr>
<td>Ullrich congenital muscular dystrophy</td>
<td>COL6A1/21q</td>
<td>COL6A2/1q</td>
<td>No</td>
<td>119</td>
</tr>
<tr>
<td>Pseudoxanthoma elasticum</td>
<td>ABC6/16p</td>
<td>GGGCV2p</td>
<td>No</td>
<td>120</td>
</tr>
<tr>
<td>Hereditary motor neuropathy</td>
<td>DSC2/11q</td>
<td>16p</td>
<td>No</td>
<td>121</td>
</tr>
<tr>
<td>Cleft lip</td>
<td>1q</td>
<td>2p</td>
<td>No</td>
<td>122</td>
</tr>
<tr>
<td>Fuchs corneal dystrophy</td>
<td>ZEB1/10p</td>
<td>9p</td>
<td>No</td>
<td>29</td>
</tr>
</tbody>
</table>

Continued
Thus, the difficulties appear to be due to some unknown gap between theory and reality. One possibility is that human pedigrees have not been found where the second gene is not functional. However, some of these patients may be better classified as having recessive, monogenic inheritance.46

Deafness
Similar to RP, deafness is an excellent candidate for DI because there are dozens of known genes that cause monogenic deafness. Additionally, there are animal models of DI for either non-syndromic or syndromic deafness.14 44 Considerable information is known about protein complexes that function in the inner ear, and hence, pairs of proteins in these complexes are good candidates for DI. Finally, in some societies, there is assortative mating among deaf individuals or close relatives.45 Assortative mating may lead to pedigrees in which multiple deafness-related alleles cosegregate.46

Table 1 shows five different entries for deafness, three for Usher syndrome (deafness and blindness), and one for a form of Bartter syndrome (salt wasting) that includes deafness. Perhaps the most compelling among these is the combination of CDH23 and PCDH15 causing digenic Usher syndrome because it has been replicated and there is an animal model.27 47 However, some of these patients may be better classified as having recessive, monogenic inheritance at PCDH15; moreover, PCDH15 has additional exons that were not sequenced in those patients found to have one mutation in each of PCDH15 and CDH23, and therefore, a second PCDH15 mutation may have been missed.48 An overlapping case where the human DI matches an animal model is the combination of CDH23 and ATP2B2 in a single human pedigree.49

The most interesting example of DI in deafness is the combination of GJB2 and GJB650–52 both of which are also monogenic deafness genes encoding connexins that function in a complex. The evidence for DI among the genes encoding proteins in this connexin complex was strengthened by a report of DI in deafness with mutations in GJB2 and GJB3.53 However, Rodriguez-Paris et al54 55 have shown that the GJB2/GJB6 case is actually monogenic recessive GJB2-caused deafness at the RNA and protein levels. The GJB6 mutations are deletions that inactivate the second GJB2 allele, which is nearby on chromosome 13. Further evidence that a regulatory element outside GJB2 regulates the expression of GJB2 and GJB6 is given via allele-specific

Table 1 Continued

<table>
<thead>
<tr>
<th>Disease</th>
<th>Gene/locus1</th>
<th>Gene/locus2</th>
<th>Replication</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axenfeld–Rieger syndrome</td>
<td>PITX2/4q</td>
<td>FOXC1/6p</td>
<td>No</td>
<td>123</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>MUTYH/1p</td>
<td>DGG1/3p</td>
<td>No</td>
<td>124</td>
</tr>
<tr>
<td>Rotor syndrome (hyperbilirubinemia)</td>
<td>SLC1B1/12p</td>
<td>SLC1B1/12p</td>
<td>Internal</td>
<td>28</td>
</tr>
<tr>
<td>Dent’s disease</td>
<td>CLCN5/4p</td>
<td>OCN/3q</td>
<td>No</td>
<td>78</td>
</tr>
<tr>
<td>Facciopapulopheralumus. dystrophy</td>
<td>DUX4/4q</td>
<td>SMCHD1/18p</td>
<td>Internal</td>
<td>16</td>
</tr>
<tr>
<td>Epidermolysis bullosa simplex</td>
<td>KRTS/12q</td>
<td>KRT14/17q</td>
<td>No</td>
<td>125</td>
</tr>
<tr>
<td>Melanoma susceptibility</td>
<td>CDKN2A/9p</td>
<td>MCT1/16q</td>
<td>Yes</td>
<td>21</td>
</tr>
<tr>
<td>Hypothyroidism (nonsyndromic)</td>
<td>CDH3/16q</td>
<td>12q</td>
<td>Internal</td>
<td>75</td>
</tr>
</tbody>
</table>

A larger table including more rows with overlapping findings and more columns, such as the mode of inheritance at each locus, can be found in online supplementary table S1. Gene names are currently Hug Genome Nomenclature Committee-approved names, not necessarily the gene names in the original publication. In some cases, the original study (eg, reference 17) reported multiple pairs of genes, and there is one representative pair in the table. Options for the Replication column (and their meanings) include: Yes (replicated in a later study), Internal (multiple pedigrees in the original study but no later study), Partial (one of the genes participates in some other documented human DI), and their relatives.

be done using a monogenic linkage analysis, conditional on the mutation status or haplotype status at the first locus.6 29 Two examples where the GLA design succeeded to find the gene at one of two loci are in deafness30 31 and pheochromocytomas.

Since GLA has been repeatedly successful in setting up the identification of genes causing monogenic diseases, the failure of the GLA design in DI merits investigation. There exist at least three software packages that can do two-locus linkage analysis: TLINKAGE,34 SUPERLINK35 and GENEHUNTER-TWOLOCUS.5 Most of the studies that used two-locus linkage analysis used GENEHUNTER-TWOLOCUS. The mathematical basis for two-locus linkage analysis is that the test statistic, such as a Logarithm of Odds (LOD) score or an NonParametric Linkage (NPL) score typically used to find single disease loci, can be generalised to simultaneous analysis of two disease loci and the genotypes at unlinked marker loci can be combined.3 There has been considerable research concerning penetrance models for two-locus linkage analysis, which are needed when the test statistic is the two-locus LOD score.1 4 36–39 Thus, the difficulty appears to be due to some unknown gap between theory and reality. One possibility is that human pedigrees with adequate power are hard to find.

The advent of HTS facilitates a third study design (HT): 1. Sequence the exomes or the genomes of a series of patients and their relatives. 2. Identify pairs of genes (g1, g2) that are recurrently mutated in patients. 3. Compare the sequences of g1 and g2 in patients and their unaffected relatives.

Since the mutated genes g1 and g2 may not be functional candidates, some functional experiments would be needed to show the molecular basis of the DI. A major difficulty in human studies is that one has to sample the relatives who are available. Animal models can mitigate this difficulty. One advantage of the HT design is that one can reconstruct haplotypes to see if multiple nearby mutations are on the same or opposite alleles40–42 which is relevant below in the section on three loci.

I could not find a DI study in which the HT design had identified both genes. Gallinane et al43 found two disease-causing mutations in a single experiment, but that patient had two monogenic diseases. In the example of FSHD, the first gene, DUX4, had already been found before HTS was applied to find the second gene.16

FIVE EXAMPLES WITH POSSIBLE REPLICATION
In this section, I summarise the understanding of possible digenic inheritance of five phenotypes where one could consider that the original claim has been replicated. Some are selected to foreshadow later sections. All five phenotypes occur usually in monogenic forms with locus heterogeneity. Via the CG paradigm, patients with mutations in pairs of the known genes were identified. Surprisingly, I could not find replications of the seminal finding of DI in non-syndromic RP, though there are many known RP genes. BBS, which is the most studied phenotype with DI does include retinal disease in the phenotype.
expression assays of a unique deafness-associated haplotype on 13q. The \textit{GJB2/GJB3} example cannot be similarly refuted because \textit{GJB3} is on chromosome 1.

**Long QT syndrome**

LQTS is a disease in which patients may suffer cardiac arrhythmias and sudden death. Inheritance is often autosomal dominant, but many pedigrees have incomplete penetrance. LQTS has substantial locus heterogeneity and several pairs of the protein products of LQTS genes interact. The combination of locus heterogeneity, PPIs, and variable expressivity of single gene mutations makes LQTS a good candidate for the model that what looks like reduced penetrance under monogenic inheritance masks DI. The way to follow-up is to compare DNA sequences of affected and unaffected relatives sharing the disease-associated mutation in the first gene. The follow-up can be done by looking either at a few CGs by Sanger sequencing, or many genes by HTS.

For LQTS, the CG design led to the finding that many LQTS patients have mutations in two of the known genes, such as \textit{KCNQ1/KCNE1}, \textit{KCNQ1/KCNH2}, \textit{KCNH2/KCNE1}, \textit{SCN5A/KCNE1} and other pairs.\textsuperscript{22–26} The penetrance could be increased either by having a second mutation in one LQTS gene or two heterozygous mutations in different genes.\textsuperscript{23 24} All patients with two mutations manifest the disease, often with earlier onset; the distinction between the two-mutation individuals and the one mutation individuals is statistically significant.\textsuperscript{23}

**BBS and other ciliopathies**

The phenotype of BBS typically includes six aspects: renal anomalies, polydactyly, obesity, retinal defects, developmental...
delay and hypogonadism. Patients are usually diagnosed when at least four symptoms are detected. There is phenotypic overlap between BBS and many other syndromes, including the next two examples. Considering the unusual combination of symptoms, it is surprising that there have been at least 15 genes identified that can cause monogenic BBS with AR inheritance.

When the first BBS genes were found, the function of the encoded proteins was poorly understood. Later studies have shown that these proteins are involved in the formation and function of cilia, primitive sensory organelles present in many cell types. Primary cilia are non-motile, while other cilia are motile because they have a flexible microtubule configuration. BBS and overlapping syndromes, such as Joubert syndrome and Meckel–Gruber syndrome, are called ‘ciliopathies’. The phenotypic spectrum of ciliopathies is broad and may include holoprosencephaly, for which DI has been proposed.13 Biochemical studies identified two protein complexes containing seven and three of the 15+ BBS proteins. The protein complexes increase the potential for DI as one could imagine that defects in two of the proteins would be more deleterious than a defect in only one protein.

Before the cilia function and BBS complexes were discovered, Katsanis and colleagues energised the study of BBS and DI by proposing that the inheritance of BBS is triallelic in some pedigrees. Triallelic inheritance means that any combination of three deleterious alleles at two BBS loci, but not three heterozygous mutations at three loci, are sufficient to cause BBS. Triallelic inheritance was also supposed to indicate that there would be individuals with ‘only’ a biallelic mutation at one BBS locus who would have no phenotype or a milder phenotype. The triallelic inheritance hypothesis has been controversial because few pedigrees in which three mutant alleles are necessary have been reported.20 Early attempts to test the triallelic inheritance hypothesis found that only a small minority of BBS families had exactly two mutant alleles at one locus and a third mutant allele at a second locus. The distribution of mutations is more variable, and the early studies are hard to interpret now because they could only test the subset of BBS genes known at the time of the study. The finding that many BBS patients have mutations in two or more BBS genes has been replicated many times (table 1, see online supplementary table S1). Some BBS patients have as many as five variant alleles in different BBS genes.

Some have argued that the large number of BBS genes and high carrier frequencies in some populations suffice to explain the high frequency of patients with two or more BBS genes, without claiming DI.20 The weakness in this argument is that it could be even more applicable to diseases such as blindness, deafness and heart disease that have high locus heterogeneity, but only some specific instances of DI as described above. More problematic to the argument for DI in BBS is that as more patients with mutations in two BBS genes have been discovered, no pattern has emerged to explain which pairs of genes have mutations simultaneously. One could have hypothesised either a ‘logical AND’ model (the two proteins mutated should be preferentially in the same protein complex) or a ‘logical OR’ model (the two proteins mutated should be preferentially in different ciliary protein complexes), but neither model fits the BBS mutation data.

The identification of possible DI in BBS has stimulated the search for DI in diseases with phenotypic overlap (see the next two subsections). It has also stimulated the search for modifier genes and for examples of DI in other ciliopathies.

**Nephrotic syndrome**

Nephrotic syndrome is a kidney disease in which essential proteins are lost into urine. There is phenotypic similarity with the renal aspect of BBS and other ciliopathies, such as Joubert syndrome. Two of the various monogenic forms of nephrotic syndrome are due to mutations in *NPHS1* on 19q encoding nephrin and *NPHS2* on 1q, encoding podocin. Koziell et al38 identified three families in which there is triallelic inheritance, and those individuals with three deleterious alleles have a more severe form called ‘focal segmental glomerulosclerosis’ (FSGS). The two proteins, nephrin and podocin, have a direct interaction. This finding was replicated exactly and in a more general form by finding FSGS patients with three deleterious alleles in several pairs of CGs: *NPHS1/NPHS2, CD2AP/NPHS2, WT1/NPHSA* (see online supplementary table S1). It is interesting that the initial discovery was made in a kidney disease shortly after Katsanis et al proposed triallelic inheritance for BBS. It shows, retrospectively, how one finding of DI might provide impetus for another. The nephrotic syndrome example and the next example suggest the hypothesis that diseases with weak phenotypic similarity to BBS may be good candidates to have DI.

**Hypogonadotropic hypogonadism**

Hypogonadotropic hypogonadism (HH) is diminished function of the sexual organs associated with deficient secretion or action of the hypothalamic gonadotropin-releasing hormone (GnRH), which controls the pituitary gonadotropins and, thereby, gonadal function. The non-syndromic form is called ‘idiopathic HH’ (IHH). There is also a widely studied syndromic form (different from BBS) called Kallman syndrome in which HH is combined with anosmia.

The initial report of DI in HH focused on cases with mutations in the ligand-receptor gene pair *PROKR2* and *PROKR2*, and also reported one patient with heterogeneous mutations in both *PROKR2* and the known gene Kallman syndrome gene *KAL1*.69 The pairing of PROKR2 and PROKR2 is understandable since they form a receptor-ligand pair, but the mechanism of *PROKR2/KAL1* digenic inheritance remains unclear.

Pitteloud et al60 used the CG design with additional HH genes and found more examples of DI including the gene pairs *FGFR1/NSMF* and *FGFR1/GNHR*. The general finding of DI in HH has been replicated multiple times (see online supplementary table S1). However, the number of known patients with two genes mutated is small relative to the number of CGs. Thus, as in BBS, no pattern as to which pairs of genes are mutated together is discernible.

While this manuscript was under review, two more studies showing digenic inheritance in HH were published. By a generalisation of the CG design, Miraoui et al71 showed that some non-syndromic HH patients and Kallman syndrome patients have heterogeneous mutations in two genes in an FGFR8-related pathway. Using HTS, Margolin et al67 found homozygous mutations in *RNFL216* and *OTUD4* in three consanguineous siblings with a syndromic form of HH. Using a zebrafish model, they showed that RNFL216 and OTUD4 have a functional interaction, but they could not find any functional relationship between RNFL216 and OTUD4 and the genes mutated in non-syndromic HH.

**THREE LESSONS FOR FUTURE STUDIES OF HUMAN DIGENIC INHERITANCE**

From the catalogue of examples of digenic inheritance, three lessons can be derived to inform future studies. The first two
are subtle enough that they were not explicitly highlighted in previous reviews. The third lesson is not new, but some of its implications have not been mentioned in previous reviews and are explored in the Discussion.

**Lesson 1: in the digenic inheritance examples found to date, the variant genotype at the second locus usually increases disease risk**

The definition of DI in the Introduction does not specify whether the variant genotype at each locus increases or decreases the disease risk. In the study of monogenic diseases, it is usually understood that the variant genotypes at the disease locus do increase the risk. One can extend this assumption to require that in DI, the locus designated as first also has the property that the variant genotype increases the disease risk. It is possible, however, that the variant genotype at the second locus decreases the disease risk. In some early definitions of epistasis, it was required that the second locus suppresses the (trait-causing) effect of the first locus.1

Theory differs from practice in the role of the second locus because table 1 shows only three examples of human DI in which the variant genotype at the second locus is definitively suppressive. The first example is deafness in which the first locus is recessive and on 1q, and another recessive locus on 4q cancels the effect of the first. This example was found by GLA of a large pedigree. The finding has not been replicated, and the genes underlying the two loci have not been identified. The second example is familial hypercholesterolaemia with a primary mutation in the LDLR gene on 19p and a recessive locus on 13q that mitigates the effects of the LDLR mutation. In this example, like the first, the suppressive locus was found by linkage analysis of a single pedigree, and the gene has not been reported, but there is other evidence of a cholesterol-related locus on 13q. In the third example, the disease is hypothriosis due primarily to mutations in CDH3.57 Previously reported cases of hypothriosis and mutations in CDH3 are all syndromic. In two pedigrees, a locus on 12p identified by GLA mitigates the hypothriosis to make it non-syndromic.58

Recently, Rachel and colleagues suggested a possible fourth example.66 In this example, the two genes, CEP290 and MKKS (also known as BS6), were identified by the CG method and MKKS has already been suggested to participate in DI of BBS. The disease is Leber’s Congenital Amaurosis (LCA) that is often caused by biallelic mutations in CEP290. Biallelic mutations in CEP290 cause a spectrum of ciliopathies, along which LCA is mild because it affects only the eyes. A surprising percentage of LCA patients had heterozygous mutations in MKKS.66 Rachel et al proposed that the MKKS mutations mitigate the effect of the CEP290 mutations, perhaps ‘reducing’ the disease severity. This study showed that the two proteins, CEP290 and MKKS, have a direct interaction and constructed a mouse model supporting the DI. The last piece of the proof, which is not reported in the study, would be human pedigrees in which multiple relatives have the same biallelic CEP290 mutations, and relatives with an MKKS mutation have a milder phenotype than relatives without an MKKS mutation.

The predominance of cases in which the second locus variant genotype increases risk reflects a bias of the CG design. If more cases of DI are found by HTS, then a greater percentage may have a second locus that reduces the risk. When the variant genotype at the second locus reduces the risk, that variant genotype is going to be found in unaffected or more mildly affected individuals. Therefore, when using the GLA design or HTS or other designs, it is important to sequence unaffected and mildly affected relatives.

Another approach to identifying a second locus that decreases the risk caused by the variant genotype at the first locus is to compare expression of genes in affected and unaffected relatives with the mutant genotype at the first locus, and this was attempted with some success for spinal muscular atrophy.6 One advantage of this approach is that it is unbiased as to which set of relatives will have the unusual expression that is sought. A second advantage of the expression approach is that if the differential expression is found, then that result is closer to a functional experiment than sequence differences would be.6 However, the corresponding disadvantage is that it may be difficult to determine whether the expression difference between ‘affecteds’ and ‘unaffecteds’ is due to nearby (in cis) sequence differences, or due to differences in some other unlinked gene (in trans).76

**Lesson 2: when the two loci are linked, the proof is more complicated**

A disproportionate number of the locus pairs in online supplementary table S1 are genetically linked. This includes pairs that are closely linked (SLCO1B1 and SLCO1B3 in Rotor syndrome) and examples with weaker linkage (LRP5 and FZD4 in familial exudative vitreoretinopathy). When the two mutations are on the same haplotype, the linkage is weak, one has a chance to find crossover events between the two genes. If such a crossover is present, then a close relative may have only one of the two gene mutations, and one can compare phenotypes between the individuals having one gene mutated and the individuals having both genes mutated. It is useful to divide the linked situations into four categories by inheritance.

The first category is AR inheritance at both loci (eg, Rotor syndrome). In this category, it is hard to prove that a biallelic mutation in one gene does not suffice to cause the disease. In the case of Rotor syndrome, the proof included multiple pedigrees in which all patients had biallelic mutations in both genes, animal models and identifying other human subjects who had biallelic mutations in only one of the genes and were unaffected.28

The second category is AD or XLR inheritance at both loci with the two mutations on the same haplotype (in cis). An XLR example is Dent’s disease (CLCN5 and OCRL).87 Again, it is difficult to prove that one mutation/gene does not suffice to cause the disease. Another problem for AD inheritance is to evaluate whether all patients have both mutations in cis, and if so, why? It should not matter at the protein level whether the mutations are in cis or in trans, but it does matter for ascertainment. If the mutations are in cis, then they can be transmitted over multiple generations, and the inheritance will appear to be AD (figure 2A). One can ascertain large pedigrees that will achieve high LOD scores assuming a single dominant locus. For such pedigrees, HTS should help find the DI because HTS can find both mutated genes on the haplotype in a single experiment.

The third category is AD inheritance at both loci with the two mutations on opposite haplotypes (in trans). This differs from the second category because the inheritance will appear to be AR (figure 2B). The affected children would typically all be in one generation. The proof can be easier than in the in cis category because, typically, the patient(s) would inherit one mutation from each parent and the parents would be unaffected or have a milder phenotype. The proof can be harder because it is harder to find large pedigrees. If all the paired mutations are in
not been replicated in other populations, so genetic drift may have brought the two mutations together.

In a study design that combines GLA with HTS, there would be a possibility of finding two mutated genes in the interval of genetic linkage. In this circumstance, investigators may apply Occam’s razor and try to ‘pin the blame’ on one gene. Examples in table 1 show that this reductionism can be flawed in two different ways. Either the inheritance can be digenic or there can be two different diseases in the pedigree each caused by a different gene (eg, cone rod dystrophy and deafness).

Lesson 3: protein-protein interactions are an important type of evidence for digenic inheritance

Many of the entries in online supplementary table S1 for which both genes are known are associated with a direct PPI between the gene products. In some cases, the investigators did the PPI experiments themselves because the interaction was not in a database of PPIs. In principle, having two mutations in interacting proteins could be either a ‘double hit’ or compensatory. Lesson 1 is that the double hit situation is much more common than the compensatory situation.

Shoemaker and Panchenko reviewed both in vitro methods for finding new PPIs and databases for searching known PPIs. Laboratory methods include: nuclear magnetic resonance, yeast two-hybrid, coimmunoprecipitation, tandem affinity-purification mass spectroscopy (TAP-MS), protein microarrays, fluorescence resonance energy transfer, atomic force microscopy and others. Useful databases of known interactions include Biogrid, MINT, HPRD, STRING, IntAct. A useful resource that collects and organises other database resources is iRefIndex (http://irefindex.uio.no/wiki/iRefIndex), which can be searched using iRefWeb (http://wodaklab.org/iRefWeb). Two limitations of the iRefIndex data downloads are: (1) they refer to genes by UniProt records, which change over time and (2) when the same interaction is included in various sources, the duplicates are not necessarily consolidated. To address these limitations, Stojimirović and Yu developed a parser ppiTrim whose results identify the genes according to their more stable integer identifiers in NCBI’s Entrez gene database. The ppiTrim output files are available at ftp://ftp.ncbi.nih.gov/pub/gmbpmn/ppiTrim/datasets/.

The files whose names start with 9606 contain human data. One complication in evaluating PPI data is that two proteins may function together in a complex, without a direct interaction.

DISCUSSION

Because HTS sequences many genes in the same experiment, HTS leads frequently to the discovery that a patient has variants in multiple genes that are potentially disease-associated. When the number of patients is small or there is only one pedigree, even sophisticated bioinformatics filtering methods applied to HTS data can leave two or more candidate causal genes and variants. It may be advisable to consider the possibility of DI in such multicandidate cases, especially if the phenotype is novel. Geneticists using Sanger sequencing (one gene at a time) often stopped looking for mutations even if the genotype–phenotype correlation based on one mutant gene was imperfect, and the finding of an additional mutant gene could explain the observed phenotypes better via DI. In this review, we focused on the distinction between two genes mutated versus one gene mutated because in making that distinction, the methods of proving causality change. The medical geneticist is faced with the general question: Do the variants at both genes together explain the phenotype better than the variant at one gene? Sometimes, the
answer will be ‘yes, because the patient has two monogenic diseases simultaneously’. Here, the focus is on cases where the answer is ‘yes, because there is DI of one disease’. I was surprised to find only one study through 2012 where HTS led to discovery of DI. Why so few?

There are three overlapping reasons. First, it is possible that there are not that many cases of human DI. Figure 3 shows the number of new (not replication) reports per year of DI in table 1. The rate of discovery has not increased since 2001. Since the number of monogenic diseases with locus heterogeneity is increasing, and the number of genes contributing to the heterogeneity is increasing, one would expect the number of cases of DI detected by the CG design to be increasing as well, but this is not so. DI is near one end of a spectrum of mechanisms by which combinations of mutations increase disease risk; as more and more of these disease mechanisms are discovered, attention to DI may be diluted.

A second possibility is that more cases of DI involve PPIs and that should be the starting point to find the genetic evidence. Badano et al pioneered the following PPI design to find cases of DI:

1. Choose g1 encoding p1 as a gene of interest in disease, D, based on past discoveries.
2. Use extensive yeast two-hybrid assays to find protein partners of p1.
3. For each partner p1 (i>1) encoded by gene gi, sequence gi in patients who carry mutations in gi.

HTS increases the throughput at step 3 because all the genes can be sequenced in parallel. Techniques more reliable than yeast 2-hybrid assays have been developed to find protein partners. In silico databases of known and predicted protein interactions are growing rapidly, making it possible to search among all genes mutated, for pairs of genes encoding protein partners. During my formal literature search, I could not find a second instance in which the PPI design was used to find human DI, but an interesting example in HH was published while the manuscript was being refereed. Some of the later discoveries in online supplementary table S1 could have been made by the PPI design, but were made by the CG design instead.

This suggests a third possible explanation. The complexity of DI transcends the genetics. To construct a compelling proof that the inheritance is digenic rather than monogenic may require a multidisciplinary team that can apply techniques to understand the two genes and proteins specifically and their interaction. If we consider two of the more exciting findings of 2012, the techniques they used include: double knockout mice, morpholino studies in zebrafish, genotyping and haplotype analysis, expression and RNA interference experiments, methylation studies, splicing experiments, chromatin immunoprecipitation, electron microscopy, yeast two-hybrid experiments, transfection of genes and so on. It is challenging to assemble a scientific team with expertise in all these procedures. The need to assemble these multidisciplinary teams could explain the predominance of the ciliopathies among the studies of DI. Several research groups studying ciliopathies have combined animal models and extensive cell biology experiments in the same study. Once such a team is assembled, there may be other exciting problems to work on instead of identifying new examples of human DI, such as defining the PPIs and biochemical pathways that underlie the DI. In particular, we seem to be closer to a consensus about syllogisms needed to prove PPIs than to a consensus on the syllogisms to give a strong proof of DI. In the case of GJB2/GJB6 and deafness, a statistical method of proof (high rate of co-occurrence of heterozygous mutations), and evidence for interaction of the two proteins turned out to be incomplete proof.

Development of commonly accepted rules of proof in medical genetics has been a slow process. Even in areas such as GLA and genome-wide association studies, in which rigorous statistics can be applied, it took years to establish standards (LOD score thresholds, NPL score thresholds, association p values and q values) of proof. Establishing standards of proof may be at least twice as hard for DI. For example, one concept that was not applied consistently among the studies cited in online supplementary table S1 is that a proof of DI can be strengthened by comparing the genotypes at the two disease-associated loci genes of the affected individuals to the genotypes of as many unaffected first-degree relatives as possible.

In conclusion, the selection here of known human DI examples provides a basis to identify new examples. Key ingredients to a convincing proof of DI include: evidence of protein–protein or protein–DNA interaction for the two proteins or genes, pedigree data, animal models or very specific functional experiments. HTS is a tool to identify quickly the possible genes in a case of DI, especially when the genes are not obvious candidates, but HTS alone does not provide these three key ingredients to proving the mode of inheritance.

Acknowledgements This research was supported by the Intramural Research Program of the National Institutes of Health (NIH), NLM. Thanks to my NIH colleagues Drs. Andrew Cullinane, Thomas Friedman, Marjan Huizing, Anna Parchenko and Aleksandar Stoimirow for useful suggestions. Thanks to Daniel Schäffer (Takoma Park Middle School Magnet Program) for help with the figures. Thanks to two anonymous referees who made numerous insightful suggestions, including several additional pertinent references, which tangibly improved this review.

Contributors AA5 did the research and wrote the manuscript.
Funding National Institutes of Health, Intramural Research Program.
Competing interests None.
Provenance and peer review Commissioned; externally peer reviewed.
Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

REFERENCES
May also participate in triallelic inheritance.