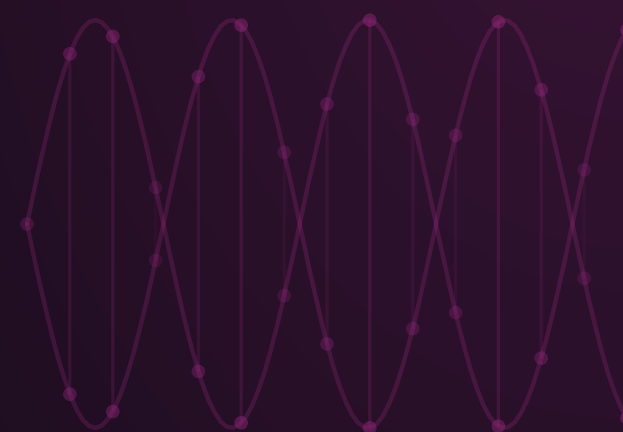


Lesson: why one broken gene copy costs 94% of an enzyme



Chemistry

Biology

Health

Age 16+

Intermediate

A free chemistry lesson with one coin per student. Ask a class what one broken copy of a gene out of two costs you and they will say half. For this enzyme the answer is 94%, the class can compute it in twelve minutes, and the reason is the shape of the protein rather than anything in the DNA.

Free to read, use and adapt. No account, nothing to buy, nothing collected from anyone.

Full lesson, figures and updates:

<https://www.mygenelog.com/updates/lesson-one-letter-four-subunits>

What this is. A ready-to-run chemistry lesson on protein structure, free to use and adapt in any school, college or university. Nothing to buy from us, no account, no data collected from anyone.

- Level: from about age 16 through undergraduate. Assumes the words gene, allele and genotype; assumes nothing about enzymes beyond “proteins that make reactions go faster”.
- Time: 45–55 minutes. The activity itself takes about twelve.
- You need: one coin per student and something to write on. That is all.
- It covers: enzymes as multi-subunit assemblies, why dominant and recessive are statements about protein and not about DNA, rate-limiting steps and why a metabolite accumulates, dominant-negative variants, allele frequencies, and why one protein has two different residue numbers in the literature.

Start here: the arithmetic that did not work

By the 1980s the alcohol flush reaction — the face and neck going red, the heart racing, the headache, after very little to drink — was well known and clearly ran in families. The chemistry was known too. Drink, and the body does two things in order: it turns ethanol into acetaldehyde, and then acetaldehyde into acetate. Acetaldehyde

is the unpleasant one. People who flush accumulate it.

But the inheritance did not add up. The enzyme that clears acetaldehyde is ALDH2, and people who flushed often had one perfectly good copy of the ALDH2 gene. One good copy should give you half the enzyme, and for most enzymes half is comfortably enough. These people should have been fine. They were not.

In 1989 a group in Indianapolis sequenced the protein subunit, found the fault — a lysine where a glutamate should be, at position 487 — genotyped 24 livers, matched genotype to enzyme activity, and reached a conclusion that sounded wrong for a broken enzyme: the broken allele is dominant.

The reason is not in the gene. It is in the shape of the protein, and your class can compute it in twelve minutes with a coin.

Do this

1. Take a prediction first, before explaining anything. “Someone has one working copy of a gene and one broken copy. What fraction of their working enzyme do they have?” Write the answers on the board. Almost every class says 50%, and almost every class is wrong by a factor of eight.
2. Now give them the one fact they were missing. ALDH2 does not work as a single protein chain. Four identical chains assemble into one molecule — a homotetramer — and the assembled molecule needs all four to be intact. One faulty subunit in the four spoils the whole assembly. Crucially, a cell with both copies of the gene makes both kinds of subunit, and does not sort them: each of the four slots is filled at random.
3. Build molecules. One molecule is four coin flips. Heads = a subunit from the working copy. Tails = a subunit from the broken copy. The molecule works only if all four come up heads. Each student builds sixteen molecules — 64 flips, about ten minutes.
4. Run the control before you count anything. Ask what the same exercise looks like for a person with two working copies. There is nothing to flip: every subunit is a working subunit, so every molecule works, 100%. This matters — it shows the model is not rigged to produce a small number. It produces a small number only when there is something to go wrong.
5. Count the class total. Working molecules over total molecules. Thirty students building sixteen each gives 480 molecules and should yield about 30 working ones.
6. Do the arithmetic against the board. $(\frac{1}{2})^4 = 1/16 = 6.25\%$. Not 50%. One broken copy out of two costs about 94% of the enzyme.
7. Then ask the question that makes it chemistry rather than probability. If ALDH2 is running at 6% and the enzyme in front of it is running at 100%, what happens to the molecule in between? Let them answer. The answer is the flush reaction.

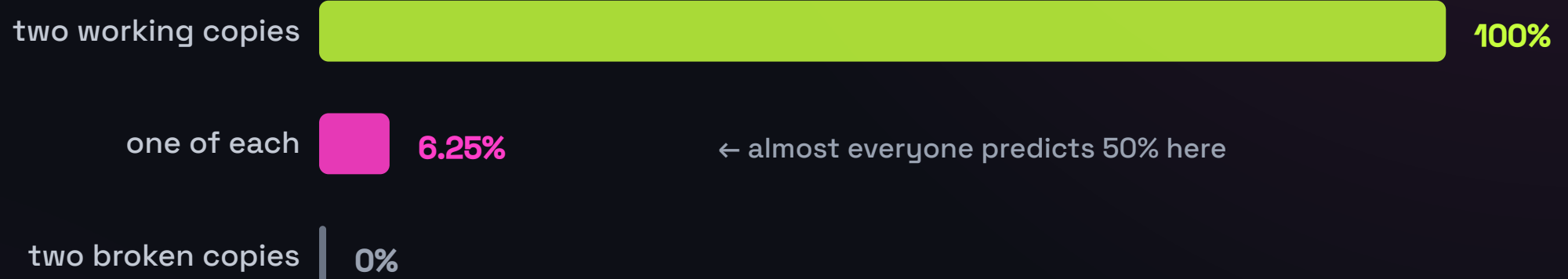
Boundaries for this one, and they matter more than usual.

Do not ask students whether they flush, whether their family does, or whether they drink. The flush reaction is visible, which makes it exactly the kind of trait a class can turn on one person — and unlike most genetics, everyone can see who has it. Keep it about the protein. Count nothing about the people in the room, collect no genetic or health data, and make no name lists.

And say the health part plainly, because a lesson about alcohol metabolism that leaves it out is doing harm: this is not information that makes drinking safer for anyone. For people who carry this variant it runs the other way, and the risk of oesophageal cancer with regular drinking is substantially raised. Nothing here is a reason to test anybody.

What the class just computed

share of assembled four-subunit molecules with no broken subunit in them



$(\frac{1}{2})^4 = 1/16 = 6.25\%$ — exact arithmetic, not a measurement

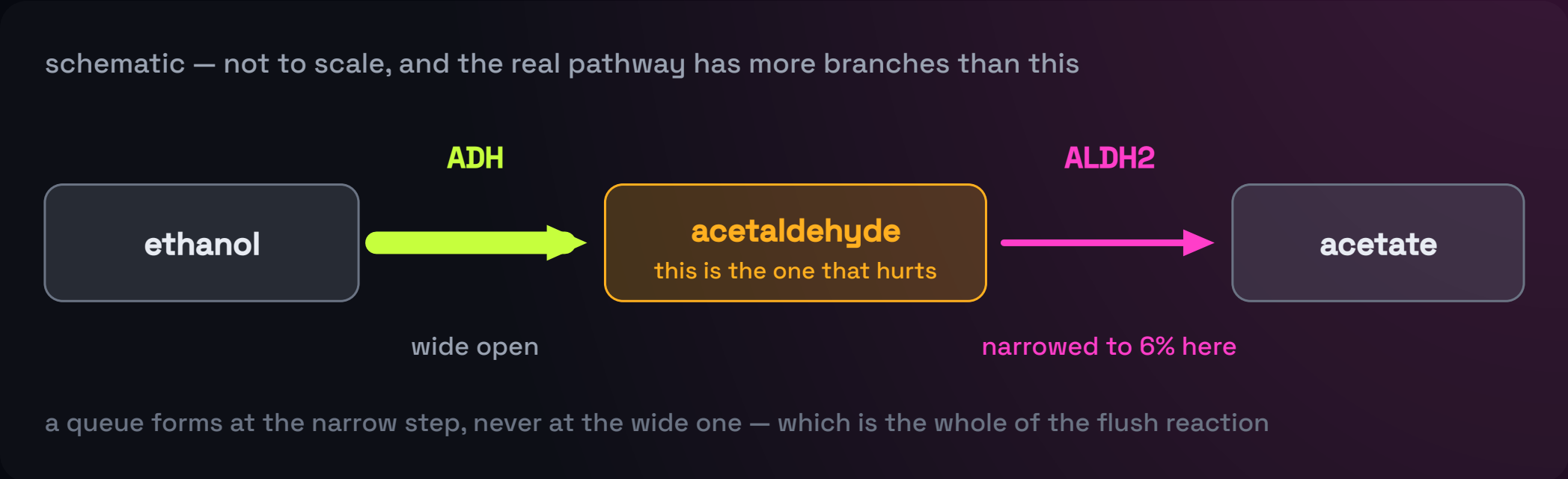
The middle bar is the lesson. One broken copy out of two does not cost half the enzyme; it costs almost all of it, because a molecule needs four good subunits and gets its subunits at random from both copies.

This is what a dominant-negative variant is, and it is worth naming for students who have only ever met dominant and recessive as labels on a Punnett square. The broken subunit is not merely useless. It is actively destructive, because it ruins every assembly it joins — including assemblies in which the other three subunits are perfect.

It also settles something students often half-believe: dominant and recessive are statements about proteins, not about DNA. Nothing about this variant is dominant at the level of the gene. It becomes dominant because the protein happens to be built out of four parts. A different enzyme with the same mutation, working as a single chain, would be recessive.



Why it is acetaldehyde that piles up



A schematic, drawn to make one point: when a two-step pathway jams, the thing that piles up is whatever sits in front of the slow step. Alcohol does not build up. Acetaldehyde does.

The second point of the lesson is a general one about pathways, and it is worth drawing out because students routinely get it backwards.

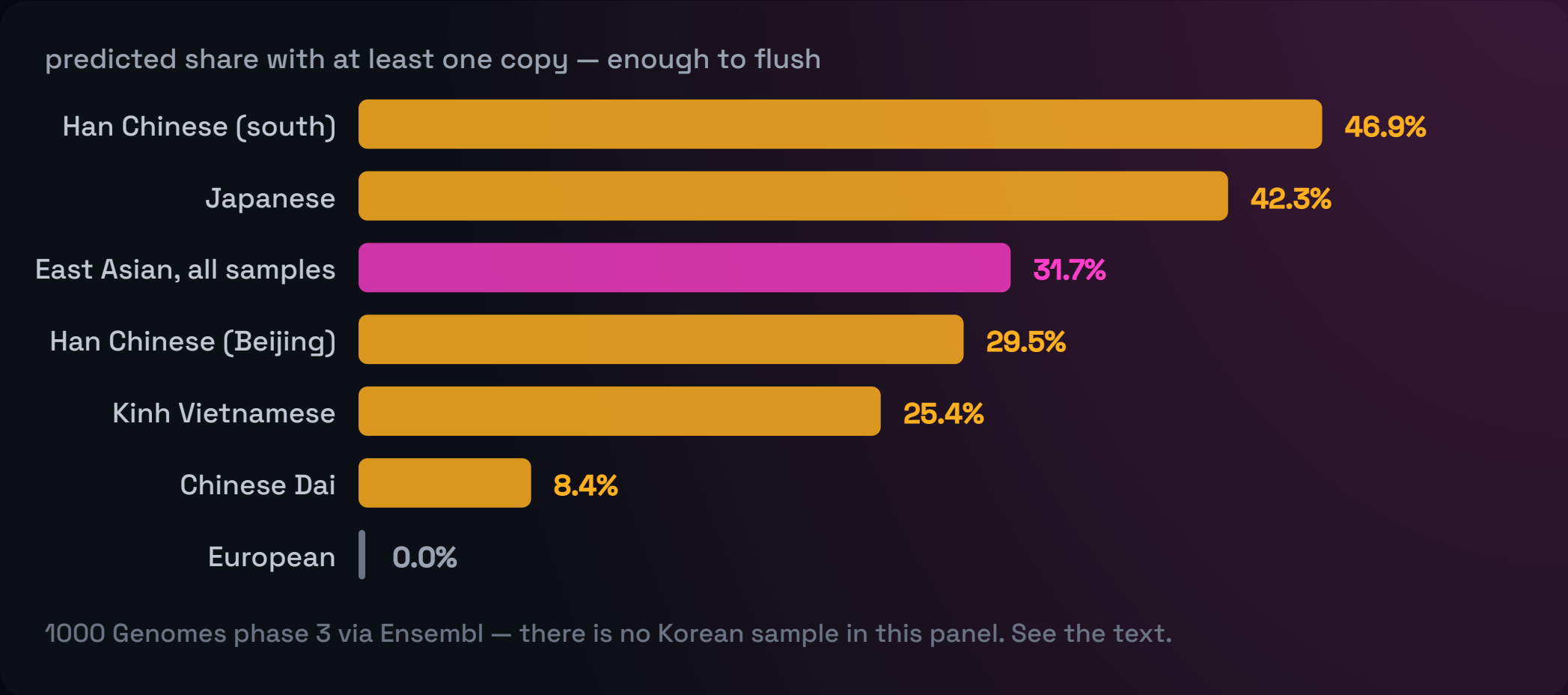
Alcohol does not accumulate. The step that converts ethanol to acetaldehyde is not the damaged one. What accumulates is whatever is waiting in front of the slow step, and here that is acetaldehyde — a reactive compound that forms adducts with proteins and DNA. The flush, the racing heart and the headache are what that feels like.

Ask the class to generalise it: in any pipeline with one slow station, where does the queue form? Every student who has queued for anything already knows the answer, and the point is that the chemistry is not different.



Real data: who carries it, and one honest gap

The variant is rs671. Here is what the published allele frequencies predict.



Calculated, not counted: $1 - (1 - q)^2$ from the published allele frequency at rs671. Two populations that a textbook would put in the same box — Chinese Dai at 8.4%, Southern Han at 46.9% — are five times apart.

Two things in that figure are worth stopping on.

First, the spread inside one region. Chinese Dai at 8.4% and Southern Han Chinese at 46.9% are five times apart, and a textbook that says “common in East Asians” hides that completely. Population labels are containers for a lot of variation.

Second, the gap. There is no Korean sample in the 1000 Genomes panel. A class in Seoul cannot look itself up in the figure above, and no amount of arithmetic fixes that — the number simply has not been collected into this reference set. Point it out. It is the most useful thing on the page for a student who is about to spend a career reading reference data, and it is the same gap in a different place every time.

One protein, two numbers

A small thing worth five minutes with an older class, because it teaches how to read a paper without being confused by it.

The 1989 paper calls this variant Glu487Lys. Modern clinical databases call it Glu504Lys. Both are correct and they describe the same single change.

ALDH2 works inside mitochondria, and proteins destined for mitochondria are made with an extra stretch at the front — a targeting sequence that acts as an address label and is cut off on arrival. Count from the start of the protein as it is made and the residue is 504. Count from the start of the protein as it ends up and it is 487. The difference is the length of the address label: $504 - 487 = 17$ residues.

The argument that is still open

Is this allele good or bad? The honest answer is that it depends entirely on what the person carrying it does, and that makes it a genuinely awkward case.

The argument for “protective”: flushing is unpleasant, so people who flush drink less on average, and drinking less protects against everything that heavy drinking causes. A variant that makes a harmful behaviour unpleasant is doing something no drug does.

The argument for “harmful”: for those who drink anyway — socially, at work, under pressure — acetaldehyde clears slowly, exposure is prolonged, and the risk of oesophageal cancer rises substantially. The variant is protective only if it successfully changes the behaviour, and whether it does is partly a matter of culture rather than chemistry.

Then the harder question, which older classes are more than capable of: should anyone be tested for this? The result would change advice for a person who drinks. It would also label people, in cultures where drinking is a work obligation, with something an employer or a family could ask about. Argue it both ways. There is no consensus to hand them.

Questions

To warm up

1. Why does alcohol not build up in someone whose ALDH2 is impaired, when the problem is a step in the pathway for breaking down alcohol?
2. Your class built molecules and about one in sixteen worked. Where does the number 16 come from?
3. What was the control condition in this activity, and what would you have concluded without it?

Core

1. Explain why this variant is dominant, in a way that never mentions DNA.
2. Suppose ALDH2 were a two-subunit enzyme rather than four. What fraction of molecules would work in someone with one broken copy? What if it were a single chain? What does that tell you about where dominance comes from?
3. The Chinese Dai figure is 8.4% and the Southern Han figure is 46.9%. Both are labelled East Asian. What should you do with a sentence in a textbook that says a variant is “common in East Asians”?
4. The same change is called Glu487Lys in 1989 and Glu504Lys today. Explain the difference, and say why 17 is the number that connects them.

To stretch

1. The 6.25% figure assumes the cell makes equal amounts of both subunits and assembles them at random. Which of those two assumptions do you think is more likely to be wrong, and how would you find out? What would it do to the prediction if the broken subunit were made slightly faster?
2. A variant that reduces an enzyme to 6% could be lethal, mildly inconvenient, or protective, depending on the enzyme. What does it depend on? Name the properties of a pathway that decide how much damage a bottleneck does.
3. Is this allele under natural selection today? Consider both directions — it reduces alcohol-related disease by reducing drinking, and it increases cancer risk in those who drink anyway — and say what you would need to measure to answer it.
4. The frequency panel has no Korean sample. Design the smallest study that would responsibly fill that gap: who would you need to recruit, what would you tell them, and what would you not do with the samples afterwards?

Teacher notes — where each question is going

1. The queue forms in front of the slow step. The ethanol step is unaffected and runs at full speed, so ethanol is consumed normally; acetaldehyde is produced normally and removed slowly. If a student says “alcohol builds up”, this is the misconception the figure exists to fix.

2. Four independent slots, two outcomes each, one favourable combination: $2^4 = 16$. Some students will get there through the tree diagram and some through the multiplication rule; both are worth hearing out loud.

3. The two-working-copies case, which yields 100%. Without it the model looks rigged to produce a small answer. Say explicitly that this is what a control does in general: it shows the method can produce the boring result when the boring result is true.

4. Looking for: the faulty subunit spoils any assembly it joins, so a cell making both kinds ruins most of its molecules. Dominance is a property of how the protein is built. Students who reach for “the allele is stronger” have not got it yet.

5. Dimer: $(\frac{1}{2})^2 = 25\%$. Single chain: 50%, and the variant would be recessive. The subunit count is the mechanism of dominance here, which is the generalisable idea.

6. Treat it as a rough grouping and go and find the number for the population actually in question. Good students will notice this is the same problem as a reference range built on one population being applied to another.

7. The mitochondrial targeting presequence, 17 residues, cut off on import. Precursor numbering versus mature numbering. This is the sort of thing that makes a student think two papers disagree when they agree exactly.
8. Random assembly is the safer assumption; equal expression is the shakier one, and it is measurable. If the broken subunit were more abundant, the working fraction would fall below 6.25%. Strong students may realise the observed activity in heterozygotes is a test of the model rather than an illustration of it.
9. Depends on how much spare capacity the enzyme has, whether the substrate is toxic, whether there is a second route around the step, and how fast the substrate arrives. This is the general theory of why some “broken” enzymes do nothing at all to a person.
10. No settled answer, and that is the point. Push them towards what would have to be measured: reproductive-age mortality, family size, and whether either differs by genotype in a population where drinking is common. Frequency alone cannot answer it.
11. Consent, purpose limitation, what happens to the samples, who gets to re-use them and for what. If a class writes a study that solves the science and ignores the consent, that is the discussion to have. This site’s answer, for what it is worth, is that we collect nobody’s genome at all.

Take it further

- ALDH2 deficiency — the full page, with the evidence on oesophageal cancer risk and the sources behind it.
- rs671 and rs1229984 — the ALDH2 variant and a variant in the enzyme upstream of it, which changes the same pathway from the other end. Good extension for question 9.
- Lesson: the threshold — the companion lesson on why a published association might not be real. Runs on coins too.
- What is a SNP — if the class needs the vocabulary first.

Use this freely. Print it, copy it, cut it up, put it on your own worksheet, translate it, change the questions. No permission needed and nothing to pay. If you credit it, mygenelog.com is enough.

If you teach with it and something in it does not work, tell us — that is worth more to us than a thank-you.