

New analysis suggests 'small' CAG length doesn't matter after all

Size isn't everything: new research suggests that only a person's big CAG count, not the small one, affects HD onset



By [Professor Ed Wild](#)

February 18, 2012

Edited by [Dr Jeff Carroll](#)

Huntington's disease happens when one of our two copies of the HD gene is bigger than normal. The role of the smaller copy has been much debated. Now a fresh analysis of a huge data set suggests the small 'CAG repeat length' doesn't influence when HD symptoms begin.

What's repeat length?

When the genetic abnormality that causes Huntington's disease was discovered in 1993, one of the things that stood out was that it wasn't just a run-of-the-mill spelling mistake.



Don't sweat the small stuff: the smaller of a person's two HD 'CAG repeat lengths' is no longer thought to affect when they'll get symptoms.

Most genetic diseases are caused by single-letter errors in our genetic code - just one chemical 'bases' that makes up our DNA is altered, added or deleted.

But in Huntington's disease, the alteration is more like a chemical 'stutter'. At the start of the HD gene, a sequence of letters - CAG - is repeated several times - usually between ten and twenty. The team that found the mutation spotted that everyone with Huntington's disease had an unusually large number of CAGs in a row - thirty-six or more in every case.

Everyone has two HD genes

In fact we all have **two** copies of the HD gene - one inherited from mom, the other from dad. And it only takes one expanded copy to cause HD.

We call the number of CAGs in each copy of the HD gene the **CAG repeat length**, and each person has two HD CAG repeat lengths.

Most people have two 'normal' repeat lengths. Most people with HD, or who are going to get HD, have one 'normal' and one expanded repeat length. And a very small number of people actually have two expanded repeat lengths.

Size matters

Before we dive into what's new here, let's look briefly about what **hasn't** changed.

Shortly after the mutation was discovered, researchers realized that people who developed HD at a young age tended to have longer repeat lengths in their big HD gene.

After careful study, it emerged that the larger repeat length was a major factor in determining both when symptoms began and how rapidly they progressed. The bigger the CAG count, the younger the disease was likely to begin.

The relationship wasn't perfect, though - for most people, the repeat length couldn't be used to predict when symptoms might begin. There was still a lot of variation that wasn't due to the bigger of the two CAG counts.

For years now, we've been trying to identify what causes that variation. Is it diet, lifestyle, drugs, or effects of genes other than the HD gene? So far, we're still not sure.

The small repeat length

"We're back to a simple situation: a person's bigger CAG count affects onset, but the smaller one doesn't seem to matter. "

Naturally, researchers have wondered whether differences in the **smaller** of a person's two CAG counts might explain why people with the same 'big' CAG length might get symptoms at totally different ages. But when different teams looked at the effect of the small CAG count, they got different results.

In 2009 a Dutch team looked at data from nearly a thousand patients enrolled in the huge REGISTRY study. As expected, they found that the larger CAG repeat length was the major factor that determined when a person developed symptoms of HD. No surprise there.

But when they examined the effect of the smaller CAG count, they found something unusual. For most people, it appeared to be good for the brain if the smaller CAG count was particularly small. But for people with a particularly high 'big CAG', the opposite was true - it was better if the other CAG count was at the high end of normal.

So, if a person's larger CAG number was 41, it seemed to be better if their other CAG number was 12 instead of 20. But if their larger CAG was very high - 60 or 70, say - then for some reason it seemed to be better if the other CAG number was 20 rather than 12.

Weird - but apparently compelling evidence that both CAG counts were important.

Not so fast!

If you're struggling to get your head round all this small number, big number business - relax! Because thanks to a new study just published in the journal *Neurology*, everything just got much easier to figure out.

A team of researchers led by Prof Jim Gusella of Massachusetts General Hospital in Boston carried out an even larger study, involving over 4,000 people pooled from the REGISTRY, COHORT and PREDICT studies. This new study included all the data from the 2009 study - and load of new data, too.

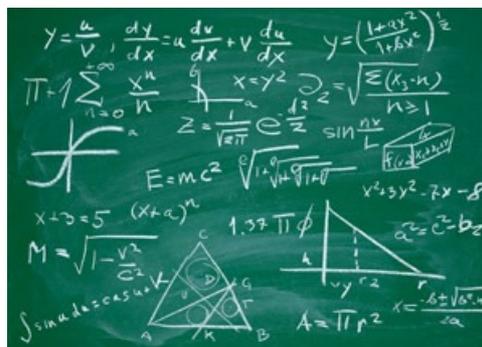
Gusella wanted to go right back to the drawing board, so he had his team question everything about the statistical models that had been used previously.

What they found is a bit geeky but pretty interesting. When statistics boffins analyze data, they have to make certain assumptions, so that they can use mathematical formulas to make predictions. Usually that's OK, because large amounts of data tend to behave as expected.

But on this occasion, they found that one assumption they'd made wasn't correct. In particular, they realized that a single unusual patient - with one very large CAG count of 120, and one very small one of 11, had been to blame for the apparent overall effect of the small CAG count!

When they analyzed the data again with that single person taken out, they found **no effect** of the small CAG count. The only factor affecting symptom onset was the larger CAG repeat length.

Starting from scratch



Gusella's team went back to the drawing board to produce new, reliable ways of studying the effect of genetic factors on HD.

Concerned that one person had had such a misleading effect on a sample of nearly a thousand subjects, Gusella's team set about designing a better statistical model to look at their big data set, that would be less affected by single extreme cases.

What they found was actually very reassuring. There was no effect of the small CAG repeat length, nor any evidence that the small and big repeat lengths can interact.

Even in the ten subjects with two abnormally expanded CAG counts, the only factor affecting onset age was the larger of the two counts.

So we're back to a relatively simple situation: the big CAG repeat length does affect onset, but not in a way that's good at making predictions for individual patients. Meanwhile, the small repeat length doesn't seem to matter at all.

Setback or progress?

This new analysis could be seen as a setback: something we thought we knew no longer holds true.

But we view it differently. We think that finding the truth about what causes HD is the most important thing, even if that means questioning our most basic assumptions.

In fact, the 2009 suggestion that the small and large CAG repeats interact was a bit awkward, and had proved quite difficult to explain in terms of what we know about the mutant huntingtin protein.

So, now that we know that the small allele is back to its original state of obscurity, we've actually got one less thing to worry about. And we can be confident that the statistics behind our understanding are sound.

Another major plus of this study is that it has given us new, more reliable mathematical ways of looking at the effect of genetic differences on symptom onset.

Since large studies are underway, scanning the entire human genome for genes that can influence HD, those methods will likely prove highly valuable in the near future.

This is a great example of what we've said before: science is cumulative. Every day, we know a little bit more about HD. And each day, we're one day closer to an effective treatment.

The authors have no conflicts of interest to declare. [For more information about our disclosure policy see our FAQ...](#)

GLOSSARY

huntingtin protein The protein produced by the HD gene.

CAG repeat The stretch of DNA at the beginning of the HD gene, which contains the

sequence CAG repeated many times, and is abnormally long in people who will develop HD

Allele one of the two copies of a gene

genome the name given to all the genes that contain the complete instructions for making a person or other organism

cohort a group of participants in a clinical research study

© HDBuzz 2011-2024. HDBuzz content is free to share, under a Creative Commons Attribution-ShareAlike 3.0 Unported License.

HDBuzz is not a source of medical advice. For more information visit hdbuzz.net

Generated on May 06, 2024 — Downloaded from <https://en.hdbuzz.net/073>