

## Articles In This Section



# Raw Genotype Data Technical Details

*The raw data provided by 23andMe has undergone a general quality review however only a subset of markers have been individually validated for accuracy. The data from 23andMe's Browse Raw Data feature is suitable only for informational use and not for medical, diagnostic or other use. Consult with a healthcare professional before making any major lifestyle changes.*

The [Browse Raw Data](#) feature is provided for customers who are interested in additional research into their genome, but it may be of limited utility for many. The raw data provided by 23andMe is an advanced view of all your uninterpreted raw genotype data, including data that is not used in 23andMe reports. This data has undergone a general quality review however only a subset of markers have been individually validated for accuracy. The data from 23andMe's Browse Raw Data feature is suitable only for informational use and not for medical, diagnostic or other use. Consult with a healthcare professional before making any major lifestyle changes.

This article will address the following questions:

- [How does 23andMe Report Genotypes?](#)
- [Which Reference Genome and Strand Does 23andMe Use?](#)
- [What Does Not Determined/ Not Genotyped Mean?](#)
- [What Are RS Numbers \(rsids\)?](#)
- [Why Don't my Raw Data Results Match Another Source?](#)

## How 23andMe Reports Genotypes

The 23andMe genotyping platform detects single nucleotide polymorphisms (SNPs) and some more complex variations such as insertions and deletions at a predetermined set of locations in the genome that have been shown to vary between individuals. These locations are known as “markers,” and the set of possible outcomes is known as “variants” or “alleles.”

### Base Pairs (A, C, T, G)

There are four DNA bases: adenine (A), thymine (T), guanine (G), and cytosine (C). At a given genomic location, you might have a C and someone else might have a T.

Your genotype will usually be reported as a pair of alleles (e.g. "A/G.") because you have two sets of autosomes (chromosomes 1-22), one from your mother and one from your father.

For markers genotyped by 23andMe, the Raw Data feature reports:

- The marker name (an rsID or internal ID number)
- The marker's exact genomic location
- The possible alleles at that marker (usually A, C, G, or T)
- The variants detected in your saliva sample (i.e. your genotype)

In some cases, your genotype will be reported as a single allele because not all DNA is inherited in chromosome pairs. Notably, this applies to mitochondrial DNA and, for the most part, the X and Y chromosomes in males.

### Insertions and Deletions

Occasionally, one or more bases may be inserted into or deleted from the genetic code at a particular location. In this case, your genotype may be reported as an insertion or deletion ('--') instead of an allele pair.

Depending on where in the genome the change is located, either an insertion or a deletion could represent the normal version of the variant. In other words, there are some places in the genome where having an extra base (insertion) is the normal variant and having a deletion is the rare variant. Conversely, there are some places in the genome where having an insertion is rare, making a deletion the normal variant at that location.

23andMe does not report on all possible insertions or deletions. In general, the ones reported on are small, spanning only one or a few bases.

### Not Determined

In order to return highly accurate data to customers, we use a stringent algorithm to make genotype calls. Occasionally, a person's data may not allow us to determine their genotype confidently at a particular marker. When the algorithm cannot make a confident genotype call, it gives a "not determined" result instead. In downloaded data, the entry for any uncalled SNP displays '--' instead of a two-letter genotype.

A number of "not determined" results throughout the raw data are expected, and your data would not have been returned to you if it had not met our quality standards. However, it's important to keep in mind that only a subset of markers have been individually validated for accuracy.

A small portion of markers, including those on the sex chromosomes (X and Y) and the mitochondrial DNA, are difficult to analyze because of biological issues (e.g. pseudogenes, DNA structure, and highly variable regions). These markers are more likely to have a "not determined" result.

## Reference Genome and Strandedness

By default the genotypes displayed on the 23andMe website refer to the plus (+) strand of the Genome Reference Consortium Human Build 37 (GRCh37 or "Build 37") genome assembly. In Browse Raw Data, genotypes are also reported on the plus (+) strand of the subsequent GRCh38 ("Build 38") genome assembly.

## Reference Genome

A reference genome is assembled by scientists as a representative example of the nucleotide sequence of the genome for a species. The reference human genome was first published in 2004, but it is occasionally revised to account for new discoveries and fix errors. When these regions are updated or corrected, a new version of the genome, known as a "reference assembly," is released.

## Strand

Each chromosome consists of two strands of DNA that are complementary to each other. The DNA nucleotide base adenine (A) always pairs with thymine (T) and the base guanine (G) always pairs with cytosine (C) across these two strands. One strand is called the positive (+) strand and the other is called the negative (-) strand. 23andMe always reports genotypes on the positive strand of the specified reference genome assembly.

## Not Determined

In some cases, we are not able to provide a genotype result for a particular SNP. If results cannot be provided, you will see a **"not determined"** message. In the downloaded raw data file, the entry

for any uncalled SNP displays '--' instead of a two-letter genotype. If you see this result, our algorithm may not have been able to confidently determine your genotype at that marker. This can be caused by random test error or other factors that interfere with the test. Some “not determined” variants are expected in the raw data and are not a cause for concern.

## RS Numbers (rsids)

The rsID number is a unique label ("rs" followed by a number) used by researchers and databases to identify a specific SNP (Single Nucleotide Polymorphism). It stands for Reference SNP cluster ID and is the naming convention used for most SNPs.

When researchers identify a SNP, they send a report (which includes the sequence immediately surrounding the SNP) to the [dbSNP](#) database. If overlapping reports are sent in, they are merged into the same, non-redundant Reference SNP cluster, which is assigned a unique rsID.

If a probe on our genotyping platform doesn't correspond to a SNP with a clear rsID, or the probe is assaying a DNA change that is not a known SNP (i.e. it doesn't have an rsID), then that marker is usually assigned an "internal" id ("i" followed by a number). Our researchers may have included some of these "custom" SNPs on our genotyping platform in order to maximize the number of 23andMe features available to customers, as well as to offer flexibility for future research.

In general, many SNPs labeled with an "internal" id in the Raw Data feature may not have a corresponding rsID in outside scientific literature or other third party services.

Genome-wide association studies linking SNPs to traits or conditions usually report their results by rsID. The rsID numbers for SNPs in Health and Traits reports can be found in the Scientific Details section.

## Why Don't my Raw Data Results Match Another Source?

Each chromosome consists of two strands of DNA that are complementary to each other. The DNA nucleotide base adenine (A) always pairs with thymine (T) and the base guanine (G) always pairs with cytosine (C) across these two strands. One strand is called the positive (+) strand and the other is called the negative (-) strand.

By default, the Browse Raw Data tool reports genotypes based on the positive (+) strand of the GRCh37 assembly ("Build 37"), and you can optionally view your raw data based on the positive (+) strand of the NCBI GRCh38 assembly ("Build 38"). Other websites or publications might refer to the negative (-) strand and/or a different genome assembly instead. This would cause a mismatch between the genotypes reported by 23andMe and that source.

You need to know if the other source is referring to the positive (+) or negative (-) strand to find out if your genotype at 23andMe is the same as your genotype at the other source. If the other source is referring to the negative (-) strand, then your genotype from the other source and your genotype from 23andMe should be complementary. For example, if your 23andMe genotype at a given SNP is GG, that is the same as a CC genotype at that SNP from the other source.

---

## **Related articles**

[Navigating Your Raw Data](#)

[Accessing Your Raw Genetic Data](#)

[What Is 23andMe Raw Data?](#)

[How Do I Upgrade My Genotyping Chip?](#)

[Contact 23andMe Customer Care](#)