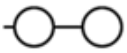
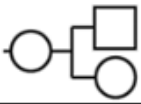
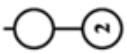
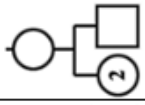
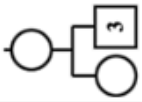
Supplementary material 1

Method for Estimating the Penetrance Rate K in female carriers of MTM1 mutations

The estimation of the penetrance rate was achieved by using or modifying (to deal with X-linked dominant mechanism) standardized methods (Rogatko et al., 1986; Horimoto et al., 2010; Otto and Horimoto, 2012, whose papers can be consulted for more details). In the particular case of myotubular myopathy (**MTM**), the parameter was estimated under the hypothesis that the condition is determined by monogenic dominant X-linked mechanism, with a penetrance rate complete ($\mathbf{K_m = 1}$) among hemizygous male subjects and having an incomplete value $\mathbf{K_f = K}$ among heterozygous females.

The first step of the method consists in trimming or filtering the information from pedigrees with affected individuals, that is, replacing the original pedigrees with the ones containing only individuals that are informative or relevant with respect to the penetrance rate estimation. In the present case we used the genealogical data from ten different families with affected individuals, two of them studied by us and described in the main article (pedigrees depicted in Figures 1 and 2), and eight families from the literature (papers 10, 12, 13, and 21–25 listed in the references list of our paper).

The set of ten pedigrees analyzed by us and considered for the estimation of the penetrance value **K** contained 18 affected female subjects, 35 obligate non-penetrant heterozygous carriers (molecularly detected or with genealogical evidence), 11 normal offspring of obligate heterozygous, and a total of 9 trees of normal individuals descendants from obligate carriers of the gene. These genealogical occurrences are listed (**i = 1, ..., 10**) in the table below, together with their corresponding probabilities $\mathbf{P_1}$ to $\mathbf{P_{10}}$ and absolute observed numbers $\mathbf{N_1}$ to $\mathbf{N_{10}}$.

i	pedigree structure	P_i	N_i
1		K	18
2		$1-K$	35
3		$\frac{2-K}{2}$	11
4		$\frac{1}{2}+\frac{1-K}{4}$	1
5		$\frac{1}{2}+\frac{1-K}{8}$	2
6		$\frac{1}{2}+\frac{(1-K)(2-K)}{4}$	1
7		$\frac{1}{2}+\frac{(1-K)(2-K)}{8}$	2
8		$\frac{1}{2}+\frac{(1-K)(2-K)^2}{8}$	1
9		$\frac{1}{2}+\frac{(1-K)(2-K)^2}{16}$	1
10		$\frac{1}{2}+\frac{(1-K)(2-K)}{32}$	1

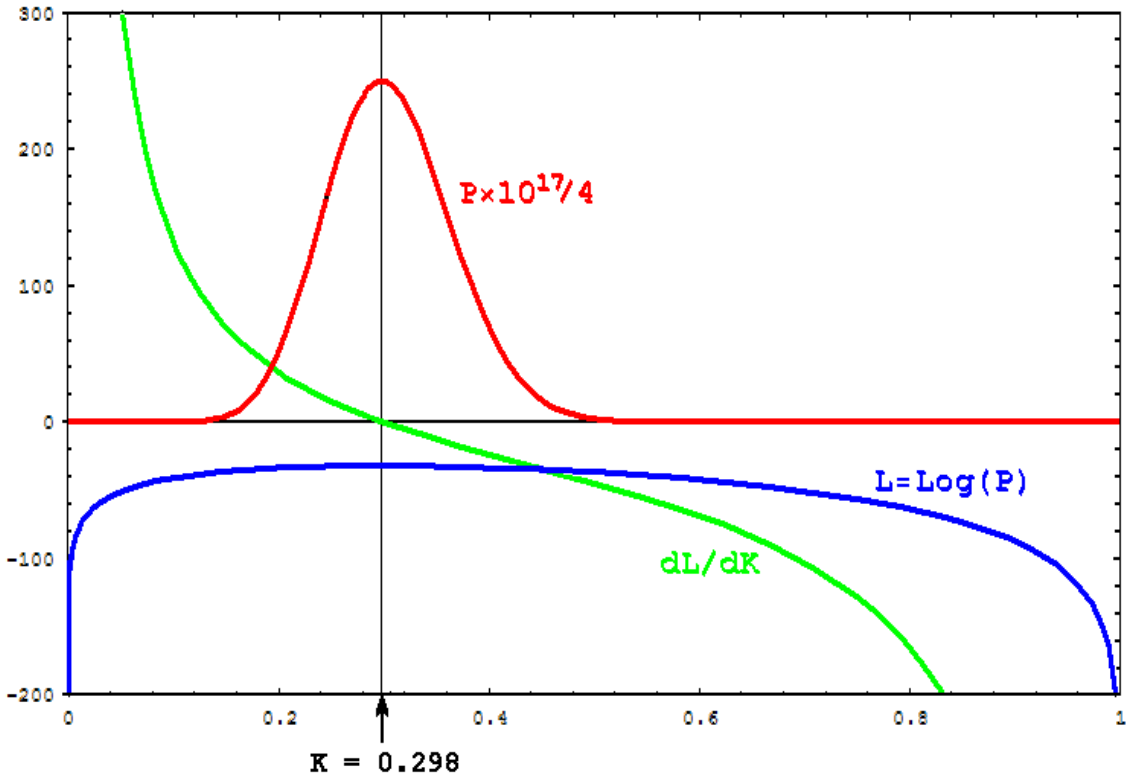
The likelihood function, that is, the probability of occurrence of the set of our ten pedigrees, conditional to the observed structures occurring in them, is derived from the quantities ($\mathbf{P_1}$ to $\mathbf{P_{10}}$) associated with these structures. In the present case, by neglecting constant values unimportant in the maximization procedure that will follow, the likelihood function takes the form $\mathbf{P = P_1 \times P_2 \times ... \times P_{10} = \prod P_i}$. By solving the equation $\mathbf{dP/dK = 0}$ (or, more conveniently, $\mathbf{dL/dK = dlog(P)/dK}$, we obtain the maximum likelihood estimate of the penetrance value K, which for this

set of families with affected cases of **MTM** takes the value of **K = 0.298**, as the following *Mathematica* (© Wolfram Inc.) script and the corresponding graph show.

```
(*
inc_pen_1.ma
*)

p01 = K^18; p02 = (1-K)^35; p03 = (2-K)^11;
p04 = 1/2 + (1-K)/4; p05 = (1/2 + (1-K)/8)^2;
p06 = 1/2 + (1-K)*(2-K)/4; p07 = (1/2 + (1-K)*(2-K)/8)^2;
p08 = 1/2 + (1-K)*(2-K)^2/8; p09 = 1/2 + (1-K)*(2-K)^2/16;
p10 = 1/2 + (1-K)*(2-K)/32;
P = p01*p02*p03*p04*p05*p06*p07*p08*p09*p10;
P1 = P*2.5*10^16;
L = Log[P]; dLdK = D[L,K];
FindRoot[dLdK==0,{K,0.5}]
Plot[{P1,L,dLdK},{K,0.000001,0.999999},PlotRange->{-200,300},
    Frame->True]
```

{K -> 0.298239}



It is also possible to obtain an exact credible interval associated with a given **K** estimate. This interval can be obtained by finding the area that corresponds to a given proportion (v.g., **95%**) of the total area under the graph of the likelihood function. Mathematically, the problem is reduced to integrating the function **y = f(K)** between two limits **a** and **b** with the same ordinate value [**f(K = a) = f(K = b)**], so that

$$\int_{a,b}[f(K)dK] / \int_{0,1}[f(K)dK] = 0.95,$$

an operation which can be accomplished by simple computer programs using numerical integration techniques such as *Romberg's oscullatory method*. The lower and upper limits of the exact **95%** credible interval we obtained for the estimate **K = 0.298** were respectively **0.192** and **0.423**.

References

Horimoto ARVR, Onodera MT, Otto PA. Pencalc: a program for penetrance estimation in autosomal dominant diseases. **Genet. Mol. Biol. 33: 455-459, 2010**.

Otto PA, Horimoto ARVR. Penetrance rate estimation in autosomal dominant conditions. **Genet. Mol. Biol. 35: 583-588, 2012**.

Rogatko A, Pereira CAB, Frota-Pessoa O. A Bayesian method for the estimation of penetrance: application to mandibulofacial and frontonasal dysostoses. **Am. J. Med. Genet. 24: 231-246, 1986**.