

# EXCEL HAS BAD GENES

---

AUGUST 11, 2025 DAN LEWIS

---

Here's a joke I learned recently:

An optimist looks at a partially filled glass of water and believes it is 1/2 full. A pessimist sees the same glass and thinks it is 1/2 empty. Microsoft Excel looks at the same glass and concludes that it is the second day of January.

If you don't get the joke, you probably haven't used Excel that often — and you've never felt the frustration of Excel turning numbers into something other than numbers. It's a feature, not a bug. When you're putting together a table of data, oftentimes one of the columns is a date — say, January 2. But it'd be cumbersome to type that out over and over again, so when Excel sees you enter “1/2” (instead of, say, “0.5”), it assumes you meant “2-Jan” and makes the swap for you automatically. That can be annoying, so Microsoft has documentation available [here](#) for how to get around it,

For most of us, that's an annoying but minor inconvenience. But most of us aren't genetics researchers. For them, it's a disaster.

Geneticists use a lot of labels to do their work. And it turns out that the Excel autocorrect feature didn't appreciate those labels. In 2016, a team of researchers led by Mark Ziemann, now the head of Bioinformatics at the Burnet Institute in Australia, authored a [paper](#) outlining just how bad the problem was. Per Ziemann and team, “Microsoft Excel, when used with default settings, is known to convert gene names to dates and floating-point numbers. A programmatic scan of leading genomics journals reveals that approximately one-fifth of papers with supplementary Excel gene lists contain erroneous gene name conversions.” Twenty percent of the papers — hardly a drop in the bucket — had an avoidable error. The problem was so bad that, [as Ziemann wrote in \*The Conversation\*](#), “the Human Gene Name Consortium, the official body responsible for naming human genes, renamed the most problematic genes. MARCH1 and SEPT1 were changed to MARCHF1 and [SEPTIN1](#) respectively, and others had similar changes.” (That last link goes to the gene's Wikipedia entry, which reflects the change.)

But unfortunately, that didn't solve the problem, because the issue extended beyond gene names. [As How Stuff Works explains](#), one of the major genetic research labs in Japan, RIKEN runs large-scale genome projects that involve making “clones” of DNA—cDNA sequences copied from living organisms—that can be stored, shared, and studied. Every clone gets a code, like 2310009E13. In the lab, that's just a

straightforward ID: a string of numbers, a letter, and more numbers. Think of it as a barcode for a gene. But in Microsoft Excel, 2310009E13 looks suspiciously like something else: scientific notation. Excel assumes you meant “ $2.31 \times 10^{13}$ ” and “helpfully” converts it to 2.31E+13. Now, instead of a clone ID that points to a specific mouse gene in a global database, you have... a giant number.

New examples of this problem kept popping up. According to a follow-up study by Ziemann and team, in 2021, the problem had gotten worse — roughly 30% of genetic research papers suffered from an Excel autocorrection error. The problem kept getting worse.

Thankfully, Microsoft seems to have intervened. Even though the help documentation at the top of this story notes that “unfortunately there is no way to turn [date-to-number auto-conversion] off,” that seems to be — pardon the pun — out of date. In October 2023, likely in response to the issue geneticists were having, Microsoft *rolled out a feature* to “disable specific types of automatic data conversions as needed.”

<https://nowiknow.com/excel-has-bad-genes/>