COMMENTARY

Rethinking genotype and phenotype correlations in polyglutamine expansion disorders

Susan E. Andrew\(^1\), Y. Paul Goldberg\(^2\) and Michael R. Hayden\(^{1,2,*}\)

\(^1\)Centre for Molecular Medicine and Therapeutics, and \(^2\)Department of Medical Genetics, 416–2125 East Mall, Vancouver, B.C., Canada

Received August 11, 1997; Revised and Accepted August 27, 1997

INTRODUCTION

Disorders caused by triplet expansions present unique challenges for understanding and correlating the genotype with the clinical phenotype. Currently, there is some confusion over ranges for normal and disease alleles and regarding understanding of penetrance and intermediate alleles. Various centres have adopted different definitions, resulting in varying interpretations of allele sizes and their relationship to phenotype by a particular age. Furthermore, the observation in some trinucleotide repeat (TNR) diseases of new mutations arising from ‘intermediate’ sized alleles has resulted in different interpretations over what is a ‘normal’ and what is a ‘disease’ allele. This has also been accompanied by an emerging recognition that many factors, including genetic changes outside of the causative gene and potentially even environmental features, might influence these classical ‘single gene’ disorders. We have collected data from >80 peer-reviewed manuscripts on the seven disorders associated with expansion of a CAG repeat, as well as consulted with experts in the field on unpublished data, in order to define the normal and disease range of alleles, determine the zone of reduced penetrance, if any, and define the size range for intermediate sized alleles proven molecularly to expand to a CAG size in the disease range. Here we present an approach for the interpretation of allele sizes and their relationship to a phenotype for this class of disorders.

CAG SIZE RANGES ASSOCIATED WITH DISEASE

The ranges of allele sizes in the seven disorders is based on the total number of individuals, normal and affected, that have been assessed. This in turn is influenced by the frequency of the disorder as well as, to a lesser extent, the length of time since the gene has been cloned (Table 1) (1–88). The affected range is defined as the range of CAGs detected in persons with the clinical phenotype of a particular disorder caused by mutations in the gene in question. The data available to date are shown in Figure 1 (1–88). For spinocerebellar ataxia type 1 and 2 (SCA1, SCA2), spinal and bulbar muscular atrophy (SBMA), dentatorubropallidoluysian atrophy (DRPLA) and especially spinocerebellar ataxia type 6 (SCA6), the small numbers of affected chromosomes examined means that the ranges will likely change at their margins, as more individuals are assessed (Table 1).

Except for SCA6, the similarity between the ranges of disease (>40 CAGs) and normal size alleles at each of the seven TNR diseases suggests a common mechanism of expansion. However, the interval between normal and disease range can also be used to divide these seven disorders into two groups. For example, SBMA, SCA6, DRPLA and SCA3 clearly have mutually exclusive normal and disease range CAG alleles. On the other hand, there is no obvious gap between normal and disease ranges for Huntington’s disease (HD), SCA1 and SCA2 (Fig. 1).

PENETRANCE, INDIVIDUALITY AND DISEASE

Despite our extensive genetic variability and exposure to different environments, it is remarkable that, for all these disorders, that CAG size is associated so strongly with age of onset—a criteria itself subject to significant differences in interpretation. A disease gene has reduced penetrance when a fraction of patients do not manifest the disorder despite inheriting a known disease-causing mutation (89). However, for many individuals with TNR diseases, the onset of illness is often late and highly variable. Thus penetrance is age-dependent. In these instances it is appropriate to define reduced penetrance as occurring if persons have inherited a mutation in the disease range but lived beyond the expected lifespan without manifesting any signs of the disease.

Reduced penetrance is the one end of the spectrum of expression of a disease-causing mutation and is clear evidence of how unique individuals accommodate to a particular mutation in the context of the whole genome, environment and experience. The delineation of the molecular basis of these diseases has allowed new and definitive approaches to assessment of penetrance in this class of disorders. For all of the CAG TNR diseases combined, the reported instances of non-penetrance are few, given the large number of individuals molecularly assessed (8,30,44,90–92).

*To whom correspondence should be addressed. Tel: +1 604 822 9240; Fax: +1 604 822 9238; Email: mrh@ulam.generes.ca
Figure 1. CAG allele sizes from unaffected and affected individuals for each of the seven CAG expansion disorders (1–88).

Table 1. Comparison of CAG expansion disorders

<table>
<thead>
<tr>
<th>Year cloned</th>
<th>Disease frequency</th>
<th>No of persons analyzed</th>
<th>Molecularly proven new mutations</th>
<th>Zone of reduced penetrance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Normal</td>
<td>Affected</td>
<td>(IA size)</td>
</tr>
<tr>
<td>SBMA</td>
<td>1991</td>
<td>&lt;1/50 000</td>
<td>&gt;300</td>
<td>&gt;200</td>
</tr>
<tr>
<td>DRPLA</td>
<td>1994</td>
<td>rare</td>
<td>&gt;900</td>
<td>&gt;200</td>
</tr>
<tr>
<td>HD</td>
<td>1993</td>
<td>1/10 000</td>
<td>&gt;2000</td>
<td>&gt;2000</td>
</tr>
<tr>
<td>SCA1</td>
<td>1993</td>
<td>&lt;1/100 000</td>
<td>&gt;600</td>
<td>&gt;150</td>
</tr>
<tr>
<td>SCA2</td>
<td>1996</td>
<td>&lt;1/100 000</td>
<td>&gt;1000</td>
<td>&gt;200</td>
</tr>
<tr>
<td>SCA3</td>
<td>1994</td>
<td>&lt;1/100 000</td>
<td>&gt;1000</td>
<td>&gt;600</td>
</tr>
<tr>
<td>SCA6</td>
<td>1997</td>
<td>&lt;1/100 000</td>
<td>&gt;1000</td>
<td>&lt;90</td>
</tr>
</tbody>
</table>

In DRPLA, two cases of possible non-penetrance have been reported (76,90). In the first, an individual homozygous for an allele with a CAG size of 57 had onset at age 17; however, neither of his parents, each harbouring 57 repeats, showed any signs of DRPLA at the ages of 71 and 73 (76). In another instance, an abstract detailing a single family (90) is referred to by Nance as a second case of reduced penetrance (91). However, this family demonstrated autosomal recessive hereditary spastic paraplegia, and this is not consistent with an example of reduced penetrance for DRPLA.

In one SCA1 family, an individual had a CAG size of 44 and was healthy at age 66, two standard deviations beyond the average age of risk (30). If continued until the expected normal lifespan, this would be compatible with reduced penetrance in this case (30).

Of all the CAG expansion diseases, HD demonstrates the greatest number of individuals with reduced penetrance. Several different reports have indicated that a small percentage of people with a CAG size in the affected range may show reduced penetrance (8,17,56,91). The likelihood of reduced penetrance is CAG size-related. The CAG size range, where the majority of persons manifest within their expected lifetimes with clinical signs but a few remain asymptomatic, is the zone of reduced penetrance (ZRP) (Fig. 1). In HD, this spans from 36 to 41 CAG repeats. At the lower range (36–38 CAGs), it appears that non-penetrance occurs with greatest frequency. Accurate delineation of the ZRP is critical and has practical and biological importance. For example, any individual who inherits a CAG size >41 (80% of affected persons) should not be given false hope by reports of reduced penetrance as this has not been noted to occur in these CAG size ranges.

While penetrance is clearly CAG size-dependent, factors other than CAG size contribute to the age of onset and also influence penetrance rates. For example, for the same CAG size between 36 and 41, different proportions of people manifest with signs and symptoms of this illness. Understanding of the unique factors, both genetic and environmental, which may differentiate between these patients, may be particularly important. It is potentially exciting to contemplate that diseases associated with triplet repeats which have been regarded as diseases solely influenced by mutations in a putative gene may eventually be viewed as diseases with both genetic and other contributing factors such as particular environments, acting over time to result in a clinical phenotype. The emergence of non-penetrance for these disorders places the phenotype seen in the present, as the outcome of numerous
influences over time. Childs has referred to this as 'hierarchies of
diseases derived from three symptoms of organization—the
genes, the environment, and development—each with its own
imperatives' (92), acting over time.

**INTERMEDIATE ALLELES**

The determination of the molecular basis of these diseases has
allowed delineation of the sequence of molecular events in sporadic
cases. HD is the only one of the seven CAG TNR disorders for
which there are molecularly proven examples of new mutations
(15,29,58). In DRPLA and SCA3, there are a few cases of potential
sporadic cases, but none have been proven molecularly (42,78). In
HD, ‘intermediate alleles’ (IAs) are defined on the basis of the
following criteria: (i) the CAG size range from which new mutations
have arisen; (ii) allele sizes larger than commonly seen in the general
population but smaller than that seen in patients with HD; and (iii)
are found in clinically normal persons (28,29). All three criteria
need to be present for the allele to be termed an IA (Fig. 1).

A patient with an IA is not at risk for developing signs and
symptoms of the disorder. However, there is a small but significant
risk that the offspring will inherit an expanded CAG allele and
eventually manifest with the disease. Thus an IA may behave as a
‘premutation’. However, ‘premutation’ suggests that it is highly
unstable and that the offspring are at high risk of developing the
disorder, which is not the case. For HD, an IA has never been shown
to expand into the disease range through the female germline, while
for male transmission of an IA the risk of expansion into the disease
range is ∼2.5% for every offspring.

Recently, different approaches to the designation of the IA have
been suggested. These included replacing IA with the term ‘mutator’
(91). However, this would appear to be unwarranted as it implies a
‘mutator phenotype’ associated with significant genetic instability,
as observed in many sporadic tumours as well as hereditary
non-polyposis colorectal tumours, which is not the case.

IAs clearly are different from CAG sizes in the ZRP. It is
inappropriate to term CAG allele sizes that have never been
associated with disease as part of a range of reduced penetrance.
The definition of penetrance requires that persons in this CAG
range have inherited a known disease-causing mutation. IAs and
alleles with reduced penetrance are distinguished by the fact that
there are some affected persons in the ZRP while there are no
affected individuals harboring IAs. These terms are therefore both
useful but are mutually exclusive.

The use of the term ‘indeterminate allele’ should also not be used
interchangeably with IA. The use of ‘indeterminate allele’ suggests
the importance of the CAG size in that range is not known. However,
analysis of large numbers of HD patients has provided reasonable
understanding of the behavior and significance of CAG sizes in
different ranges.

**GAPS, LIFESPAN, RULES AND EXCEPTIONS**

Since penetrance is CAG-size-related, the CAG size of the
interval between disease and normal state may serve as a predictor
of the likelihood of reduced penetrance. For instance, for
disorders with no discernible interval between allele sizes on
normal and disease chromosomes, such as HD, reduced
penetrance may be more frequent, especially in the lowest range
of CAGs seen in affected persons. Similarly, in SCA1 and SCA2
where normal and disease ranges are continuous, some examples
of reduced penetrance would be expected in the low range of
CAG allele sizes seen in affected persons.

By contrast, in DRPLA and SCA3, which both demonstrate a
large interval between the normal and disease ranges, there would
be less likelihood of observing reduced penetrance, owing to the
broad distinction between normal and disease genotypes. We
would therefore predict relatively fewer cases of reduced
penetrance in DRPLA and SCA3 than in HD and SCA2.

The frequency of detecting persons with reduced penetrance is
also related to the expected lifespan. If the expected lifespan for
HD is reduced, the likelihood of finding persons with reduced penetrance
is increased. For example, in Russia, where the life expectancy for
males has recently decreased to 58, the zone of reduced penetrance
for HD would increase to 44 CAG repeats (8) and more unaffected
persons in this range destined for later onset will be detected.
Conversely, as the expected lifespan increases, the zone of reduced
penetrance will decrease and the number of persons with reduced
penetrance will decrease as more persons who have late onset will
develop symptoms in their expected lifespan.

If the mechanism of expansion is one of small incremental
increases, with greater instability associated with increasing allele
length, the probability of finding new mutations (sporadic cases)
would also be predicted to be greater for those disorders where
clinically normal persons with CAG sizes close to the disease
range are found. However, new mutations have not been
described in SCA1 and SCA2 (small or no gap) because in all
likelihood, in these instances, interruptions of the CAG tract
confers stability on normal sized alleles (Table 2). An additional
step, namely loss of the interruption is needed to confer increased
instability on normal sized alleles (Table 2).

| Table 2. Sequence comparison of the CAG tract for the seven CAG expansion disorders |
|---------------------------------|----------------|
| Disease | Normal |
| HD interrupted by penultimate CAA | interrupted by penultimate CAA |
| SCA1 pure CAG | 1–3 CAT interruptions in middle of (CAG)n |
| SBMA pure CAG | pure CAG |
| SCA2 pure CAG | CAA interruptions |
| DRPLA pure CAG | (CAG)nCAA(CAG)nCAA(CAG)n |
| SCA3 pure CAG | (CAG)nCAA(CAG)n |
| SCA6 not described | not described |
REFERENCES


