I'm a Scientist Get me out of here

BBSRC 20 Years of Pioneering Great British Bioscience

SFI Discover Science Foundation Ireland

Science Debate Kit: Big Data

For more kits and activities go to debate.imascientist.org.uk

Debate Kit: Big Data

Should we sequence the genomes of one million people, to find out more about living longer and healthier?

A structured practice debate on a controversial topic. The different 'rounds' of the debate help students think through the issues and reconsider their opinions. The structure also shows them how to build a discussion and back up their opinions with facts.

You can use all eight characters, or fewer, as you wish.

The minimum is the four essential characters (**in bold**), this gives two for and two against.

Characters

For genome sequencing	Against genome sequencing
• **Andy Eaves – Healthy ageing researcher**	• **Josh Hoplin – Anti-drugs companies campaigner**
• **Nuala McNicol – Epidemiologist**	• **Katya Jandziol – Agriculture researcher**
• Ben McMillan – Politician	• Bayram Parmuk – Physicist
• Siobhan Weaver – Poet	• Chrissie Tabor – Social worker

Facilitation tips

Ensure pupils know there is no right or wrong answer. Be observant of those who want to speak and are not getting a chance. Encourage students to give a reason for their opinions.

Designed for 14-16 years but can be used by 11-18 years

For groups who may need extra support you can put the following prompt sentences upon the board:

"I think we should/shouldn't sequence a million people's genomes because……."

"I think …… is the most important point to think about."

Learning notes

Learning objective:

- To practise discussing and debating issues and expressing an opinion.
- Understand more of the technical, social and ethical issues around genomics.

Other learning outcomes:

- Consider social, ethical and factual issues in an integrated way.
- Think about different points of view.
- Learn to back up their opinions with facts.

Curriculum points covered:

Working scientifically
- Societal aspects of scientific evidence
- Developing an argument

Substantive
- Consider the potential impact of genomics on health, and the practical and ethical issues it raises.
- Consider what science can understand and contribute about healthy living and living longer.

Teacher Notes

***Question:* Should we sequence the genomes of one million people, to find out more about living longer and healthier?**

Lesson plan

The different 'rounds' of the debate help students think through the issues and reconsider their opinions. The structure also shows them how to build a discussion and back up their opinions with facts.

Starter: 5 minutes.
Do they know what genome sequencing is? What do they think you can tell about someone by sequencing their genome? Would they want to have their genome sequenced if they had the chance? What would be reasons to say yes or no to that question?

Main Activity: 35 minutes.
1) **Split students into** as many **groups** as characters you want to cover.
2) **Give** them their **character cards** – one per group, and give them a few minutes to read them over.

KS4: Designed for 14-16 years but can be used by 11-18 years

I'm a Scientist Get me out of here

3) Get one student in each group to **read out** their **first section** to the rest of the class. What are the class's initial thoughts? Is there one position they identify with or reject?
4) Take it in turns to read out their fact. Does it change the way they think?
5) **Read** the **issue**. Any different feelings?
6) Each team **asks** their **question to** the **character of their choice.**

Support: To help students you can put the following prompt sentences up on the board:
"I think we should/shouldn't sequence a million people's genomes because……."

"I think …… is the most important point to think about."

Plenary: 10 minutes
Vote for which position they agree with most (if there is one). Why? Which arguments were the most persuasive?

Note – Pupils can stay in roles all the way through the debate, or only for the first round if you prefer. If it's all the way through, give them a chance to express their own opinion at the end and in the plenary.

For groups who are not confident at class discussion, it might help to have them start by discussing the question and/or their character's position in pairs, and then compare notes in fours. They've then had chance to rehearse some of what they want to say before having to do it in front of the whole class.

Background notes

What is Big Data?

Big Data is 'anything where crunching the numbers is too big for you to handle alone'. By that definition, Big Data has been with us for a long time. The Human Genome Project involved thousands of researchers at dozens of institutions across the world. And there were certainly multi-site collaborations in science well before that. But advances in data collection and computer power mean that what was extraordinary then is becoming commonplace now.

Data-driven science

We also now have the computing power to crunch much bigger datasets than we could previously – looking for associations or shared patterns. This means that we can gather huge quantities of data and then go on 'fishing expeditions' to see what pops out. Some people describe this as data-driven science, rather than theory-driven science.

Instead of starting with a hypothesis – e.g., a particular mutation in gene X causes Y – and gathering specific data to test that hypothesis, we can now gather huge amounts of data – e.g. the genomes of thousands of individuals – and then look for correlations. The huge size of the data set means you can pick up smaller associations than you'd be able to with a smaller sample of the population.

However, this also means that you can find chance correlations that mean nothing. And some argue that because of the huge numbers involved, we put too much faith in Big Data's results.

Sequencing and genomics

It took eleven years to first sequence the human genome (from 1990 to 2001). Now an individual's genome can be sequenced in a couple of days. Scientists are sequencing many new organisms, giving us insights into everything from the genetics of the crops we depend upon, to the interdependence of ecosystems. We are also sequencing more individual people, giving us better insights into genetic variation.

Sequencing technology is improving all the time. Although it's not likely at the moment that we would sequence a million people's genomes, it could become a real possibility in the not too distant future, and the time to think about the issues this might raise is now. In fact, some think that in the future **all** babies will be sequenced at birth.

Too much of a good thing!

This means we are producing massive quantities of data, much of which we can't interpret.

Another issue faced by health bodies and research councils as we enter the era of Big Data is that because it's so new, we don't have large numbers of biologists with the maths and computer skills to handle this sort of data. (Or, we don't have the mathematicians and computer experts with the biological knowledge to make sense of it.) These are questions that today's teenagers may grow up to answer. So if there's a student in your class who loves biology, computer science and visual art, then encourage them to carry on with all three!

In 2011, the McKinsey Global Institute estimated that between 140,000 and 190,000 additional data scientists will need to be trained by 2018 in order to meet the increased demand in academia and industry in the United States alone.

Healthy ageing

Research in Iceland (where they have genealogical records going back to 874 A.D.!) has started to identify some genes that seem to contribute to longevity. The obvious next step in that research would be to sequence a much larger number of people - perhaps the entire population of Iceland. But as we mention in the kit, the Icelandic population is unusually homogeneous, and science might be better served by sequencing a more diverse population. Or a large sample of them. Large-scale population sampling in that way would raise various questions about privacy, and about representative samples, some of which are raised in the kit.

Suggested homework:

1. Draw up a list of pros and cons for sequencing a million people for healthy ageing research.
2. Imagine a new character in the debate and draw up a character card for them - each character needs a **name**, a **summary** of who they are and their position, a **fact**, an **issue**, and a **question**. If you think your class will have trouble with imagining a new character from scratch, you could brainstorm what other issues came up in the debate, and what other people might have strong opinions on it, as part of the plenary.

All facts in this kit have been researched. References can be found online at: debate.imascientist.org.uk/big-data

Special thanks to Dr Richard Emes, from the University of Nottingham; Dr Caroline Relton from Newcastle University; Dr Nicolas Le Novère, from the Babraham Institute; Dr Nick Schurch, from University of Dundee; and Dr Anna Middleton and the Wellcome Sanger Institute Outreach Team.

This kit has been produced by the award-winning I'm a Scientist team and funded by BBSRC and Science Foundation Ireland.

BBSRC 20 Years of Pioneering Great British Bioscience

SFI Discover Science Foundation Ireland

Debate

Nuala McNicol – Epidemiologist

I study the patterns of health within a whole population. Some diseases – like cancer or osteoporosis – will affect some people much more than others. We could use the data from the healthy ageing study too. With a sample of the population this big, we could make much better models of who is likely to get a disease.

Fact: With a small sample of the population there's always a risk you'll find associations that aren't really there, because of sampling error. Or you may miss small associations that are there, because your sample isn't big enough to capture them.

Issue: Technology is advancing so rapidly that soon we'll be able to sequence the genomes of everyone at birth.

Question: Now we are able to handle this much data, shouldn't we be using it? Instead of small samples that we can find out much less with?

I'm a Scientist Get me out of here

Debate

Josh Hoplin – Anti-drugs companies campaigner

I don't trust the big drugs companies, they've got no morals. They also swoop on data that we, the tax-payers, have paid for, and then they do an extra little bit of research, invent a drug then they make all the money off it! If there was a big database like this, who gets to access it? I bet the government would sell it to drugs companies.

Fact: Once genetic data is combined with medical data it can't ever really be anonymous, because it contains so much information about you

Issue: I don't want drugs companies getting to look at MY data about ME, and I definitely don't want them making money off it!

Question: Once the information is there, who gets to access it? And how do we know it can be kept secure?

I'm a Scientist Get me out of here

Debate

Chrissie Tabor – Social worker

I work with some of the poorest and most vulnerable people in our society. Immigrants, vulnerable adults, people with mental health issues. Some of these people don't trust our health system. Some are unable and unwilling to give informed consent. Your big sample wouldn't include them.

Fact: Research shows that some groups of people are harder to make contact with and sample.

Issue: I think a skewed data set is worse than none, because it will SEEM like we've got the whole picture, but actually we'll be missing stuff.

Question: Are you happy to base your research on a sample that's probably biased?

I'm a Scientist Get me out of here

Debate

Ben McMillan – Politician

I think the Human Genome Project was a triumph of humanity, like putting a man on the moon. It took thousands of people working together, and it was a huge step forwards for science. But in a way, that just gave us a rough map. Sequencing a million people's genomes could be the next leap forwards as we'd learn so much about the variation between people.

Fact: The Human Genome Project launched a new area of biology called bioinformatics helping us understand much more about genetics and health.

Issue: If we understood genetics even better we could transform the lives of so many people. Fewer cancer sufferers, less dementia, less heart disease.

Question: This could be a great achievement for humanity. Shouldn't we spend our money on that, instead of things like nuclear weapons?

I'm a Scientist Get me out of here

Debate

Bayram Parmuk – Physicist

In a normal experiment, you ask one or two questions. When you've got a great big data set like this, you keep going back to it and asking lots and lots of questions. In physics we are used to handling huge data sets. My worry about this genome plan is that biologists aren't used to handling this much data. Most of them aren't as mathsy as we are. You have to be really careful about false positives, or you'll 'find' all sorts of stuff that's not really there.

Fact: A false positive is when you get what looks like a positive result, but actually it's just chance.

Issue: If you ask one question in an experiment, and get a 'significant' positive result, then the probability of it being a fluke is 1 in 20. But if you ask 20 questions, then odds are, one of your results looks positive, just by chance.

Question: Where are we going to find enough biologists who understand maths, to make sure they do this right?

I'm a Scientist Get me out of here

Debate

Siobhan Weaver – Poet

My mother had Alzheimer's disease. My grandfather also had dementia. I've been tested, and I have an increased risk of dementia. A healthy lifestyle and avoiding stress can help me avoid dementia. I want scientists to find out as much as possible so we can all be armed with the right knowledge.

Fact: Most genes associated with a disease aren't a death sentence, they just make it a bit more likely you'll get it – so by changing your lifestyle you can even the odds.

Issue: I think having information about your own genome gives you more control over your own destiny.

Question: Don't we want to face life with as much information as possible?

I'm a Scientist Get me out of here

Debate

Andy Eaves – Healthy ageing researcher

It's my team that would be leading the research. From studies of all 270,000 people living in Iceland we already know some genes that are associated with living longer. But Iceland is small and not very diverse. We want to sequence a big section of our more diverse population. This will give us loads of data to look at. We can then try to work out what helps people stay healthy and live longer.

Fact: There are now thousands of people who've had their genome sequenced, but they must undergo a rigorous screening process that limits the range of volunteers.

Issue: If you only look at a subsection of the population, you'll get the picture all wrong. We want to know the full picture!

Question: Don't we all want to live longer, healthier lives?

I'm a Scientist Get me out of here

Debate

Katya Jandziol – Agriculture researcher

The world is running out of land to grow enough food for everyone, and we plan to spend billions sequencing a million people just to tinker with our understanding of our genes? I think the money would be much better spent on sequencing food crops to help us work out how to feed everyone in the future.

Fact: We've already sequenced the genomes of thousands of people, but over 99% of plant species haven't even been sequenced once.

Issue: Human genomics might create more news headlines, but plant genomics, helping us feed the world in future, could be far more beneficial.

Question: Aren't there other, more important things we should be spending our money on?

I'm a Scientist Get me out of here