Published in final edited form as:
Mol Genet Metab. 2012 April ; 105(4): 658–664. doi:10.1016/j.ymgme.2012.01.005.

Utilizing prospective sequence analysis of *SHH, ZIC2, SIX3* and *TGIF* in holoprosencephaly probands to describe the parameters limiting the observed frequency of mutant gene x gene interactions

Erich Roessler[a,b], **Jorge I. Vélez**[a,b], **Nan Zhou**[a], and **Maximilian Muenke**[a,*]
[a]Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

Abstract

Clinical molecular diagnostic centers routinely screen *SHH, ZIC2, SIX3* and *TGIF* for mutations that can help to explain holoprosencephaly and related brain malformations. Here we report a prospective Sanger sequence analysis of 189 unrelated probands referred to our diagnostic lab for genetic testing. We identified 28 novel unique mutations in this group (15%) and no instances of deleterious mutations in two genes in the same subject. Our result extends that of other diagnostic centers and suggests that among the aggregate 475 prospectively sequenced holoprosencephaly probands there is negligible evidence for direct gene-gene interactions among these tested genes. We model the predictions of the observed mutation frequency in the context of the hypothesis that gene x gene interactions are a prerequisite for forebrain malformations, i.e. the “multiple-hit” hypothesis. We conclude that such a direct interaction would be expected to be rare and that more subtle genetic and environmental interactions are a better explanation for the clinically observed inter- and intra-familial variability.

Keywords

Holoprosencephaly; digenic inheritance; mutation spectrum; *SHH*; *ZIC2*; *SIX3*; *TGIF*.

INTRODUCTION

Holoprosencephaly (HPE) is the most common malformation of the forebrain in humans and is considered clinically and molecularly to be an autosomal dominant disorder [1,2]. However, there has been considerable discussion in the published literature of the requirement for additional genetic and environmental co-morbid factors in order to understand and explain the extensive inter- and intra-familial variability [e.g. 3,4,5]. Alterations in the *SHH* gene [3,6,7, reviewed in 2; MIM *600725; HPE3 #142945] attributable to chromosomal rearrangements [8], microdeletions [9–12; Pineda-Alvarez and

*Correspondence to Maximilian Muenke, Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, 35 Convent Drive, MSC 3717, Building 35, Room 1B-203, Bethesda, MD 20892-3717, Tel.:(301) 402-8167, Fax.:(301) 480-7876, mmuenke@nhgri.nih.gov.
[b]These authors contributed equally to this work

The authors have indicated they have no financial interest to disclose.

Muenke, unpublished], or more commonly point mutations account for at least 6–8% of heterozygous sequence-detectable HPE-related genetic variation and serve as the prototypical type of HPE mutation [13–15; reviewed in 1, 2]. Additionally, in many cases the molecular alteration presents as a familial autosomal dominant form, often involving many generations, where its clinically relevant phenotypic spectrum is highly variable between mutation carriers. The basis of this clinical variability is largely unknown, but has been attributed to several possible mechanisms including modifying factors consisting of gene variants and co-morbid environmental factors (e.g. gestational diabetes, teratogen exposure), as well as to a speculative digenic model called the "multiple hit" hypothesis [16]. Clinical findings in HPE can extend from cyclopia, at the severe extreme, to less severe conditions that include several types of recognizable microforms such as the solitary median maxillary central incisor syndrome [SMMCI; see 17–20], or even incomplete penetrance of obligate mutation carriers [2, 5, reviewed in 21]. Such extreme clinical presentations are encompassed in the term the "HPE spectrum". Mutations have also been detected in related pathway factors such as *PTCH1* [22; MIM *601309; HPE 7 #610828], the transcription factor gene *GLI2* [23; MIM *165230; HPE 9 #610829] and the putative ligand transporter *DISP1* [24; MIM *607502]. *SHH* is presently one of four genes, which also include *SIX3* [25–27; MIM *603714; HPE2 #157170], *ZIC2* [28–31; MIM *603073; HPE5 #609637] and *TGIF* [32–34; MIM *602630; HPE4 #142946], routinely screened as part of the molecular evaluation of new sporadic or familial HPE or SMMCI cases [reviewed in 1, 2, 35–37]. Consequently, a large collection of mutations is presently available for comparative analysis.

Here we draw together the molecular findings of three prospective series of unrelated HPE probands in order to formulate guidelines for the recognition of the types of mutations most likely to contribute to disease and the likelihood of combinatorial gene x gene interactions for any particular proband. We believe that the model we have developed can be adapted to additional HPE genes in the future and is not dependent on the mode of mutation (e.g. works with non-synonymous, nonsense, deletion, duplication, copy number variants, non-coding regulatory element mutations, etc.) provided there is good control data for healthy individuals. Hence, as we begin to confront the challenge of next generation sequencing data we feel that there are several important lessons gleaned from the past decade of diagnostic work that will be instructive as research moves forward using new technologies.

MATERIALS AND METHODS

The NIH Study population

We analyzed 189 HPE patients (collectively comprising the entire spectrum of HPE brain malformations and prospectively collected by passive referral over 5 years) for potential sequence variations in *SHH*, *ZIC2*, *SIX3* and *TGIF* (the "four gene" screen) under our brain research protocol and newly established CLIA laboratory. All subjects provided informed consent in accordance with the guidelines of our NHGRI protocol and IRB review process and could opt in or out of receiving genetic information results.

Mutation screening, PCR amplification and DNA sequencing

A strategy for the "four gene" screen has been modified from previously described ones [3,6,25,28,32] and is available upon request. Descriptions of common variations (Supplementary Table 1) and NIH mutations (Supplementary Table 2 and 6) including data from Dubourg et al., 2004 [15] and Paulussen et al., 2010 [38] (Supplementary Tables 3 and 4) are based on the *SHH* NM_000193.2, *ZIC2* NM_007129.2, *SIX3* NM_005413.2 and *TGIF* NM_003244.2 reference sequences (Supplementary Data 7). Guidelines for the naming of the sequence variants conform to the recommendations of the human

nomenclature committee (www.hgvs.org/mutnomen) and were checked using the Mutalyzer software for the predicted effects on the reference gene (http://www.mutalyzer.nl/2.0/). Position +1 refers to the A position of the ATG initiation codon for that gene. Mutation descriptions requiring revision according to Mutalyzer nomenclature recommendations are indicated with the symbol "#". Family-specific mutations (shown in green in the Supplementary Tables) meet the following criteria needed to be included in Table 1: are (1) they are not known to be described in other families, (2) are unique to the indicated study group, and (3) are not described as variations detectable in dbSNP and/or the 1000 genomes data, nor 4) detected among the common variations in our NIH cohort (Supplementary Table 1). For consistency, synonymous changes are indicated by italic font, non-coding variations are underlined and italicized, while non-synonymous variations are in normal font (red for rare variants and green for family-specific mutations). Note that 3 of the rare variants in our series actually describe non-sense changes in apparently unrelated families: one case of [*SHH* c.525C>G (p.Tyr175*)] and two instances of [*ZIC2* c.1096_1096del (p.Cys365*)] that although they formally do not meet the criteria we use for family-specific alterations, they probably should. These three special cases are not included in the calculations for the remaining Tables. Unlike the likely *SHH* c.995T>C (p.Val332Ala) benign variant that was also detected in the French series [15] we conclude that the reason that the same molecular alteration is detected in unrelated families reflects the type of change that is needed to produce a stop codon. Similarly, we have identified two apparently different families with *SIX3* c.770G>C (p.Arg257Pro) and therefore we choose to label this as a benign rare variation (see [15] and this study). Interestingly, when tested functionally this allele tested as normal [26].

Amplification of human genomic DNA is performed in a 35 μl reaction volume, using 60–100 ng of DNA template, 50 μM each of deoxynucleotide triphosphate, 20 pmol of each primer, 3.5 μl of 10X PCR Amplification buffer (Invitrogen, CA), 1.75 μl 10X PCR Enhancer solution (Invitrogen, CA), 1.5mM MgSO4 (Invitrogen, CA) and 2.5U of AmpliTaq DNA polymerase (Roche, IN). All reactions were performed using a PTC-255 thermocycler (MJ Research, MA; now supported by Biorad). Typical PCR cycling parameters were 95°C for 4 minutes followed by 30 cycles at 95°C for 30 seconds, annealing at 60°C for 30 seconds, extension at 72°C for 1 minute, and a final extension step of 72°C for 7 minutes. Amplicons were purified using a high pure PCR purification kit (Roche, IN) and bi-directionally sequenced using the BigDye™ version 3.1 terminator cycle sequencing kit according to the manufacturer's protocol (Applied Biosystems, CA) on an ABI 3100 and 3130 automated sequencers.

Data collected for statistical modeling

Holoprosencephaly (HPE) patients are screened in the indicated CLIA lab for at least one of k genes. Once the genetic testing has been completed, the results are summarized in Table 1. For the N_i patients screened for gene i (i =1,2,…,k) we detect a total of m_i clinically and molecularly proven mutations. Furthermore, the observed mutation frequency 1) by gene and 2) by base pair (bp) are then calculated as shown in columns 5 and 6 in Table 2, respectively.

Statistical Model description

Mutations are often defined as sudden and spontaneous changes in the genomic sequence of a living organism; in statistical terms, this is equivalent to "rare" events (changes in the DNA sequence) occurring in the genome. The Poisson distribution is a discrete probability distribution for describing occurrences of rare events [39] in a fixed interval of time and/or space [40], which expresses the probability that these events occur with a known average rate. Some examples include the statistical modelling of the number of electronic devices

that fail before their guarantee period [41], the estimation of cancer incidence [42], the description and further modelling of the number of motor vehicle crashes [43], genetic association studies in which the disease trait corresponds to count data [44,45] and, more recently, testing RNA-Seq data [46]

Let X_i be a random variable counting the number of mutations per patient in gene *i*. Then, the probability that there are exactly m_i mutations is [47]

$$P(X_i=m_i|\lambda_i)=\lambda_i^{m_i}\exp\{-\lambda_i\}/m_i! \quad m_i=0,1,2,\ldots \quad i=1,2,\ldots,k \tag{1}$$

where λ_i is the expected value and variance of X_i. In our context, λ_i can be determined as the mutation frequency per gene, per base pair or per amplicon, or as a function of a set of covariates of interest $Y_1,Y_2,\ldots,Y_m$, by changing the definition of X_i accordingly. If the latter approach were chosen, then a Poisson regression [48] setting would be more appropriate. Several methods are available for estimating λ_i; these methods include, among others, maximum likelihood (ML) estimation and estimation using the method of moments [47]. We estimated λ_i using the ML method (columns 5–6, Table 2) because of its computational simplicity and its well-know asymptotic properties [40, 47].

Now, suppose that two independent genes G_1 and G_2 are screened in HPE patients. Then, the joint distribution of X_1 and X_2 is

$$P(X_1=m_1,X_2=m_2|\lambda_1,\lambda_2)=\lambda_1^{m_1}\lambda_2^{m_2}\exp\{-(\lambda_1+\lambda_2)\}/(m!m_2!), \quad m_1=0,1,2,\ldots;m_2=0,1,2,\ldots \tag{2}$$

The expected value of $Y = X_1X_2$ is $E[Y] = \lambda_1\lambda_2$ and its variance $V[Y] = \lambda_1\lambda_2(1+\lambda_1+\lambda_2)$. On the other hand, the probability that none and one unique mutation per patient occurs in the G_1 and G_2 genes (a double hit) are $p_0 = \exp\{-(\lambda_1+\lambda_2)\}$ and $p_1 = p_0\lambda_1\lambda_2$, respectively.

Comparison of mutation rates per base pair—Three independent research groups (NIH, French and Dutch; Supplementary Tables 2–4) have studied HPE over the last 15 years. In their respective labs, they have performed gene screening for a total of n=475 HPE patients including the *SHH, ZIC2 SIX3,* and *TGIF* HPE genes (Table 3, details in Supplementary Table 5).

Recently, it was reported that, on average, one mutation is found every 85 million base pairs [49, 50]. Using this figure as our baseline mutation rate per base pair for comparison proposes, we tested whether the mutation frequency per base pair in each of the *k* HPE genes was statistically different from the reference, i.e.,

$$H_{0,i}:\lambda_i=\lambda_0 \text{ vs. } H_{1,i}:\lambda_i\neq\lambda_0 \quad i=1,2,\ldots,k \tag{4}$$

with λ_0 being the reference mutation rate per base pair in [50]. For testing (4), we used a Poisson test as implemented in R [51].

Sample size for detecting double hit mutations—Here we describe a methodology for determining the minimum number of patients that would need to be screened in order to detect a statistically significant deviation from the expected number of patients with mutations in genes G_1 and G_2 based on the empirically measured number of detected mutations per gene. Although we only consider the case in which patients have unique mutations in two genes, our methodology can be easily extended to detect "multiple hits", i.e., *k* unique mutations, one per gene, when *k* genes are screened.

Suppose that N patients with unique mutations in genes G_1 and G_2 are observed and that it is of interest to determine how many patients M need to be screened in order to detect any deviation from the expected number of HPE patients with unique mutations in both genes, denoted as E, when a type I error probability α is used. Even thought it is not said explicitly, E is a function of M. For instance, if M patients would have been screened for genes G_1 and G_2, the expected number of them with a "double hit" under model (2) is $E=p_1M$, with p_1 as previously defined. The test statistic comparing the N and M is given by

$$T=E^{-1}(N-E)^2 \sim \chi^2_{(1)} \quad (3)$$

We will say that the number of observed and expected number of patients with mutations in both genes are statistically different if $T > \chi^2_{(1,\alpha)}$, with $\chi^2_{(1,\alpha)}$ the 1-α percentile of a χ^2 with one degree of freedom. Using our model, it follows that the number of patients M is such that

$$(N-p_1M)^2 - p_1\chi^2_{(1,\alpha)}M>0 \quad (3)$$

Note that in this expression all quantities but M are known. To obtain M we solve (3) numerically.

RESULTS

For our own dataset (Supplementary Table 2, see details by amplicon in Supplementary Table 6) we used our complete exonic Sanger sequencing from the CLIA lab to address a basic research question, namely, are there any recognizable molecular features that would be helpful for medical geneticists and genetic counselors when approaching a new case of HPE? Furthermore, do independent referral centers detect similar findings? Here we show that both statements are true.

The first generalization of value is that when testing is limited to the "four gene" screen in a new HPE proband the result is frequently entirely normal (85% in our cohort). Using our model the probability of finding no mutations (in G_1 and G_2) approaches unity at the observed mutation frequencies we found for even the most commonly mutated HPE genes (Figure 1a). Since the reason for referral often varies among clinicians, and the spectrum of HPE is so broad, we find that this number is similar to those of other reports [1,2]. The next most common result is the detection of single nucleotide polymorphisms (SNP) that we detect at frequencies similar to public databases (Supplementary Table 1). It is noteworthy that most of these SNPS are synonymous changes and do not predict even subtle changes in protein function. The most common exception is the single codon (p.Pro163) in *TGIF* and secondarily the relatively uncommon length variation in the His-tract of *ZIC2* (Supplementary Table 6). Perhaps the most striking, and potentially useful, feature of the remaining sequence changes is that at least 92% of them are family-specific alterations (see Methods for a definition) and that at least one third of them can be demonstrated to be *de novo* in the pooled datasets (Table 1). We suggest that this empiric observation is tightly linked to the mechanism of HPE pathogenesis (i.e. that novel translocations, deletions or point mutations are the proximate cause) and can be used to model the chance that two or more classical mutations in any of these categories might be detected in the same individual.

The measurable parameters needed for our model are described in Table 2 and allow for comparison between different diagnostic centers (Table 3). We focused in on the observed mutation frequency in the four genes to distinguish this parameter from the mutation rate

more generally (calculated to be 1 in 85 million base pairs; [49,50]). This measurement of 30–50 novel mutations per proband defines a parameter that limits the number of genes that could likely participate in mutant x mutant genetic interactions even assuming that they all would have the potential to do so.

Since we note that virtually all of the observed mutations are family-specific we can derive an estimate of relative risk of a *de novo* mutation in these genes causing HPE based on this expected mutation rate of healthy individuals. As shown in Table 4, the observed mutation frequency is several orders of magnitude higher than one would expect for a region of the genome not previously implicated in HPE. Nevertheless, despite this highly statistically significant finding the frequency of deleterious mutations (λ_i) is rather low (ranges between 1% and 10%, Table 3) and the probability of detecting mutations in two genes (G_1 and G_2) in the same proband would be expected to be quite uncommon (see Figure 1b). Therefore, we should not be surprised that we failed to observe a single case of a "double hit" among 475 prospectively evaluated cases.

Figure 2 provides a graphical representation of the number of individuals that we would need to sequence (for G_1 and G_2) based on the experimentally determined mutation frequencies per gene (λ_1 and λ_2). Note that as λ_i falls below 1%, the number of patients needed to detect even a single hit grows exponentially, and instances of a "double hit" are even more remote. Similarly, we can model the number of cases *M* needed to detect a "double hit" for any desired threshold of *N* observed patients. As shown in Table 5, it is indeed unlikely to detect even a single patient (*N*=1) with a "double hit" and thousands of unrelated HPE individuals would be required to detect such a case.

DISCUSSION

Holoprosencephaly is the most common structural malformation of the developing forebrain in humans and its causation includes genetic and environmental components typical for a complex trait that must integrate both types of factors [35,36,52]. Its genetic heterogeneity is supported by the fact that nearly 75% of cases are wild-type in their DNA sequence of the coding regions of the nine currently known HPE genes in patients with normal chromosomes [1,2]. Here we have shown that the detection of novel mutation in an HPE gene is one salient feature shared among the four genes that are routinely screened that this type of mutation is likely to be both necessary and sufficient. By extrapolation to additional HPE genes, we expect that this trend should continue (see [55–58]; and the candidate gene described by Dupé et al., 2011 [54]). Our current testing recommendations for HPE are consistent with our model in that candidate genes where λ_i is less than 1% are extremely difficult to establish as clinically significant factors [37] as are low frequency copy number variations [59]. Although the ability to discriminate between ultra-rare variations and *bone fide* mutations is a challenge without functional analysis, nevertheless, the vast improvements in normative comparative datasets of healthy individuals in publically available databases makes the recognition of novel alterations far simpler than in the past. With similar improvements in databases of copy number variations (CNV) this mechanism can also be incorporated into the interaction analysis.

We also suggest that an obligate digenic model of HPE (caused by mutations in two genes) is unlikely to apply in the vast majority of cases. In fact this is the observed result for these cases. Furthermore, if this obligatory digenic model were correct (with an unknown interacting factor λ_i) this requirement would likely confound any meaningful genotype/phenotype correlation. We certainly do not mean to imply that gene x gene interactions are unimportant. Rather, our model predicts that common genetic variants are the likely substrate to influence the clinical manifestations of a family-specific mutation (i.e. $\lambda_2 > \lambda_1$).

This model is entirely consistent with the known strain-specific modifier effects seen in mouse studies (reviewed in [53]). These genetic modifiers are generally silent (and are fixed in the respective classical inbred strains of mice) unless an experimental genetic lesion is introduced. Extrapolating to humans, the modifiers of known HPE genes would also be silent variations that can be either coding or non-coding in nature that influence the function or expression of key developmental pathways when a critical component is compromised. Taken together, we deduce from the theoretical and empirical absence of instances of a "double mutation" (including recessive cases of HPE were λ_1 and λ_2 are the same) that the autosomal dominant with modifier model is the best description of how these and future HPE genes disturb forebrain development in humans.

We illustrate this notion (Fig. 1) with the most extreme case: namely the case of two *de novo* base-pair changes in the same individual. This is indeed unlikely, and marginally worthy of mentioning. Furthermore, we stipulate that there is little evidence that any of these genes fundamentally changes the mutation rate; this is influenced by DNA synthesis and repair mechanisms, or recombination factors, etc. However, if one of these functional variants is inherited ($\lambda_2 = 0.5$) it is no longer so rare, nor is it readily distinguishable from the incidence of a single "hit". This second modifier "hit" can be a common variant in a relevant developmental pathway, or a common environmental insult, e.g. alcohol, maternal diabetes, etc. Our model is particularly useful because different clinically relevant alterations occur at intrinsically different rates (e.g. CNV>translocation≫base-pair changes). What our model provides is a way to estimate the likelihood of two classes of alterations occurring in the same individual.

We call for a revised view of the "multiple hit hypothesis" in order to satisfy both empirical and theoretical realities. We now propose an autosomal dominant with modifier model that is superior in its utility and predictive power. Furthermore, we argue that just such a model is needed and can apply to similar genetic conditions where a single genetic alteration (translocation, deletion, duplication, point mutation) serves as the final insult leading to malformation and disease – *within* the context of otherwise silent genetic variants that serve as modifiers. Finally, we conclude that the best way to identify such intrinsic modifier loci is to select a group of subjects with a lesion in one of the key HPE genes and perform classical association analysis of SNP arrays looking for interacting loci.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank the patients who participated in this research and the many clinicians who referred them. We thank BD Solomon and D Pineda-Alvarez for constructive comments. JIV would like to thank Prof. Juan Carlos Correa for helpful comments and suggestions on the statistical model. This research was supported by the Division of Intramural Research of the National Human Genome Research Institute, National Institutes of Health.

References

1. Dubourg C, Bendavid C, Pasquier L, Henry C, Odent S, David V. Holoprosencephaly. Orphant J Rare Dis. 2007; 2:8.
2. Roessler E, Muenke M. The molecular genetics of holoprosencephaly. Am J Med Genet Part C, Semin Med Genet. 2010; 154C:52–61. [PubMed: 20104595]
3. Roessler E, Belloni E, Gaudenz K, Jay P, Berta P, Scherer SW, Tsui L-C, Muenke M. Mutations in the human *Sonic Hedgehog* gene cause holoprosencephaly. Nat Genet. 1996; 14:357–360. [PubMed: 8896572]

4. Heussler HS, Suri M, Muenke M. Extreme variability of expression of a Sonic Hedgehog mutation: attention difficulties and holoprosencephly. Arch Dis Child. 2002; 86:293–296. [PubMed: 11919111]
5. Hehr U, Gross C, Diebold U, Wahl D, Beudt U, Heidemann P, Hehr A, Mueller D. Wide phenotypic variability in families with holoprosencephly and a sonic hedghog mutation. Eur J Pediatr. 2004; 163:347–352. [PubMed: 15107988]
6. Roessler E, Belloni E, Gaudenz K, Vargas F, Scherer SW, Tsui L-C, Muenke M. Mutations in the C-terminal domain of Sonic Hedgehog cause holoprosencephaly. Hum Mol Genet. 1997; 6:1847–1853. [PubMed: 9302262]
7. Roessler E, El-Jaick K, Dubourg C, Vélez JI, Solomon BD, Pineda-Álvarez DE, Lacbawan F, Zhou N, Ouspenskaia M, Paulussen A, Smeets HJ, Hehr U, Bendavid C, Bale S, Odent S, David V, Muenke M. The mutational spectrum of holoprosencephaly-associated mutations within the SHH gene in humans predict loss-of-function through either key structural alterations of the ligand or its altered synthesis. Hum Mutat. 2009; 30:E921–935. [PubMed: 19603532]
8. Roessler E, Ward DE, Gaudenz K, Belloni E, Scherer SW, Donnai D, Siegel-Bartelt J, Tsui L-C, Muenke M. Cytogenetic rearrangements involving the loss of the Sonic Hedgehog gene at 7q36 cause holoprosencephaly. Hum Genet. 1997; 100:172–181. [PubMed: 9254845]
9. Bendavid C, Haddad BR, Griffin A, Huizing M, Dubourg C, Gicquel I, Cavalli LR, Pasquier L, Long R, Ouspenskaia M, Odent S, Lacbawan F, David V, Muenke M. Multicolor FISH and quantitative PCR can detect submicroscopic deletions in holoprosencephaly patients with a normal karyotype. J Med Genet. 2005; 43:496–500. [PubMed: 16199538]
10. Bendavid C, Dubourg C, Gicquel I, Pasquier L, Saugler-Veber P, Durou M-R, Jaillard S, Frebourg T, Haddad BR, Henry C, Odent S, David V. Molecular evaluation of foetuses with holoprosencephaly shows high incidence of microdeletions in the HPE genes. Hum Genet. 2005; 119:1–8. [PubMed: 16323008]
11. Bendavid C, Rochard L, Dubourg C, Sequin J, Gicquel I, Pasquier L, Vigneron J, Laquerriére A, Marcorelles P, Jeanne-Pasquier C, Rouleau C, Jaillard S, Mosser J, Odent S, David V. Array-CGH analysis indicates a high prevalence of genomic rearrangements in holoprosencephaly: an updated map of candidate loci. Hum Mut. 2009; 30:1175–1182. [PubMed: 19431187]
12. Bendavid C, Dupé V, Rochard L, Gicquel I, Dubourg C, David V. Holoprosencephaly: an update on cytogenetic abnormalities. Amer J Med Genet, Part C. 2010; 154C:86–92. [PubMed: 20104602]
13. Orioli IM, Castilla EE, Ming JE, Nazer J, Burle de Aguiar MJ, Lerena JC, Muenke M. Identification of novel mutations in SHH and ZIC2 in a South American (ECLAMC) population with holoprosencephaly. Hum Genet. 2001; 109:1–6. [PubMed: 11479728]
14. Traiffort E, Dubourg C, Faure H, Rognan D, Odent S, Durou M-R, David V, Ruat M. Functional characterization of Sonic Hedgehog mutations associated with holoprosencephaly. J Biol Chem. 2004; 279:42889–42897. [PubMed: 15292211]
15. Dubourg C, Lazaro L, Pasquier L, Bendavid C, Blayau M, Le Duff F, Durou M-R, Odent S, David V. Molecular screening of SHH, ZIC2, SIX3 and TGIF genes in patients with features of holoprosencephaly spectrum: mutation review and genotype-phenotype correlations. Hum Mut. 2004; 24:43–51. [PubMed: 15221788]
16. Ming JE, Muenke M. Multiple hits during early embryonic development: digenic disease and holoprosencephaly. Am J Hum Genet. 2002; 71:1017–1032. [PubMed: 12395298]
17. Nanni L, Ming JE, Du Y, Hall RK, Aldred M, Bankier A, Muenke M. Shh mutation is associated with solitary median maxillary central incisor: A study of 13 patients and review of the literature. Am J Med Genet. 2001; 102:1–10. [PubMed: 11471164]
18. Marini M, Cusano R, De Biasio P, Caroli F, Lerone M, Silengo M, Ravazzolo R, Seri M, Camera G. Previously undescribed nonsense mutation in SHH caused autosomal dominant holoprosencephaly with wide intrafamilial variability. Am J Med Genet. 2003; 117A:112–115. [PubMed: 12567406]
19. Ribeiro LA, Richieri-Costa A. Single median maxillary incisor, hypophyseal tumor, and SHH mutation. Am J Med Genet. 2005; 136A:346–347. [PubMed: 15942953]

20. El-Jaick KB, Fonseca RF, Moreira MA, Ribeiro MG, Bolognese AM, Dias SO, Pereira ET, Castilla EE, Orioli IM. Single median maxillary central incisor: new data and mutation review. Birth Defects Res (PartA). 2007; 79:573–580.
21. Solomon BD, Mercier S, Vélez JI, Pineda-Alvarez DE, Wyllie A, Zhou N, Dubourg C, David V, Odent S, Roessler E, Muenke M. Analysis of genotype-phenotype correlations in human holoprosencephaly. Am J Med Genet Part C, Semin Med Genet. 2010; 154C:133–141. [PubMed: 20104608]
22. Ming JE, Kaupas ME, Roessler E, Brunner HG, Golabi M, Tekin M, Stratton RF, Sujansky E, Bale SJ, Muenke M. Mutations in PATCHED-1, the receptor for SONIC HEDGEHOG, are associated with holoprosencephly. Hum Genet. 2002; 110:297–301. [PubMed: 11941477]
23. Roessler E, Ermilov AN, Grange DK, Wang A, Grachtchouk M, Dlugosz AA, Muenke M. A previously unidentified amino terminal domain regulates transcriptional activity of wild-type and disease-associated human *GLI2*. Hum Mol Genet. 2005; 14:2181–2188. [PubMed: 15994174]
24. Roessler E, Ma Y, Ouspenskaia MV, Lacbawan F, Bendavid C, Dubourg C, Beachy PA, Muenke M. Truncating loss-of function mutations of DISP1 contribute to holoprosencephly-like microform features in humans. Hum Genet. 2009 Jan 31. [Epub ahead of print] PMID: 19184110.
25. Wallis DE, Roessler E, Hehr U, Nanni L, Wiltshire T, Richieri-Costa A, Gillessen-Kaesbach G, Zackai EH, Rommens J, Muenke M. Missense mutations in the homeodomain of the human *SIX3* gene cause holoprosencephaly. Nat Genet. 1999; 22:196–198. [PubMed: 10369266]
26. Domené S, Roessler E, El-Jaick KB, Snir M, Brown JL, Vélez JI, Bale S, Lacbawan F, Muenke M, Feldman B. Mutations in the human SIX3 gene in holoprosencephaly are loss of function. Hum Mol Genet. 2008; 17:3919–3928. [PubMed: 18791198]
27. Lacbawan F, Solomon B, Roessler E, El-Jaick K, Domené S, Vélez JI, Zhou N, Hadley D, Balog J, Long R, Fryer A, Smith W, Omar S, McLean SD, Clarkson K, Lichty A, Clegg N, Delgado M, Levey E, Stashinko E, Potocki L, VanAllen M, Clayton-Smith J, Donnai D, Bianchi D, Juliusson P, Njølstad PR, Brunner H, Carey J, Hehr U, Müsebeck J, Wieacker PF, Hennekam RCM, van den Boogaard M-JH, van Haeringen A, Paulussen A, Herbergs J, Schrander-Stumpel C, Janecke A, Chitayat D, Hahn J, McDonald-McGinn DM, Zackai EH, Dobyns WB, Muenke M. Clinical spectrum of SIX3-associated mutations in holoprosencephaly: correlation between genotype, phenotype and function. J Med Genet. 2009; 46:389–398. [PubMed: 19346217]
28. Brown SA, Warburton D, Brown LY, Yu C-y, Roeder ER, Stengel-Rutkowski S, Hennekam RCM, Muenke M. Holoprosencephaly due to mutations in ZIC2, a homologue of Drosophila odd-paired. Nat Genet. 1998; 20:180–183. [PubMed: 9771712]
29. Warr N, Powles-Glover N, Chappell A, Robson J, Norris D, Arkell RM. Zic2-associated holoprosencephaly is caused by a transient defect in the organizer region during gastrulation. Hum Mol Genet. 2008; 17:2986–2996. [PubMed: 18617531]
30. Roessler E, Lacbawan F, Dubourg C, Paulussen A, Herbergs J, Hehr U, Bendavid C, Zhou N, Ouspenskaia M, Bale S, Odent S, David V, Muenke M. The full spectrum of holoprosencephaly-associated mutations within the *ZIC2* gene in humans predict loss-of-function as the predominant disease mechanism. Hum Mutat. 2009; 30:E541–544. [PubMed: 19177455]
31. Solomon BD, Lacbawan F, Mercier S, Clegg NJ, Delgado MR, Dubourg C, David V, Rosenbaum K, Olney AH, Wehner L-E, Hehr U, Bale S, Paulussen A, Smeets HJ, Hardisty E, Tylki-Szymanka A, Pronicka E, Clemens M, McPherson E, Hennekam RCM, Hahn J, Stashinko E, Levey E, Wieczorek D, Majewski F, Roeder E, Imaizumi K, Schell-Apacik C, Booth C, Keaton A, Balog JZ, Hadley D, Zhou N, Long B, Pineda-Alvarez DE, Odent S, Roessler E, Muenke M. Mutations in ZIC2 in human holoprosencephaly: comprehensive analysis of 141 individuals and description of a novel ZIC2-specific phenotype. J Med Genet. 2010; 47:513–524. [PubMed: 19955556]
32. Gripp KW, Wotton D, Edwards MC, Roessler E, Ades L, Meinecke P, Richeri-Costa A, Zackai EH, Massague J, Muenke M, Elledge SJ. Mutations in TGIF cause holoprosencephaly and link NODAL signaling to human neural axis determination. Nat Genet. 2000; 25:205–208. [PubMed: 10835638]
33. El-Jaick KB, Powers SE, Bartholin L, Myers KR, Hahn J, Orioli IM, Ouspenskaia M, Lacbawan F, Roessler E, Wotton D, Muenke M. Functional analysis of mutations in TGIF associated with holoprosencephaly. Mol Genet Metab. 2007; 90:97–111. [PubMed: 16962354]

34. Keaton A, Solomon BD, Kauvar EF, El-Jaick KB, Gropman AL, Zafer Y, Meck JM, Bale SJ, Grange DK, Haddad BR, Gowans GC, Clegg NJ, Delgado MR, Hahn JS, Pineda-Alvarez DE, Lacbawan F, Vélez JI, Roessler E, Muenke M. Mutations in TGIF in human holoprosencephaly: correlation between genotype and phenotype. Mol Syndromol. 2010; 1:211–222. [PubMed: 22125506]
35. Muenke, M.; Beachy, PA. Holoprosencephaly. In: Scriver, CR., et al., editors. The Metabolic & Molecular Bases of Inherited Disease. McGraw-Hill; New York: 2001. p. 6203-6230.
36. Cohen MM Jr. Holoprosencephaly: clinical, anatomic, and molecular dimensions. Birth Defects Res Part A Clin Mol Teratol. 2006; 76:658–673. [PubMed: 17001700]
37. Pineda-Alvarez DE, Dubourg C, David V, Roessler E, Muenke M. Current recommendations for the evaluation of newly diagnosed holoprosencephaly patients. Am J Med Genet Part C, Semin Med Genet. 2010; 154C:93–101. [PubMed: 20104604]
38. Paulussen ADC, Schrander-Stumpel CT, Tserpelis DCJ, Spee MKM, Stegmann APA, Mancini GM, Brooks AS, Collée M, Maat-Kievit A, Simon MEH, van Bever Y, Stolte-Dijkstra I, Kerstjens-Frederikse WS, Herkert JC, van Essen AJ, Lichtenbelt KD, van Haeringen A, Kwee ML, Lachmeijer AMA, Tan-Sindhunata GMB, van Maarle MC, Arens YHJM, Smeets EEJGL, de Die-Smulders CE, Engelen JJM, Smeets HJ, Herbergs J. The unfolding clinical spectrum of holoprosencephaly due to mutations in *SHH*, *ZIC2*, *SIX3* and *TGIF* genes. Euro J Med Genet. 2010; 18:999–1005.
39. van Belle, G. Statistical Rules of Thumb. 2. John Wiley & Sons, Inc; 2008.
40. Correa JC, Castrillón F. Comparación por intervalos entre diferentes métodos de estimación de la media de la distribución Poisson. Revista EAFIT. 2006; 42:81–96.
41. Meeker, W.; Escobar, L. Statistical Methods for Reliability Data. New York: John Wiley & Sons; 1998.
42. Frome EL, Checkoway H. Use of Poisson Regression Models in Estimating Incidence Rates and Ratios. Am J Epidemiol. 1985; 121:309–323. [PubMed: 3839345]
43. Lord D, Mannering F. The Statistical Analysis of Crash-frequency Data: A Review and Assessment of Methodological Alternatives. Transportation Research Part A: Policy and Practice. 2010; 44:291–305.
44. Cui Y, Kim DY, Zhu J. On the generalized poisson regression mixture model for mapping quantitative trait loci with count data. Genetics. 2006; 174:2159–2172. [PubMed: 17028335]
45. Rodrigues-Motta M, Gianola D, Heringstad B, Rosa GJ, Chang YM. A zero-inflated poisson model for genetic analysis of the number of mastitis cases in Norwegian Red cows. J Dairy Sci. 2007; 90:5306–5315. [PubMed: 17954771]
46. Auer PL, Doerge RW. A Two-Stage Poisson Model for Testing RNA-Seq Data. Statistical Applications in Genetics and Molecular Biology. 2011; 10
47. Casella, G.; Berger, R. Statistical Inference. Duxbury Press; 2001.
48. Myers, R.; Montgomery, D.; Vining, G. Generalized Linear Models: With Applications in Engineering and the Sciences. John Wiley & Sons; Netherlands: 2001.
49. Conrad DF, et al. Variation in genome-wide mutation rates within and between human families. Nat Genet. Published online June 12, 2011. 10.1038/ng.86
50. Hesman-Saey, T. [Accesed September 21, 2011,] Human mutation rate slower than thought. Science News. http://tinyurlcom/3rhtwzx
51. R Development Core Team. A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2011. URL http://www.R-project.org/
52. Monuki ES. The morphogen signaling network in forebrain development and holoprosencephaly. J Neuropath Exp Neurosci. 2007; 66:566–575.
53. Krauss RS. Holoprosencephaly: new models, new insights. Expert Rev Mol Med. 2007; 9:1–17. [PubMed: 17888203]
54. Dupé V, Rochard L, Mercier S, Le Pétillon Y, Gicquel I, Bendavid C, Bourrouillou G, Kini U, Thauvin-Robinet C, Bohan TP, Odent S, Dubourg C, David V. NOTCH, a new signaling pathway implicated in holoprosencephaly. Hum Mol Genet. 2011; 20:1122–31. [PubMed: 21196490]

55. Arauz RF, Solomon BD, Pineda-Alvarez DE, Gropman AL, Parsons JA, Roessler E, Muenke M. A hypomorphic allele in the FGF8 gene contributes to holoprosencephaly and is allelic to gonadotropin releasing hormone deficiency in humans. Molec Syndrom. 2010; 1:59–66.

56. Bae G, Domené S, Roessler E, Schachter K, Kang J-S, Muenke M, Krauss R. Holoprosencephaly-associated mutations in CDON result in defective interactions with other Hedgehog receptors. Am J Hum Genet. 2011; 8:231–240. [PubMed: 21802063]

57. Pineda-Alvarez DE, Hu P, Roessler E, Srivastava K, Siple CE, Fan C-M, Muenke M. Missense substitutions in the GAS1 protein present in holoprosencephaly patients reduce the affinity for its, ligand, SHH. Hum Genet. (Epub ahead of print, August 13, 2011).

58. Ribeiro LA, Quiezi RG, Nascimento A, Bertolacini CP, Richieri-Costa A. Holoprosencephaly and holoprosencephaly-like phenotype and GAS1 DNA sequence changes: Report of four Brazilian patients. Am J Med Genet A. 2010; 152A:1688–94. [PubMed: 20583177]

59. Mercier S, Dubourg C, Garcelon N, Campillo-Gimenez B, Gicquel I, Bellewuic M, Ratié L, Pasquier L, Loget P, Bendavid C, Jailard S, Rochard L, Quélin C, Dupé V, David V, Odent S. New findings for phenotype-genotype correlations in a large european series of holoprosencephaly cases. J Med Genet. 2011 [e pub ahead of print; PMID 21940735].

Highlights

- Holoprosencephaly (HPE) is genetically heterogeneous and variable in expressivity.
- We compare digenic vs. autosomal dominant with modifier models of HPE.
- Multiple prospective molecular studies do not support a digenic model for HPE.
- An autosomal dominant with modifier model is now recommended.

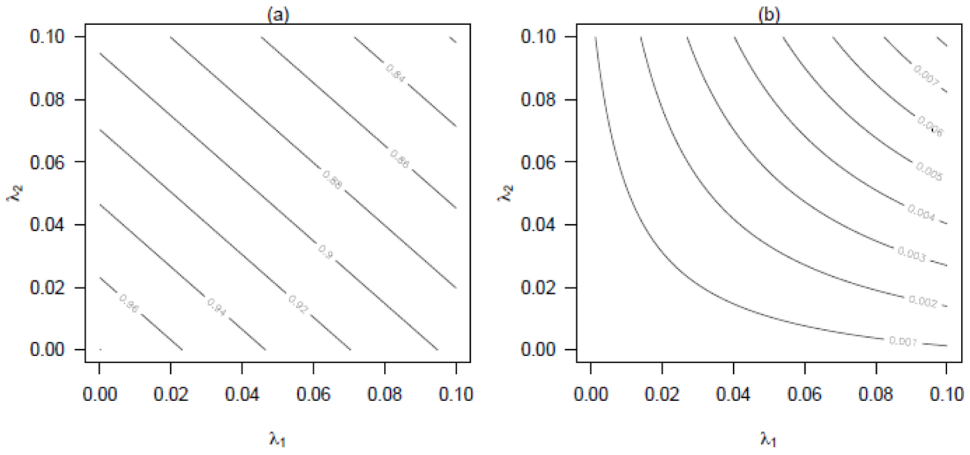

Figure 1.

Probability that one observes (a) none mutations in genes G_1 and G_2, and (b) one mutation in both genes G_1 and G_2 (i.e., a "double hit") when the mutation frequency per gene varies from 0% to 10%.

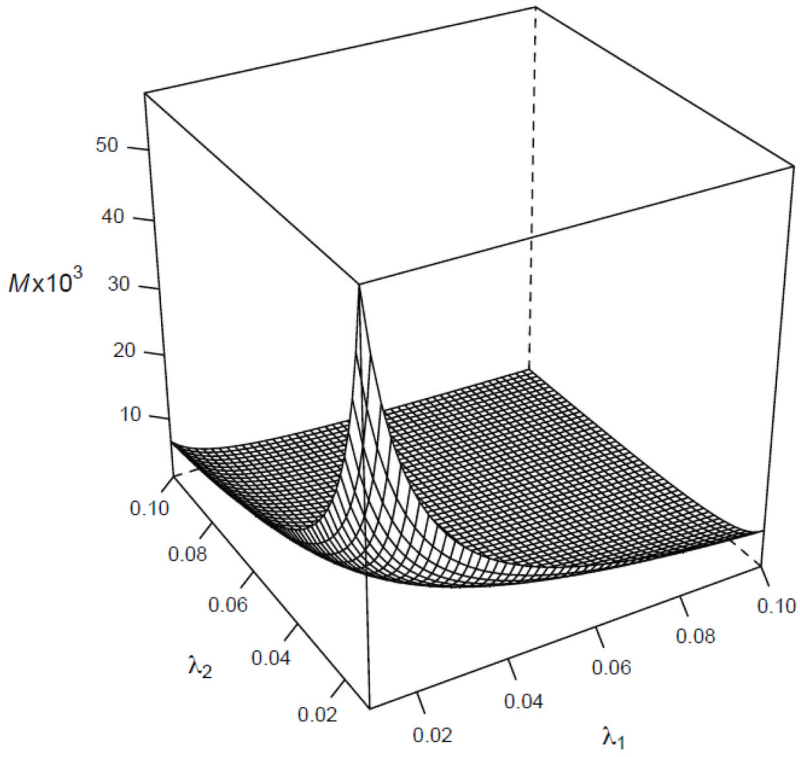

Figure 2.
Total number of HPE individuals *M* to be screened to detect a "double hit" (e.g. individual mutations in genes G_1 and G_2) when the observed mutation frequency per gene varies from 1% to 10%.

Table 1

Molecular characteristics of mutations in HPE genes.

Cohort	Family-specific	*De novo*	Length variants	Rare variants (deleterious)	Rare variants (benign)	Deletion	Number probands
NIH	28	11	1	3	2	ND	189
Dutch	16	11	3	ND	ND	2	86
French	32	1	0	ND	2	ND	200
Totals	76	23	4	3	4	2	475

ND: Not detected

Table 2

Data collected for the screening of *k*-genes among HPE patients.

Gene	Length (bp[a])	Patients	Mutations	Mutation frequency (per patient)	Mutation frequency (per bp[a])
1	l_1	N_1	m_1	m_1/N_1	$m_1(N_1\ l_1)^{-1}$
2	l_2	N_2	m_2	m_2/N_2	$m_2(N_2\ l_2)^{-1}$
3	l_3	N_3	m_3	m_3/N_3	$m_3(N_3\ l_3)^{-1}$
4	l_4	N_4	m_4	m_4/N_4	$m_4(N_4\ l_4)^{-1}$
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
k	l_k	N_k	m_k	m_k/N_k	$m_k(N_k\ l_k)^{-1}$

[a]bp = base pair

Table 3

Number of family-specific mutations and mutation frequencies per gene and base pair (bp) in four HPE genes by the respective research group.

Gene	NIH (*n*=189)	French (*n*=200)	Dutch (*n*=86)	Aggregate (*n*=475)
Mutations detected				
SHH	10	15	3	28
ZIC2	8	7	8	23
SIX3	9	8	8	25
TGIF	1	2	0	3
Mutation frequency per gene				
SHH	5.29%	7.50%	3.49%	5.89%
ZIC2	4.23%	3.50%	9.30%	4.84%
SIX3	4.76%	4.00%	9.30%	5.26%
TGIF	0.53%	1.00%	0.00%	0.63%
Mutation frequency per bp[a]				
SHH	3.81×10^{-5}	5.40×10^{-5}	2.51×10^{-5}	4.24×10^{-5}
ZIC2	1.58×10^{-5}	1.30×10^{-5}	3.47×10^{-5}	1.81×10^{-5}
SIX3	4.77×10^{-5}	4.00×10^{-5}	9.31×10^{-5}	5.27×10^{-5}
TGIF	6.48×10^{-6}	1.23×10^{-5}	0	1.95×10^{-5}

[a] *SHH*, *ZIC2*, *SIX3* and *TGIF* have, respectively, 1389 bp, 2682 bp, 999 bp and 816 bp in length.

Table 4

Results from the Poisson test comparing the observed and expected mutation rates by base pair in all HPE genes.

Gene	Group	Ratio[a]	95%CI[b]	*P*-value[c]
SHH	NIH	3237.8	(1552.6, 5954.6)	2.17×10^{-32}
	French	4589.6	(2568.7, 7570)	3.95×10^{-50}
	Dutch	2134.7	(440.2, 6238.7)	4.62×10^{-10}
	Aggregate	3607.3	(2397, 5213.6)	2.71×10^{-89}
ZIC2	NIH	4051.7	(1852.7, 7691.5)	3.62×10^{-30}
	French	3403.4	(1469.3, 6706.2)	2.31×10^{-26}
	Dutch	7914.9	(3417, 15596.2)	2.70×10^{-29}
	Aggregate	4478.2	(2898, 6610.7)	3.01×10^{-82}
SIX3	NIH	551.1	(14, 3070.8)	1.81×10^{-3}
	French	1041.7	(126.1, 3762.9)	1.84×10^{-6}
	Dutch	*d*	*d*	*d*
	Aggregate	657.9	(135.7, 1922.7)	1.57×10^{-8}
TGIF	NIH	1341.5	(579.2, 2643.3)	3.95×10^{-23}
	French	1109.2	(446, 2285.5)	7.86×10^{-20}
	Dutch	2948.2	(1272.8, 5809.2)	7.27×10^{-26}
	Aggregate	1534.6	(972.8, 2302.7)	4.20×10^{-65}

[a] Quotient between the observed mutation rate per base pair and a baseline mutation rate per base pair estimated as 1 in 85 million [Hesman-Saey, 2011].

[b] Confidence Interval.

[c] Two-tailed *P*-value comparing the observed and expected mutation rates.

[d] Statistical test could not be performed as zero events occurred.

Table 5

Number of HPE patients M to be screened as a function of N, the number of observed patients with one "double hit" in genes G_1 and G_2 when a type I error probability α=0.05 is used. Mutation frequencies per gene were calculated using the aggregate data in Table 1.

N	Pair of genes					
	SHH/ZIC2	***SHH/SIX3***	***SHH/TGIF***	***ZIC2/SIX3***	***ZIC2/TGIF***	***SIX3/TGIF***
1	2211	2042	16243	2460	19567	18078
2	2846	2629	20911	3167	25190	23273
3	3442	3180	25293	3830	30469	28149
4	4013	3708	29492	4466	35528	32823
5	4567	4219	33563	5083	40432	37354
6	5108	4719	37537	5684	45218	41776
7	5638	5209	41433	6274	49912	46113
8	6159	5690	45267	6855	54530	50379
9	6674	6166	49048	7427	59085	54587
10	7182	6635	52784	7993	63586	58746
11	7685	7100	56481	8553	68039	62861
12	8183	7560	60144	9107	72452	66938
13	8677	8017	63778	9657	76829	70981
14	9168	8470	67384	10203	81173	74995
15	9655	8920	70966	10746	85488	78981