

Science AMA Series: I'm Nancy Cox, I study the genetic and environmental causes of diseases like diabetes, asthma, cancer, and heart disease, AMA!

Dr_{NancyCox}¹and/ScienceAMAs¹

¹Affiliation not available

April 17, 2023

[REDDIT](#)

Science AMA Series: I'm Nancy Cox, I study the genetic and environmental causes of diseases like diabetes, asthma, cancer, and heart disease, AMA!

DR_NANCY_COX [R/SCIENCE](#)

ABSTRACT

[removed]

[READ REVIEWS](#)

[WRITE A REVIEW](#)

CORRESPONDENCE:

DATE RECEIVED:

November 19, 2015

DOI:

10.15200/winn.144784.49008

ARCHIVED:

November 18, 2015

CITATION:

Dr_Nancy_Cox , r/Science , Science AMA Series: I'm Nancy Cox, I study the genetic and environmental causes of diseases like diabetes, asthma, cancer, and heart disease, AMA!, *The Winnower* 2:e144784.49008 , 2015 , DOI: [10.15200/winn.144784.49008](https://doi.org/10.15200/winn.144784.49008)

© et al. This article is distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author and source are credited.



How much does the food we eat affect the likeliness of our ending up with these diseases? I've read a lot in the last few years that it's near 100%. Are we able to confirm any number near that high?

[mmhmmyes](#)

Well, certainly, the amount of food that we eat is known to impact the risk of some common diseases through weight gain, and the metabolic consequences of weight gain. For many years, nutrition scientists stressed the overall number of calories, rather than where those calories came from, but there is emerging evidence that liquid calories -- sugar water in essence -- may contribute disproportionately to some of the bad metabolic consequences that had been generally attributed to obesity. With respect to phenotypes like allergy and asthma, there can be triggers for disease after it has developed, of course, and there is some evidence that exposure to certain kinds of environments ("dirty" environments -- barns, pets, my house) in early life reduces the risk of developing asthma and allergies. So, the short answer to the specific question is that there is some emerging evidence that what you particular things you eat and are exposed to can influence your risk of disease. But how to say exactly what proportion of total risk can be attributed to specific diet, is more difficult, and it will not be anything like 100%. Most obese people do not have and will never develop type 2 diabetes; it is true, though, that for most people with type 2 diabetes, if they could successfully lose weight and increase exercise levels, they would no longer have type 2 diabetes -- especially if they could do that near the time they are first diagnosed. Similarly, most people who live in "clean" environments never develop asthma. Most common diseases arise as a consequence of BOTH the genetic variation we have inherited and the environment we have. You cannot pick your parents, and so reducing risk is often based on altering what environmental risk factors we know will help reduce risk of disease. But it is also important to note that some of our environmental risk factors are harder to control. People exposed to higher levels of pollution are more likely to develop asthma, and people with asthma are more likely to be symptomatic when pollution is worse. You can modify exposures by staying indoors, but some people cannot change locations physically because of job / family commitments. Some kinds of environmental exposures that will be important are not ones we know yet.

Stress, smoking, poor diet... those are all risk factors that everybody knows about.

What are some risk factors that people are not familiar with but are very common?

[morenax](#)

Some exposures that can be important for development of disease are not uncommon, but may be much more common in certain professions, for example. Adult onset asthma can occur in people exposed to very fine particles in their work -- like flour in bakeries, for example. Asbestos exposure is probably more common than people appreciate, and is another example of an exposure where the underlying genetic risk factors are also very important in whether or not someone develops disease as a consequence of exposure. It is widely known that people who mine asbestos or work in asbestos factories have a higher risk of developing a form of cancer called mesothelioma. But fewer than 10% of the most heavily exposed workers develop disease, and sometimes disease arises not in the person who is most exposed, but in one or more family members who are exposed only to the particles that travel on that persons clothes. Here again, it is having exposure when you also have genetic risk factors that make you most vulnerable to the consequences of that exposure.

A) Do you have an opinion on the hygiene hypothesis?*

B) Do you have any accepted techniques for analyzing large genetic data sets and their change with time? I am thinking about how Fourier Transform IR spectroscopy accomplished this for IR spectroscopy and if similar approaches are reasonable using genetics to monitor the "spectrum" of the health of a population.

*The idea that certain disease like asthma and cancer are higher in the western world because we are less exposed to bacteria, etc so our immune systems run a bit wild.

[TheHumanPiranha](#)

There is a lot of support in many different dimensions for the hygiene hypothesis. But as with any fairly general hypothesis, it may be too often stretched to fit stories for which there are not yet supporting data. Our immune system is clearly developing in a very different environment than humans had for most of their evolution, and so it cannot be surprising that this would lead to some consequences. There is now a lot of research on how to modulate immune system development to nudge it in a more favored direction if it seems to be going off in the wrong way, but taking that research into human studies requires that we be very certain we can measure when the immune system is developing in a way that will not be optimal for health, and that we can confidently move it only as much as we need to for improving health outcomes. It will be challenging, but the recent successes in using our own immune system to fight cancer suggests that we are learning rapidly and all of what we learn should help improve our understanding of how to do this in other contexts.

With respect to big data science, we have worked in large-scale prediction modeling, and in ways of truly integrating data (integrating genome function with genome variation, for example). This is really exciting science and among the most intellectually stimulating research being done in the lab now. To see how the large scale predictive models perform in large numbers of individuals who have agreed to these large studies is going to be very important in moving personalized medicine forward rapidly. It is often many, many years between when scientific discoveries are made and when they actually benefit patients, but with the size studies that are being conducted now and the even larger ones being planned for the near term, I hope we can bend that curve for genetic discovery. The key is building use cases for diseases that we can do something about -- where prediction will matter. And one of the most challenging questions about making this cost effective is whether people -- and physicians -- will be willing to trade the kinds of generic screening we have now (eg mammograms for women over 40, or 50,) for more targeted screening, where some people might be told that they don't really need much in the way of cancer screening, but should have more rigorous and regular exams for cardiovascular health. I'm curious to know how you think people would react to that. How would you react to that kind

of advice?

As a long time asthma sufferer, I would firstly like to thank you. My question however is, what kind of treatments are they looking to for the future?

[charfahl](#)

This is a great general question -- curing as opposed to treating disease. To be honest, I think it is likely that we will be able to prevent some diseases faster than we can cure them. That is because the disease process itself alters key tissues in ways that -- at least right now -- we do not know how to fix. When we really understand the biological basis of disease and how genetic and environmental risk factors interact to increase risk of disease, we can work to modify environmental risk factors particularly in those at highest genetic risk to reduce risk of disease. Preventing disease is of course an outstanding outcome, but curing diseases that people have today is something that will likely be possible for some diseases. For others, the disease may move from something that is very difficult for patients and may reduce the length of life, to something that is a very manageable chronic disease. For asthma, one possibility is that treatments will become much more specific to the actual gene-environment interactions that appear to be most important in driving a given person's disease. And combination therapies are also likely, where one part of the treatment is to reduce the likelihood that the particular genetic factors that a person has will be triggered, while another treatment is directed at reducing the consequences of the damage that has been done by having the disease for a lifetime.

Hi Dr. Cox, thank you for doing this AMA. My question is more about the education of our future scientists. With large data sets becoming increasingly available and valuable to our scientific inquiries, does this mean that all aspiring scientists should become comfortable with computer coding so they can process these large amounts of data? If not, is there some other arrangement where computer programmers work with scientists in order to help in this regard?

I ask because in my physics undergraduate program we were not encouraged to have any background with computer science and I feel that this hurt our ability to be effective after graduation.

[OneBildoNation](#)

This is one of my personal passions! I believe we do undergraduates a disservice when we do not explicitly recognize that the world they are going out into is so different than the one for which we were educated. Information and computation are a ubiquitous part of the modern world. There is really no profession that is untouched by this revolution and the explosion in data science. I truly believe that we need to incorporate "computational thinking" more deeply into all of undergraduate curriculum. Done correctly, this would enhance the education of students in every discipline from the classics to hard-core science, and empower students to evaluate more critically both the tendency for more and more data to be collected on us all, and the claims being made about the analysis of that data. My trainees who have gone into data science in non-academic arenas are genuinely surprised by how little the people running companies often understand about the science (beyond the fact that everyone needs big data science now), and clearly that is something that we need to work to change. Moreover, the opportunities for data science to create a better evidence base for social science and urban planning, and to bring us new ways of understanding history and culture are really exciting. Computational thinking has real value across disciplines, and I think we need to work more on ways of updating curriculum to reflect where the world is headed.

From a bioinformatics perspective, what interesting cross-correlates have popped up in diseases? For

example, a number of endosomal sorting genes have been implicated in familial forms of neurodegenerative diseases, which wouldn't have necessarily made intuitive sense to many researchers.

Personally, I think applying powerful computational modeling to data sets like this is a fantastic way of identifying new research avenues!

[lzawwlgood](#)

I think this is an absolutely fascinating area! We have been looking at how the genetically determined part of gene expression is associated with risk of disease, and that concept comes with a direction of effect: it is either the reduced genetically determined expression of a gene that increases the risk of some disease, or the decreased genetically determined expression of a gene that increases the risk of that disease. When I round up all the genes that for which either increased or decreased expression is associated with a single disease, acute myeloid leukemia for example, I noticed that the flip side for those genes was also associated with a discrete set of diseases related to infection (cellulitis, sepsis, abscess, etc). So it has really gotten me thinking about whether trying to understand the genetic "opposite" of a disease might help us understand more about disease mechanisms and pathways. We hope to extend these studies by looking at whether the diseases that we see in these "flip" studies actually show negative correlations in electronic medical records from about 130 million individuals that one of my colleagues has assembled. It will be fascinating to see if this will be a concept that "has legs"!

I got Type 1 Diabetes two years ago. We didn't have any family history at the time, other than my mom getting thyroid cancer and celiac a little bit before I was diagnosed. Recently, my cousin was diagnosed with it too. I've been round 130-140 pounds for about the whole time, and 5'11".

So was it more likely I got it from genetics (even though I was the first person in my family to get T1D) or was it my environment or something? I asked the doctors why I would get it, and they just said that they don't know a lot about the causes.

Thanks!

[blabel3](#)

Most people who develop type 1 diabetes have no positive family history of the disease, so in that sense you are in the same boat as most other patients. That said, there is some evidence that the risk of type 1 diabetes is rising in at least some countries. Type 1 diabetes is strongly genetic. If we could devise an effective prevention for type 1 diabetes, we know how to find the people who are at highest genetic risk who could be targeted for the prevention strategy, but right now, we do not have an effective prevention. And it has always been suspected that there are one or more important environmental exposures that trigger type 1 diabetes in susceptible individuals, but these are likely to be fairly common exposures. In this case, it may be the combination of a common exposures (or even a particular order of a set of common exposures), genetic susceptibility, and aspects of immune system development that combine to make people more vulnerable. Doctors are also more sophisticated about diagnosing type 1 diabetes now than they once were, and more adults are being diagnosed with type 1 diabetes than they once were. Type 1 diabetes often has onset in childhood, and because of that, many physicians did not consider it as a diagnosis when diabetes developed in adulthood, but it is clear that adults can and do develop type 1 diabetes.

How can I avoid getting Type 2 diabetes? It's common on my dad's side of the family once they get over 60 and gain weight. Obviously I'm keeping my weight down (sort of), but I can't do much about

turning 60 some day.

Also, does intake of sugar or high-fructose corn syrup have a negative effect?

[DerPanzerfaust](#)

The best advice we can give you is to maintain a normal weight and be as active as you can. Both independently reduce risk of developing type 2 diabetes but there is some evidence of a kind of synergistic effect in that doing both reduces risk even more than you might expect for how much either reduces risk by itself.

Is Alzheimer's an inevitability for someone with 3 of 4 grandparents who had the disease? At what point is it too late to affect the probability of diagnosis?

[CaptCurmudgeon](#)

One thing we all have to live (or rather die) with now, is that when we decrease death rates due to cardiovascular disease or cancer (and we have), death rates for other things will rise, because we all die of something. A huge reason for the increased rate of Alzheimers disease as a cause of death is the reduction in death rates due to cardiovascular disease and cancer. Someone who would once have died of leukemia in childhood may now die of Alzheimers at age 87. Someone who would once have succumbed to a fatal MI at 61 may now die of Alzheimers at 82. In many ways it is a great trade off, with many years of additional productive life. But we all die of something, and if your grandparents had the good fortune (and good genes) to escape death from early coronary disease and cancer, good for them, and good for you. Some people believe that we would all develop Alzheimer's disease if we lived long enough and did not die from something else first. So, your risk of Alzheimer's may be increased, but if the primary reason for this is your reduced risk of other diseases that would kill you earlier, that is not such a bad thing. More effort is being made to find effective treatments for Alzheimers. Not much is concretely known about prevention -- but it is never bad to eat sensibly and exercise as much as you can.

Is there any data on pesticides in foods causing any long-term effects?

Is there any data on grass-fed vs. grain-fed animal products?

How about any other interesting environmental factors that we wouldn't think of?

Follow-up question: is there any way for a layperson like myself to follow interesting developments in your research or your field? I find this type of information fascinating and would love to stay up to date.

[alabaster1](#)

High exposure to pesticide (like in farm workers or in chemical manufacturing plants) has been shown to be a risk factor for some diseases, but it is challenging to show effects for the kinds of low levels of exposures we would have from diet. I know of no scientific studies on large numbers of individuals contrasting grass and grain diet in cattle. For an example of exposures that we may not think about as much, but might matter, consider the chemicals used to treat fabrics for flammability. This has been a subject of some controversy, but fabrics treated to reduce flammability are used in most furniture -- where we sit way too much. The highly processed nature of many of the foods we eat may turn out to be a bigger problem than the foods themselves. That is, not so much that we eat too much bread and starch, but that we eat bread that is devoid of most of the fiber (and flavor) that grains we once used to make bread had in abundance. Everything -- including convenience -- has its price.

I think all of us in science need to find ways to communicate our science in more and better ways to

the public. I would love to hear ideas on how we might do that.

What's your diet like? What's your opinion on utilizing a ketogenic diet for treatment / avoidance of illnesses such as diabetes or Alzheimer's?

[The King Of Nothing](#)

I would like to think my diet is improving. Like everyone else, I get frustrated with how rapidly the conventional wisdom on how to eat right changes (eat more margarine => margarine is worse than butter, eat no fat => eat good fat only, etc). And I would like to think that starving yourself doesn't really make you live longer, it just makes you feel like you have been living forever. I love hearing about the diversity of lives that centenarians have lived. There are those who lived a life of moderation, but plenty who did not. My husband and I have been trying to make simple meals with fresh ingredients, and to have plenty of fish and somewhat less red meat. I love eggs as a food that is a great source of protein for few calories and am glad to see that it is now considered less problematic from a dietary perspective. But I am sure things will continue to evolve on what is considered to be a healthy diet for at least the rest of my life, and I think evaluating whatever the current guidelines are with common sense and the recognition that eating right is a lifetime goal not a short-term fad is about the best we can do.

Hi Nancy! How can use of electronic medical records change research in genetic medicine? What are some of the pros/cons of using medical records this way? What kind of privacy concerns do you have about their use?

[p1percub](#)

There were so many good questions that I didn't have time to answer them all in my 2-hour time slot, so I will answer as many more as I can now. Electronic medical records (EMR for short) provide data for medical research of all kinds. One of the challenges is that the records were usually developed for billing and compliance, and therefore the accuracy reflects what is needed for billing and compliance, and the level of diagnostic accuracy we often use for research is higher (or at least thought to be higher). But I see this as an opportunity as well, since it is real data -- we need to be able to show that things we discover can be seen in this context, and that we can show them to matter in this context. Most of the ways that scientists use EMR data are completely decoupled from any identifying information about an individual patient. In large-scale records, scientists may look at diseases that seem to be co-morbid, seen together more commonly than expected based on their individual frequencies, or perhaps negatively correlated, raising the question of whether having one might protect against having the other. On the pro side, the sample sizes can be much larger than we are able to achieve in even large epidemiological studies, but the data quality and the ability to control for unknown differences among individuals (confounding factors) are not nearly as good as for epidemiological studies. The value of EMR studies in the context of genetics is the opportunity to look across the whole medical spectrum for associations between DNA variants and disease. Geneticists are used to studying one disease, and good studies on a single disease may include a number of additional measures that are thought to be related to that disease. But by being able to conduct studies over the whole medical record, it is possible to get an entirely new perspective on what the genetic variation you are looking at actually does. Finding a strong association between genetic variation and bone fractures, for example, might have a completely different meaning if you found that the same genetic variation was also highly significantly associated with dizziness and fainting, as you might believe that the fractures were secondary to the dizziness, rather than a consequence of fragile or brittle bones.

When patients have consented to genetic studies linked to their health records, the primary concerns

about privacy have to do with someone who is not authorized gaining information about their health or data. This is a concern for all medically related studies, and scientists take great pains to safeguard data. Moreover, although it is theoretically possible to hack computers and access data inappropriately, the kinds of data scientists use have no commercial value in terms of credit card or social security numbers. The information stored is about DNA variants and codes reflecting diseases that have been diagnosed in a person. So while such a hack is always a concern, there is no reported instance of this ever having occurred, or even attempted, in relation to genetic data (although some attempts have been made by unauthorized persons to gain access to health records of celebrities, for example). I worry much less about that kind of privacy concern than about the consequences of holding onto data so tightly that the full value of the data to the whole scientific community is never realized. My understanding from people who have agreed to participate in medical research (myself included) is that they agree to do so for genuinely altruistic reasons and want the samples and data that they agree to give to be used for the greatest possible good. Thoughtful stewardship of data is essential, but doing so in a way that permits maximal use of the data by qualified investigators is likely to move science forward faster, and I think that is what we all would like to see.

For a History of Science seminar, I am currently researching the connection between genetic science and Big Data in the time leading up to the Human Genome Project. Since you deal in large-scale data, there are a couple of questions you could help me with if you still have some time:

1. How did geneticists come to favour the approach of finding correlation in large data sets?
2. Genetic code is often imagined similarly to computer code, with overlapping terms like "code" or "programming". Obviously it's more complicated, but would you say there is a connection between genetic science and computer science since both are interested in "data"?
3. How would you rate the importance of advances in computing for the development of genetic science up to now? Were advances in storage or processing power more important? Lastly, are you optimistic about new opportunities that further advances might bring to your discipline?

[ze_Void](#)

Geneticists, like all scientists, value what works. When it became possible to look at the association of DNA variation across the entire genome with disease, it was disappointing to realize that many of the genetic effects we were finding were relatively modest (compared with our initial optimistic expectations), and that the larger the sample sizes we used for discovery, the more association we found that were both highly significant and reproducible in new samples. That still remains true and we have not actually seen a plateau in the number of new discoveries even as sample sizes have gotten very large for some phenotypes (e.g. hundreds of thousands of individuals for height or BMI).

Genetics has the huge virtue of having clear (and easily programmable) rules. It is very straightforward to do genetics research through computer simulations, for example, that reproduce analytically expectations on genes in populations and consequences of selection originally derived from theory. That is actually a great way to help students understand the science. There is also really interesting work related to using DNA encoding to store information. Genetics is one of the most quantitative of the biological sciences and computing has been hugely important in driving the science forward, through both the technology advances that would not be possible without high speed computing (e.g. image processing is key to many genome technologies), but also the direct analysis of all of the data being generated.

Hi Nancy,

First, forgive me for copying my questions from Manolis Dermitzakis' AMA yesterday, but he didn't get around to them, and I'd love to hear your take on them!

I'm curious what you make of the discrepancies between twin/family-derived heritability estimates, and genotype-derived estimates (from e.g. GCTA). I'll keep it generic, as I think the questions apply to a range of traits/diseases:

- 1) Do you think there actually is missing heritability, or do you think uncertainty in both estimates are to blame?
- 2) Do you know much about the statistical properties of the genotype-derived estimates and do you trust them?
- 3) Do you think that additive genetic variance is predominant, or do you expect substantial non-additive effects?
- 4) How big a deal is it that "environment" is often ignored, either as a confounder or effect modifier of genetic associations?

And finally, on a different note:

- 5) For my MPH class I'd love to have a decent (and freely available) genetic data set as an option for their dissertation work. Something phenotype-rich, with demographic/clinical covariates perhaps. Any suggestions?

[red_concrete](#)

1) Yes, to all. Some heritability remains "missing" because we have not yet adequately assayed all DNA variation that may contribute to disease. We miss some areas of the genome that are hard to sequence and genotype, and there is no reason to think these regions don't have variation that contributes to disease. We have barely begun to assay rare variation for common disease, and that will contribute some to missing heritability. Structural variation is another source of variation that we assay poorly with current technologies but that we are getting better and better at studying. And even simple repeat polymorphisms may contribute more to heritability of common diseases that we have appreciated because it is also difficult to assay and caught only imperfectly by the variation that is easily surveyed. Heritability from twin and family studies is easy to overestimate because it is hard to avoid the confounds of common family environment, the influence of cultural factors on at least some kinds of traits, and so forth. The statistical methods we use (GCTA, for example) can estimate heritability for only the DNA variants that are included in the analysis, which is not all of the variation we have in our genome, and depending on details of analysis and quality control of the data can under- or over-estimate heritability. 2) There are a number of methods that have been proposed, and while you have to do very careful quality control of input data, the methods are all reasonably robust. And of course can only tell you about the variation that you include in the analysis. 3) I remain puzzled by how much epistasis has been characterized in model organisms, and how little we have been able to characterize in human studies to date. At this point, sample sizes in human studies are quite good relative to the model system studies, so it is not just that the studies in humans are underpowered to detect epistasis. Model organisms often have had very peculiar breeding histories, and it is possible that our expectation of ubiquitous epistasis based on the results of studies in model systems is just wrong, with the magnitude of epistasis seen in those studies being more reflective of these unusual breeding systems. Additive variance is predominant in humans given the data we have to date. But we probably have many more ways of examining this question, so I would not say that the last word on this has been written -- not by any means. 4) Part of the reason effect sizes are so modest are that we are estimating those effects sizes across all environments, and effect sizes are likely to differ greatly by environment. Pharmacogenomics effect sizes are, on average, twice as large as effect sizes measured for complex traits, which almost certainly reflects the fact that a key environmental factor for pharmacogenomics traits (drug exposure) is measured and appropriately taken into account in

analysis. But to be fair, it is not that we deliberately ignore important known environmental effects. We usually do not know what they are, and therefore cannot measure them. When we know of the existence of a big environmental risk factor, people usually try to measure it for genetic studies. Cigarette smoking is often measured in studies of lung cancer or chronic obstructive pulmonary disease, for example. For many common diseases, we know that "western diet and lifestyle" contributes to risk of disease, because as developing countries adopt western diet and lifestyle the risk of those diseases rise. But exactly which aspects of diet and lifestyle are the main culprits can be hard to parse because they so many things change at once. And it is notoriously difficult to get people to report accurately on what they eat, and how they live. Metabolomics measures may ultimately help finesse some of these challenges. 5) There are many datasets available through dbGaP (the database of genotype and phenotype) which is run by NCBI that can be obtained for use in analysis by qualified investigators, including graduate students who are supervised by a qualified scientist. But there are also great simulated datasets that have been created for competitions and workshops that might be even better for demonstrating key concepts. Some of the data generated for the Genetic Analysis Workshops might be available to you for this purpose.

I have had a small portion of my genome sequenced at 23 and Me. Before the FDA banned the practice, company was able to provide a fair amount of information regarding things like how well (or not well) and medication is likely to work, conditions you are more or less likely to inherit, etc. All this is provided with copious explanations, citations to medical journals, etc. The FDA believes that DTC genotyping is not good because somebody may, for example seek out and receive unneeded medical procedures, cancer treatments, etc. Do you agree with the the FDA's position?

I was taught a cardinal rule of statistics is that correlation is not causation. Are not non-large-scale correlation studies of a specific gene (I forgot the acronym) exactly that?

[da5id1](#)

One of the challenges of providing information to an individual about what their genome variation is likely to do is understanding how complex the model relating genome variation to disease really is. At the time the FDA shut down medical information from a number of commercial companies it was becoming clear that while they were making predictions on the risk of a complex disease like breast cancer or psoriasis on the basis of one or a few common variants that they tested, geneticists were discovering that such diseases were likely to derive susceptibility from hundreds or even thousands of such variants. Different companies had chosen different of the first few variants identified to report on, and so would give very different estimates of risk for exactly the same person. I think that at the beginning of large-scale association studies, we all hoped things would be simpler than they turned out to be. The FDA was absolutely correct that much of the information on disease risk being reported back to people was likely to be so incomplete as to be effectively useless. But that will not always be the case, and it raises the truly challenging question of "what level of certainty should be required for reporting risks back?" And how much DNA information do you have to use to make such a report. Many of the commercial sites genotype a set of common variants, and an individual can certainly be at low risk of disease based on common variants, but high risk of disease based on very rare DNA variation that could only be detected through sequencing of the genome.

Correlation is not causation, but genetic factors are unlike most risk factors in that there can be no confusion about the directionality. When we see that inflammation is associated with obesity and diabetes, it is unclear whether inflammation might cause obesity or diabetes, or whether being obese or having diabetes causes inflammation. Even the apparent timing (seeing inflammation first) does not fully establish causality because phenotypes may be hidden for some time in the body before we can measure them (e.g. diabetes may be present as metabolic abnormalities long before we can diagnose high blood glucose levels). But disease does not cause genome variation (cancer being a special case, but even there, the disease affects only the genome of the tumor). Geneticists must still be very careful

about confounders such as different ancestry between case and control populations, or systematic errors in data generation that could differ between case and control samples, but at least the direction from genome variation to disease risk is clear.

What can you tell us about diseases of multi factorial inheritance, specifically mental illnesses such as bipolar disorder?

[beatsdropheavy](#)

Neuropsychiatric disorders were slower to show associations in the first wave of studies. Part of the reason is that the sample sizes for those disorders were smaller, and that is partly a reflection of the fact that it is more difficult and expensive to make a research diagnosis of schizophrenia or bipolar disorder than to diagnose diabetes, which can be done with a simple blood test. As sample sizes increased, the number of associations observed for schizophrenia behaved much like what was observed for height or diabetes. There have also been studies showing that there are genetic risk factors that are shared by a number of psychiatric disorders like schizophrenia and bipolar disorder. But it is fair to say that most if not all common disorders have multi-factorial inheritance, not just mental illnesses.

Greetings, and thank you for participating in human genetics week, and for doing this AMA.

Is there any information and/or active research being conducted regarding the effects/actions of the human microbiome on the human genome/ gene expression and how that relates to common diseases?

Also, how is genome function being assessed in your studies? Does your research team utilize transcriptomic technologies to assess gene expression?

[Serpes](#)

A lot of work in many different areas are using data from micro biome studies -- this is a very active research topic. It is more challenging than simple genetics, because we are once again in the situation where distinguishing cause and effect can be a challenge. We know that having a disease will often lead to changes in the micro biome, so when we observe differences in the micro biome between cases and controls, we cannot know simply from that observation what is cause and what is effect. But since the micro biome can be manipulated, it is possible to get to cause and effect, at least in animal models (and in humans when we are confident that we know the right manipulation).

Yes, we use data from many large public datasets with measured transcript levels and DNA variation in the same subjects for our data integration studies.

In the US, most people's exposure to the study of genetics is in K-12 education and focuses on simple Mendelian genetics, which is not the best mindset for understanding how complex (most) traits work. What are common misconceptions that you hear regarding complex trait genetics?

[foodisfood](#)

The single most common misconception about genetics is that genetic variation is deterministic in leading to disease. Even simple Mendelian diseases are quite complex, and while there is "a" gene for cystic fibrosis, there are many genetic factors that influence how severe the disease is, when it is likely to be diagnosed, and what kind of course the disease will take. And for common diseases, genetic variation at many different genes make contributions to disease risk, and there are very few common

diseases that do not have major environmental risk factors as well (whether or not we know with certainty what they are). More subtle misconceptions run in the other direction. Many scientists outside the field of genetics will talk about environmental risk factors for disease like BMI, or age at menarche, which from a genetics perspective are highly heritable phenotypes, even though they do clearly have both genetic and non-genetic factors affecting them.

Greetings, I would like to know what methodology do you use to determine the effect of environment factors on dna of the diseases that you work with. Do you examine a patient's gene sequence or gene mutation? I'm very curious on how do you evaluate that variation I don't even know how to ask. P.S. I'm a biology major (but more orientated to physiology) and I'd love if you didn't restrain from using any technical terms.

[landrobert](#)

There are ways of trying to study things that we know reflect how the environment "talks" to DNA. Methylation status at DNA bases that are sometimes methylated and sometimes not methylated can be a key way that the environment helps to determine what happens to DNA, because how much a nearby gene is expressed can be in part determined by how much of the DNA near the gene is methylated. But there is also a great deal of work in the field of genetic epidemiology to collect data on both genome variation and environmental exposures to better understand the main effects and potential interactions between the two. So some studies are done at the molecular level (i.e. assays related to methylation), and some are done at the level of the organism (i.e. the genetic epidemiology).

Hi Dr. Cox,

I have a theory that "family history" concerning long term illness is really just confused with "Family" eating habits. You are part of a family, so more than likely you eat like your family. Any studies to support or dispute this? I guess this goes along with the nature/ nurture argument. I sincerely feel that if you buck your family trend by eating and acting/exercising differently, you can limit your exposure to the so called "family history".

[bac0neggcheese](#)

See the comments on missing heritability below -- not everything that is familial is genetic!

Hi Dr. Cox,

I have a very serious question for you. I've never had seasonal allergies, never had asthma, and suddenly partway through my fourth tour in Iraq (was living in Baghdad in squalor very near the Iraqi troops with open burn pits and trash everywhere) I developed allergies and asthma. The asthma can be triggered by allergies, exercise, and even laughing. It's pretty bad, it put me in the ER once. The allergies are manageable with zyrtec and flonase, but without them I can't function.

My big questions are twofold: 1 - What caused this? and 2 - Should I be concerned about something bigger that I haven't noticed, like lung cancer or anything like that?

Thanks!

[cloverskull](#)

There are some forms of allergy that develop in adulthood as a consequence of having completely novel exposures. I never had allergies until I was 28 and living in the Philadelphia area for the first

time. It was horrible. Some people who develop allergies also develop asthma, and once you have asthma, you may notice that a variety of things can trigger an episode. The cause is likely to be a combination of genetic factors and environmental exposures and we all wish we could get to enough understanding of these things to do a better job of treating and preventing these conditions.

Hi Dr. Cox,

As someone who really believes (and knows) that quantitative genomics and EMR are fundamentally going to change healthcare, what advice can you give me and others that would want to pursue medical education in these fields? For example, sequencing and whole genome analysis are going to be more common, how do we integrate ourselves, as well as the community into being more "data-based"?

Thank you so much for your time!

[HIRITWQAEBIPTTOM](#)

Learning the basics of informatics science and computational thinking will be useful now in every field - and really valuable in all areas of genetics and genomics.

I read an article awhile back about the addition of methyl groups to parts of DNA due to environmental stresses. The article was about how these methyl groups could be inherited by the children of the individual. I was just wondering if these inherited methyl groups, or methyl groups in general, could contribute to any of the diseases you study? Sorry about the lack of a link...

[Max10255](#)

Yes, I think that is likely, and it is the subject of lots of research. But in general, I think that genome variation that is regulatory in some way does drive a lot of the genetic risk to common diseases.

A prominent physicist at UC Berkeley, sort of off-the-cuff, (and about five years ago) predicted that cancer would be cured or made manageable (something like HIV) within 25 years. This seems to be an overly optimistic timeframe. What is your opinion if you have one?

[da5id1](#)

Many cancer biologists would say that 25 years is an egregious underestimate of the time frame for making cancer a chronic disease. I hope they are right!