Laura Dean • Johanna McEntyre

Coffee Break

Tutorials for NCBI Tools

Last Updated: October 20, 2014

National Center for Biotechnology Information (US)
Bethesda (MD)

National Center for Biotechnology Information (US), Bethesda (MD)

NLM Citation: Dean L, McEntyre J, editors. Coffee Break: Tutorials for NCBI Tools [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 1999-.

Coffee Break is a resource at NCBI that combines reports on recent biomedical discoveries with use of NCBI tools. The result is an interactive tutorial that tells a biological story. Each report is based on a discovery reported in one or more articles from the recently published peer-reviewed literature. After a brief introduction that sets the work described into a broader context, the report focuses on how a molecular understanding can provide explanations of observed biology and lead to therapies for diseases.

Each vignette also highlights the NCBI tools and resources used in the research process. These tools include PubMed, PubMed Central, Entrez Gene, and MapViewer.

Coffee Break articles should be fun and informative reading for molecular biologists, clinicians, and students, and may serve as teaching aids for college and graduate students.

Editors

Laura Dean

Johanna McEntyre

Table of Contents

From the Statue of Liberty to the coin in your back pocket: The secret life of copper 1
Laura Dean

Yoda and the fountain of youth? The many hats of IGF-1 5
Laura Dean

How low can you go? The promise of a new class of cholesterol lowering drugs 9
Laura Dean

Turning the white fat brown: a new approach to obesity? 13
Laura Dean

A gut feeling: bugs are critical for good health 15
Laura Dean and Wayne Matten

DNAs of our lives: The role of pharmacogenomics in modern medicine 17
Laura Dean

Roses, noses, and underarms: how one variation in our DNA influences underarm perspiration (and ear wax) 19
Laura Dean

Neanderthal man lives on in some of us: the genome of our extinct sister species has been sequenced 21
Laura Dean

From Africa to the Arctic: how the woolly mammoth adapted to the cold 25
Laura Dean

The songbird and the chicken: how a song can change a genome 27
Laura Dean

Will malaria soon be a thing of the past? the potential of recombinant protein vaccines to control one of the world's most deadly diseases 29
Tyler Beck

Don't put anything smaller than your elbow in your ear: the genetics of ear wax 35
Laura Dean

What you see is not what you get: DNA barcoding is helping scientists unveil nature's most hidden diversity 37
Marcia Triunfol

Do brains have a freshness date? the effect of aging on the human brain 41
Victoria Sutton

Who let the dogs out? a genetic classification of dog breeds 45
Laura Dean

Microbial diversity: let's tell it how it is 47
Jo McEntyre

Talking about the genetics of talking: the discovery of the first "speech and language" gene 53
Laura Dean

Variations on a gene: investigating the genetic basis of iron overload 55
Laura Dean

A small fortune: small RNAs involved in the regulation of developmental timing 57

Finding Fanconi: the hunt for the cause of autosomal dominant renal Fanconi syndrome 59

Opening the flood gates? association of NOD2 with Crohn's disease 61

Honey, I shrunk the genome: genome reduction in the leprosy bacillus ... 63

Ready, steady, go! a two-part switch that regulates gene expression ... 65

Fluorescent timer: the E5 mutant of the red coral protein drFP583 changes its fluorescence from green to red over time ... 67

Cytosolic help for mitochondrial defects: a novel method for importing tRNA into mitochondria in order to suppress mutations ... 69

The mouse that eats less but gains weight: a neuropeptide receptor, Mc3r, is shown to play a role in regulating energy stores 71

Tuberous sclerosis complex in flies too? a fly homolog to TSC2, called gigas, plays a role in cell cycle regulation ... 73

The beginning of the END: the EAST protein assembles a nucleoskeleton between chromosomes ... 77

Mutations and blood clots: how point mutations in clotting factor genes conspire to increase the risk of thrombosis ... 79
Laura Dean

Viruses provide direction on the plant information superhighway: a viral movement protein helps to identify a counterpart in plants ... 83

How *Candida albicans* switches phenotype - and back again: the SIR2 silencing gene has a say in Candida's colony type ... 87

PTEN and the tumor suppressor balancing act: PTEN turns out to be the first tumor suppressor to have phosphatase activity ... 91

The *Salmonella* battle plan: how Salmonella gain entry into human intestinal cells to grow and divide ... 95

RNA surveillance: watching the defectives: detecting premature stop codons in mRNA halts the production of dangerous truncated proteins ... 97

The compound eye of flies divulges evolutionary secrets: analysis of fly eye development may shed light on human eye disease ... 99

The neighborhood of Alzheimer's amyloid precursor protein: new clues on Alzheimer's pathology from other proteins linked to amyloid plaque formation ... 103

What do Lyme disease and syphilis have in common? two pathogenic spirochetes unexpectedly share an ATP synthase ... 107

Dissecting the mechanism of our internal clock: how living organisms tune in to the time of day ... 111

Ubiquitin links Parkinson's disease genes: a tantalizing link between two new genes ... 115

Plant genes contribute to a sexually transmitted disease? how plant genes found their way into a human parasite ... 119

Tutorial list ... 123

From the Statue of Liberty to the coin in your back pocket

The secret life of copper

Laura Dean, MD[1]

Created: October 20, 2014.

We have been using copper for almost 10,000 years. And it proved to be such a useful metal that an historical era was named after it—The Bronze Age (bronze is an alloy consisting mainly of copper) (1).

The properties of copper mean that it is still in demand today—it is an excellent conductor of heat and electricity, and is relatively resistant to corrosion. Thus, the uses of copper range from cooking (pots and pans), electronics and appliances (wiring and circuit boards), to one of the most famous statues in the world, the Statue of Liberty (2). Copper is frequently combined with other metals, such as zinc, tin, and nickel, to form alloys. The US 5 cent coin may be called a nickel, but it is only 25% nickel, the rest of the metal is copper (3).

But copper isn't just useful; it is an essential trace mineral for most life forms. Copper is incorporated into many enzymes, such as plastocyanin, which is needed for photosynthesis in plants, algae, and cyanobacteria; and hemocyanin, which transports oxygen in species such as horse-shoed crabs and emperor scorpions, and gives them their characteristic blue-colored blood (4, 5) (Figure 1).

In humans, copper is vital for the normal development and functioning of connective tissue, the nervous system, the immune system, and metabolism. For example, copper is part of cytochrome c oxidase, the last protein in the electron transport chain that is responsible for generating energy in the form of ATP (4).

Because copper is only needed in trace amounts, most people get enough copper from their diet. But rarely, inherited conditions can lead to extremely low (Menkes disease) or high (Wilson disease) levels of copper.

The importance of copper in the body means that its levels must be carefully regulated. Copper homeostasis is tightly linked to other metabolic processes. In serum, 70%-90% of copper is bound to ceruloplasmin, an enzyme that plays an important role in iron homeostasis and inflammatory response (6). Ceruloplasmin is also present in the brain, where the loss of ceruloplasmin is associated with iron accumulation and neuronal degeneration (7, 8)

To find out more about ceruloplasmin and the *CP* gene that encodes it, take a look at the series of "Carrying Copper" tutorials below, which explain the Entrez Gene record for *CP*.

Tutorial 1: Table of contents, Summary, Genomic context

Tutorial 2: Genomic regions, transcripts and products (Part I)

Tutorial 3: Genomic regions, transcripts and products (Part II)

Author Affiliation: 1 NCBI; Email: dean@ncbi.nlm.nih.gov.


Tutorial 4: Bibliography and Phenotypes


Tutorial 5: Variation, Pathways from biosystems, and Interactions


Tutorial 6: General gene and protein information, NCBI Reference Sequences (RefSeq), Related sequences and Additional links

Acknowledgements

The author would like to thank Svetlana Lutsenko, PhD, Professor of Physiology, Johns Hopkins University.

Figure 1: Horseshoe crabs. Top: The horseshoe crab is an ancient species that predate dinosaurs. It has a large carapace (shell) that protects its internal organs, and has ten eyes and can see UV light.

Bottom: The horseshoe's crab unique blue blood protects them, and humans, again infection. It is highly valued by the pharmaceutical industry.

References

1. *Early and Middle Bronze Age*. Smithsonian National Museum of Natural History [Last accessed: 16th October 2014]. Available from: http://www.mnh.si.edu/exhibits/cyprus/early-bronze.html.
2. *Statue of Liberty*. National Park Service [Last accessed: 16th October 2014]. Available from: http://www.nps.gov/stli/faqs.htm - aboutthestatue.
3. *Coin specifications*. United States Mint [Last accessed: 16th October 2014]. Available from: http://www.usmint.gov/about_the_mint/?action=coin_specifications.
4. Gupta A., Lutsenko S. Evolution of copper transporting ATPases in eukaryotic organisms. Curr Genomics. 2012;13(2):124–33. PubMed PMID: 23024604.
5. Rehm P., Pick C., Borner J., et al. The diversity and evolution of chelicerate hemocyanins. BMC Evol Biol. 2012;12:19. PubMed PMID: 22333134.
6. Hellman N.E., Gitlin J.D. Ceruloplasmin metabolism and function. Annu Rev Nutr. 2002;22:439–58. PubMed PMID: 12055353.
7. Miyajima, H., *Aceruloplasminemia.* Neuropathology, 2014.
8. Vassiliev V., Harris Z.L., Zatta P. Ceruloplasmin in neurodegenerative diseases. Brain Res Brain Res Rev. 2005;49(3):633–40. PubMed PMID: 16269323.

Yoda and the fountain of youth?

The many hats of IGF-1

Laura Dean, MD[1]

Created: July 21, 2014.

Insulin growth factor (IGF-1) is on the World Anti-Doping Agency list of banned substances. In children, IGF-1 is important for growth, and in adults, it continues to have anabolic effects. Although data are lacking for humans, animal studies have shown that IGF-1 can increase muscle mass and speed up healing of injured tendons (1, 2).

IGF-1 also appears to play a part in the aging process. It has long been known that one of the causes of short stature, or dwarfism, is a lack of the IGF-1 hormone in childhood. And now it appears that IGF-1 deficiency also influences longevity.

The *C. elegans* worm can live for twice as long if it is genetically engineered to have a mutation in *daf-2* (equivalent to *IGF-1* receptor in humans) (3). And Yoda, a dwarf mouse, lived twice as long as normal laboratory mouse. With Princess Leia (his cage mate) for company, Yoda lived long enough to celebrate his 4th birthday (1).

Yoda, a Snell dwarf mouse (left), sniffing his cage-mate, Princess Leia (right). Dwarf mice are usually caged with normal-sized mice to help them to keep warm.
Photo credit: Richard Miller, University of Michigan Medical School

Dwarf mice, like the long-living worms, lack IGF-1. These mice are commonly used to study the effects of aging and there are three main types. The Snell and Ames dwarf mice have been bred to inherit mutations in *Pit-1* and *Prop1* genes, respectively, which disrupt the embryonic development of the pituitary gland. As a result, both types of mice lack multiple hormones, including growth hormone, prolactin, and thyroid stimulating hormone.

The third type of dwarf mouse, the Laron dwarf, is a knock-out mouse model. This mouse has a targeted gene deletion of either the growth hormone receptor (GHR-KO) or the growth hormone binding protein (GHBP-KO). So even though this mouse produces growth hormone, it is still growth-restricted because it is unable to respond to it (4).

Author Affiliation: 1 NCBI; Email: dean@ncbi.nlm.nih.gov.

Dwarf mice all share non-detectable levels of IGF-1 because one of the main actions of growth hormone is to stimulate the production of IGF-1. Interestingly, in addition to their small size, these mice also have reduced glucose and insulin levels, a lower incidence of tumors, and a longer average and maximum lifespan. In fact, their lifespan is increased by about 40% (4, 5).

Download video

Such mouse models are useful, but is there a similar link between IGF-1 deficiency and long-life seen in humans? Individuals with Laron syndrome have been helping to provide this answer. Laron syndrome, also known as primary growth hormone insufficiency, is caused by a mutation in the growth hormone receptor gene. A variety of different mutations have been identified, and most affect the extracellular region of the receptor that contains the growth hormone binding site. Affected individuals are typically less than 4 feet tall, but if they are given IGF-1 before puberty, they may grow taller (6, 7).

In southern Ecuador, there is a village where more than 250 individuals are thought to have Laron syndrome, and this community has been well-studied. The results have been surprising. For example, Laron patients appear to be protected against developing cancer, a disease that is associated with aging (8). Unfortunately, this apparent protection by IGF-1 deficiency does not translate to a longer life span compared to the general population (9). One factor may be a higher death rate from causes of death that are not related to age, such as trauma and alcoholism (10).

So, could targeting IGF-1 be a pathway to the elixir of youth? We know that a certain level of GH and IGF-1 is required for the functioning of the heart and other organs, and the level of these hormones declines with age. However, GH and IGF-1 are oncogenes, and using IGF-1 as a treatment for individuals who do not have a genetic deficiency might lead to unexpected and harmful outcomes (11).

What is clear is that extending the life of any organism, from worm to human, is an incredibly complex process, involving many pathways, and potentially many drug targets. And although IGF-1 may be one of these targets, knowing how to use this treatment safely and effectively remains a long way away.

Acknowledgments

The author would like to thank Zvi Laron, Professor Emeritus of Pediatric Endocrinology, Tel Aviv University, for reviewing this summary.

References

1. Pobojewski, S. *World's oldest mouse reaches milestone birthday*. The University Record [Online] May 1, 2014]; Available from: http://www.ur.umich.edu/0304/Apr19_04/26.shtml.
2. Kurtz C.A., et al. Insulin-like growth factor I accelerates functional recovery from Achilles tendon injury in a rat model. Am J Sports Med. 1999;27(3):363–9. PubMed PMID: 10352775.
3. Chen D., et al. Germline signaling mediates the synergistically prolonged longevity produced by double mutations in daf-2 and rsks-1 in C. elegans. Cell Rep. 2013;5(6):1600–10. PubMed PMID: 24332851.
4. Bartke A., Brown-Borg H. Life extension in the dwarf mouse. Curr Top Dev Biol. 2004;63:189–225. PubMed PMID: 15536017.
5. Flurkey K., et al. Lifespan extension and delayed immune and collagen aging in mutant mice with defects in growth hormone production. Proc Natl Acad Sci U S A. 2001;98(12):6736–41. PubMed PMID: 11371619.
6. Savage M.O., et al. Endocrine assessment, molecular characterization and treatment of growth hormone insensitivity disorders. Nat Clin Pract Endocrinol Metab. 2006;2(7):395–407. PubMed PMID: 16932322.

7. Laron, Z., Kopchick, J. (Eds.), *Laron Syndrome - From Man to Mouse. Lessons from Clinical and Experimental Experience*. 2011: Springer.
8. Steuerman R., Shevah O., Laron Z. Congenital IGF1 deficiency tends to confer protection against post-natal development of malignancies. Eur J Endocrinol. 2011;164(4):485–9. PubMed PMID: 21292919.
9. Aguiar-Oliveira M.H., et al. Longevity in untreated congenital growth hormone deficiency due to a homozygous mutation in the GHRH receptor gene. J Clin Endocrinol Metab. 2010;95(2):714–21. PubMed PMID: 19965916.
10. Laron Z. The GH-IGF1 axis and longevity. The paradigm of IGF1 deficiency. Hormones (Athens). 2008;7(1):24–7. PubMed PMID: 18359741.
11. Werner, H., Attias-Geva, Z., Bentov, I., Sarfstein, R., Schayek, H., Weinstein, D., Bruchim, I., *Cancer Genes, Tumor Suppressors, and Regulation of IGF1-R Gene Expression in Cancer*, in *Insulin-like Growth Factors and Cancer: From Basic Biology to Therapeutics*, D.E. LeRoith, Editor. 2012, Springer Science. p. p159-177.

How low can you go?

The promise of a new class of cholesterol lowering drugs

Laura Dean, MD[1]

Created: March 25, 2014.

Cardiovascular disease is the leading cause of death in the United States, accounting for 40% of all deaths (1). And cholesterol, transported in the blood in the form of LDL (Low Density Lipoprotein), is at the heart of the fatty plaques that narrow critical arteries in the heart and nervous system, increasing the risk of heart attack and stroke.

Cholesterol, however, is essential to health. It is an important structural component of cell membranes, and it is also a sterol from which steroid hormones such as testosterone, estrogen, and progesterone are made. The liver is responsible for synthesizing most of the cholesterol found in the blood via the HMG Co-A (3-hydroxy-3-methylglutaryl-coenzyme A) reductase pathway, but a diet that is high in saturated fats can increase cholesterol levels further, as can inheriting a genetic tendency towards having higher lipid levels.

LDL normally circulates in the bloodstream for 2 – 3 days before being removed. The lipid-binding protein, Apo-B-100, is the only protein found in LDL and it contains a binding domain that interacts with LDL receptors, mainly found on the surface of the liver, triggering endocytosis. Endosomes form, containing the LDL, which fuse with transport vesicles carrying lysosomal hydrolases from the Golgi apparatus to become lysosomes leading to the hydrolysis of LDL to cholesterol, fatty acids, and amino acids.

Cholesterol can then be recycled and used in the synthesis of new cell membranes, for hormone synthesis, or be transformed back in to esters and stored. But some individuals have dyslipidemia, which refers to an abnormal amount (high or low) of any or all lipids in the blood. The most common form is hyperlipidemia, an abnormally high level of any lipid, associated with an increased risk of heart disease and stroke.

The management of hyperlipidemia often begins with a recommendation to change lifestyle factors and to reduce any other cardiovascular risk factors, such as giving up smoking and treating hypertension. The next step is usually treatment with lipid-lowering drugs. By far the most commonly prescribed class of drugs are HMG-CoA reductase inhibitors, known as statins (2). Statins are effective for people with known heart disease, significantly reducing the risk of another cardiovascular event and death. However, the benefits of statins are less clear for people who do not have known heart disease, and about 1 in 5 people who take statins report side effects such as muscle pain, memory loss, sexual dysfunction, and in women, there is an increased risk of diabetes (3).

Rarely, high cholesterol levels are caused by a single genetic mutation, a condition known as familial hypercholesterolemia (FH). Cholesterol levels are usually extremely high and are less responsive to lifestyle changes or statin therapy. In affected individuals, LDL receptors are either absent or have a reduced function, and as a result, LDL is left for longer to circulate in the bloodstream. The most common causes of FH are a mutation in the LDL receptor (affects about 1 in 500), and a mutation in the Apo B protein (about 1 in 1000) (4-6).

A less common cause of FH is a mutation in *PCSK9*, (proprotein convertase subtilisin/kexin type 9). This newly identified subtilase (an enzyme that resembles a serine protease) is highly expressed in the liver where it has an important role in cholesterol homeostasis. PCSK9 binds to the LDL receptor and induces its degradation, thereby controlling the number of LDL receptors available to remove LDL from the circulation. It is generally

Author Affiliation: 1 NCBI; Email: dean@ncbi.nlm.nih.gov.

thought that polymorphisms in *PCSK9* contribute towards the natural variation in cholesterol levels seen in populations (7).

The first mutations to be discovered in *PCSK9* conferred a gain of function—a switch of a single amino acid increased the protease activity of PCSK9 so it could reduce the numbers of LDL receptors more quickly (8). This led to FH, specifically, familial hypercholesterolemia type 3 (9, 10).

More recently, several nonsense mutations of *PCSK9* have been reported. Here, the switch of a single amino acid results in the introduction of a premature stop codon in to the DNA sequence, leading to the translation of a shorter, incomplete protein product. The lucky individuals with nonsense mutations in *PCSK9* have unprecedented *low* levels of cholesterol, together with a much lower risk of cardiovascular disease (11, 12).

The discovery of these lipid-lowering variants ignited a race to find a new type of drug that could mimic the effects. And currently a number of potential drugs are progressing through clinical trials (13).

The drugs Alirocumab (REGN727/SAR236553), Evolocumab (AMG145), and PF-04950615 (RN316) have completed phase 1 trials (small, short trials enrolling healthy volunteers, to determine whether the drug is safe and to determine drug doses) and phase 2 trials (larger, longer-term trials designed to see if a drug is both safe and effective). They are currently in various stages of phase 3 trials. These trials are typically very large, involving several thousand patients, and last for several years—they aim to see how well drugs work in real patients.

New drugs that lower LDL are in clinical trials—are any of the trials near you?

All three drugs are monoclonal antibodies that have been designed to be highly specific for PCSK9 and once bound, they prevent its interaction with the LDL receptor. Another type of drug class is the small interfering RNAs (siRNAs), which work by "gene-silencing." The drug ALN-PCS is one example that is currently in Phase 1 trials; it works by binding to *PCSK9*-specific messenger RNA and prevents the production of protein (14).

One thing that all these drugs designed to target PCSK9 share in common is that they need to be taken in the form of an injection, typically 2 weeks between doses. And it will take many years to determine how well they reduce the risk of future cardiovascular events.

But starting today, what can be done to reduce these risks is very simple and extremely effective—enjoying a heart healthy diet with a daily dose of exercise.

Acknowledgments

With thanks to William H. Frishman, M.D., for his review of this article. Dr. Frishman is the Rosenthal Professor of Medicine, and a Professor of Pharmacology at New York Medical College, and Director of Medicine at Westchester Medical Center.

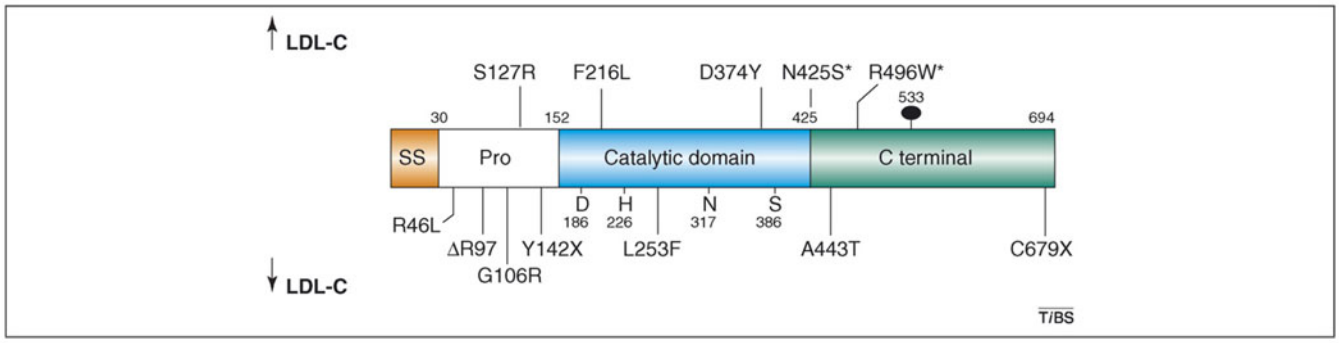


Figure 1. Mutations in *PCSK9* that result in elevated or reduced plasma levels of LDL. Abbreviations: LDL-C, Low Density Lipoprotein Cholesterol, SS, signal sequence; Pro, prodomain.
A schematic of *PCSK9* with the location of naturally occurring mutations associated with elevated (top) or reduced (bottom) plasma levels of LDL-C. The mutations included are limited to those associated with significant differences in plasma levels of LDL-C in at least two independent populations or those that co-segregate consistently with hypercholesterolemia in families. Mutations associated with elevated plasma cholesterol levels found only in families who also have mutations in the LDLR are indicated by with an asterisk (*). The major domains of *PCSK9* are delineated using different colors. The location of the aspartic acid (D), histidine (H) and serine (S) comprising the catalytic triad and the site of binding of the single N-linked sugar (Asn533) are shown. The oxyanion hole is located at Asn317.
Adapted with permission from *Trends in Biochemical Sciences*.
Author manuscript; available in PMC Jul 16, 2009.
Published in final edited form as:
Trends Biochem Sci. Feb 2007; 32(2): 71–77. Published online Jan 9, 2007. doi: 10.1016/j.tibs.2006.12.008

References

1. Go A.S., Mozaffarian D., Roger V.L., Benjamin E.J., et al. Executive summary: heart disease and stroke statistics--2014 update: a report from the american heart association. Circulation. 2014;129(3):399–410. PubMed PMID: 24446411.
2. Stone N.J., Robinson J., Lichtenstein A.H., Bairey Merz C.N., et al. ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Journal of the American College of Cardiology. 2013.:2013. PubMed PMID: 24239923.
3. Abramson J.D., Rosenberg H.G., Jewell N., Wright J.M. Should people at low risk of cardiovascular disease take a statin? BMJ. 2013;347:f6123. PubMed PMID: 24149819.
4. Youngblom E, K.J., *Familial Hypercholesterolemia*, in *GeneReviews*, A.M. Pagon RA, Bird TD, et al., Editor 2014, University of Washington: Seattle (WA).
5. Austin M.A., Hutter C.M., Zimmern R.L., Humphries S.E. Familial hypercholesterolemia and coronary heart disease: a HuGE association review. American journal of epidemiology. 2004;160(5):421–9. PubMed PMID: 15321838.
6. *Familial hypercholesterolemia captures gene test controversies.* Biotechnology healthcare, 2010. **7**(1): p. 8-9.
7. Kotowski I.K., Pertsemlidis A., Luke A., Cooper R.S., Vega G.L., Cohen J.C., Hobbs H.H. A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. American journal of human genetics. 2006;78(3):410–22. PubMed PMID: 16465619.
8. Nassoury N., Blasiole D.A., Tebon Oler A., Benjannet S., et al. The cellular trafficking of the secretory proprotein convertase PCSK9 and its dependence on the LDLR. Traffic. 2007;8(6):718–32. PubMed PMID: 17461796.
9. Abifadel M., Varret M., Rabes J.P., Allard D., et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nature genetics. 2003;34(2):154–6. PubMed PMID: 12730697.
10. OMIM. Proprotein Convertase, Subtilisin/Kexin-Type, 9; PCSK9. [cited March 2014]. Available from: http://www.omim.org/entry/607786

11. Kathiresan S. A PCSK9 missense variant associated with a reduced risk of early-onset myocardial infarction. The New England journal of medicine. 2008;358(21):2299–300. PubMed PMID: 18499582.
12. Benn M., Nordestgaard B.G., Grande P., Schnohr P., Tybjaerg-Hansen A. PCSK9 R46L, low-density lipoprotein cholesterol levels, and risk of ischemic heart disease: 3 independent studies and meta-analyses. Journal of the American College of Cardiology. 2010;55(25):2833–42. PubMed PMID: 20579540.
13. Catapano A.L., Papadopoulos N. The safety of therapeutic monoclonal antibodies: implications for cardiovascular disease and targeting the PCSK9 pathway. Atherosclerosis. 2013;228(1):18–28. PubMed PMID: 23466067.
14. Weinreich M., Frishman W.H. Antihyperlipidemic Therapies Targeting PCSK9. Cardiology. 2014. in review. PubMed PMID: 24407047.

Turning the white fat brown

a new approach to obesity?

Laura Dean, MD[1]

Created: March 3, 2014.

When temperatures fall, adults can shiver to keep warm. Babies, however, are born without the ability to shiver, and with their larger head-to-body ratio, they lose heat faster than adults do. But they have protection from the cold in the form of brown adipose tissue (BAT), otherwise known as brown fat. Unlike white fat, brown fat generates heat, and it covers the backs of infants, creating a warm shawl.

Whereas white fat cells store extra calories in the form of large fat droplets that in excess make us obese, brown fat cells contain smaller fat droplets that metabolize glucose and fat instead of storing it. Their cells have a high density of mitochondria, which contain iron that give the tissue a reddish brown color, and unique to mitochondria in BAT, they also contain an uncoupling protein 1 (UCP1, also known as thermogenin).

The oxidative phosphorylation that takes place in all mitochondria generates a proton gradient across the mitochondrial membrane that results in the production of energy in the form of ATP. But in the presence of UCP1, the proton gradient is reduced—protons that were pumped out into the intermembrane space are able to return to the mitochondrial matrix, resulting in less ATP production, and the energy from substrate oxidation is converted to heat, accounting for the thermogenic properties of BAT (1).

About 5% of the total mass of an infant is made up of brown fat, and until recently it was thought that stores of brown fat disappeared by adulthood. However, in 2009, it was found that adults still have small, but physiologically significant, reserves of brown fat in their shoulders and neck (2-4). The amount of brown fat varies—older people have less brown fat than younger people, and often no brown fat is detectable in obese people—the amount of BAT inversely correlates with body mass index.

In mice, two types of brown fat have been identified—constitutive (present from birth) and recruitable ("browning" of white fat can occur when levels of brown fat are insufficient to maintain body temperature) (5). Brown fat has been found to protect mice against obesity when they are overfed, and mice with more brown fat are leaner and healthier (5).

Several key genes have been identified as driving the production of brown fat in humans. Brown fat cells are thought to arise from myoblast precursor cells through the action of two proteins, PRDM16 and C/EBP-beta (6). Together, these proteins form a transcriptional complex, and through inducing peroxisome proliferator-activated receptor (PPAR)-gamma expression, this complex has the power to switch the lineage of specific precursor cells to brown fat cells. However, in order to do so, PRDM16 has to interact with the enzyme euchromatic histone-lysine N-methyltransferase 1 (EHMT1) (7).

EHMT1 is essential in determining the fate of brown fat cells. While EHTM1 was originally identified as a histone methyltransferase, it also has non-histone targets. It appears that EHMT1 functions as the "engine" of the PRDM16-C/EBP-b transcriptional complex, and without this engine, the PRDM16 complex can not drive precursor cells to brown adipocytes. The characteristics of brown fat cells are lost and cells are induced to differentiate in to muscle cells (7).

Looking at the loss of the "brown fat switching gene", *EHMT1*, in humans

Author Affiliation: 1 NCBI; Email: dean@ncbi.nlm.nih.gov.

Studies of *Ehmt1*^{adipo} knockout mice, in which the mouse *EHMT1* gene has been specifically deleted from brown adipocyte precursor cells only (a whole body knock out of *Ehmt1* is lethal), found the knockout mice gained more weight than normal mice, even though they ate identical diets. The knockout mice also showed higher blood glucose levels, greater insulin resistance, and increased amounts of fat in their liver—all characteristic traits of diabetes and other metabolic diseases (7).

And in patients with Kleefstra syndrome, in which a microdeletion at 9q34.3 results in the deletion of the *EHMT1* gene along with approximately 20 other genes, about 40-50% of the patients are obese. This is likely to be the first example in which the loss or alteration of a human gene has effected the development of brown fat and resulted in obesity (8, 9).

Future studies of *EHMT1* will investigate whether this gene could be a drug target—activation of *EHTM1* and increasing brown fat production could potentially be a new type of treatment for obesity. Current medicines for obesity are limited to either suppressing appetite or inhibiting intestinal fat absorption, but the treatment results and side effects mean for many, a new approach to obesity treatment cannot come quickly enough.

Acknowledgments

With thanks to Shingo Kajimura, Ph.D., Assistant Professor of Cell & Tissue Biology at the University of California, San Francisco, for his review of this article.

References

1. Fedorenko A., Lishko P.V., Kirichok Y. Mechanism of fatty-acid-dependent UCP1 uncoupling in brown fat mitochondria. Cell. 2012;151(2):400–13. PubMed PMID: 23063128.
2. Cypess A.M., Lehman S., Williams G., Tal I., et al. Identification and importance of brown adipose tissue in adult humans. The New England journal of medicine. 2009;360(15):1509–17. PubMed PMID: 19357406.
3. van Marken Lichtenbelt W.D., Vanhommerig J.W., Smulders N.M., Drossaerts J.M., et al. Cold-activated brown adipose tissue in healthy men. The New England journal of medicine. 2009;360(15):1500–8. PubMed PMID: 19357405.
4. Virtanen K.A., Lidell M.E., Orava J., Heglind M., et al. Functional brown adipose tissue in healthy adults. The New England journal of medicine. 2009;360(15):1518–25. PubMed PMID: 19357407.
5. Schulz T.J., Huang P., Huang T.L., Xue R., et al. Brown-fat paucity due to impaired BMP signalling induces compensatory browning of white fat. Nature. 2013;495(7441):379–83. PubMed PMID: 23485971.
6. Kajimura S., Seale P., Kubota K., Lunsford E., Frangioni J.V., Gygi S.P., Spiegelman B.M. Initiation of myoblast to brown fat switch by a PRDM16-C/EBP-beta transcriptional complex. Nature. 2009;460(7259):1154–8. PubMed PMID: 19641492.
7. Ohno H., Shinoda K., Ohyama K., Sharp L.Z., Kajimura S. EHMT1 controls brown adipose cell fate and thermogenesis through the PRDM16 complex. Nature. 2013;504(7478):163–7. PubMed PMID: 24196706.
8. Willemsen, M.H., A.T. Vulto-van Silfhout, W.M. Nillesen, W.M. Wissink-Lindhout, et al., *Update on Kleefstra Syndrome.* Molecular syndromology, 2012. 2(3-5): p. 202-212.
9. Kleefstra T., van Zelst-Stams W.A., Nillesen W.M., Cormier-Daire V., et al. Further clinical and molecular delineation of the 9q subtelomeric deletion syndrome supports a major contribution of EHMT1 haploinsufficiency to the core phenotype. Journal of medical genetics. 2009;46(9):598–606. PubMed PMID: 19264732.

A gut feeling

bugs are critical for good health

Laura Dean, MD[1] and Wayne Matten, PhD[1]

Created: July 24, 2013.

While still in the womb, the gut of an infant is completely sterile. But as the birthing process begins, bacteria and other microbes soon start to colonize the surfaces of the infant's body—the skin, nose, mouth, and by far the most densely colonized, the gut. Babies born by caesarean section are initially populated with different microbes (from the hospital and skin microbiota) than babies delivered vaginally (vaginal microbiota) (1, 2).

During the first few years of a host's life, communities of microbes become established in it. They adapt and flourish in different parts of the body. Eventually, every human cell is outnumbered by about 10 microbial cells, and every human gene is outnumbered by about 100 microbial genes. The collective genomes of these microbes are known as the microbiome.

The microbiome performs many important functions. For example, the bacteria that line the host's intestines aid digestion and synthesize vitamins. They also protect against infection—they prevent potential pathogens from gaining a foothold. And there is growing evidence that these microbes shape the immune system. Germ-free mice, which have been raised to have intestines without bacteria, are very different than mice with a normal microbiome. Their immune systems are so poorly developed that they are extremely susceptible to pathogens.

The composition of the microbiome is shaped by many different factors. The microbes are mobile—they move between a host's body and the environment. Changes in the microbiome are associated with who lives in the same household as the host, even the presence of a pet dog. Diet also plays a role—a diet limited to processed, uncontaminated food results in a less diverse microbiome. Obese individuals have a simpler microbiome than lean people (4). And when food is scarce, evidence from animal models suggests that a "bad" microbiome may conspire with a poor diet to cause and perpetuate malnutrition (5).

Having a BLAST! Comparing DNA sequences from an obese mouse gut microbiome and a lean mouse gut microbiome against the BLAST microbial representative genomes database.

The host's gut microbiota can be disturbed by a course of antibiotics. It has been observed that infants who were given antibiotics before the age of 5 months, a time that is critical to the microbiome being established, became slightly heavier over time. By 38 months of age, they were 22% more likely to be overweight (9). Antibiotics kill harmful bacteria, but they also kill healthy bacteria, which may influence energy extraction and help keep us lean. A shift in bacterial demographics may also influence the risk of disease—from intestinal disorders (e.g., ulcerative colitis, Crohn's disease, and irritable bowel disease), to diseases such as eczema, asthma, obesity, and heart disease. Microbes may even modulate sex hormones and influence the risk of autoimmune diseases, such as type 1 diabetes (6).

Because there are more genes in the host's gut microbiota alone than there are in the rest of the host's human body, these microbes are said to hold the "second genome". But whereas a human genome is inherited, a human microbiome is acquired—and therefore can be manipulated. If germ-free mice are given the microbiome of obese mice, they too gain body fat. This is probably because the bacteria in the obese microbiome were extracting more energy from food, thus influencing the metabolism to store rather than break down fat (7). Another study found that if young normal mice were fed with one type of bacterium, it led to an altering of those

Author Affiliation: 1 NCBI; Email: dean@ncbi.nlm.nih.gov.

mice's brain neurotransmitters, causing changes in their behavior that are analogous to reduced levels of stress, anxiety, or depression (8).

Many of the microbes we carry have never been studied because they cannot be grown in microbial culture—they do not survive in laboratories. However, by sequencing their genomes, their genes can be studied. In 2008, the Human Microbiome Project was launched by the NIH, with the goal to identify and characterize the microorganisms that are associated with humans. It studied over 240 healthy volunteers, and took samples from five body areas or "habitats" (nose, mouth, skin, urogenital tract, and gastrointestinal tract) (10).

In total, over 5000 samples were collected, purified, and sequenced. Bacterial sequences were identified by sequencing the hypervariable regions of bacterial ribosomal RNA (16s rRNA), which contain signature sequences specific to a species. 5177 unique microbial taxonomic profiles were found. By sequencing whole-community DNA (whole genome-shotgun or metagenomic sequencing), the researchers found that while the composition of the microbes varied widely among individuals, there was remarkable functional stability. In other words, there are many ways to construct microbial communities to perform similar functions, such as aiding the digestion of lipids. Other key findings included an association between ethnicity and microbiome composition, and a correlation between vaginal pH and microbial diversity.

Now that the Human Microbiome Project has provided the first reference data for microbes living in healthy adults, the avenues for research include how these microbes fare over time—in health, disease, and treatments, from infancy to old age. Drugs may be developed to specifically target a microbiome—perhaps to aid weight loss, combat infection, inflammation, or even treat cancer.

References

1. Azad, M.B., T. Konya, H. Maughan, D.S. Guttman, et al., *Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months.* CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne, 2013. 185(5): p. 385-94.
2. Dominguez-Bello M.G., Costello E.K., Contreras M., Magris M., Hidalgo G., Fierer N., Knight R. *Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns.* .Proceedings of the National Academy of Sciences of the United States of America. 2010;107(26):11971–5. PubMed PMID: 20566857.
3. Song, S.J., C. Lauber, E.K. Costello, C.A. Lozupone, et al., *Cohabiting family members share microbiota with one another and with their dogs.* eLife, 2013. 2: p. e00458.
4. Ley R.E., Turnbaugh P.J., Klein S., Gordon J.I. *Microbial ecology: human gut microbes associated with obesity.* .Nature. 2006;444(7122):1022–3. PubMed PMID: 17183309.
5. Smith M.I., Yatsunenko T., Manary M.J., Trehan I.et al. *Gut microbiomes of Malawian twin pairs discordant for kwashiorkor.* .Science. 2013;339(6119):548–54. PubMed PMID: 23363771.
6. Markle J.G., Frank D.N., Mortin-Toth S., Robertson C.E.et al. *Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity.* .Science. 2013;339(6123):1084–8. PubMed PMID: 23328391.
7. Turnbaugh P.J., Ley R.E., Mahowald M.A., Magrini V., Mardis E.R., Gordon J.I. *An obesity-associated gut microbiome with increased capacity for energy harvest.* .Nature. 2006;444(7122):1027–31. PubMed PMID: 17183312.
8. Neufeld, K.M., N. Kang, J. Bienenstock, and J.A. Foster, *Reduced anxiety-like behavior and central neurochemical change in germ-free mice.* Neurogastroenterology and motility : the official journal of the European Gastrointestinal Motility Society, 2011. 23(3): p. 255-64, e119.
9. Trasande L., Blustein J., Liu M., Corwin E., Cox L.M., Blaser M.J. *Infant antibiotic exposures and early-life body mass.* .International journal of obesity. 2013;37(1):16–23. PubMed PMID: 22907693.
10. *A framework for human microbiome research.* .Nature. 2012;486(7402):215–21. PubMed PMID: 22699610.

DNAs of our lives

The role of pharmacogenomics in modern medicine

Laura Dean, MD[1]

Created: April 25, 2013.

Different people can react quite differently to the same medicine. For one person, the medicine might be extremely effective, improving or eradicating symptoms. For another, it might fail to work or require a dose adjustment. For yet another, it could cause severe side effects.

Many factors account for such differences, including age, gender, body mass index (BMI), ethnicity, other medicines, and co-existing medical conditions. But even if all the above factors are the same, patients may still react quite differently to the same medication. This is because of our DNA—the way a person responds to a drug is a complex trait that is influenced by many different genes.

More than 100 drugs now mention pharmacogenomic biomarkers on their FDA-approved drug label. Pharmacogenomics is the study of the inherited variations in genes that determine drug response. Often the variations affect a small region of DNA, as small as a single nucleotide, known as a single nucleotide polymorphism (SNP). Many genetic tests are based on genotyping to determine which SNPs the patient has. Alternatively, sequencing may be used—a more complex process which determines the exact sequence of the relevant part of the genome.

At least three SNPs have an important effect on how individuals respond to warfarin—an anticoagulant given to patients with an increased risk of developing blood clots. Two of these SNPs occur in the gene that contributes to the metabolism of warfarin, *CYP2C9*. The third SNP occurs in the target enzyme of warfarin, encoded by *VKORC1*. Carriers of these SNPs are more sensitive to warfarin and require lower doses.

The aim of warfarin therapy is to keep the patient's INR (International Normalized Ratio—a measure of blood coagulation) within a target range, usually between 2 and 3. If the INR is too low, the risk of blood clotting remains, but, if the INR is too high, there is a new risk of bleeding. Many factors influence what dose of warfarin a patient will need and are taken into account when calculating what the first dose should be. This is important because optimizing the starting dose of warfarin shortens the time it takes before the INR is safely within range.

It is now possible to include the patient's *CYP2C9* and *VKORC1* genotype to further optimize the warfarin starting dose. But currently, genetic testing for a drug response has yet to become an integrated part of routine clinical practice, and warfarin therapy is nearly always started without knowing the patient's genotype. This is partly because in many clinical scenarios, treatment has to be started quickly and there is no time to wait for the results of gene testing to come from a lab, which may take days to weeks.

However, this is changing. In a recent trial, genotype results were delivered within 60 minutes of a cheek swab being taken, by nurses performing the genetic test near the patient. Patients who were randomized to the group that received genotyping and were found to carry a particular variant of the *CYPC19* gene were given an alternative drug that worked better for them (1). And perhaps in the not so distant future, all patients will pre-emptively have their genomes sequenced so that the most effective drugs are given, and the drugs with an increased risk of adverse events avoided (2, 3).

Currently, there are genetic tests for about 2,500 diseases, and a growing number of tests for variations in genes involved in drug responses. The NCBI's online tool, the Genetic Testing Registry (GTR), helps clinicians and patients navigate through the increasing use of genetic testing, as its role in improving drug safety and effectiveness becomes more commonplace.

Author Affiliation: 1 NCBI; Email: dean@ncbi.nlm.nih.gov.


Take a tour of the Genetic Testing Registry.

References

1. Scott S.A. *Clinical pharmacogenomics: opportunities and challenges at point of care.* .Clinical pharmacology and therapeutics. 2013;93(1):33–5. PubMed PMID: 23212102.
2. Schildcrout J.S., Denny J.C., Bowton E., Gregg W.et al. *Optimizing drug outcomes through pharmacogenetics: a case for preemptive genotyping.* .Clinical pharmacology and therapeutics. 2012;92(2):235–42. PubMed PMID: 22739144.
3. Koch B.C., van Schaik R.H., van Gelder T., Mathijssen R.H. *Doubt about the feasibility of preemptive genotyping.* .Clinical pharmacology and therapeutics. 2013;93(3):233. PubMed PMID: 23232547.

Roses, noses, and underarms

how one variation in our DNA influences underarm perspiration (and ear wax)

Laura Dean, MD[1]

Created: March 25, 2013.

Sweating is useful. It helps to maintain body temperature, and hydrates and protects the skin. Sweating also helps regulate the body's levels of fluid and electrolytes such as sodium and calcium.

Skin contains two types of sweat gland—eccrine and apocrine. Eccrine glands are found all over the body, especially in the soles of feet and palms of hands. These glands discharge a salty perspiration on to the surface of the skin, which has a cooling effect as it evaporates.

Apocrine glands are found in the armpit and groin where they secrete an oily sweat into the canal of the hair follicle. These glands start secreting at puberty, and the sweat they produce may contain pheromones. But once this oily sweat reaches the surface of the skin, bacteria start to break it down, a process that can produce a noticeable odor and drives us to use deodorant.

But not everyone produces underarm (axillary) odor. Individuals who have a certain variant of the *ABCC11* gene produce less odorous sweat. They also happen to have dry earwax, as discussed in an earlier Coffee Break.

Axillary odor and the type of earwax (wet or dry) are genetically determined by the variant rs17822931, which is a single nucleotide polymorphism (SNP) in the *ABCC11* gene. One of the functions of the transporter protein encoded by the *ABCC11* gene is to secrete amino acid conjugates of human odorants that lead to the production of axillary odor. The SNP is a 538G>A substitution and at least one functional G allele is needed for the transporter to be active. Therefore, individuals who have the GA or GG genotype are "genetically odorous", whereas the "genetically non-odorous" have the AA genotype.

In light of intriguing new research, the *ABCC11* gene is now being brandished as the "deodorant gene". A large study (n~6500) (1) has found that how often people used deodorant is strongly associated with which variant of the rs17822931 SNP they have. The non-odorous individuals (AA genotype) were five times *less* likely to use deodorant than the odorous (GG and GA genotypes).

The frequencies of the A and G alleles vary markedly across different ethnic groups. The A allele is very common in East Asians, and as expected, most people in this population don't need to use deodorant. And so they don't use it—it's estimated that only 7% of North East Asians regularly use deodorant. In contrast, the G allele is by far the more common allele in African and European populations.

In the UK, where the study took place, only 2% of the population (just over 1 million) is estimated to have the AA genotype. Interestingly though, the study found that only a quarter of the non-odorous individuals seemed to recognize that they don't produce odor and chose not to use deodorant. The remaining three quarters regularly used deodorant, perhaps only because it is the social norm. The study goes on to predict that around ~$14 million is wasted each year in the UK by non-odorous people buying deodorants, not to mention the needless exposure to chemicals and possible skin irritation that could be avoided.

So, are you one of the lucky individuals who can stop buying deodorant? A genetic test of your *ABCC11* gene would give you the answer, or much simpler, a quick check inside your ear to look for dry (gray and flakey) earwax. Or simpler still, just do the sniff test!

Author Affiliation: 1 NCBI; Email: dean@ncbi.nlm.nih.gov.

However, genotypes and earwax aside, the use of deodorant will most likely continue as it is now—an entirely personal choice.

This tutorial highlights some of the NCBI resources that provide information about the "deodorant gene".

References

1. Rodriguez S, Steer C.D., Farrow A, Golding J, Day I.N. Dependence of Deodorant Usage on ABCC11 Genotype: Scope for Personalized Genetics in Personal Hygiene. J Invest Dermatol. 2013. PubMed PMID: 23325016.

Neanderthal man lives on in some of us

the genome of our extinct sister species has been sequenced

Laura Dean, MD

Created: August 2, 2010.

Just over 150 years ago, the first fossils of our closest relative, the Neanderthal, were discovered in Germany. From analyzing these bones and comparing them to ours, we have predicted what the Neanderthal might look like, when the Neanderthal emerged, roamed, and eventually went extinct about 30,000 years ago. During their travels, they would have encountered the ancestors of modern day humans. But much controversy has existed on whether there was any interbreeding between Neanderthals and the ancestors of modern humans. Fossils have not given us a definitive answer and the Neanderthal mitochondrial genome suggested no gene mixing had occurred. But now the complete nuclear genome has been sequenced, and it has given us some surprising answers.

Green et al. published the first draft of the Neanderthal genome in May 2010. It is a composite genome, based on ancient DNA samples collected from the bones of three female Neanderthals (figure 1), who lived at different times in a cave in Croatia, about 40,000 years ago (figure 2).

The team compared the Neanderthal genome to other samples of Neanderthal DNA collected from fossils in Spain, Germany and Russia. This confirmed that their composite genome is a fair representation of the Neanderthal species. They then compared the genome to that of the chimpanzee and to modern human genomes. For this comparison, they also sequenced the genomes of five living humans who originate from different parts of the world: Southern Africa, West Africa, France, Papua New Guinea, and China. These comparisons revealed which parts of the current human genomes have evolved recently, since Neandertals and humans diverged.

To perhaps everyone's surprise, the genome comparisons revealed interbreeding had taken place between Neanderthals and our ancestors, but not as we expected. Despite there being no evidence that Neanderthals lived in China and Papua New Guinea, the two individuals from these areas are as closely related to the Neanderthals as the French individual is. In fact, these three people share 1% to 4% of their nuclear DNA with Neanderthals. In contrast, the comparisons do not show any evidence of interbreeding with the ancestors of Africans. One theory to explain this is that gene mixing took place much earlier than thought, after early humans had migrated out of Africa and into regions such as the Middle East (at least 100,000 years ago) but before they migrated to Europe and western Asia. The genetic contribution from Neandertals would then have been carried with early humans as the colonized all of Eurasia.

The comparisons with modern human genomes also revealed 78 individual genetic changes that result in protein differences between humans and Neanderthals. The number of changes is remarkably small given the 300,000 years which separate humans and Neanderthals from their common ancestor. But these differences do occur in interesting genes.

For example, three out of six genes that had multiple substitutions were in skin, suggesting the importance of skin form and function to the evolution of humans. One of these is the TRPM1 Gene, which code for melastatin, an ion channel important for maintaining melanocyte pigmentation in the skin. The same substitution found in some humans removes the start codon. The RPTN gene encodes repetin, an extracellular epidermal matrix protein found in the epidermis, and at high levels in sweat glands, hair roots, and the tongue.

Other differences are in genes important in cognitive development, and mutations in some of these genes in modern humans lead to a variety of disorders. For example, mutations of DYRK1A contribute to Down syndrome, CADPS2 and AUTS2 to autism, and NRG3 to schizophrenia. Other interesting genes affected include SPAG17, which encodes a protein that is important for the beating of the sperm tail, while RUNX2 is the only

gene in the genome known to cause cleidocranial dysplasia. Features of this disease include a bell-shaped rib cage, a protruding frontal bone, and a small or absent collar bone. As Neanderthal rib cage is typically bell-shaped, the frontal bone is prominent, and the collar bone is different in shape to human (figure 3), it is plausible that changes in RUNX2 were important in human evolution.

Such genetic changes are important to our understanding of the most recent evolution of humans. And it also brings us closer to understanding what led to the mysterious extinction of one of our closets relatives.

This Coffee Break was reviewed by Professor Richard E. (Ed) Green.

Figure 1. Neanderthal bone fragments. The researchers obtained the majority of the DNA used for their study from the bone fragments of three female Neandertals who were excavated in the Vindija Cave in Croatia.

Image copyright: Max Planck Institute for Evolutionary Anthropology.

Figure 2. Entrance of the Vindija Cave, Croatia. (Copyright: Johannes Krause, Max Planck Institute for Evolutionary Anthropology)

Figure 3. Reconstruction of a Neandertal group. Image copyright: Johannes Krause, Neandertal group by Atelier Daynes, Paris, France. In: Museum of the Krapina Neanderthals, Krapina, Croatia. Project and realization of the Museum: Zeljko Kovacic and Jakov Radovcic.

References

1. Green Richard E, Krause Johannes, Briggs Adrian W, Maricic Tomislav, Stenzel Udo, Kircher Martin, Patterson Nick, Li Heng, Zhai Weiwei, Fritz Markus Hsi-Yang, Hansen Nancy F, Durand Eric Y, Malaspinas Anna-Sapfo, Jensen Jeffrey D, Marques-Bonet Tomas, Alkan Can, Prüfer Kay, Meyer Matthias, Burbano Hern´n A, Good Jeffrey M, Schultz Rigo, Aximu-Petri Ayinuer, Butthof Anne, Höber Barbara, Höffner Barbara, Siegemund Madlen, Weihmann Antje, Nusbaum Chad, Lander Eric S, Russ Carsten, Novod Nathaniel, Affourtit Jason, Egholm Michael, Verna Christine, Rudan Pavao, Brajkovic Dejana, Kucan Zeljko, Gusic Ivan, Doronichev Vladimir B, Golovanova Liubov V, Lalueza-Fox Carles, de la Rasilla Marco, Fortea Javier, Rosas Antonio, Schmitz Ralf W, Johnson Philip L F, Eichler Evan E, Falush Daniel, Birney Ewan, Mullikin James C, Slatkin Montgomery, Nielsen Rasmus, Kelso Janet, Lachmann Michael, Reich David, Pääbo Svante. A draft sequence of the Neandertal genome. Science. 2010;328:710–722. PubMed PMID: 20448178.

From Africa to the Arctic

how the woolly mammoth adapted to the cold

Laura Dean, MD

Created: May 31, 2010.

Mammoths, African elephants, and Asian elephants diverged from one another about 7 million years ago in Africa. These members of the elephantid family were adapted to living in the hot tropical climates. This is why elephants today have large ears to maximize heat loss.

Figure 1

But mammoths migrated North at the same time as a sudden climate change saw temperatures plunge in the Arctic. And so 2 million years ago they adapted to survive the Arctic cold. To minimize heat loss, they developed small ears and tails, and a woolly undercoat that led to their common name of woolly mammoth.

On a molecular level, scientists have found that specific mammoth proteins contain amino acid changes that are not found in elephants. But it had not been possible to link these genetic changes to how the mammoth might have physiologically adapted to the cold. Until now.

Campbell et al. have found that amino acid changes in hemoglobin would have given the mammoth a unique genotypic adaptation to the cold. Hemoglobin is the oxygen-carrying protein found in red blood cells. In humans and in most mammals, an increase in temperature decreases the stickiness of oxygen to hemoglobin. This helps release additional oxygen where it is needed, to warm exercising muscles. However, Arctic mammals such as reindeer keep their limbs cold to minimize heat loss. Their hemoglobin is much less sensitive to changes in temperature as oxygen needs to be delivered to the tissues even when they are cold.

The genes encoding woolly mammoth hemoglobin acquired three nucleotide mutations that similarly helped oxygen delivery in its new Arctic environment. These three changes occurred on the HBB/HBD globin gene such that the chimeric β/δ globin chain had three amino acid substitutions: T12A, A86S and E101Q. African and Asian elephants, who remained in warmer climes, do not have these changes.

All three amino acid substitutions are on the same side of the hemoglobin protein. The T12A and A86S substitutions are exposed on the outer surface of the protein. Molecular modelling indicated that the E101Q substitution may be structurally important, as it occurs at the highly conserved interface between the two sliding αβ dimer subunits. To confirm their function, the three changes were spliced into cDNA of Asian elephants and the mammoth hemoglobin protein faithfully synthesized in *E. coli*. As expected, the substitutions had large phenotypic effect, thereby reducing the energy (heat) requirements for releasing oxygen.

The mammoth became extinct from mainland areas about 10,000 years ago, and the last mammoths died about 3700 years ago on Wrangel Island in the Arctic Ocean. Their extinction was possibly caused by hunting by man, further climate change, or a combination of the two. By sequencing and resurrecting phenotypic attributes of extinct species such as the woolly mammoth, we can discover functional differences that are not attainable by studying fossils. But just as we are discovering genetic factors that are beneficial in evolution, we may also discover genetic factors that played a part in the extinction of the woolly mammoth and other species.

This Coffee Break was reviewed by Kevin L. Campbell, Ph.D.

***Mammuthus primigenius* (woolly mammoth).**
The woolly mammoth lived in the late Pleistocene era across Asia, Europe, and North America. By using fossil DNA from preserved remains of woolly mammoth samples, scientists have decoded its mitochondrial genome, and are now sequencing other mammoth genes.

Image source: What Killed the Woolly Mammoth? Sedwick C *PLoS Biology* Vol. 6, No. 4, e99

References

1. Campbell Kevin L, Roberts Jason E E, Watson Laura N, Stetefeld Jörg, Sloan Angela M, Signore Anthony V, Howatt Jesse W, Tame Jeremy R H, Rohland Nadin, Shen Tong-Jian, Austin Jeremy J, Hofreiter Michael, Ho Chien, Weber Roy E, Cooper Alan. Substitutions in woolly mammoth hemoglobin confer biochemical properties adaptive for cold tolerance. Nature genetics. 2010. PubMed PMID: 20436470.

The songbird and the chicken

how a song can change a genome

Laura Dean, MD

Created: May 3, 2010.

The zebra finch is a tiny Australian songbird. Only the males sing, and they learn a particular song by listening to their fathers.

As in all songbirds, there is a discrete area of the forebrain called the "song control center". When the young finch is learning to sing, he almost "babbles" like a human baby would. During this time, nuclei in the song control center change in size and organization, and during adulthood, the act of singing alters the expression of genes. Listening to a song also activates many genes in their auditory centers, also located in the bird's forebrain.

The zebra finch recently had its genome decoded by Warren et al. Once they identified the positions of the song-responsive genes, they looked for similar genes in the chicken, the only other bird to have its genome fully sequenced. The chicken does not sing, it merely clucks, and its clucking is instinctive, not learned. Therefore by comparing the genomes of the two birds, the scientists could identify the genes that are most likely to be important in singing and learning to sing.

Warren et al. also compared complimentary DNA from a juvenille songbird (learning to sing) and an adult one (able to sing). The finch's communication system turned out much more complex than previously thought. Singing changes expression for over 800 genes in the song control center, including ones that alter the expression of transcription factors and their targets.

Recent research by Dong et al. found that RNAs in the finch's auditory centers respond in different ways as the bird learns a song. Warren et al. found that many of the RNA transcripts (40%) are non-coding, and of the RNAs that are suppressed in response to song, the majority are non-coding. Noncoding RNAs (ncRNAs) were once thought to be "junk", but they have increasingly been found to have important functions in regulating other genes. It has even been proposed that these molecules aid the evolution of complexity in humans. It is therefore fitting that ncRNAs are so extensively involved in the complex ability of the songbirds learning to communicate.

Hearing song also activates small ncRNAs called microRNAs. These regulate gene expression by binding to target messenger RNAs. A potential site of action for the microRNAs aligns to an area in a human gene, NR4A3, that encodes a transcription factor protein. The fact that this transcription factor appears to be conserved in humans and songbirds suggests it has an important function in microRNA regulatory pathways.

Learning to sing is crucial for the finch. Without a song, he is unable to attract a mate and reproduce. The unique neurobiological apparatus that enables the finch to sing developed after the lineage of the zebra finch and chicken diverged, about 100 million years ago. It is likely that the genes involved, which include genes that encode neuronal ion channels, have been subject to a positive selection pressure leading to an accelerated evolution of genetic changes. Another possible channel of evolution is gene duplications. For example the genes PHF7 and PAK3 are found in mammals, but in the zebra finch, these genes have been duplicated and many variants are present in the finch's brain.

This tiny songbird has shown us how a behavior, such as learning a song, can change a genome. Studying these changes may bring us closer to understanding how babies learn to speak by listening to their parents, and ultimately provide insights into speech disorders, autism, and neurological disorders such as Parkinson's disease.

This Coffee Break was reviewed by Wes Warren, Ph.D.

***Taeniopygia Guttata* (zebra finch).**
The zebra finch is the second bird to have its genome sequenced, the first being the chicken. Studying and comparing these two avian genomes may help us understand how humans learn to speak their first words.

References

1. Warren Wesley C, Clayton David F, Ellegren Hans, Arnold Arthur P, Hillier Ladeana W, Künstner Axel, Searle Steve, White Simon, Vilella Albert J, Fairley Susan, Heger Andreas, Kong Lesheng, Ponting Chris P, Jarvis Erich D, Mello Claudio V, Minx Pat, Lovell Peter, Velho Tarciso A F, Ferris Margaret, Balakrishnan Christopher N, Sinha Saurabh, Blatti Charles, London Sarah E, Li Yun, Lin Ya-Chi, George Julia, Sweedler Jonathan, Southey Bruce, Gunaratne Preethi, Watson Michael, Nam Kiwoong, Backström Niclas, Smeds Linnea, Nabholz Benoit, Itoh Yuichiro, Whitney Osceola, Pfenning Andreas R, Howard Jason, Völker Martin, Skinner Bejamin M, Griffin Darren K, Ye Liang, McLaren William M, Flicek Paul, Quesada Victor, Velasco Gloria, Lopez-Otin Carlos, Puente Xose S, Olender Tsviya, Lancet Doron, Smit Arian F A, Hubley Robert, Konkel Miriam K, Walker Jerilyn A, Batzer Mark A, Gu Wanjun, Pollock David D, Chen Lin, Cheng Ze, Eichler Evan E, Stapley Jessica, Slate Jon, Ekblom Robert, Birkhead Tim, Burke Terry, Burt David, Scharff Constance, Adam Iris, Richard Hugues, Sultan Marc, Soldatov Alexey, Lehrach Hans, Edwards Scott V, Yang Shiaw-Pyng, Li Xiaoching, Graves Tina, Fulton Lucinda, Nelson Joanne, Chinwalla Asif, Hou Shunfeng, Mardis Elaine R, Wilson Richard K. The genome of a songbird. Nature. 2010;464:757–762. PubMed PMID: 20360741.
2. Dong Shu, Replogle Kirstin L, Hasadsri Linda, Imai Brian S, Yau Peter M, Rodriguez-Zas Sandra, Southey Bruce R, Sweedler Jonathan V, Clayton David F. Discrete molecular states in the brain accompany changing responses to a vocal signal. Proceedings of the National Academy of Sciences of the United States of America. 2009;106:11364–11369. PubMed PMID: 19541599.

Will malaria soon be a thing of the past?

the potential of recombinant protein vaccines to control one of the world's most deadly diseases

Tyler Beck

Created: June 26, 2006.

An effective malaria vaccine may well be on the way. These days, vaccines are a normal part of our lives. Most United States children are vaccinated for some form of hepatitis, along with a slew of other diseases including polio, rubella, measles, and mumps. Some of us take these things for granted, but in other parts of the world, vaccines are not so easily obtained.

Malaria, globally one of the most devastating diseases affecting humans, is caused by four species of the *Plasmodium* genus. Of the four species that can cause malaria in humans, *Plasmodium falciparum* is the one that causes most deaths (1). Malaria could be made significantly less deadly by using a cheap, easy-to-produce vaccine.

Traditionally, vaccines have been made from attenuated viruses or bacteria, or by creating a virus-like particle, in the case of some types of flu vaccine(2). Another method for creating vaccines, recently becoming more popular, is to create a recombinant protein. Vaccines made with recombinant proteins offer an advantage over other types of vaccines in that there is no need to handle the actual disease-causing agent, which can be costly and sometimes dangerous. Instead, one or more proteins are expressed and purified, then formulated to be injected into the subject to cause an antibody response against the foreign protein. If, in the future, that person is again exposed to the same protein, it is hoped that his or her immune system will recognize it more quickly as a threat. Recombinant protein vaccines are currently being researched and tested for a variety of diseases, including ricin toxin exposure, pneumococcus infection, and malaria (3-5). There are currently a few different recombinant protein vaccines against *P. falciparum* being tested in clinical trials, including apical membrane antigen 1 (AMA1) (6).

AMA1 appears on the surface of the merozoite during the blood-stage of *P. falciparum* parasites (Figure 1). Studies suggest that AMA1 is a necessary component for invasion of red blood cells by merozoites (7). Vaccination with recombinant AMA1 has been shown to elicit antibody responses that provide protection against homologous parasite challenges in both rodent and monkey models of malaria infection, and a derivative vaccine has been in a Phase I human trial in Mali, West Africa (7-9).

How did we arrive at this point? As early as 1997, scientists were testing some form of AMA1 for its antibody response against *Plasmodium* species (7). But to use a protein as a vaccine, it must be economically feasible to create large amounts of the protein. The *Plasmodium* genome is highly A+T rich, which makes it hard to express *P. falciparum* proteins in sufficient yields for commercial use in classic expression systems such as *Escherichia coli* and *Pichia pastoris*. One way to augment expression in such species is to recode the gene to match the host's tRNA pool. DNA codons that are rare in the target species are replaced with those that are used more often, while keeping the amino acid sequence unchanged. This raises protein yield because more tRNA molecules exist in the cell for those codons, making protein synthesis easier.

Genes that are recoded, or "synthetic" have been used for years to raise yields and reduce costs for many medically and industrially useful proteins, such as insulin (10). For AMA1, *Pichia pastoris* is the most widely used expression system, because the protein can be expressed in much greater quantities than in the original host organism (8, 9).

One of the problems with AMA1, however, is that it is strain specific. This means that an AMA1 protein cloned from one strain of *P. falciparum*, the FVO strain, for example, may not protect against other strains of *P. falciparum*, such as the 3D7 strain (8). This is because of a highly polymorphic cluster of amino acids surrounding the interior of the protein (Figure 2).

To remedy this, some AMA1-derived vaccines, such as AMA1-C1, are mixtures of AMA1 cloned from different strains of *P. falciparum* (9). These combination vaccines are intended to elicit better antibody responses against diverse strains of *P. falciparum* than any one strain-specific AMA1 protein (11).

Manufacturing a cheap, effective vaccine for malaria will depend on many factors. A large-scale method for preparing AMA1-derived vaccines is still far from a reality, and limited human trial data is available (9). There are still three other species of *Plasmodium* that can cause malaria, so effective vaccines must be considered for these, especially because it has been shown that *P. vivax* can replace *P. falciparum* in areas in which the *falciparum* species has been contained (6). Interest has also been shown in certain oligodeoxynucleotide (ODN) molecules that, when added to the vaccine formulation, may strengthen the immune response against AMA1-derived vaccines (11).

Protein vaccines, as compared to other types of vaccines, could potentially be cheap, easy to produce vaccine candidates against malaria, one of the world's deadliest diseases. Promising early research results have been shown, but much research must still be done to make a malaria vaccine a reality. Clinicians and researchers have their work cut out for them, as always. For now, effective treatment is still extremely important in the fight against devastating diseases such as malaria, but in the next few years, that may change.

This Coffee Break was contributed by Tyler Beck, during an internship at the National Center for Biotechnology Information, while on sabbatical from the University of Maryland, Baltimore County.

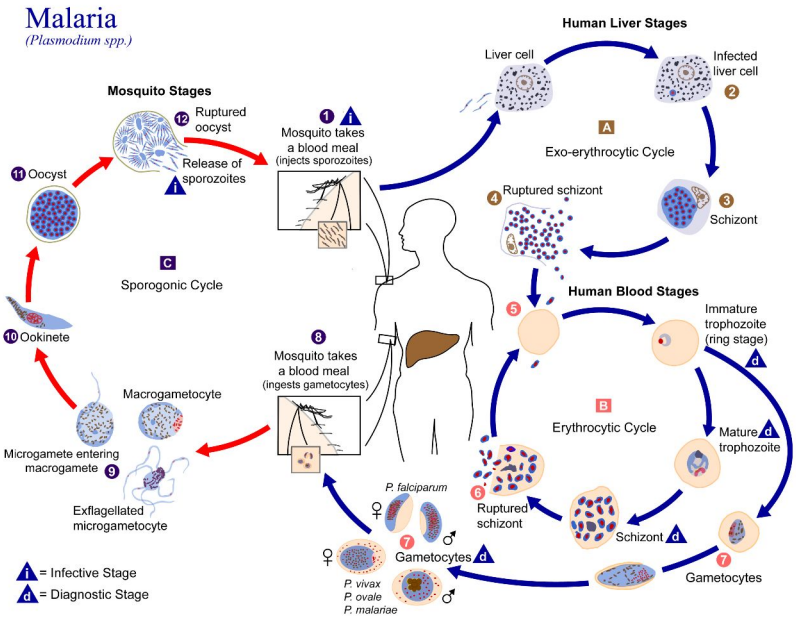


Figure 1. The malaria parasite life cycle involves two hosts. During a blood meal, a malaria-infected female Anopheles mosquito inoculates sporozoites into the human host ❶. Sporozoites infect liver cells ❷ and mature into schizonts ❸, which rupture and release merozoites ❹. (Of note, in P. vivax and P. ovale a dormant stage [hypnozoites] can persist in the liver and cause relapses by invading the bloodstream weeks, or even years later.) After this initial replication in the liver (exo-erythrocytic schizogony A), the parasites undergo asexual multiplication in the erythrocytes (erythrocytic schizogony B). Merozoites infect red blood cells ❺. The ring stage trophozoites mature into schizonts, which rupture releasing merozoites ❻. Some parasites differentiate into sexual erythrocytic stages (gametocytes) ❼. Blood stage parasites are responsible for the clinical manifestations of the disease.

The gametocytes, male (microgametocytes) and female (macrogametocytes), are ingested by an Anopheles mosquito during a blood meal ❽. The parasites' multiplication in the mosquito is known as the sporogonic cycle C. While in the mosquito's stomach, the microgametes penetrate the macrogametes generating zygotes ❾. The zygotes in turn become motile and elongated (ookinetes) ❿ which invade the midgut wall of the mosquito where they develop into oocysts ⓫. The oocysts grow, rupture, and release sporozoites ⓬, which make their way to the mosquito's salivary glands. Inoculation of the sporozoites ❶ into a new human host perpetuates the malaria life cycle.

Figure and legend reproduced, with permission, from the Centers for Disease Control and Prevention website.

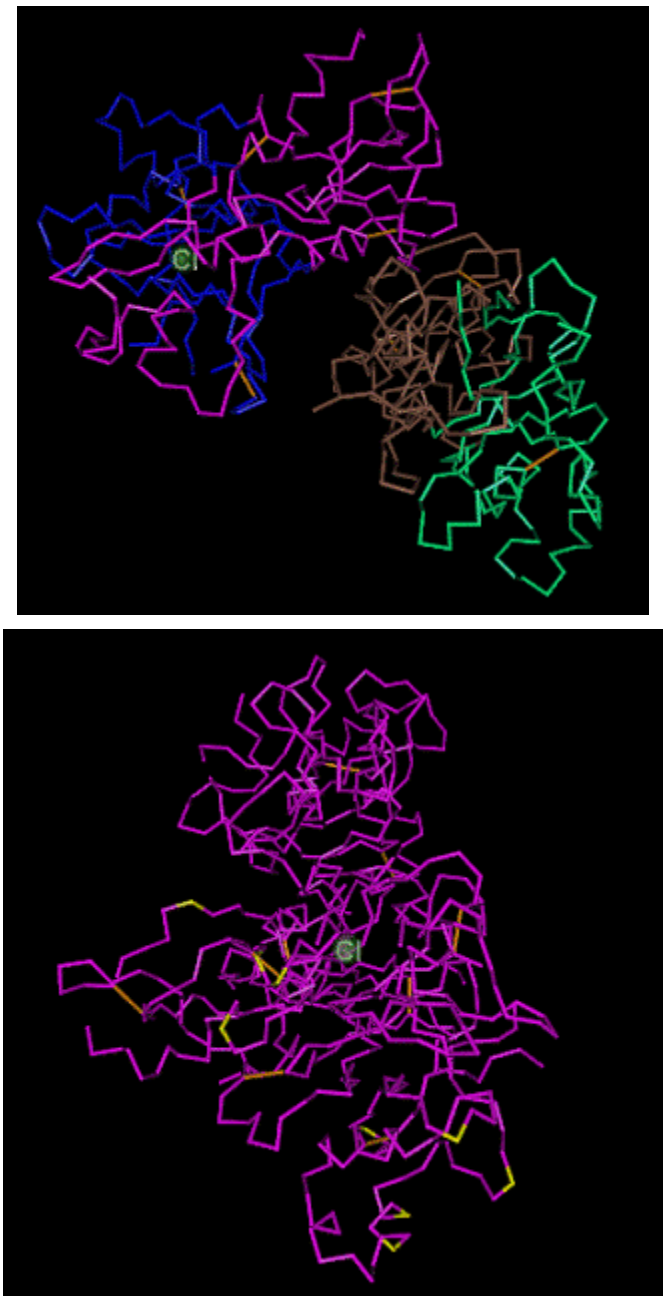


Figure 2. (Top) Three-dimensional view of the AMA1 protein with domains colored differently. You can download the 3-dimensional model in .

(Bottom) Three-dimensional view of the AMA1 protein with polymorphic residues highlighted.

References

1. Jamison DTBreman JGMeasham ARAlleyne GClaeson MEvans DBJha PMills AMusgrove P , editors.21. Conquering Malaria. In: Disease control priorities in developing countries. Washington (DC):IBRD/The World Book and Oxford University Press; 2006 (Bookshelf)
2. Galarza JM, Latham T, Cupo A. Virus-like particle (VLP) vaccine conferred complete protection against a lethal influenza virus challenge. Viral Immunol. 2005;18(1):244–251. PubMed PMID: 15802970.

3. Jomaa M, Terry S, Hale C, Jones C, Dougan G, Brown J. Immunization with the iron uptake ABC transporter PiaA and PiuA prevents respiratory infecation with Streptococcus pneumoniae. Vaccine. 2006;24(24):5133–5139. PubMed PMID: 16707196.
4. Vitetta ES, Smallshaw JE, Coleman E, Jafri H, Foster C, Munford R, Schindler J. A pilot clinical trial of a recombinant ricin vaccine in normal humans. Proc Natl Acad Sci U S A. 2006;103(7):2268–2273. PubMed PMID: 16461456.
5. James S, Miller L. Malaria vaccine development: status report.[monograph online]. Bethesda (MD): National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH); 2001. [updated 2001 Jan. 5; cited 2006 Jun 21]. Available from: http://www.niaid.nih.gov/dmid/malaria/malariavac.html
6. Matuschewski K. Vaccine development against malaria. Curr Opin Immunol. 2006;18(4):449–457. PubMed PMID: 16765576.
7. Amante FH, Crewther PE, Anders RF, Good MF. A cryptic T cell epitope on the apical membrane antigen 1 of *Plasmodium chabaudi adami* can prime for an anamnestic antibody response. J Immunol. 1997;159(11):5535–5544. PubMed PMID: 9548494.
8. Kennedy MC, Wang J, Zhang Y, Miles AP, Chitsaz F, Saul A, Long CA, Miller LH, Stowers AW. In vitro studies with recombinant *Plasmodium falciparum* apical membrane antigen 1 (AMA1): production and activity of an AMA1 vaccine and generation of a multiallelic response. Infect Immun. 2002;70(12):6948–6960. PubMed PMID: 12438374.
9. Malkin EM, Diemert DJ, McArthur JH, Perreault JR, Miles AP, Giersing BK, Mullen GE, Orcutt A, Muratova O, Awkal M, Zhou H, Wang J, Stowers A, Long CA, Mahanty S, Miller LH, Saul A, Durbin AP. Phase 1 clinical trial of apical membrane antigen 1: an asexual blood-stage vaccine for *Plasmodium falciparum* malaria. Infect Immun. 2005;73(6):3677–3685. PubMed PMID: 15908397.
10. Dunn CJ, Plosker GI, Keating GM, McKeage K, Scott LJ. Insulin glargine: an updated review of its use in the management of diabetes mellitus. Drugs. 2003;63(16):1743–1778. PubMed PMID: 12904090.
11. Mullen GED, Giersing BK, Ajose-Popoola O, Davis HL, Kothe C, Zhou H, Aebig J, Dobrescu G, Saul A, Long CA. Enhancement of functional antibody responses to AMA1-C1/Alhydrogel, a *Plasmodium falciparum* malaria vaccine, with CpG oligodeoxynucleotide. Vaccine. 2006;24(14):2497–2505. PubMed PMID: 16434128.

Don't put anything smaller than your elbow in your ear

the genetics of ear wax

Laura Dean, MD

Created: October 11, 2006.

Recently, an exciting genetic discovery was made in the field of ear wax. It appears that a change in a single nucleotide of your DNA can determine whether your ear wax is wet or dry. This marks the first time that a single-nucleotide polymorphism (SNP) has been found to determine a visible genetic trait.

Known medically as cerumen, ear wax is found in the ears of humans and many other mammals. Secreted by apocrine glands in the ear canal, the exact function of ear wax is not clear. Proposed benefits range from keeping the ear canal clean and lubricated, to acting as an insect trap.

In contrast, many have experienced the disadvantages of ear wax. In those who are hard of hearing, ear wax can damage their hearing aids. Ear wax is also a cause of hearing loss—a build up of wax can block the ear canal, hindering the passage of sound waves to the ear drum.

But it has long been noted that there are 2 types of ear wax, "wet" and "dry". Wet ear wax is common in Caucasians and African-Americans, it tends to be honey-to-brown in color and sticky in nature. In contrast, dry ear wax is common in East Asians and is gray in color and more brittle and flakey. And now we know the underlying genetic basis.

A major clue came in 2002 when scientists reported the case of a Japanese woman who had a rare genetic disorder that caused her arms and legs to twist uncontrollably (paroxysmal kinesigenic choreoathetosis). Uncommon in the Japanese, she also happened to have wet ear wax, as did several of her relatives who also had choreoathetosis. This suggested that the inheritance of the rare neurological disorder and the inheritance of the dry ear wax type were linked.

Linkage analysis of these 2 traits in affected family members pointed to the pericentromeric region of chromosome 16 as the site that contained both the "choreoathetosis locus" and the "ear wax locus". To isolate the specific genetic change that determined ear wax type, scientists compared the sequence of this region in Japanese people with dry ear wax with the sequence of Japanese people with wet ear wax. They found 3 SNPs which were associated with ear wax type, spread over a 5-gene region of DNA.

But only 1 of the SNPs, found in the ABCC11 gene, resulted in a nonsynonymous change in the protein product. The SNP, found in the ABCC11 gene, is a G538A substitution (in the reference sequence rs17822931) and results in a G180R substitution in the transporter protein it encodes. Individuals who inherit at least 1 copy of guanine at position 538 (GG homozygotes and GA heterozygotes) have wet ear wax, a dominant trait. Individuals with dry ear wax , a recessive trait, are AA homozygotes.

It is possible that the ABCC11 protein transports some of the lipids and granules found in wet ear wax, and a change of just one of its amino acids results in the production of dry ear wax that lacks some of these molecules.

Around the world, the "dry" A allele is more common in Asia, being most common in the North and East of Asia (100% in Northern Han Chinese and Koreans, less high in Mongolians, other areas of China, and Japan). Having apocrine glands that secrete dry ear wax may also be linked to decreased sweating from apocrine glands under the arms, and a decrease in bodily secretions in colder climates could be a survival advantage.

The A allele is also found in Native Americans, possibly reflecting the migration of their Ancestors from Siberia to North America.

In contrast, the "wet" G allele is more common in Sub-Saharan Africans and Caucasians.

Regardless of whether the ear wax is wet or dry, the ear has the same built-in conveyer-belt mechanism for ridding itself of cerumen—it uses epithelial cells that migrate away from the ear drum and up the ear canal to transport ear wax, and any debris it contains, out of the ear. These cells migrate at speeds of about 0.05 mm/day (similar to the rate of nail growth) .

Many problems with ear wax result from people using cotton tips in an attempt to remove the wax. But what they actually do is push the wax further into the ears, against the tide of migrating cells, causing it to impact and accumulate. So, to quote the words said by many Ear, Nose, and Throat (ENT) surgeons, "Do not put anything smaller than your elbow in your ear!".

Let nature take its course, or see a doctor instead.

References

1. Yoshiura K, Kinoshita A, Ishida T, Ninokata A, Ishikawa T, Kaname T, Bannai M, Tokunaga K, Sonoda S, Komaki R, Ihara M, Saenko V A, Alipov G K, Sekine I, Komatsu K, Takahashi H, Nakashima M, Sosonkina N, Mapendano C K, Ghadami M, Nomura M, Liang D S, Miwa N, Kim D K, Garidkhuu A, Natsume N, Ohta T, Tomita H, Kaneko A, Kikuchi M, Russomando G, Hirayama K, Ishibashi M, Takahashi A, Saitou N, Murray J C, Saito S, Nakamura Y, Niikawa N. A SNP in the ABCC11 gene is the determinant of human earwax type. Nature Genetics. 2006;38(3):324–330. PubMed PMID: 16444273.
2. Tomita H, Yamada K, Ghadami M, Ogura T, Yanai Y, Nakatomi K, Sadamatsu M, Masui A, Kato N, Niikawa N. Mapping of the wet/dry earwax locus to the pericentromeric region of chromosome 16. Lancet. 2002;359(9322):2000–2002. PubMed PMID: 12076558.

What you see is not what you get

DNA barcoding is helping scientists unveil nature's most hidden diversity

Marcia Triunfol, PhD

Created: August 1, 2005.

Any chef in town knows what *Lutjanus campechanus*, also known as the red snapper, looks like. It is a pricey, medium-sized, light-red fish that weighs about 10 pounds. You may enjoy it fried, baked, broiled, poached, or sautéed. But what the chefs don't know is that about 77% of red snappers sold in the United States are actually a mix of different species that includes some not-so-appreciated and rare reef fish (1). That's because some snappers look very much alike, making it difficult to sort them based on appearance. Besides being a possible theft against consumers who pay too much for ordinary reef fish, the mislabeling results in deceptive stock sizes of *L. campechanus,* with several implications for resource management.

Over two centuries ago, the great taxonomist Carl Linnaeus introduced the binomial species nomenclature that is still in use today. Focused mainly on morphology, Linnaeus' pioneer work was a milestone toward a classification system of the species. In 1942, the ornithologist Ernst Mayr proposed the biological species concept (BSC) (2) by which a species is defined as a group of organisms reproductively isolated. The BSC considers the organism's behavior, geographical location, reproductive attributes, genetic data, ecology, and physical appearance. Now scientists want to add a new wrinkle to the species definition: a DNA bar code.

According to Paul Hebert, one of the fathers of DNA barcoding, both phenotypic plasticity and genetic variability in a given key character can lead to a faulty classification. Additionally, most morphological keys are particular to a life stage or gender, preventing many organisms from being identified this way. A barcode, in contrast, can be applied to all stages of life (including eggs) and used for routine identification as well as detection of hidden species, with no expertise required.

Scientists believe that mitochondrial DNA is the best choice for DNA barcoding in animals, for many reasons. First, mitochondrial DNA is maternally inherited, which avoids recombination among individuals of the same species. Second, the low frequency of DNA deletions and insertions makes sequence alignments of different species easier because abrupt gaps are rare. Third, mitochondrial DNA is present in many copies in the cell and therefore easier to detect.

Recent studies (3) have shown that a 650-bp stretch of the mitochondrial cytochrome c oxidase I gene (*COI*) is very powerful in discriminating species and phylogeographic groups within species. Scientists attribute that to *COI* fast evolutionary rate and high incidence of base substitutions in third-position nucleotides. The existence of robust primers that enables routine PCR of the *COI* locus in most species has also contributed to make *COI* the number one choice for DNA barcoding.

In reality, any locus that has evolved fast and is abundant in the cell could be used for this purpose. In the past, scientists have inferred phylogenetic relationships using the 12S rRNA and 16S rRNA mitochondrial genes (4-6). Other possible choices include genomic rRNAs such as 18S and 28S, although their slow rates of divergence might represent a problem for species delineation. The real power of DNA barcoding lays on the ability of using the same locus to classify all species.

Hebert and his team (7) have used the *COI* system to identify 200 species of lepidopterans as well as several new specimens. More recently, the same group was able to identify 1,500 species by using this system (8). In another study (9), the application of DNA barcoding to a museum collection revealed that what has been known as the *Astraptes fulgerator* is actually composed of 10 previously unknown species. More recently, scientists have tested 260 species of North America birds and found that each species has a unique *COI* DNA barcode (10). Several other projects are underway, including a project to barcode Costa Rica's plants and Fish-BOL, an initiative that will collect 15,000 marine and 8,000 freshwater species, with the hope of assembling DNA barcodes for all fish.

The *COI*-based identification system does not work with all species, though. In some groups that have evolved apart recently, such as the stony corals and the cichlid fish, the 650-bp stretch has not accumulated enough variation. In other groups such as amphibians, *COI* primers have shown lower success rates because of highly variable *COI* priming sites present among groups and closely related species (11).

In the next 20 years, The Consortium for the Barcode of Life (CBOL), an international initiative that aims to accelerate compiling of DNA barcodes of known and newly discovered species, expects to have a *COI* profile for most of the estimated 5–10 million animal species on the planet. Using DNA barcoding, it will be possible to identify a species using as little as a single cell, allowing the detection of minute amounts of undesirable or regulated species in processed foods. Other applications include the identification of bird species that strike aircraft and of mosquitoes' eggs and larvae, which represent health threats to people around the world. DNA barcoding might also help protect endangered and threatened populations and prevent the mislabeling of commercial species.

Scientists hope that DNA barcoding will be a master key in precisely identifying every single species on the planet, and that includes the pricey and so-called red snapper sitting on your dinner plate

DNA Barcoding.

Using NCBI resources to find *COI* DNA sequences

Download file

For this tutorial, you will need the latest version of Flash player installed on your computer.

References

1. Marko PB,et al. Fisheries: mislabelling of a depleted reef fish. Nature. 2004;430:309–310. PubMed PMID: 15254528.
2. Mayr E. Systematics and the origin of species from the viewpoint of a zoologist. New York: Columbia University Press. 1942.
3. Hebert PD, Ratnasingham S, deWaard JR. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc Biol Sci. 2003;270:S96–S99. PubMed PMID: 12952648.
4. De Giorgi C,et al. Mitochondrial DNA in the sea urchin Arbacia lixula: evolutionary inferences from nucleotide sequence analysis. Mol Biol Evol. 1991;8:515–529. PubMed PMID: 1681410.
5. Ruttkay H, Solignac M, Sperlich D. Nuclear and mitochondrial ribosomal RNA variability in the obscura group of Drosophila. Genetica. 1992;85:131–138. PubMed PMID: 1378037.
6. Edwards SV, Wilson AC. Phylogenetically informative length polymorphism and sequence variability in mitochondrial DNA of Australian songbirds (Pomatostomus). Genetics. 1990;126:695–711. PubMed PMID: 1979038.
7. Hebert PD,et al. Biological identifications through DNA barcodes. Proc Biol Sci. 2003;270:313–321. PubMed PMID: 12614582.
8. deWaard JR, Ivanova NV, Topan JC, Landry J-F, Hebert PDN. DNA Barcoding of North America's Lepidoptera: The Story After 5000. in Consortium for the Barcode of Life. 2005.
9. Hebert PD,et al. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. Proc Natl Acad Sci U S A. 2004;101:14812–14817. PubMed PMID: 15465915.
10. Hebert PDN,et al. Identification of Birds through DNA Barcodes. PLoS Biol. 2004;2:e312. PubMed PMID: 15455034.

11. Vences M, Thomas M, van der Meijden A, Chiari Y, Vieites DR. Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. Front Zool. 2005;2:5. PubMed PMID: 15771783.

Do brains have a freshness date?

the effect of aging on the human brain

Victoria Sutton, PhD

Created: January 12, 2005.

As humans age, many bodily functions and abilities change. Scientists have long been interested in discovering how aging produces a decline in brain function and contributes to the development of certain diseases. To investigate the molecular shifts that occur in the brain during aging, some recent studies have focused on changes in gene expression patterns, both in humans and in other animals.

A recent study by Lu et al. (1) compared the levels of gene expression in postmortem brain samples from young and old human subjects. The gene expression profiles were clustered into three broad groups: young (below 40 years), intermediate (40-70 years), and old (70+ years) (see data at the Gene Expression Omnibus) (2). Whereas the young and old subjects were relatively homogeneous in their gene expression profiles, the intermediate group was much more heterogeneous, with vastly different rates of progression through the transition from the young gene expression profile to the old profile.

In older brains, the expression of genes involved in learning and memory, neuronal survival, and maintenance was decreased. Among these were the genes coding for microtubule-associated protein MAP1B, which stabilizes microtubules and promotes axonal growth (3,4) (see Entrez Gene entry, GEO data); MEF2C (see Entrez Gene entry, GEO data), which promotes the survival of neurons (1); subunit 2A of the glutamate receptor NMDAR (see Entrez Gene entry, GEO data), which is centrally involved in synaptic plasticity, the brain-restructuring process used in learning and memory; and calmodulin 1 (see Entrez Gene entry, GEO data), which is a central regulator of calcium-mediated signaling and plays an important role in memory. In contrast, the expression of genes coding for proteins involved in sensing and responding to cellular stresses increased, suggesting that older brains are exposed to higher levels of damaging stressors than are young brains (1). For example, one of the genes more highly expressed in older brains was the DNA repair enzyme OGG1 (see Entrez Gene entry, GEO data), which targets oxidatively damaged DNA.

Similar studies have been performed with a range of other animals, including *Caenorhabditis elegans*, *Drosophila melogaster* (5), and mice (6,7). Although the affected genes were not the same among all of the systems, the functional systems of the cells were affected in similar ways across species. For example, in the brains of both mice and humans, the expression of genes involved in synaptic function was decreased, whereas those involved in stress responses increased.

Further investigation of the genes whose expression was decreased in older human subjects showed that the promoters of these genes were more susceptible to oxidative damage than other genes tested (1), suggesting a mechanism for the observed decreased gene expression. Oxidative damage is caused by oxidants such as superoxide and hydrogen peroxide, which are produced as natural by-products of cell metabolism (8). The cell has response systems specifically dedicated to sensing and destroying such oxidants, as well as systems to repair any damage caused by those oxidants. It has long been hypothesized that oxidants play a role in aging processes, but this has proved difficult to demonstrate directly (9). The variable gene expression profiles (1) of the intermediate-age group suggest that oxidative damage may accumulate over a long period of time, with the effects occurring long after the initial oxidative damage is inflicted.

Perhaps some of the effects of aging can be slowed or lessened by controlling the level of oxidative stress in the cell; many methods are currently under investigation for reducing the creation of oxidants during metabolism. These include calorie restriction (currently being tested in clinical trials, see news coverage) (10), the use of antioxidant supplements, and a wide range of methods designed to target various cellular and molecular processes to safely and effectively reduce the production of oxidants during metabolism (10). Another

potentially fruitful therapeutic strategy is to enhance DNA repair in cells to slow the accumulation of oxidative damage in the DNA. Because the results outlined by Lu et al. (10) show that after the age of about 40 years, humans may begin to exhibit age-related changes in gene expression, some treatments might be more effective in young patients than in those already experiencing the effects of aging. The challenge for future scientific investigation will be to find effective treatments for increasing life expectancy while ensuring high quality of life.

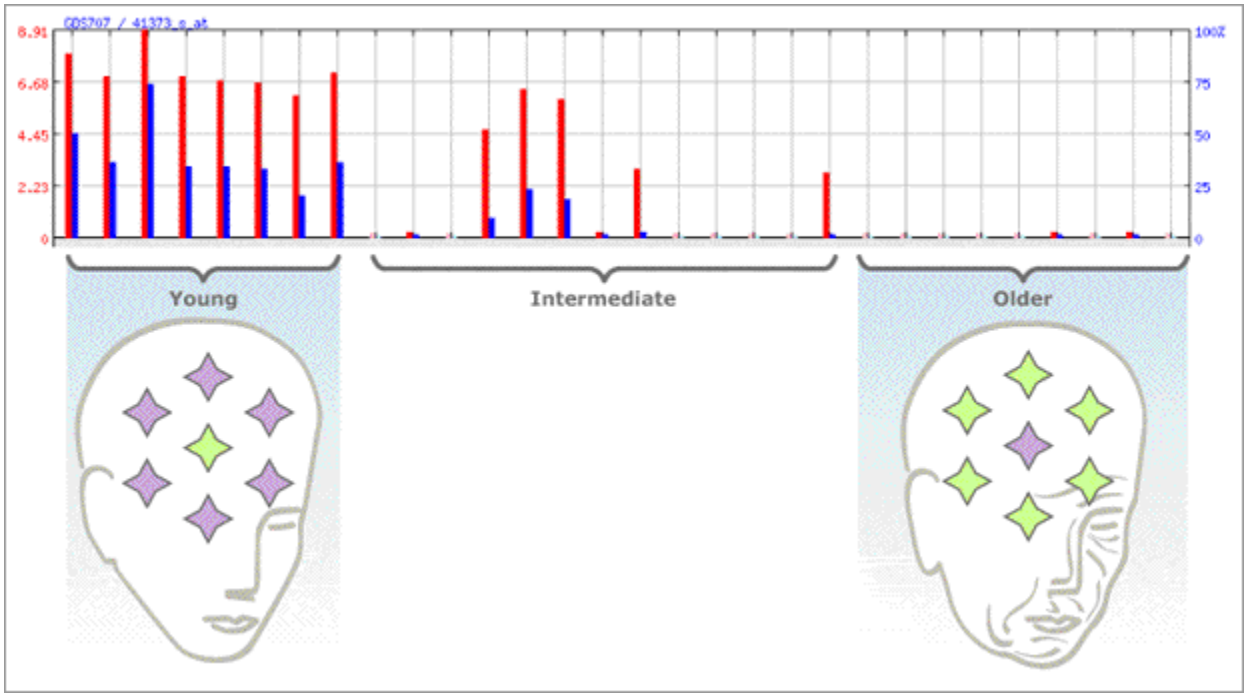
This Coffee Break was contributed by Victoria Sutton, PhD, while on rotation at the National Center for Biotechnology Information as a part of the Emerging Leaders Program from the Department of Health and Human Services (DHSS).

Aging and the Human Brain.

The changes in gene expression that underlie the effects of aging in humans

Download file

For this tutorial, you will need the latest version of Flash player installed on your computer.



The gene expression in human brains changes with age, with some genes decreasing expression while others increase. This is illustrated by a chart of the expression of the gene MAP1B, which shows high levels expression in the young, variable levels of expression at intermediate ages, and greatly reduced expression in older people. For more information about the chart, please visit the tutorial on "Aging and the Human Brain".

References

1. Lu T,et al. Gene regulation and DNA damage in the ageing human brain. Nature. 2004;429:883–891. PubMed PMID: 15190254.
2. Barrett T,et al. NCBI GEO: Mining millions of expression profiles - database and tools. Nucleic Acids Research. 2005;33:D562–D566. PubMed PMID: 15608262.
3. Black M M, Slaughter T, Fischer I.et al. Microtubule-associated protein 1b (MAP1b) is concentrated in the distal region of growing axons. J Neurosci. 1994;14:857–870. PubMed PMID: 8301365.
4. Gonzalez-Billault C,et al. Microtubule-associated protein 1B function during normal development, regeneration, and pathological conditions in the nervous system. J Neurobiol. 2004;58:48–59. PubMed PMID: 14598369.
5. McCarroll S A,et al. Comparing genomic expression patterns across species identifies shared transcriptional profile in aging. Nat Genet. 2004;36:197–204. PubMed PMID: 14730301.
6. Lee C K, Weindruch R, Prolla T A.et al. Gene-expression profile of the ageing brain in mice. Nat Genet. 2000;25:294–297. PubMed PMID: 10888876.
7. Jiang C H, Tsien J Z, Schultz P G.et al. The effects of aging on gene expression in the hypothalamus and cortex of mice. Proc Natl Acad Sci. 2001;98:1930–1934. PubMed PMID: 11172053.
8. Fridovich I,et al. Mitochondria: are they the seat of senescence? Aging Cell. 2004;3:13–16. PubMed PMID: 14965350.
9. Van Voorhies W A,et al. Live fast--live long? A commentary on a recent paper by Speakman et al. Aging Cell. 2004;3:327–330. PubMed PMID: 15379856.
10. Merry B J,et al. Oxidative stress and mitochondrial function with aging--the effects of calorie restriction. Aging Cell. 2004;3:7–12. PubMed PMID: 14965349.

Who let the dogs out?

a genetic classification of dog breeds

Laura Dean, MD

Created: October 13, 2004.

Are you surprised that the tiny chihuahua belongs to the same species as the imposing great dane? The domestic dog species (*Canis familiaris*) includes more than 400 breeds that differ, for example in their appearance (size, coat length, and color) and behavior, (guarding, herding, and hunting).

More than 150 breeds are officially recognized by the American Kennel Club, which assigns each breed to one of seven groups or a miscellaneous class, based on the uses for which the breeds were originally developed.

The domestic dog has evolved into a diverse number of breeds that are genetically distinct.

In a recent study, Parker et al. studied the genetic relationships among a diverse range of dog breeds. They found that most breeds of dog fall into four groups—three "modern" categories and one "ancient" group that may date back to antiquity.

The modern categories include breeds that have been around for fewer than 400 years:

- Working dogs/guard dogs, e.g., mastiff, bulldog, boxer
- Herding dogs, e.g., Belgian sheepdog, collie, shetland sheepdog
- Hunting dogs, e.g., scent hounds, terriers, spaniels, pointers, retrievers

This genetic classification of breeds grouped dogs together in a way that matched similarities in morphology and geographical origin. However, there were some surprises. For example, the oldest of all dog breeds are

commonly believed to be the Pharaoh Hound and Ibizan Hound, which resemble the ancient Egyptian dogs drawn on tomb walls more than 5000 years ago. However, this study failed to detect their ancient lineage. This may be because they are modern recreations of old breeds or because current tools are unable to detect their ancient genes.

In contrast, a diverse group of dog breeds appears to be most related to the dog's ancient ancestor, the grey wolf. These breeds include dogs whose appearances resemble the wolf (e.g., the Siberian Husky) and dogs that do not (e.g., the cuddlely Sharpei). Breeds that belong to this ancient grouping are diverse and originate from different continents, e.g., the Afghan from the Middle East, the Basenji from Africa, the Tibetan Terrier from Tibet, the Pekingese from China, and the Alaskan Malamute from the Arctic.

Parker et al. looked at microsatellites to find the genetic differences between breeds of domestic dogs. Microsatellites are short segments of DNA that contain repeats of DNA sequence. The repeats usually occur in a noncoding part of the gene, and their number is highly variable. Analysis of the microsatellites of 414 dogs representing 85 different breeds revealed that the degree of genetic differentiation between dog breeds is much higher than that found between human populations on different continents!

Given that most modern dog breeds have existed for fewer than 400 years, it is surprising that dog breeds are genetically distinct. But a dog can be matched to its breed by its individual genotype. Of 414 dogs tested, only 4 dogs were assigned to the wrong breed.

This apparent genetic isolation of dog breeds through selective breeding was reinforced by the formation of breed clubs in the mid-19th century. Rules such as the "Breed Barrier Rule" states that "no dog may become a registered member of a breed unless its dam and sire are registered members". Such selective breeding generates not only genetically diverse breeds of dog but also leads to the accumulation of mutations and inherited diseases. By using a genetic classification of dog breeds, scientists will be able to select breeds of dogs that share the same ancient mutations and genetic predisposition to diseases that some humans have. Analysis of this DNA is more likely to yield information about the diseases and the mutations responsible for them.

References

1. Parker H G, Kim L V, Sutter N B, Carlson S, Lorentzen T D, Malek T B, Johnson G S, DeFrance H B, Ostrander E A, Kruglyak L. Genetic Structure of the Purebred Domestic Dog. Science. 2004;304:1160–1164. PubMed PMID: 15155949.

Microbial diversity

let's tell it how it is

Jo McEntyre, PhD

Created: March 4, 2004; Updated: March 26, 2004.

An impressive number of bacteria—about 30,000 species—are represented in GenBank. However, our view of the microbial world is both scant and skewed. A recent estimate suggests that the sea may support as many as 2 million different bacteria, and a ton of soil might contain 4 million (1). Less than half of the bacteria represented in GenBank—about 13,000—have been formally described, and almost all of these (90%) lie within 4 of the 40 bacterial divisions (2). Similar or greater paucity of knowledge also exists for archaea and viruses (3).

Collecting sea water samples from the Sargasso Sea for the whole genome shotgun sequencing of microbial populations. (Photo supplied by The Center for the Advancement of Genomics.)

Sampling "wild" microorganisms leads to the discovery of new species and novel metabolisms, which may be important from both a basic science and a practical perspective (for example, see Refs 4,5 [search PubMed]). For example, if we characterized the community in the human gut, it would be easier to spot non-native organisms in food poisoning outbreaks. Pathogens that may underlie neurological syndromes that present with features of infection would stand out against the background flora (1). Engineered communities of microorganisms might also be able to assist clean up of environmental disasters or create sustainable energy sources.

Exploring bacterial diversity is typically done by amplifying rRNA genes, in particular 16S rRNA genes, from DNA samples isolated from a habitat. The sequences are then compared to each other and to the 16S rRNA sequences from known species. If no close match to an existing 16S rRNA gene sequence is found, then the test sequence is thought to represent a new bacterium and is listed in GenBank as "uncultured bacterium". Even in well-studied, discrete places like the human mouth, new groups of uncultured bacteria continue to be discovered all the time. A newly identified organism has to be isolated and cultured in the lab to be described further; but many bugs are just not amenable to monoculture—they have adapted to living in a specific environment and may need to be part of a complex community to survive (1-3).

16S rRNA genes are considered standard because they are thought to be conserved across vast taxonomic distance (they are critical for protein translation), yet show some sequence variation between closely related species. However, one problem with using rRNA genes is that they are often present in multiple copy numbers; therefore, other representative genes may be used for sampling specific populations.

Whole Genome Shotgun Sequencing of Environmental Samples

New approaches to environmental sampling are emerging (6–9). One of these used a microarray to discover and assist in the isolation of new viruses (6); another used a shotgun clone and sequencing method to explore marine viral communities (9). Two others have used whole genome shotgun (WGS) sequencing on a population of bacteria, obviating the need to isolate each organism before sequencing can begin (7,8). These methods, used in combination with existing methods, may provide shortcuts to the discovery of new genes and give a holistic persective to microbial populations.

One recent study used a WGS approach to explore a sample from an acid mine drainage biofilm (7; AADL00000000). These investigators report that near-complete genomes for *Leptospirillum* Group II and *Ferroplasma* Type II were assembled, along with more fragmentory assemblies for *Leptospirillum* Group III, *Thermoplasmatales archaeon gpl*, and *Ferroplasma acidarmanus Type I*. Analysis of the results provided some insight into how such organisms survive in an extreme environment.

In another test case of the WGS method, Venter *et al.* (8) sampled water from the Sargasso Sea—one of the most well-characterized regions of ocean in the world. The major set of samples produced 1.66 million short sequences, some of which could be grouped together into larger genomic pieces. There remained about 400,000 paired-end reads and singleton reads.

Finding the Data

Using a WGS method to sequence an undefined population as opposed to a single organism adds significant complexity to the assembly process and to the identification of genes. About 25% of the assembled data from the Sargasso Sea had 3X coverage or greater; these well-sampled portions were used to cluster the sequence by "organism".

The assembled sequences have been deposited in the WGS division of GenBank, with the project Accession number AACY01000000; thus, there are 811,372 WGS contigs in GenBank with the Accession numbers AACY01000001–AACY01811372. 498,641 of the WGS contigs are assembled into 232,442 scaffolds, the rest remain "singleton" WGS contigs; all but 10,685 of the scaffolds are made up of two contigs only. For the organism genomes listed in Table 1, 301 of the total scaffolds plus 36 singleton WGS contigs were used; the remainder have not been associated with any particular organism.

All of the short sequence reads, including those that were not included in the assembly, can be found in the Trace Archive.

The assemblies were then further clustered into 30 tentative organism "bins" based on depth of coverage, oligonucleotide frequencies and similarities to previously sequenced genomes. Of these, 12 are of sufficient size to be considered a genome assembly, while the remaining 16 are relatively small single scaffolds (Table 1). All organism bins have been assigned a taxonomy ID, and have been placed in the taxonomic tree. Figure 1 shows the graphical representation of the cf. Shewanella SAR-1 "genome" sequence.

A variety of approaches suggested that there are at least 1000 species represented in the Sargasso Sea samples (8). *Burkholderia* species were represented in a high proportion (a genus that includes human and plant pathogens and some environmentally important bacteria), as were two distinct species closely related to *Shewanella oneidensis*. Both of these genera require a more nutrient-rich environment than the open ocean can offer, suggesting that they originated from microhabitats such as marine snow. The cyanobacterium *Prochlorococcus* was also relatively abundant in some samples.

Although the primary focus of this study was on bacterial populations, WGS environmental sampling may be an equally valid approach for exploring plasmids (Table 2), phage, viruses, and eukaryotic microbes.

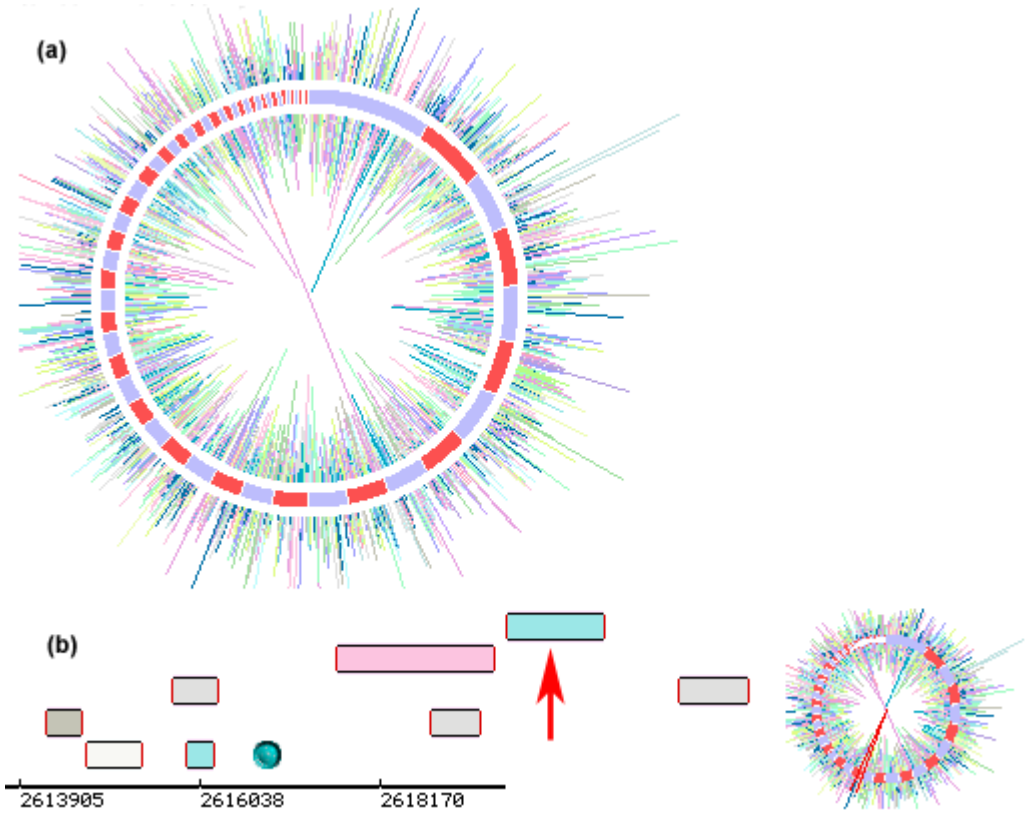


Figure 1. **(a)** Genome view of cf. Shewanella SAR-1, constructed from the whole genome shotgun sequence derived from Sargasso Sea environmental samples (8). Genes have been classified according to the COG functional categories of the protein products, and color-coded accordingly. note that the actual order of the scaffolds is unknown, so in this representation they have been ordered by size. Clicking on the image reveals the gene sequences and approximate location. **(b)** Selecting one of the genes (in this case, the blue gene around position 2619000) shows the results of an automated BLAST search (BLink). This gene is similar to L-sorbosone dehydrogenase from a variety of bacteria, archaea, and fungi. L-sorbosone dehydrogenase is an enzyme required for the biosynthesis of L-ascorbic acid, a product widely used in the food industry as a vitamin and antioxidant.

Each of the 28 organism "genomes" can be viewed in a similar mannner (see Table 1).

Table 1. The organism bins assembled from the Sargasso Sea WGS environmental sample dataset (8).

Organism Bin	Description	Data	Further Reading
Genome Assemblies			
cf. Alphaproteobacteria SAR-1	Oligotrophic Typical of marine bacterioplankton	Genome GenBank	PubMed Books
cf. Archaea SAR-1	One of the three major domains of life Often inhabit extreme environments	Genome GenBank	PubMed Books
cf. Bacteria SAR-1	One of the three major domains of life	Genome GenBank	PubMed Books
cf. Burkholderia SAR-1	Gram-negative bacilli Aerobic Found in a variety of aquatic environments	Genome GenBank	PubMed Books
cf. Gammaproteobacteria SAR-1	Purple bacteria Some plant pathogens	Genome GenBank	PubMed Books
cf. Microbulbifer SAR-1	Marine bacteria that degrade and recycle complex carbohydrates	Genome GenBank	PubMed Books

Table 1 continued from previous page.

Organism Bin	Description	Data	Further Reading
cf. Prochlorococcus SAR-1	Smallest known photosynthetic organism The most abundant in the ocean	Genome GenBank	PubMed Books
cf. Proteobacteria SAR-1	Phylum includes nitrogen-fixing bacteria and enteric bacteria	Genome GenBank	PubMed Books
cf. Pseudomonadaceae SAR-1	Gram-negative rods Often motile Includes many plant and a few animal pathogens	Genome GenBank	PubMed Books
cf. Shewanella SAR-1	Versatile metabolism Potential biotech applications such as heavy metal or chlorinated solvent reduction	Genome GenBank	PubMed Books
cf. Shewanella SAR-2*	Versatile metabolism Potential biotech applications such as heavy metal or chlorinated solvent reduction	Genome GenBank	PubMed Books
cf. Streptomyces SAR-1	Superficially similar to fungi (filaments and spores) Common in many habitats	Genome GenBank	PubMed Books
Single Scaffolds			
cf. Actinobacteria SAR-1	High G+C group of Gram-positive bacteria Most found in soil Some pathogens	GenBank	PubMed Books
cf. Bordetella SAR-1	Gram-negative coccobacilli Strict aerobes	GenBank	PubMed Books
cf. Burkholderiaceae SAR-1	Occupy diverse ecological niches May have potential for biotech applications but also involved in human infections	GenBank	PubMed Books
cf. Caulobacter SAR-1	Found in oligotrophic environments Prosthecate (having appendages)	GenBank	PubMed Books
cf. Crenarchaeota SAR-1	Archaeal Most species are motile Tolerant of extreme acidity and temperature	GenBank	PubMed Books
cf. Cyanobacteria SAR-1	Aquatic and photosynthetic Often called ”blue-green algae”	GenBank	PubMed Books
cf. Enterobacteriaceae SAR-1	Large Gram-negative rods Facultative anaerobes	GenBank	PubMed Books
cf. Haemophilus SAR-1	Gram-negative rods Like to grow on blood agar Some pathogens	GenBank	PubMed Books
cf. Magnetococcus SAR-1	Gram-negative coccus Magnetic bacteria Usually located at sediment-water interface	GenBank	PubMed Books
cf. Magnetospirillum SAR-1	Magnetic bacteria	GenBank	PubMed Books
cf. Ralstonia SAR-1	Includes medically and economically important plant and animal pathogens	GenBank	PubMed Books
cf. Rhizobiales SAR-1	Involved in nitrogen fixation, often in symbiotic relationships with plants	GenBank	PubMed Books

Table 1 continued from previous page.

Organism Bin	Description	Data	Further Reading
cf. Sinorhizobium SAR-1	Symbiotic nitrogen fixation in plant root nodules	GenBank	PubMed Books
cf. Spirochaetales SAR-1	Spiral rods Some pathogens (e.g. *Borrelia burgdorferi* - Lyme disease)	GenBank	PubMed Books
cf. Streptomycetaceae SAR-1	Typically aerobic and found in soil Some parasitic forms	GenBank	PubMed Books
cf. Vibrionaceae SAR-1	Gram-negative, non-sporing rods Generally motile Many strains of *Vibrio* genus cause infection	GenBank	PubMed Books

cf. is used to designate an unidentified species of the genus. Therefore, "cf. Burkholderia" means "something that is like the genus Burkholderia" (in this case, by sequence similarity).
As each organism bin could actually represent several different unidentified species, a strain name cannot be assigned, so instead, the suffix "SAR-#" identifies each bin as a "Sargasso Sea cyber-species".
* cf. Shewanella SAR-2: two distinct Shewanella genomes were constructed from the dataset.

Table 2. The plasmid bins assembled from the Sargasso Sea WGS environmental sample dataset (8).

Plasmid Bin	Data
Plasmid pSAR-1	GenBank
Plasmid pSAR-2	GenBank
Plasmid pSAR-3	GenBank
Plasmid pSAR-4	GenBank
Plasmid pSAR-5	GenBank
Plasmid pSAR-6	GenBank
Plasmid pSAR-7	GenBank
Plasmid pSAR-8	GenBank
Plasmid pSAR-9	GenBank

References

1. Curtis T P, Sloan W T, Scannell J W. Estimating prokaryotic diversity and its limits. Proc Natl Acad Sci USA. 2002;99:10494–10499. PubMed PMID: 12097644.
2. DeLong E F. Microbial seascapes revisited. Curr Opin Microbiol. 2001;4:290–295. PubMed PMID: 11378481.
3. Roossinck M J. Plant RNA virus evolution. Curr Opin Microbiol. 2003;6:406–409. PubMed PMID: 12941413.
4. Kazor C E, Mitchell P M, Lee A M, Stokes L N, Loesche W J, Dewhirst F E, Paster B J. Diversity of bacterial populations on the tongue dorsa of patients with halitosis and healthy patients. J Clin Microbiol. 2003;41:558–563. PubMed PMID: 12574246.
5. Béejà O, Aravind L, Koonin E V.et al. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science. 2000;289:1902–1906. PubMed PMID: 10988064.
6. Wang D, Urisman A, Liu Y T.et al. Viral discovery and sequence recovery using DNA microarrays. PLoS Biol. 2003;1 PubMed PMID: 14624234.
7. Tyson G W, Chapman J, Hugenholtz P, Allen E E, Ram R J, Richardson P M, Solovyev V V, Rubin E M, Rokhsar D S, Banfield J F. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature. 2004;428:25–26. PubMed PMID: 14961025.

8. Venter J C, Remington K, Heidelberg J.et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science.[Epub ahead of print] PubMed PMID: 15001713.
9. Breitbart M, Salamon P, Andresen B.et al. Genomic analysis of uncultured marine viral communities. Proc Natl Acad Sci USA. 2002;99:14250–14255. PubMed PMID: 12384570.

Talking about the genetics of talking

the discovery of the first "speech and language" gene

Laura Dean, MD

Created: November 1, 2003.

Language has played a fundamental role in human cultural evolution. The ability of humans to speak involves fine motor control of the larynx and mouth; this control is absent in other primates.

The ability to articulate speech includes performing complex orofacial movements.

Chimpanzees and other great apes lack this ability.

Individuals with specific language impairment (SLI) have a developmental disorder of speech and language that is not attributable to deafness, autism, or any other general causes. Examples of symptoms include problems in articulation, identification of basic speech sounds, and understanding sentences and grammar. Although SLI does run in families, the inheritance patterns are usually complex, and until recently, little could be said about its genetic basis.

This changed with the study of a British family known as "KE". Several generations of this family were affected by SLI which was inherited in an autosomal dominant fashion. In 1998, this disorder was linked to a small segment of chromosome 7, labeled SPCH1. Then came the discovery of CS, an unrelated person who had a SLI, similar to that of the KE family and had a chromosomal translocation involving the SPCH1 interval. The discovery of CS helped to narrow the language disorder to a specific gene, FOXP2 (forkhead box P2).

FOXP2 belongs to a family of genes that encodes transcription factors (proteins that regulate the expression of genes). FOXP2 contains a forkhead domain that binds to DNA. The forkhead domain is disrupted in CS by translocation breakpoint, and in the KE family by a point mutation. Lai *et al.* propose that such a disruption in this region alters the DNA-binding and/or transactivation properties of FOXP2.

There is some evidence that the FOXP2 gene is a key part of human evolution. For example, the protein product is almost identical in mice, monkeys, and apes, which are separated by 130 million years of evolution. But humans differ from these species in two or three amino acids, through mutations estimated to have occurred within the last 200,000 years.

What is the role of FOXP2? We know that forkhead domain transcription factors have many important roles in the regulation of the development of embryos. Indeed, FOXP2 is expressed in fetal tissue, including the developing brain.

Enard *et al.* suggest that FOXP2 is needed in the development of the normal brain circuitry that underlies language and speech. They propose that at a critical point in fetal brain development, affected individuals have only half the normal amount of functioning transcription factor, which is not enough for normal early brain development.

FOXP2 is the first gene to be directly linked to speech and language disorders. Whatever the exact function of the gene turns out to be, it is unlikely that only one gene is responsible for speech. However, this discovery of a genetic component fuels the search for genetic causes of other cognitive and learning disorders.

A Gene involved in Speech?

Taking a closer look at the forkhead domain.

Download file

For this tutorial, you will need to have flash installed on your computer.

Variations on a gene

investigating the genetic basis of iron overload

Laura Dean, MD

Created: August 25, 2003.

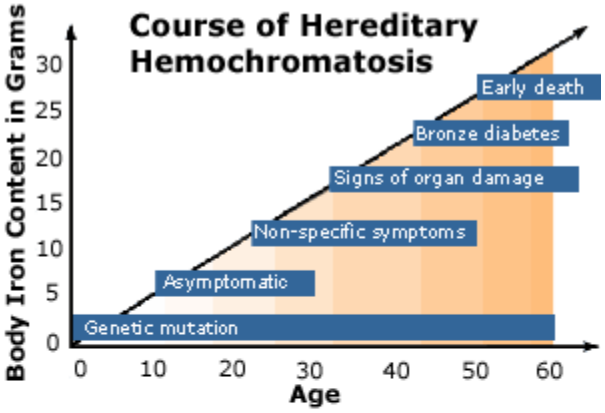


Chart reflecting the course of hereditary hemochromatosis This graph depicts the clinical course of iron overload due to hereditary hemochromatosis. The Y axis reflects increasing amounts of iron concentration in the body and the X axis reflects increasing age in years from birth to age 60. The diagonal line shows that as a person's iron concentration increases over time, the effects of iron on the body are initially seen as non-specific symptoms, followed by more serious signs of organ damage, bronze diabetes, and ultimately early death.

Slope depends on individual.

Hereditary hemochromatosis is a common disease that results in the accumulation of iron in the body. Approximately 5 per 1000 Americans inherit this disorder, but symptoms may never appear, leaving many people undiagnosed.

Mutations of the *HFE* gene are commonly at fault. In this tutorial, we examine the link between SNPs of the *HFE* gene and the development of disease. In the course of deciphering whether a SNP is significant, we look to see whether the SNP causes an amino acid change, the nature of this change (e.g., a switch of cysteine to another amino acid may disrupt a disulfide bond), and whether this change takes place in a conserved domain. We also take a look at the amino acid change on a three-dimensional representation of the HFE protein structure.

So... can a switch of a single nucleotide be responsible for a mutant HFE protein that allows iron regulation to go astray?

Search the SNPs database for HFE.

Do single nucleotide polymorphisms in the *HFE* gene have functional consequences?

Download file

For this tutorial, you will need to have flash installed on your computer.

A small fortune

small RNAs involved in the regulation of developmental timing

Created: June 5, 2002.

The world of small RNAs just got bigger. In *Caenorhabditis elegans*, two small temporal RNAs produced by the ribonuclease Dicer have previously been shown to be involved in regulating developmental timing. These RNAs — lin-4 and let-7, 22 and 21 nucleotides in length, respectively — act as antisense repressors of messenger RNA translation and, until recently, they were the only known RNAs of this class. But three papers published in *Science* now show that *lin-4* and *let-7* probably belong to a large family of single-stranded RNAs, 20 - 24 nucleotides in length, called microRNAs (miRNAs). These results indicate that post-transcriptional regulation by small RNAs is more common than previously believed.

Lagos-Quintana and colleagues used complementary DNA libraries constructed from a size-fractionated RNA population to identify 14 new miRNAs in *Drosophila melanogaster* and 19 new miRNAs in humans. Lau *et al*. created a cDNA library enriched for Dicer products, distinguished from other oligonucleotides by their small size, 5'-monophosphate group and 3'-hydroxyl group, to identify 54 novel miRNAs in *C. elegans*. Finally, using size-selected cDNA cloning together with computational methods, Lee and Ambros identified 15 miRNAs in *C. elegans*, 11 of which matched those identified by Lau and co-workers. In all cases, they showed the miRNAs were not simply the breakdown products of mRNAs or structural RNAs.

These papers identified 91 different miRNAs in total, about 12% of which have been conserved through evolution. Moreover, Lau and colleagues found that ~85% of the miRNAs identified in *C. elegans* had homologues in the 90%-complete *C. briggsae* genome sequence.

miRNAs are produced through processing, probably by Dicer, of a ~70-nucleotide precursor stem-loop structure. Either the 5' or the 3' arm of the precursor can be released to form the miRNA, with one exception. miR-56, identified by Lau *et al*., exists in two forms, resulting from processing of both the 5' and 3' arms of the precursor stem. How miRNA excision occurs has yet to be defined.

The *mir* genes often cluster together in the genome; for example, Lagos-Quintana and colleagues showed that *mir-3*, *-4*, *-5*, and *-6* form a gene cluster in the *Drosophila* genome. The *mir* gene clusters investigated so far are co-expressed, and Lau and co-workers predicted that, in *C. elegans*, the gene cluster *mir-35* to *-41* is transcribed to form a single RNA precursor, which is processed to produce miR-35 to -41. Some *mir* genes have multiple genomic copies, and some miRNAs are highly homologous.

All three groups investigated the expression of miRNAs and found that, in some cases, it was both stage- and tissue-specific. For example, Lee and Ambros found that *mir-1* is expressed stage-specifically in mouse embryogenesis and tissue-specifically in the human heart. These regulated expression patterns indicate an involvement in developmental control.

miRNAs have been proposed to function as "riboregulators", regulating gene expression by binding sequence-specifically to mRNAs, thereby blocking translation. The challenge now is to define the potential targets of miRNAs and their exact functions. There are probably many miRNAs yet to be identified and, if they are found to be as numerous and diverse as the miRNAs identified in these papers, they could have a range of regulatory functions. These authors seem to have discovered a small fortune, and the world of small RNAs could turn out to be very big indeed.

Story by Rachel Smallridge, *Nature Reviews Molecular Cell Biology*

Search the Bookshelf.

Created: June 5, 2002

Click on the link below to start an html tutorial.

can additional resources be researched?

Finding Fanconi

the hunt for the cause of autosomal dominant renal Fanconi syndrome

Created: October 22, 2001.

The human genome lays down the blueprint for our physiology and thus provides a framework by which to study genetic-based diseases. Researchers have focused recently on a region that may be responsible for a debilitating kidney disease — autosomal dominant renal Fanconi syndrome (RFS). In RFS, the proximal tubules of the kidney are functionally impaired. This causes many essential compounds that would normally be returned to the bloodstream to instead be excreted into the urine and removed from the body.

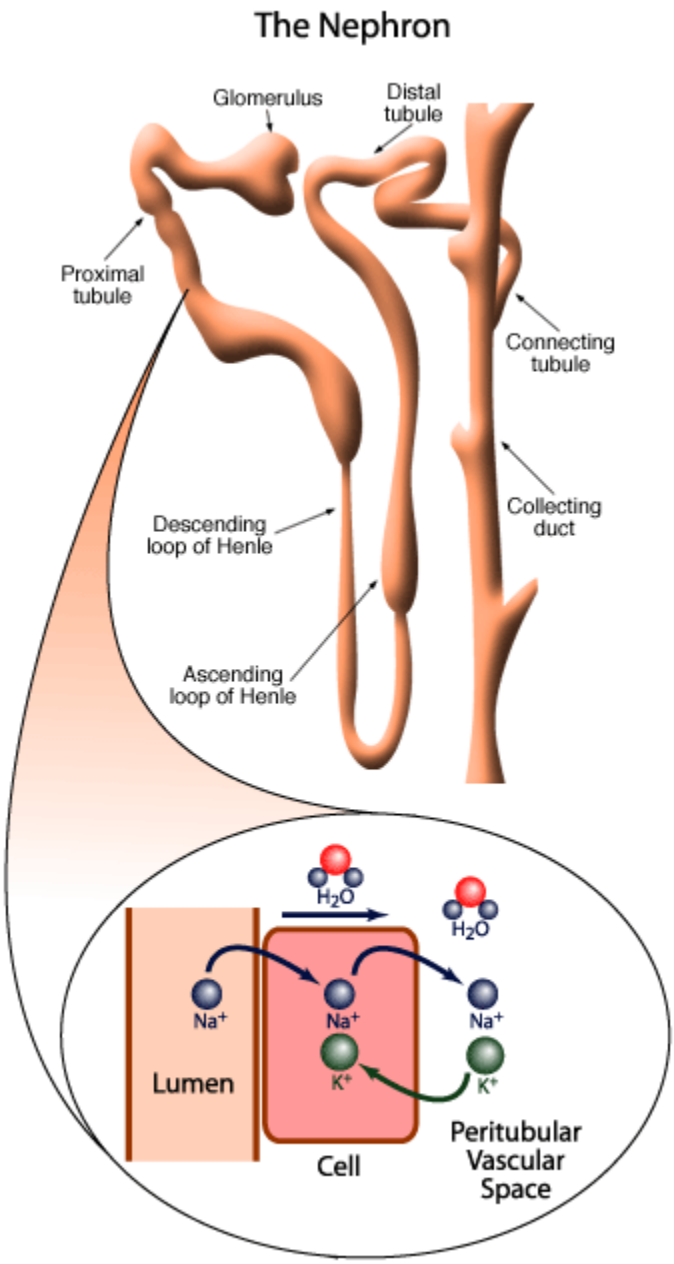
Genetic as well as environmental factors can lead to the development of RFS. An autosomal dominant form of RFS has been observed in several families; one of these was used as the basis for an attempt to find a genetic locus to the disease. The inheritable form of the syndrome is of particular interest for researchers because it can potentially provide insight into the workings of the proximal tubules.

Correlating a disease with a genetic mutation is not an easy task. To successfully map a disease gene on the human genome, it is necessary to have a series of genomic landmarks. This has been one of the major accomplishments of the Human Genome Project; over 10,000 polymorphic markers have been identified and contextually placed onto framework maps.

To find the region associated with RFS, DNA samples from the afflicted family were initially scanned using polymorphic markers that were distributed throughout the genome. Linkage analysis implicated marker D15S659, which is found on the long arm of chromosome 15. This initial hit gave researchers a rough area to further scrutinize by conducting a more detailed screen with 24 localized markers. From the secondary screen, two markers—D15S182 and D15S537 —were determined to exhibit the greatest correlation with RFS.

Genes in the 15q15.3 region are now being considered as candidates for association with RFS. By definitively associating autosomal dominant RFS with a gene, new insights into the pathology of the disease can be gained. One possible candidate gene, HSPC129, codes for a hypothetical protein with unknown function.

Further clues as to the possible function of an uncharacterized protein can be discerned by comparing it with other, characterized proteins (e.g., using NCBI's Related Sequences feature). In the case of HSPC129, one of the proteins it shares similarity with is the yeast protein Psr2P. This protein is involved in the indirect regulation of transmembrane sodium transportation. Active transport of ions such as sodium takes place in the proximal tubules of the kidneys and is a key component of healthy kidney function. Could HSPC129 also be involved in the regulation of ion transport in the kidneys similar to Psr2P regulation of sodium transportation in yeast cells? Although the presence of HSPC129 in the proximal tubules of the kidneys remains to be determined, studies on Psr2P in yeast may give insight into human HSPC129 function and possibly lead to a treatment for autosomal dominant RFS.



Active sodium transportation in the proximal tubule.
The kidneys are responsible for a number of important regulatory functions such as the maintenance of ion levels in the body, water retention/removal, waste excretion, blood pressure regulation, and maintenance of proper blood acidity. Nephrons are the functional unit of the kidney. Within the proximal tubule portion of the nephron is found the highest concentration of sodium transporters. These transporters are responsible for the active reabsorption of sodium ions from filtrate present in the lumen of the nephron. Water is absorbed passively during this process due to the accumulation of sodium in the peritubular spaces. Any disruption to this system would result in the loss of large amounts of water, sodium, and other ions.

Opening the flood gates?

association of NOD2 with Crohn's disease

Created: August 6, 2001.

It's been a long time coming, but now two papers report a clear cut identification by linkage mapping of a gene involved in a common human disorder — Crohn's disease (CD). Importantly, they also indicate how the innate immune system might be involved in the aetiology of CD, because the identified gene — *NOD2* — encodes an intracellular receptor for bacterial lipopolysaccharides (LPS) that activates NFκB, a target of the innate immune signalling pathway and a transcriptional regulator of inflammatory genes.

CD is a chronic inflammatory gut disorder, thought to be caused by an abnormal inflammatory response to enteric microbes. In 1996, a CD susceptibility locus, *IBD1*, was identified on chromosome 16. Little progress has been made since, but it is this locus that the two research teams — one European, the other US-based — tackled in their studies, using positional-cloning and candidate-gene strategies, respectively.

Hugot *et al.* took a decisive step when they identified association of CD to an allele of a chromosome-16 microsatellite marker. Despite the borderline significance of this association, the authors went on to identify putative transcripts in the region of this marker, and identified over 30 single nucleotide polymorphisms (SNPs) by sequencing the region from affected and unaffected individuals. Several turned out to be non-synonymous variants in a chromosome — 16 gene, *NOD2*. Three of these SNPs — each independently associated with disease susceptibility — altered the leucine-rich repeat (LRR) region of NOD2, which is required for LPS recognition.

Having previously identified *NOD2*, Ogura *et al.* considered it a candidate for CD because of its chromosome-16 location. On sequencing the gene from CD individuals, they identified an insertion that caused two frameshift mutations in the LRR region and the premature truncation of NOD2. In *in vitro* assays, this mutant NOD2 produced considerably diminished levels of NFκB activation in response to bacterial LPS compared to wild-type NOD2.

So how could *NOD2* contribute to susceptibility to CD? The innate immune system regulates the immediate immune response to bacterial pathogens, components of which are recognized in host immune cells by specific receptors, such as NOD2. A defect in this recognition might lead to an exaggerated inflammatory reaction being mediated by the adaptive immune system. Alternatively, NOD2 might act to trigger cytokines that dampen inflammatory responses. Although NOD2 does not account for all susceptibility to CD, it does provide a first glimpse into the aetiology of the disease and should speed the discovery of other CD loci and future therapies, and improve its diagnosis. These papers are hopefully the first of many such successes in grappling with the genetic basis of multifactorial, common disease.

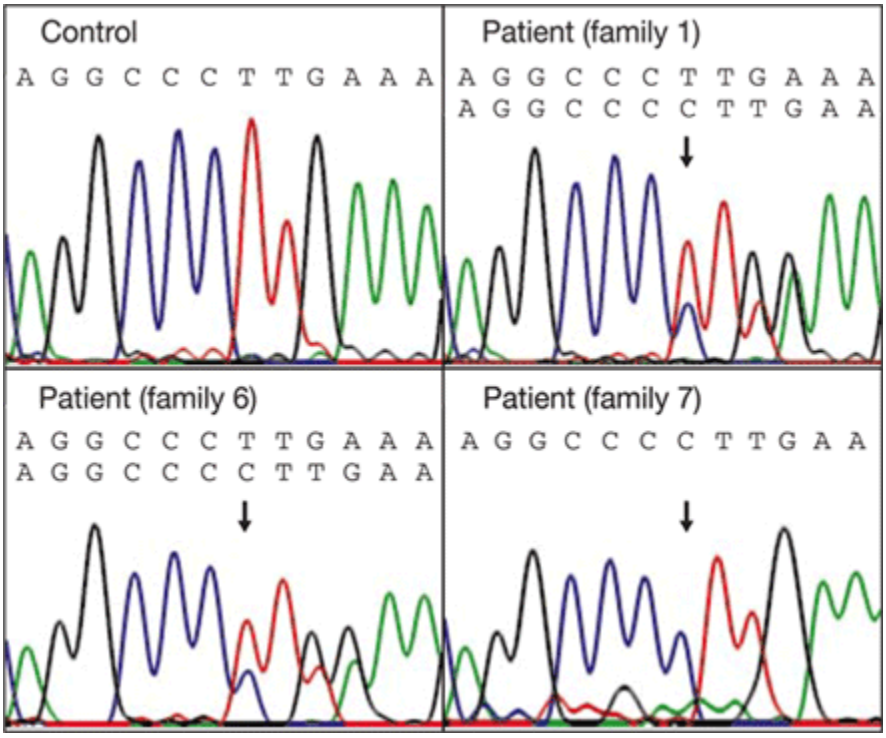
Story by Jane Alfred, *Nature Reviews Genetics*

Search the genome for the *NOD2* gene polymorphisms.

Created: August 6, 2001

Click on the link below to start an html tutorial.

Are there additional polymorphisms in the *NOD2* gene?



DNA sequence electropherograms of the *NOD2* gene.
A portion of *NOD2* exon 11 DNA sequence from control and three CD-affected individuals. The control sequence codes for full-length NOD2 protein. The patients from families 1 and 6 are heterozygous for a cytosine insertion at position 3020 in the *NOD2* gene. The wild-type sequence in these panels is in the upper position and is read GCC-CTT-GAA. The sequence containing the cytosine insert is in the lower position and is read GCC-CCT-TGA. The extra cytosine base (marked by the arrows) causes a framshift mutation to occur, and the TGA sequence immediately downstream is recognized as a stop codon, causing the NOD2 protein to be truncated. The patient from family 7 is homozygous for the same cytosine insertion.

Honey, I shrunk the genome

genome reduction in the leprosy bacillus

Created: April 30, 2001.

Getting by with the bare minimum seems to be the *modus operandi* of *Mycobacterium leprae* — the causal agent of leprosy. Its genome sequence reveals that it has undergone massive genome 'downsizing' over time, discarding more than half its genes and rendering it the most striking example of genome reduction in a microbial pathogen.

The leprosy bacillus is famed for being the first microorganism definitively shown to be associated with human disease. It evades the host's immune response by invading and propagating inside the vacuoles of macrophages called phagosomes. From there, it infects the Schwann cells of the peripheral nervous system, where it disrupts myelin production, thus leading to the characteristic features of leprosy, which include skin lesions and sensory loss.

Attempts to study *M. leprae* have been thwarted by fruitless efforts to grow it in the laboratory. This is due in part to its excruciatingly slow growth — it divides only once every two weeks, slower than any other bacterium — and its requirement for an elusive cocktail of nutrients. But the sequencing of its genome by Stewart Cole and colleagues now uncovers clues to its unique metabolism and unusual 'reductive evolution'.

M. leprae seems to have completely dispensed with or substantively reduced certain metabolic pathways, including oxidative and anaerobic respiratory chains. The enzymes for breaking down host-derived lipids, a means by which many mycobacterial pathogens derive their energy, are also drastically reduced in *M. leprae*. By contrast, most anabolic pathways seem to be intact, indicating that *M. leprae* depends on these pathways to survive in the nutrient-poor microenvironment of phagosomes.

Comparison of the genome sequence of *M. leprae* with that of its cousin *Mycobacterium tuberculosis* indicates that the former has undergone substantial downsizing, losing more than 2,000 genes since its divergence from a common mycobacterial ancestor. Presumably its genes became inactivated once their functions were no longer essential to survival, and this was followed by genome shrinkage through rearrangements and deletions.

The availability of the genome sequence will aid efforts to define virulence factors of the leprosy bacillus, of which few are presently known, and will provide inroads for developing new vaccines and diagnostic tests. Also, comparative genomic analyses with *M. tuberculosis* are likely to yield insight into the pathogenicity of the causal agent of tuberculosis and into potential drug targets based on the set of genes crucial to the survival of both organisms.

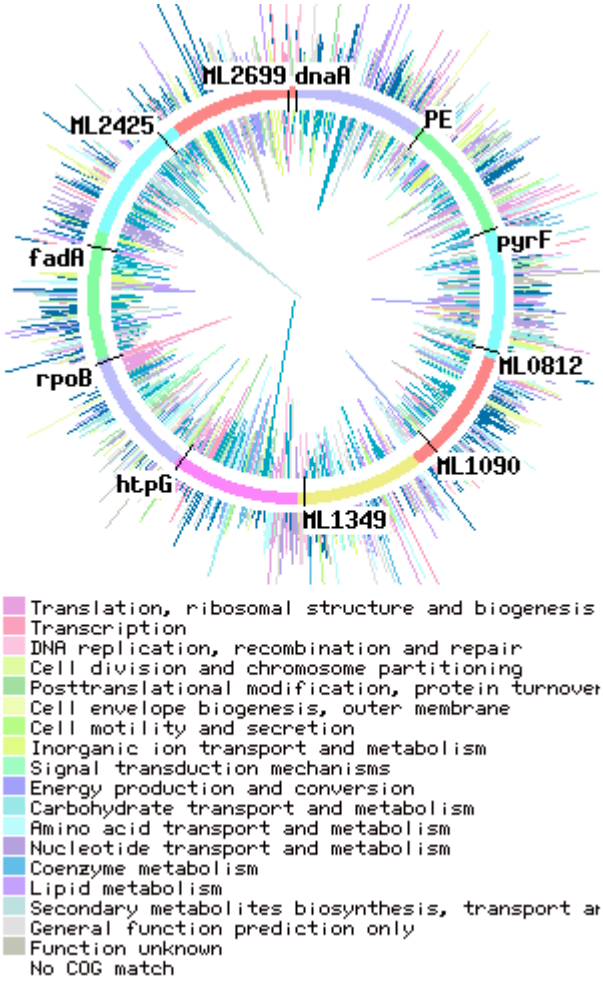
Story by Carina Dennis, *Nature Reviews Genetics*

Search TaxPlot for structures similar to E5.

Created: April 30, 2001

Click on the link below to start an html tutorial.

Search for structures similar to E5



Protein coding genes distribution map for *Mycobacterium leprae*.
The leprosy bacillus genome contains numerous examples of gene deletion and decay. The relative locations of various genes in the genome are depicted in the map above. Protein coding genes are color coded in the map according to their classification within clusters of orthologous groups (COGs) functional categories. COGs represent proteins or groups of paralogs that are found in at least 3 phylogenetically-distant genomes. For more information about COGs, see *Science* 1997 Oct 24:278(5338):631-7.

Ready, steady, go!

a two-part switch that regulates gene expression

Created: March 12, 2001.

Traffic lights regulate the movement of vehicles on roads by transmitting 'stop', 'get ready' and 'go' signals to drivers. Similarly, antigen-presenting cells use cytokines as stop and go signals for lymphocytes. But what's the switch that changes the signal? In the January issue of *Nature Immunology,* Amy Weinmann and colleagues describe a two-part switch for regulating transcription of a cytokine gene: one signalling pathway leads to chromatin remodelling, and a second, independent pathway activates transcription.

An important element in the initiation of inflammatory responses is the activation of macrophages, resulting in the production of pro-inflammatory cytokines such as interleukin 12 (IL-12), a heterodimeric protein comprising p40 and p35 subunits. Toll-like receptors (TLRs), which are expressed on macrophages, recognize microbial molecules and transmit signals that initiate transcription of cytokine genes; TLR4 recognizes the Gram-negative bacterial product lipopolysaccharide (LPS). TLRs use several signalling pathways, including the nuclear factor κB (NF-κB) and Jun N-terminal kinase pathways, to initiate gene transcription. Which of these pathways stimulates macrophages to produce IL-12?

Using restriction enzyme accessibility assays, Weinmann and colleagues found that TLR4 signalling in response to LPS activation results in nucleosome remodelling at the p40 promoter. Curiously, although active NF-κB is essential for transcription of p40, remodelling was not dependent on NF-κB or another transcription factor, CCAAT enhancer-binding protein β. It seems that other TLR4-inducible factors can stimulate remodelling, perhaps making the p40 promoter more accessible to transcription factors such as NF-κB.

So chromatin remodelling — a previously unrecognized endpoint of TLR signalling — behaves like an amber signal that prepares the chromatin for NF-κB, the green light for transcription of p40. But what is the identity of the protein that recruits the remodelling complex, and what exactly is this complex? Further work in this area should enhance our understanding of TLR signalling and the regulatory mechanisms controlling induction of the inflammatory response.

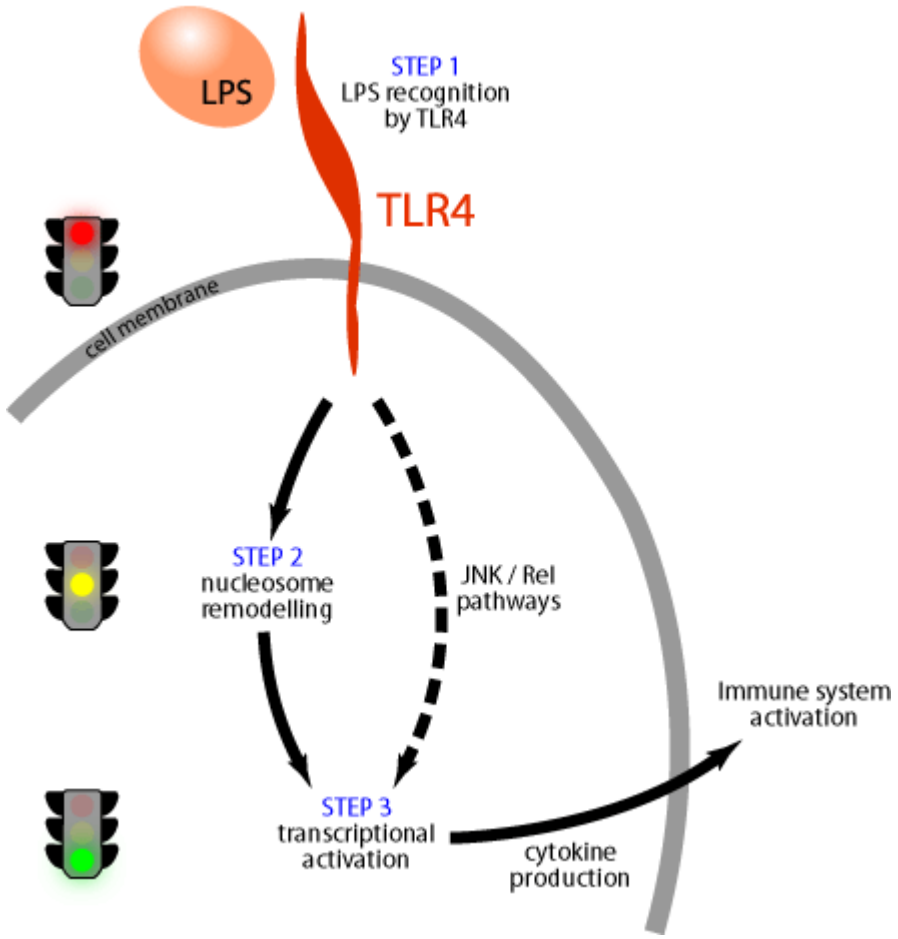
Story by Elaine Bell, Nature Reviews Molecular Cell Biology

Use BLAST to search for TLR4.

Created: March 12, 2001

Click on the link below to start an html tutorial.

Blast the human genome with mouse TLR4



TLR4 signalling pathway.
Activation of IL-12 subunit p40 transcription involves several coordinated steps. In the first step, Toll-like receptor 4 (TLR4) recognizes the bacterial surface molecule lipopolysaccharide (LPS). TLR4 then triggers the p40 nucleosome to undergo a remodelling event that is directed by an as yet unknown factor. TLR4 also activates other pathways such as Rel and JNK which act upon the remodelled nucleosome in the third step. This series of events provides a "green light" that allows the transcription of the IL-12 p40 gene to take place. The cytokine IL-12 is an important regulator of immune functions such as inflammation and Th1 development.

Fluorescent timer

the E5 mutant of the red coral protein drFP583 changes its fluorescence from green to red over time

Created: January 22, 2001.

There are many ways to monitor the onset of gene expression, but so far it has been impossible to detect its down-regulation. This problem might have been solved now, as Terskikh and colleagues report in *Science* a simple method to follow promoter activity.

Last year, a red fluorescent protein (drFP583) was identified in tropical corals, further increasing the wide spectrum of possibilities to light up cells in different colors. Not satisfied with just one color, Terskikh and colleagues introduced random mutations into drFP583, and found one mutant (called E5) that changes its fluorescence from green to red in a time-dependent manner. As E5 switches from green to red fluorescence over time, it can be used as a timer for gene expression. During the first hours of activity of a promoter, green fluorescence is predominant, whereas sustained activity of the promoter leads to a mixture of green and red fluorescence. A few hours after the promoter is turned off, only red fluorescence remains.

Terskikh and colleagues verified these predictions in three experimental systems. First they monitored up- and down-regulation of E5 expression in Tet-on and Tet-off mammalian expression systems. Then they followed the activity of a heat-shock promoter during heat-induced stress of*Caenorhabditis elegans.* Last, they traced the expression of a homeobox gene involved in the patterning of anterior structures in *Xenopus laevis.* In all cases, green fluorescence correctly indicated the onset of gene expression and was replaced with red fluorescence when expression ceased.

So after decades of blue-stained embryos, we'll now have to get used to seeing gene expression in green and red.

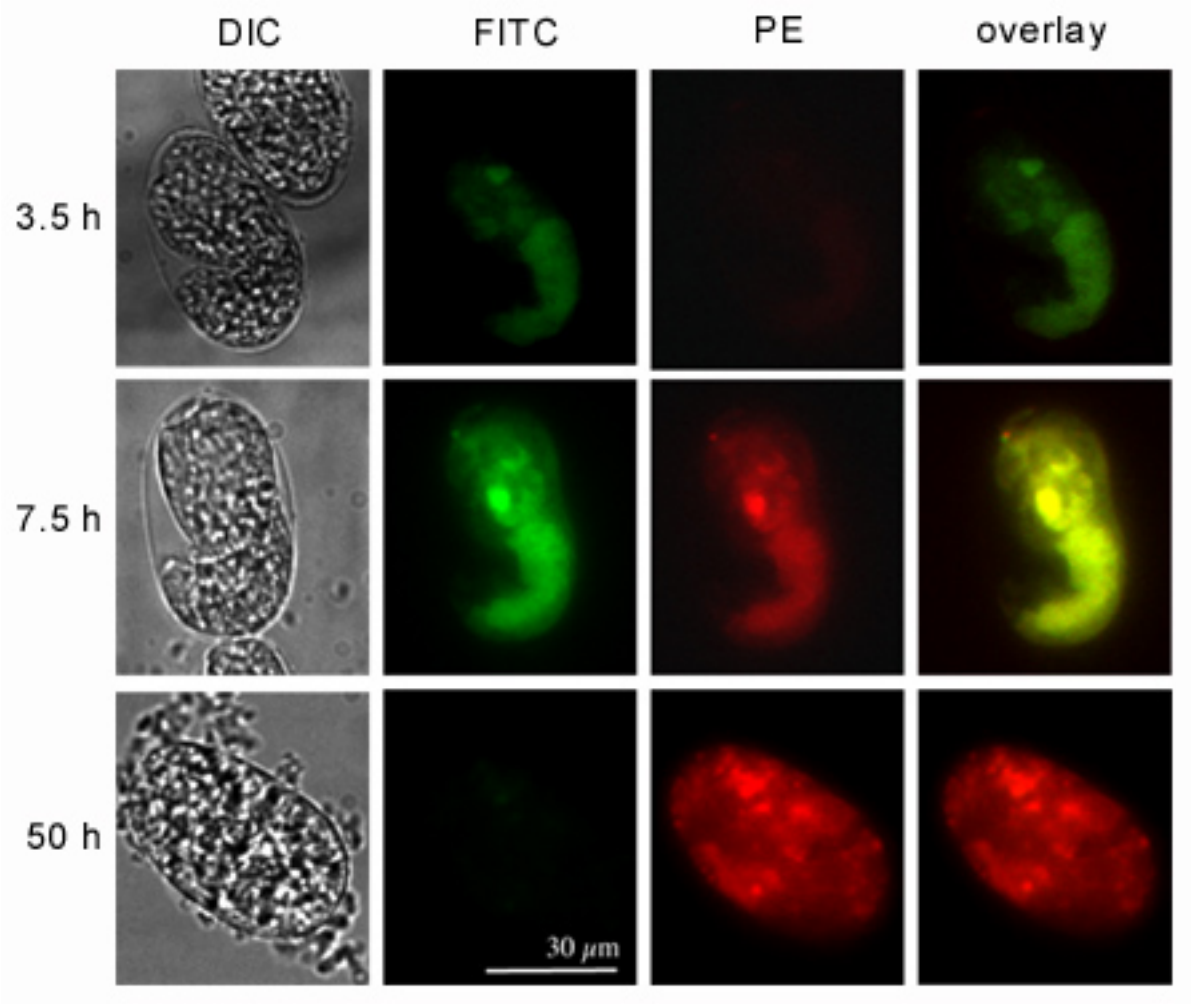
Story contributed by Raluca Gagescu, Nature Reviews Molecular Cell Biology

VAST search for structures similar to E5.

Created: January 22, 2001

Click on the link below to start an html tutorial.

Search for structures similar to E5



The change in fluorescence of E5 over time in *C. elegans*.
The *E5* mutant was placed under the control of a heat shock promoter and injected into *C. elegans* embryos. Green fluoresence was detected 2 hours into the recovery phase following a standard heat shock treatment (1 hour incubation at 33°). The embryos were documented under bright field (DIC), with a FITC filter, with a PE filter, and with an overlay at 3.5, 7.5, and 50 hours following heat shock. Yellow fluorescence, as seen in the overlay column at 7.5 hours, indicates a combination of green and red fluorescence.

Cytosolic help for mitochondrial defects

a novel method for importing tRNA into mitochondria in order to suppress mutations

Created: December 4, 2000.

The mitochondrion has cut back its genome substantially since taking up residence in cells as a symbiont 1.5 billion years ago, but it retains its personal transcription, translation and protein-assembling systems, including its tRNA genes. Even so, the mitochondrion is not fully self-sufficient — to varying extents yeast, plants and protozoan cells can borrow nuclear-encoded tRNA molecules to ease the task of translating transcripts of their mitochondrial genes. New data indicate that nuclear-encoded tRNAs can even be used to salvage errors in mitochondrial transcripts.

Figure 1

In the yeast *Saccharomyces cerevisiae,* only one tRNA ($^{Lys}_{CUU}$) is carried into the mitochondrion, something it can do only if charged with an amino acid, and only if aided by cytosolic import factors. Among these factors is the precursor of the mitochondrial lysyl-tRNA synthetase (pre-MSK).

Figure 2

In a recent publication, researchers altered the aminoacylation identity of $tRNA^{Lys}_{CUU}$ so that it was charged with methionine rather than lysine. Both in live yeast cells and in isolated mitochondria, the engineered tRNA could enter the mitochondrion, where the radiolabelled methionine charged on the imported tRNA was incorporated normally into mitochondrial proteins. A second, modified $tRNA^{Lys}$ version with alanine identity was also successfully used *in vivo* to suppress an *amber* (UAG) stop codon (a nonsense mutation) in the mitochondrial *COX2* gene.

Defects in mitochondrial (mt) DNA, caused by base substitutions or rearrangements in genes that encode proteins or tRNAs underlie a range of human pathologies (as discussed in the previous highlight).

Could the technique used to modify mitochondrial mutations be adapted for use in humans, given that import of nuclear-encoded tRNAs into mammalian mitochondria has never been seen? It seems so, because isolated human mitochondria imported the yeast $tRNA^{Lys}_{CUU}$ and its derivatives, provided that the human cytosolic extracts were supplemented with the yeast pre-MSK. The foreign tRNA was functional on the translational apparatus of human mitochondria, just as in yeast.

This recent innovation might be useful for replacing non-functional tRNAs or for suppressing nonsense mutations in mtDNA.

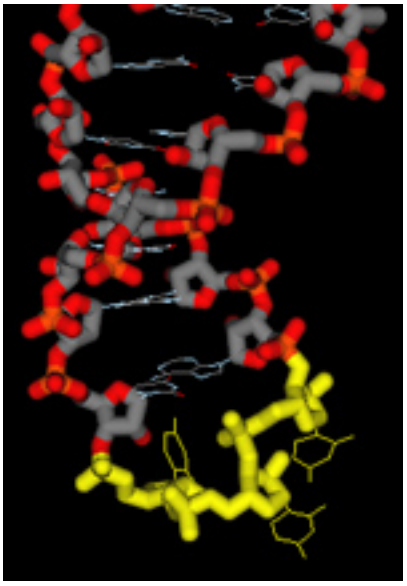
Story contributed by Tanita Casci, Nature Reviews Genetics

Search Organelle Genome Resources.

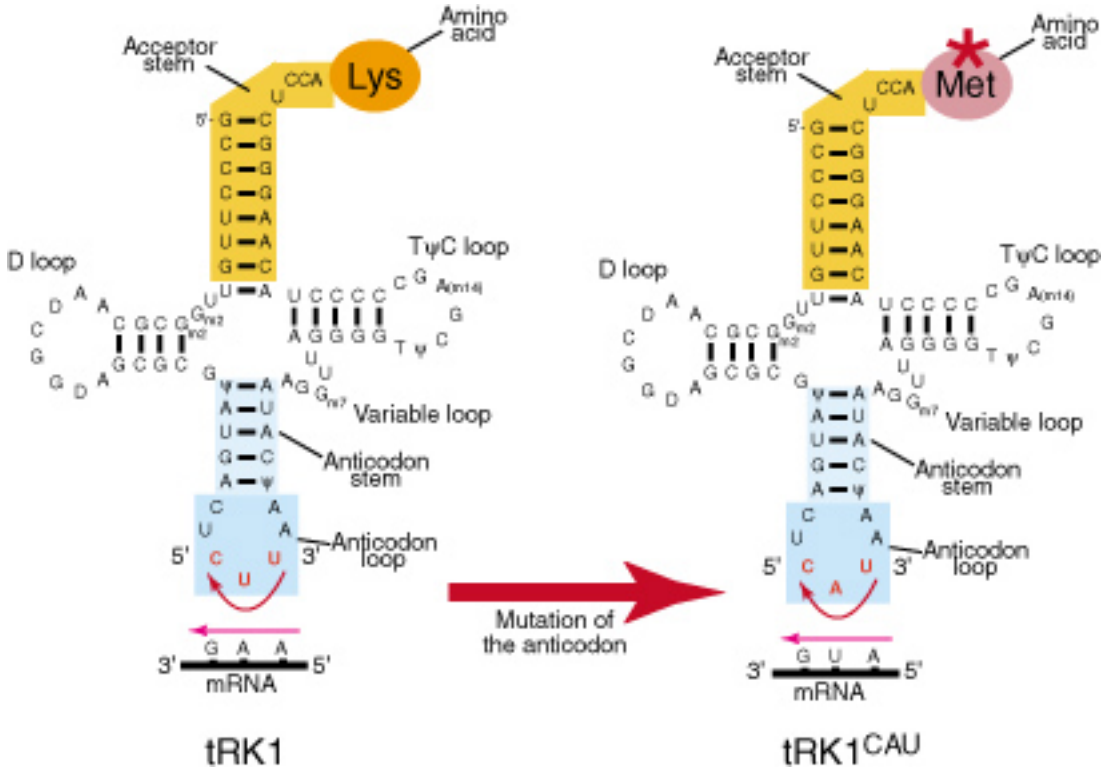
Created: December 4, 2000

Click on the link below to start an html tutorial.

Evolution and the mitochondrial gene order of tRNAs



Tertiary structure of tRNA.
The anticodon loop and stem of $tRNA^{Lys}$ is depicted in the image to the left. The three bases that compose the anticodon (in this example, "U-U-U") are highlighted in yellow.



Mutation of the anticodon of tRNA.
Outside of the mitochondria, $tRNA^{Lys}_{CUU}$ is charged with lysine, its cognate amino acid. However, by changing the anticodon from C-U-U to C-A-U, the mutated tRNA is subsequently charged with methionine. To be able to track mutant tRNAs, an ^{35}S-radiolabelled methionine was used. This experiment demonstrated that aminoacylation was necessary for transport of tRNA into the mitochondria, although the identity of the amino acid that was charged onto the tRNA was less important.

The mouse that eats less but gains weight

a neuropeptide receptor, Mc3r, is shown to play a role in regulating energy stores

Created: November 6, 2000.

Obesity contributes to poor public health in many Western populations, underlying illnesses that range from cardiovascular disease to hypertension and stroke. Over the past years, research into animal models of obesity has teased apart some of the endocrinological pathways that mammals have evolved to regulate the body's fat content in times of feast and famine. Now, new insight into the subtle workings of these pathways comes from a paper in the September issue of *Nature Genetics*, which reports the phenotypic effects of inactivating the gene that encodes the mouse melanocortin receptor 3 (Mc3r).

Mice that possess a defective melanocortin 3-receptor (*Mc3r*$^{-/-}$) are unable to respond to the neuropeptide α-MSH. Weight gain in *Mc3r*$^{-/-}$ mice is apparent after 26 weeks of age. Littermates that express functional Mc3r do not exhibit increased adiposity, even though the level of food intake is equivalent between the two groups.

Mc3r functions in a feedback loop that lies downstream of leptin, an adipocyte-derived hormone that circulates in the blood in proportion to body adiposity. In the brain, leptin elicits neuropeptide responses that stabilize the body's fat content by decreasing food intake and increasing energy expenditure. One such neuropeptide is α-melanocyte stimulating hormone (α-Msh), which acts on several melanocortin receptors, including Mc3r and Mc4r. Until now, the relative importance of each receptor in this feedback loop has been unknown, but the study by Chen *et al.* shows that inactivating Mc3r has different effects on food intake and adiposity to inactivating Mc4r.

Mc3r$^{-/-}$ mice appear to grow normally up to 26 weeks of age, but although at this age they are not overtly obese, their fat mass is almost double that of wild-type and heterozygote littermates. This is because their increased fat mass is initially obscured by a compensatory decrease in lean muscle mass. It is only after 26 weeks that their increased weight gain becomes more obvious (see picture). Unexpectedly, this increased adiposity is not caused by increased food intake. Instead, *Mc3r*$^{-/-}$ mice gain more fat per calorie of food consumed, apparently at the

expense of their lean body mass. This so-called increased feed efficiency means that the mutant mice store more fat despite eating less than normal mice do, and they become obese if fed a high-fat diet. The mechanism behind these responses is unclear because the *Mc3r*$^{-/-}$ mice have normal metabolic rates, body temperatures, and thyroid function. However, they are less active than wild-type mice, which might contribute to their tendency to obesity, and they also show a transient reduction in neuropeptide Y levels. Because this hypothalamic neuropeptide has been implicated in feeding-control mechanisms, Chen *et al.* suggest that its reduction in *Mc3r*$^{-/-}$ mice may contribute to their reduced food intake.

So how do *Mc3r*$^{-/-}$ mice differ from *Mc4r*$^{-/-}$ mice, and what does this tell us about the different functions of the two receptors in the control of food intake and energy expenditure? *Mc4r*$^{-/-}$ mice eat more than normal mice and are obese. They also have altered metabolic rates and a normal lean body mass. When Chen *et al.* treated *Mc3r*$^{-/-}$ mice with a non-selective melanocortin agonist, it reduced food consumption in both mutant and normal mice to a similar degree, indicating that α-Msh probably inhibits food intake by acting through Mc4r. Further evidence supporting distinct functions for these two receptors came when Chen and colleagues crossed the two knock-out mice to produce double homozygote mutants, which were more obese than mice lacking just Mc4r. In an accompanying News and Views article, David Cummings and Michael Schwartz speculate that this phenotype occurs because the double mutants eat excessively, because of the loss of Mc4r signalling, and store consumed calories more efficiently, because of the absence of both receptors.

These new insights into the functions of Mc3r could contribute to the development of new diagnostic and therapeutic approaches to treating obesity disorders in humans, and further research should clarify whether drugs that act through Mc3r and Mc4r could be used therapeutically to reduce food intake and its storage as fat.

Story contributed by Jane Alfred, Nature Reviews Genetics

Search for proteins similar to leptin.

Created: November 6, 2000

Click on the link below to start an html tutorial.

Search for proteins with tertiary structure similar to leptin

Tuberous sclerosis complex in flies too?

a fly homolog to TSC2, called gigas, plays a role in cell cycle regulation

Created: July 27, 2000.

Tuberous sclerosis complex (TSC) affects as many as 1 in 6000 newborns. Although named in 1880 for the firm, potato-like nodules that form in the cerebral cortex, few organs escape the benign tumors called hamartomas that also sprout in the kidneys, lungs, heart, eyes, and skin. The first clinical signs of TSC include seizures, mental retardation, and skin lesions. However, symptoms can be so subtle that they go undetected for many years.

One-third of TSC patients inherit a defective copy of either the *TSC1* gene (chromosome 9q34) or the *TSC2* gene (chromosome 16p13.3). The remaining two-thirds acquire the disease by spontaneous mutation of either gene, probably very early in embryonic development. Because no individual lacking functional copies of both genes has ever been found, it is believed that such a condition is lethal to the embryo.

The product of the *TSC1* gene, hamartin, shows no overall sequence similarity to any known protein, although it contains an extensive coiled-coil region near its carboxyl end. Coiled-coil domains often provide a binding surface for protein-protein interactions. The TSC2 product, tuberin, is also novel but shows homology to the GTPase-activating protein (GAP) for Rap1 protein, a Ras superfamily member. Rap1's cellular function is not known, however, many Ras-related proteins help to pass stimulatory signals, from the plasma membrane to the nucleus, that tell a cell when to divide. The GTP-dependent proteins within these signaling cascades can only transmit a signal when in a GTP-bound state. The role of the GAP is to inactivate the signaling protein by hydrolyzing its bound GTP to GDP. GAPs are therefore critical negative regulators of the Ras-like proteins and thus prevent cells from dividing unchecked.

The discovery last year of *TSC1* and *TSC2* homologs in *Drosophila* strengthens the idea that they are fundamental to cell division control. Mutant flies lacking tuberin possess enlarged cells that contain on average ten times the normal amount of DNA, indicating that the cells repeatedly replicated their DNA without going through intervening cell divisions. A similar process may explain the giant, multinucleated cells seen in certain neuronal tumors of TSC patients. In addition, experiments in human cells have shown that when tuberin protein levels are reduced, the cells prematurely enter the cell division cycle.

What about hamartin? It may be telling that TSC1 patients are clinically indistinguishable from TSC2 patients. One idea is that tuberin and hamartin bind to each other in cells, facilitated by hamartin's coiled-coil domain. Perhaps hamartin is there to keep tuberin fully active, or the tuberin-hamartin complex itself is the functional tumor suppressor. Although the precise cellular actions of tuberin and hamartin remain to be discovered, the strongly homologous proteins found in *Drosophila* offer the potential for further experimental investigation.

Search PubMed for the outcomes of surgery.

Created: July 27, 2000

Click on the link below to start an html tutorial.

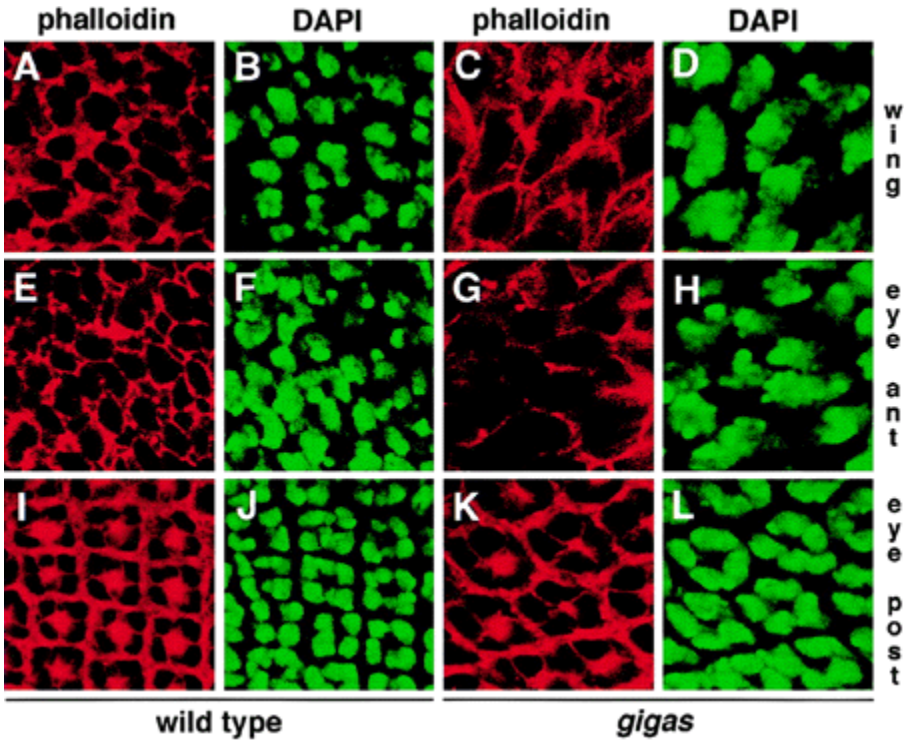
Outcomes of surgery for TSC

Use BLAST to search for proteins similar to human TSC2.

Created: July 27, 2000

Click on the link below to start an html tutorial.

Search for proteins similar to human TSC2



In *Drosophila*, the *gigas* gene codes for a protein homologous to human *TSC2*.
Fly cells from the wing and eye imaginal discs that contain mutant *gigas* are enlarged, suggesting that the cells failed to undergo cell division at the appropriate time. The figure shows confocal images of DNA-stained (green; DAPI) and phalloidin-stained (red — phalloidin is a toxin that tightly binds actin) wild-type (left-most two columns) and mutant (right-most two columns) cells. **(A-D)** are wing disc cells, **(E-H)** are cells from the eye disc anterior to the morphogenetic furrow (MF), while **(I-L)** are from the eye disc posterior to the MF. The red actin stain shows that the size of the whole cell increases in mutant cells, while the green DNA stain shows that the size of the nucleus also increases in mutant cells.

The beginning of the END

the EAST protein assembles a nucleoskeleton between chromosomes

Created: June 16, 2000.

What do traditional office skills like filing and book-keeping have in common with cell biology? The answer lies in the organization of essential components — be they documents or proteins — into distinct compartments. Take the nucleus, for example, where the proteins involved in processes such as transcription and RNA metabolism are physically clustered, and chromosomes are partitioned into discrete territories.

Reporting in *Nature Cell Biology*, Martin Wasser and William Chia from the National University of Singapore describe a molecular secretary that may organize this nuclear filing system in the fruit fly Drosophila melanogaster. Known as EAST (for 'enhanced adult sensory threshold'), it contains 12 potential proteolytic sites as well as seven putative nuclear-localization signals. These characteristics led the authors to propose that EAST is an unstable protein targeted specifically to the nucleus.

To test this prediction, Wasser and Chia studied the expression pattern of EAST in giant nuclei from Drosophila salivary glands. Confocal microscopy captured images of EAST in a region that the authors refer to as the extrachromosomal nuclear domain (END), which, as its name suggests, is the area around and between the chromosomes. Similar compartments have previously been detected in other organisms, where they are thought to belong to a putative nuclear endoskeleton.

What might be the function of EAST within this 'nucleoskeleton'? The authors reasoned that, if it is involved in forming a structural backbone, EAST may recruit other factors to the END. So they studied the effect of disrupting the east gene on distribution of a protein called CP60 that normally co-localizes with EAST. Consistent with their ideas, the CP60 expression pattern was destroyed. What's more, Wasser and Chia found overlapping expression patterns between EAST and nuclear actin — a satisfying discovery given that, outside the nucleus, actin is a core structural component of the cytoskeleton.

Overexpression of EAST gave equally striking results - an expansion of the END. This effect, which can also be induced by heat-shocking Drosophila, results from accumulation of the extra EAST. But as the figure shows, the consequences of expansion vary depending where the END expands: if EAST accumulates mainly between the chromosomes (left), the effect will be to increase the spacing between them; expansion between the chromosomes and nuclear membrane (right), however, will compress the chromosomes. In either case, this could prevent the random collision of neighbouring chromosome arms when cells are under stress.

Wasser and Chia have shown, then, that EAST is a nuclear architect involved in the assembly of an expandable nucleoskeleton between chromosomes. The next steps will be to work out how increased levels of EAST cause the END to expand, and to take a closer look at how the EAST protein is regulated.

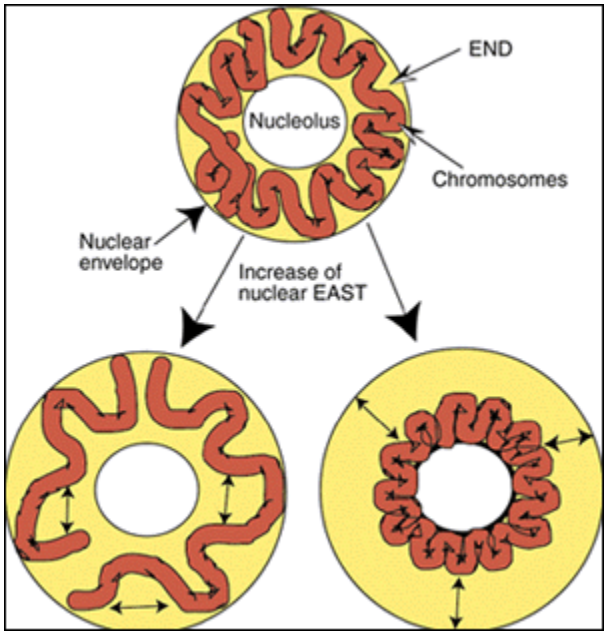
Story contributed by Alison Mitchell, Nature Reviews Molecular Cell Biology

Search the Drosophila genome for CP60.

Created: June 16, 2000

Click on the link below to start an html tutorial.

Find CP60 in the Drosophila genome



The EAST protein from *Drosophila* may be involved in the assembly of a nucleoskeleton.
When overexpressed, EAST expands the extrachromosomal nuclear domain (END) - the space between chromosomes. The consequences of expansion vary depending where the END expands: if EAST accumulates mainly between the chromosomes (left), the effect will be to increase the spacing between them; expansion between the chromosomes and nuclear membrane (right), however, will compress the chromosomes. In either case, this could prevent the random collision of neighbouring chromosome arms when cells are under stress.

Mutations and blood clots

how point mutations in clotting factor genes conspire to increase the risk of thrombosis

Laura Dean, MD

Created: April 26, 2000.

Clotting is essential, yet can be fatal. Pathological activation of the clotting cascade can lead to the formation of a blood clot, typically a deep vein thrombosis (DVT) in the legs. This blood clot may then be carried in the bloodstream to the lungs. This is known as a pulmonary embolism and is a medical emergency, being one of the leading causes of sudden death.

After trauma, the formation of a thrombus is essential to stem bleeding. A cascade of pro-enzymes, enzymes, and cofactors interact with damaged vessel endothelium to converge on a common pathway with the formation of a fibrin clot. The clot acts as a mechanical plug to prevent bleeding and is vital for normal vascular function. Disturbance of this pathway can be deadly; too little clotting results in bleeding disorders such as hemophilia, whereas excessive clotting produces blood clots that can block the lungs.

There are many factors that lead to an excessive propensity to clot, or thrombophilia. These can be classified by: (1) changes in blood vessel wall (2) changes in blood flow and (3) changes in blood constituents. Among the genetic components that underlie problems with blood constituents are mutations of clotting factor genes. These cause a deficiency of the body's natural anticoagulants, such as protein C, protein S, or antithrombin III (see figure). However, the most common inherited mutation that predisposes to thrombosis is the factor V Leiden mutation.

Factor V acts toward the end of the clotting cascade, where it is a co–factor for the Xa-dependent proteolytic cleavage of prothrombin to thrombin. Thrombin then catalyzes the conversion of soluble fibrinogen to a solid fibrin clot. Activated factor V (Va) is kept in check by a serine protease called activated protein C (APC). APC stops factor V from working by cleaving sites on its heavy chain; in particular at the sites Arg 506 and Arg 306. Thus, APC is important in limiting clot formation.

Factor V Leiden is a single point mutation resulting in an amino acid substitution of arginine for glutamine at Arg 506. The mutation affects factor V's APC-binding site, therefore preventing factor V inactivation. Carriers of this APC-resistant factor V suffer from a propensity to inappropriate clot formation.

What if you are a carrier of factor V Leiden? It is a common mutation, with a prevalence of 2% in Caucasian populations. It is especially found in patients with DVTs and increases the risk of thrombosis during pregnancy or while taking oral contraceptives. It is also associated with an increased risk of miscarriage. Although it is the most important genetic risk factor that we know of, the overall probability of thrombosis is still low with a single mutation. However, with the co-inheritance of other clotting factor polymorphisms, such as that of prothrombin which increases levels of prothrombin in the blood, the risk of thrombosis now becomes more significant.

Further investigation of the clotting factor mutations will help explain the hereditary basis of thrombophilia. Most importantly, however, the main causes of DVT are not inherited but are acquired. Despite our genetic make-up, a healthy lifestyle is our most important weapon for keeping thrombosis at bay.

Search PubMed for factor V mutations in pregnancy.

Created: April 26, 2000

Click on the link below to start an html tutorial.

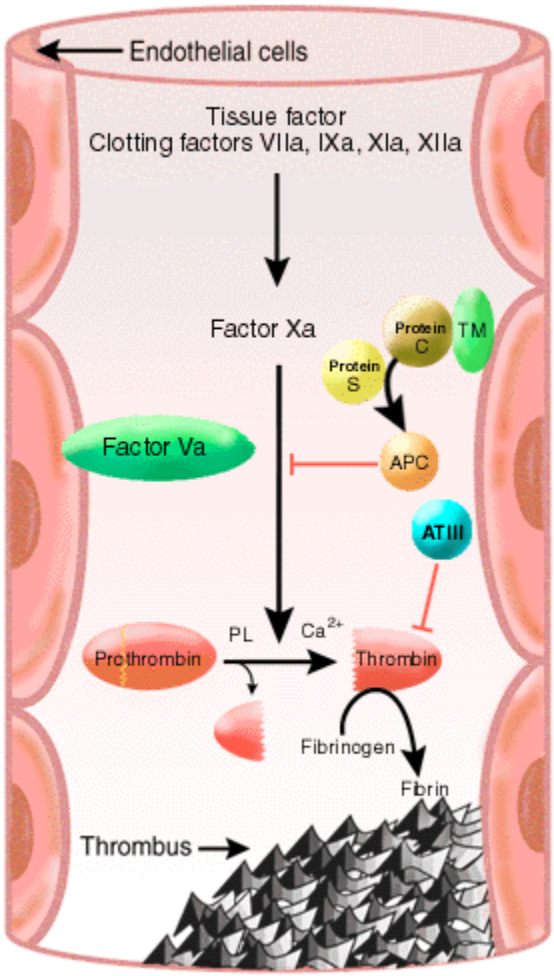
What are the consequences of factor V mutations in pregnancy?

Search UniGene for proteins similar to factor V.

Created: April 26, 2000

Click on the link below to start an html tutorial.

Find proteins similar to factor V



Coagulation cascade.
Key:

Black arrows = activation

Red arrows = inactivation

APC = activated protein C

TM = thrombomodulin, a protein bound to endothelial cell membranes to which protein C binds

PL = phospholipid

Ca^{2+} = calcium

Each reaction in the coagulation cascade involves the conversion of a clotting factor precursor into an active protease by proteolysis, regulated by cofactors and calcium. The end point is the generation of enough thrombin to catalyze the formation of fibrin, which then polymerizes and crosslinks to form a clot. Under pathological conditions, the mutation in **factor V** renders it resistant to inactivation by APC. Hence mutated factor V pushes the cascade towards excessive blood clot formation. Mutations in the upstream region of the **prothrombin** gene result in increased levels of prothrombin in the blood, again encouraging the formation of a thrombus. Protein C, protein S and antithrombin III all have anti-coagulant action. Deficiencies of proteins C and S usually result in a syndrome of recurrent venous thrombosis and pulmonary embolism. Deficiency of antithrombin III is usually mild.

Viruses provide direction on the plant information superhighway

a viral movement protein helps to identify a counterpart in plants

Created: December 8, 1999.

The phloem is the long-distance transport system of plants. In addition to distributing nutrients, the phloem plays a role in transporting hormones and signaling proteins. The phloem, together with the xylem, which transports water and dissolved ions, provide the plant with its internal trafficking system.

Phloem tubes are made up of aligned cells, called sieve elements, which lack nuclei and therefore must rely on associated companion cells for physiological support and nourishment. Sieve elements and companion cells are connected by plasmodesmata — a network of tiny channels through which molecules such as sugars, hormones and amino acids travel. How these molecules are ushered via plasmodesmata into the long-distance transport systems has eluded researchers until only recently, when viruses have provided some vital clues.

Plant viral movement proteins (VMPs) have been known to exist for some time. These are proteins encoded by the virus that assist in the transportation of the viral nucleic acid around the plant. Their exact mode of action, plus the requirement (or otherwise) for coat proteins for systemic infection can vary from virus to virus. For example, cytoplasmically replicating viruses (e.g. tomato spotted wilt virus) require one or more movement proteins and a coat protein, while bipartite geminiviruses (e.g. squash leaf curl virus) require two movement proteins, but no coat protein.

The virus that has recently shed light on the plant transportation mechanism is red clover necrotic mosaic virus (RCNMV), which uses a single virus-encoded movement protein to move between cells. This VMP binds to viral RNA, and, using host-cell microfilaments, chaperones the viral RNA to the plasmodesmata. Once at the plasmodesmata, the VMP acts to somehow increase the diameter of these channels, permitting the viral nucleic acid to enter the adjoining cell. However, in order to infect a plant systemically, the red clover necrotic mosaic virus RNA must first be encapsulated by its protein coat.

Recently, a plant protein, CmPP16, isolated from *Cucurbita maxima* (winter squash), was discovered that was reported to share sequence similarity with the red clover necrotic mosaic virus VMP. If so, this suggests that CmPP16 may assist in the long-distance transport of plant RNA, using a similar mechanism to the red clover necrotic mosaic virus VMP.

Why transport RNA around the plant? One possibility is that the RNA is being used as a long-distance signaling molecule, helping to coordinate developmental processes with physiological signals. Looking at how viruses spread throughout a plant could assist in the fascinating puzzle of how different parts of the plant talk to each other.

Search PubMed for RNA transport in plants.

Created: December 8, 1999

Click on the link below to start an html tutorial.

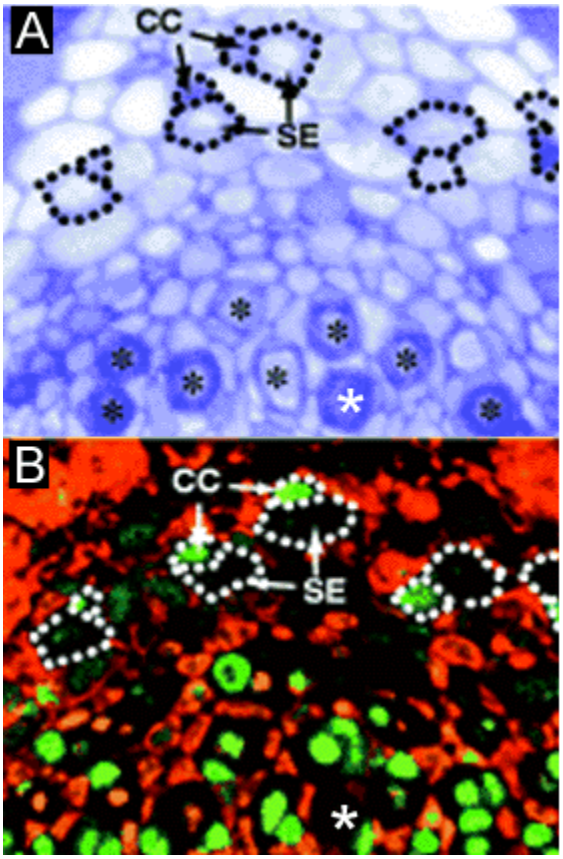
Why is RNA transported around plants?

Use BLAST to search for proteins similar to CmPP16.

Created: December 8, 1999

Click on the link below to start an html tutorial.

Search for proteins similar to CmPP16



mRNA was detected within the vascular tissue of *Cucurbita maxima* or winter squash.
(A) This image of a transverse section of winter squash depicts the various components of the phloem. Black dots outline companion cells (CC) and sieve elements (SE) joined by fine, branched plasmodesmata. Black asterisks identify immature sieve elements; the white asterisk reflects the identical cell in images (A) and (B).

(B) *CmPP16* mRNA (green fluorescent signal) is shown to have moved within the phloem. mRNA was found mostly in companion cells, but also in mature, functional sieve elements suggesting its movement through plasmodesmata.

(Reproduced with permission from: Xoconostle-Cazares, B., Xiang, Y., Ruiz-Medrano, R., Wang, H.-L., Monzer, J., Yoo, B.-C., McFarland, K.C., Franceschi, V.R. and Lucas, W.J. (1999) Plant paralog to viral movement protein that potentiates transport of mRNA into the phloem. Science 283, 94-98.)

How *Candida albicans* switches phenotype - and back again

the SIR2 silencing gene has a say in Candida's colony type

Created: November 24, 1999.

Among the many bugs that live in the mouth and digestive tract is the yeast *Candida albicans,* which under normal circumstances lives in 80% of the human population with no harmful effects. However, overgrowth results in thrush, a condition often observed in immunocompromised individuals such as HIV-positive patients. Candidiasis also occurs in the blood and in the genital tract. To infect host tissue, the usual unicellular yeast-like form of *C. albicans* reacts to environmental cues and switches into an invasive, multicellular filamentous form. This switching between two cell types is known as dimorphism.

In a process that superficially resembles dimorphism, *C. albicans* undergoes a process called "phenotypic switching", in which different cellular morphologies are generated spontaneously. One of the classically studied strains that undergoes phenotypic switching is WO-1, which consists of two phases, one that grows as smooth white colonies and one that is rod-like and grows as flat gray colonies. The other strain known to undergo switching is 3153A; this strain produces at least seven different colony morphologies. In both the WO-1 and 3153A strains, the different phases convert spontaneously to the other(s) at a low frequency. The switching is reversible, and colony type can be inherited from one generation to another. Although several genes that are expressed differently in different colony morphologies have been identified, some recent efforts have focused on what might be controling these changes. Further, whether there is a potential molecular link between dimorphism and phenotypic switching is a tantalizing question.

In the 3153A strain, a gene called *SIR2* (for silent information regulator) has been found that seems to be important for phenotypic switching. *SIR2* was originally found in *Saccharomyces cerevisiae* (brewer's yeast), where it is involved in chromosomal silencing, a form of transcriptional regulation in which regions of the genome are reversibly inactivated by changes in chromatin structure (chromatin is the complex of DNA and proteins that make chromosomes). In yeast, genes involved in the control of mating type are found in these silent regions, and SIR2 represses their expression by maintaining a silent-competent chromatin structure in this region. The discovery of a *C. albicans SIR2* that is implicated in phenotypic switching suggests that it too has silent regions controlled by *SIR2* in which the phenotype-specific genes may perhaps reside.

Another potential regulatory molecule is Efg1p, a transcription factor found in the WO-1 strain that regulates dimorphism, and more recently has been suggested to help regulate phenotypic switching. Efg1p is expressed only in the white and not in the gray cell type, and overexpression of Efg1p in the gray form causes a rapid conversion to the white form.

Thus far there are few data that dimorphism and phenotypic switching use common molecular components. However, it is not inconceivable that phenotypic switching may occur in response to some change in the environment as well as being a spontaneous event. How SIR2 itself is regulated in *S. cerevisiae* may yet provide clues as to the switching mechanisms of *C. albicans.*

Search PubMed for gene silencing in humans.

Created: November 24, 1999

Click on the link below to start an html tutorial.

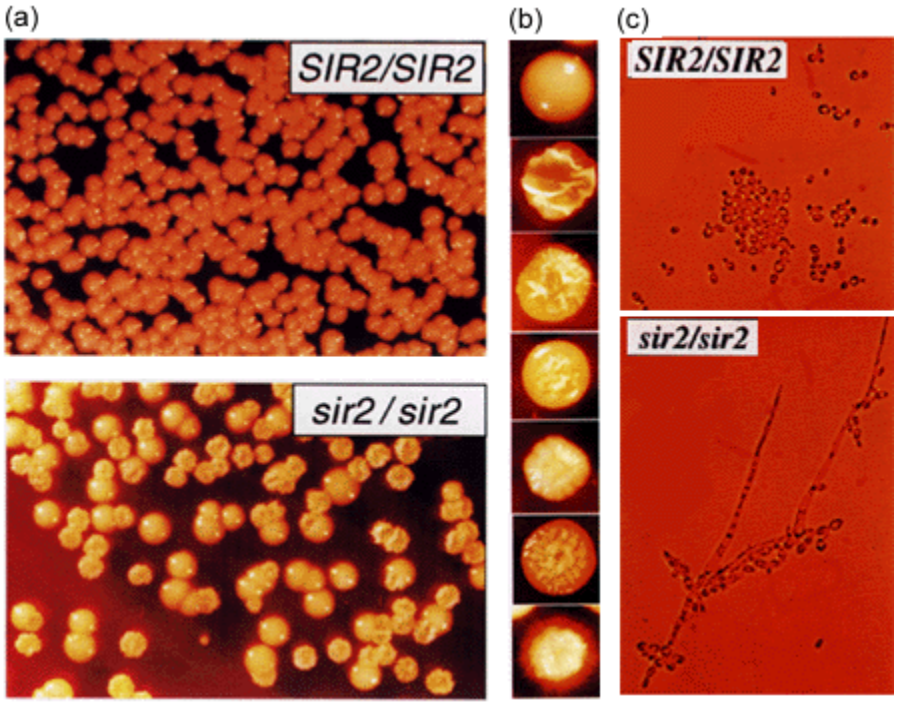
Does gene silencing happen in humans?

Use BLAST to search for proteins similar to Candida SIR2.

Created: November 24, 1999

Click on the link below to start an html tutorial.

Search for proteins similar to Candida SIR2



Phenotypic switching in the yeast *Candida albicans* can be controlled by a *SIR2* gene.
(a) Yeast colonies that are homozygous positive for the *SIR2* gene *(SIR2/SIR2)* have a uniform colony morphology, while those that lack completely *SIR2* (homozygous negative - *sir2/sir2)* have variant colony morphologies.

(b) Several different colony types are possible for cells of the *sir2/sir2* phenotype. Many of the seven colony types shown here can arise from a single colony of one of the other morphologies, indicating that the phenotypic switching is heritable.

(c) Filamentous growth is enhanced for *sir2/sir2* cells over*SIR2/SIR2,* which suggests that there may be common ground between the mechanisms of phenotypic switching and dimorphism in *Candida albicans.*

PTEN and the tumor suppressor balancing act

PTEN turns out to be the first tumor suppressor to have phosphatase activity

Created: November 10, 1999.

Tumors are formed by an abnormal proliferation of undifferentiated cells. At the molecular level, this represents a failure to adequately control cell growth and division.

In normal cells, there are many genes that code for regulatory proteins, which are responsible for maintaining the delicate balance required for cell division to proceed at the right time and in the right place. Among these, proto-oncogenes stimulate the cell division cycle, whereas tumor suppressor genes act as brakes. When these types of genes fail to do their jobs, perhaps as a result of a mutation, the control mechanisms break down and cancerous growth can ensue.

Because many proto-oncogenes are kinases (enzymes that have a stimulatory effect in cell signaling pathways), the existence of a tumor suppressor gene that acts as a phosphatase (an enzyme that counteracts the action of kinases) was predicted. However, it was almost 10 years after the discovery of the retinoblastoma gene, the first tumor suppressor to be described, that a gene product answering to this phosphatase description was found.

The *PTEN* gene, located on chromosome 10q23, is missing or mutated in a variety of human cancers, including glioblastoma (a type of brain cancer), endometrial (uterine) tumors, and prostate cancer, as well as in Cowden disease cells. PTEN stands for "phosphatase and tensin homolog". As well as having phosphatase activity, PTEN is similar to tensin, a protein that interacts with actin filaments at sites of intense signaling activity on the inner surface of cells known as focal adhesions.

PTEN taken from tumors often has a disabling mutation in the phosphatase domain, showing that it is important for normal PTEN function. But what might PTEN's substrate be? Although PTEN can act on both proteins and lipids *in vitro*, its favorite physiological substrate appears to be phosphatidylinositol 3,4,5-trisphosphate (PIP3), a lipid signaling molecule.

PIP3 is generated by the enzyme phosphoinositide kinase (PI3-kinase), which itself is activated by stimulatory signals emanating from the cell surface, often from focal adhesions. PIP3 activates yet another kinase called PKB/Akt, a proto-oncogene product. If PTEN fails to deactivate PIP3 because of a mutation in its phosphatase domain, downstream signals are not switched off, and therefore PKB/Akt remains in the "on" state. In this case, PKB/Akt can continue to stimulate downstream proteins such as transcription factors and glucose transporters, which could lead to enhanced cell growth.

PTEN is not exclusively a human protein. In the worm *Caenorhabditis elegans,* a PTEN homolog seems to help control lifespan and dauer formation (a hibernation state). Use of such animal models will help further investigate PTEN and could give clues to outstanding questions, such as how PTEN itself is regulated.

Search PubMed for the function of tensin.

Created: November 10, 1999

Click on the link below to start an html tutorial.

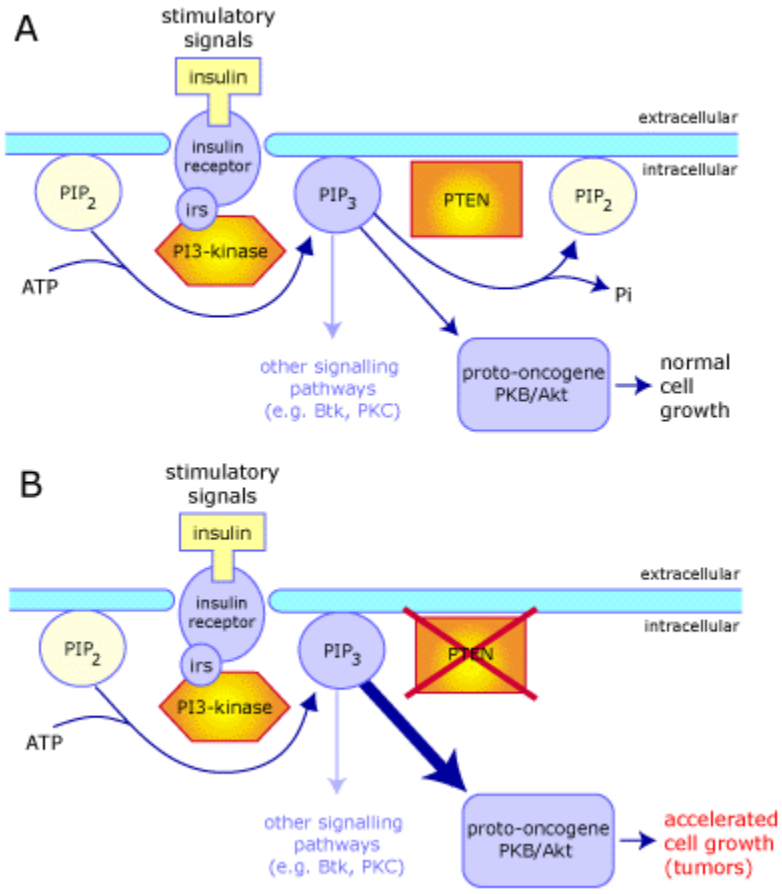
What does tensin do?

Use BLAST to search for proteins similar to PTEN.

Created: November 10, 1999

Click on the link below to start an html tutorial.

Search for proteins similar to PTEN



The PTEN tumor suppressor gene acts as a phospholipid phosphatase.
(A) Under normal growth conditions, stimulatory signals from the insulin receptor activate the enzyme phosphoinositide kinase (PI3-kinase), which phosphorylates phosphatidylinositol 4,5-bisphosphate (PIP2) to generate phosphatidylinositol 3,4,5-trisphosphate (PIP3), a lipid signaling molecule. Downstream, PIP3 activates several effectors, including the proto-oncogene product PKB/Akt. The role of PTEN is to dephosphorylate PIP3, acting as a negative control on PKB/Akt activation.

(B) If a mutation in PTEN renders it unable to carry out its phosphatase function, PIP3 can no longer be deactivated, so continues to propagate its signal downstream. This may result in the continued activation of PKB/Akt , which, in combination with other factors, could lead to increased cell growth and possible tumor development.

The *Salmonella* battle plan

how Salmonella gain entry into human intestinal cells to grow and divide

Created: October 27, 1999.

$2.3 billion a year — that's the cost of *Salmonella*-related illness to the US economy, according to new data from the US Centers for Disease Control. The figure represents 1.4 million reported salmonellosis cases, including 600 deaths. Although the numbers are down from previous estimates, the human toll and dollar cost remain significant.

So how do people get sick from *Salmonella* infection? After eating contaminated food, often egg-based products, *Salmonella* gain access to the human intestine, where there is a low concentration of oxygen. This switches on a number of *Salmonella* genes that equip the bacteria with the proteins required to invade the cells on the surface of the intestine. A cocktail of these proteins is secreted and then translocated into the host epidermis cell. This stimulates the formation of a hollow in the host cell membrane, which grows deeper and progressively encloses the bacteria (a process known as pinocytosis). Eventually, the membrane pinches off to form a vacuole in which the bacterium resides. From this sheltered environment inside the intestine cell, the bacteria grow and divide, before spreading to other cells. When this happens, usually about 2 or 3 days after ingestion, the carrier experiences the symptoms of diarrhea and dehydration that are commonly associated with food poisoning.

All of these rearrangements in the shape of the cell membrane mean that actin, the scaffolding of the cell, must undergo significant organizational change. Some of the secreted *Salmonella* proteins, whose job is to manipulate the host's actin, have now been identified, for example, the 'Sips' (for *Salmonella* invasion proteins) — SipA, SipB, SipC, and SipD. Of these, SipA has recently been implicated in organizing actin filaments to facilitate *Salmonella* entry by binding directly to actin itself or by binding to T-plastin, an actin-bundling protein. SipA lowers the concentration of actin monomers required for polymerization, allowing long filaments to be formed that extend the cell, and stabilizes those filaments so they can not break down again into actin monomers. But in reality, for the cell extensions to flex around the bacterium, actin filaments would need to break down and reform in an organized fashion. This suggests that there may be a regulatory mechanism that controls SipA action.

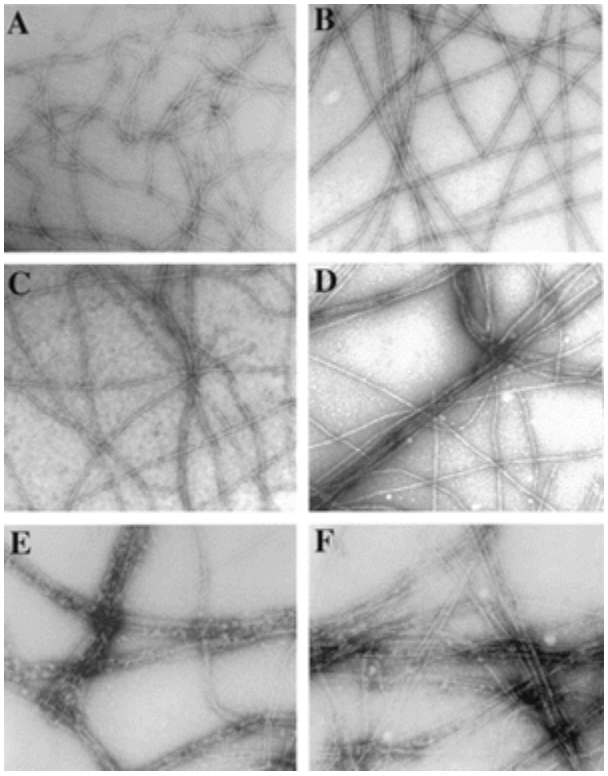
Shigella, which are found in the same family of bacteria as *Salmonella,* also invade intestine epithelial cells on ingestion and causes diarrhea, fever, nausea, and vomiting when they multiply and spread. Similar to *Salmonella, Shigella* secrete a collection of proteins that rearrange the actin cytoskeleton of the host. Among these, the proteins IpaA-D bear more than a passing resemblance to the Sips of *Salmonella,* suggesting a shared mechanism of entry between the two species. The further investigation of both bacterial invasion mechanisms may shed light on the evolutionary origins of this key virulence step, as well as provide targets for potential new therapies.

Use BLAST to search for relatives of *Salmonella* SipA.

Created: October 27, 1999

Click on the link below to start an html tutorial.

Find relatives of *Salmonella* SipA



***Salmonella* SipA interacts with T-plastin to accelerate the bundling of actin filaments.**
The figure shows an electron micrograph of the effect of *Salmonella typhinurium* SipA on the bundling activity of T-plastin.

A and B show actin with no T-plastin present; **C and D** show actin with low levels of T-plastin (0.125 micromolar), and **E and F** have a high level of T-plastin present (0.5 micromolar).

A, C and E have no SipA present, whereas **B, D and F** have SipA present as 2.0 micromoles.

The effect of SipA on actin bundling is especially clear in the pair **C and D**, where the concentration of T-plastin alone (**C**) is not sufficient to induce actin bundling. On addition of SipA (**D**), bundling occurs.

(Figure reproduced from Zhou, D., Mooseker, M.S. and Galan, J.E. (1999) An invasion-associated *Salmonella* protein modulates the actin-bundling activity of plastin *Proc. Natl Acad. Sci. USA* 96, 10176-10181 [PubMed]).

RNA surveillance: watching the defectives

detecting premature stop codons in mRNA halts the production of dangerous truncated proteins

Created: October 13, 1999.

Only those changes in DNA sequence that have functional consequences are known as disease-causing mutations. One such frequently occurring mutation causes a premature stop codon to appear in the middle of a protein-coding sequence of messenger RNA (mRNA). Stop codons (a triplet of nucleotides: UAA, UAG, or UGA) normally signal the end of the stretch of mRNA that is translated into protein so that when one appears early, the result can be a truncated protein that could have nasty consequences for the host organism.

However, a mechanism known as "nonsense-mediated decay" has evolved to detect these harmful RNAs, and sequence analysis suggests that it may have been conserved in eukaryotic organisms, including humans. In yeast, three proteins have been identified that are required to seek and destroy the partly translated RNAs: Upf1p, Upf2p, and Upf3p.

Upf1p is an RNA unwinding enzyme, a helicase, that requires ATP for activity. Unfortunately, Upf1p will unwind pretty much anything, not just the problem mRNAs. So Upf2p and Upf3p are thought to be required to help Upf1p discriminate between nonsense and "real" mRNAs.

How do the core proteins work in synergy to trigger nonsense-mediated decay? One possibility is that Upf3p, along with several other ribonuclear proteins, may first bind to an mRNA as it is being exported from the nucleus en route to the ribosome, the site of protein synthesis. If the mRNA is fully translated into protein, Upf3p and the other protein factors are displaced. However, if there is a premature stop codon, Upf3p and cohorts may sit tight and mark the mutant mRNA as one that needs to be disposed of.

Experiments have shown that Upf3p can bind Upf2p. Once bound, Upf2p could signal to the "termination complex", a mixed bag of termination factors that includes Upf1p. This results in the release of the incomplete polypeptide from the ribosome, mRNA unwinding by Upf1p and, exposure of the mRNA for total degradation by exonuclease.

Although this model is attractive, more experiments are required to show that this actually happens in a living yeast cell.

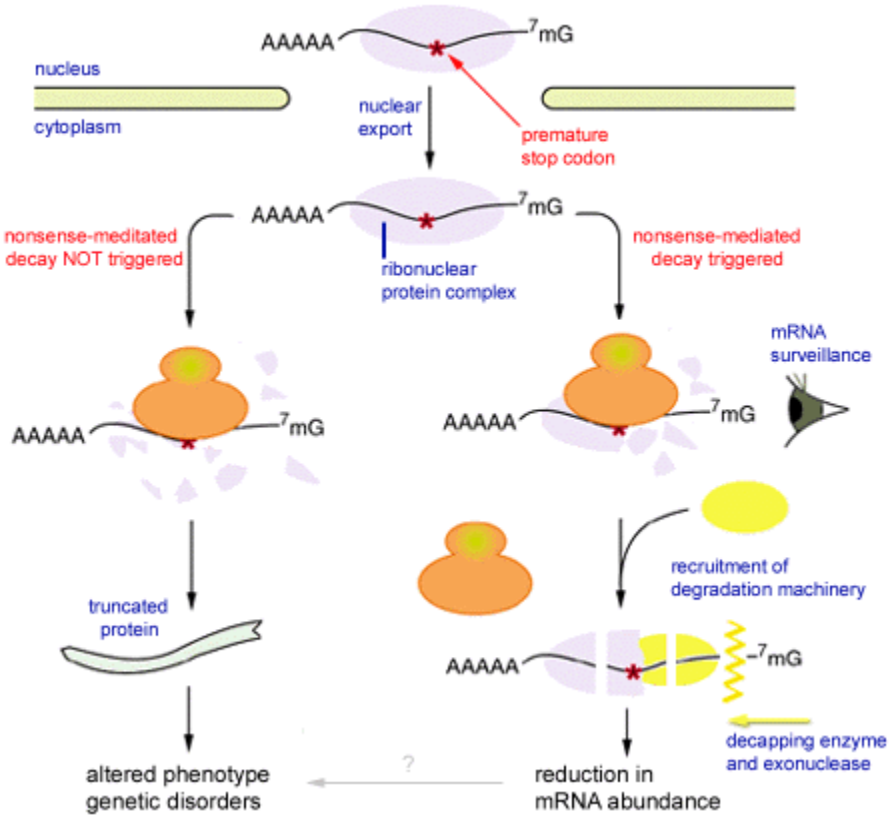
Many of the mutations that form a premature stop codon lead to human disease, for example, those in *BRCA1* that lead to breast cancer, or those in *NF1* that lead to neurofibromatosis type 1, to name just two. There are two ways by which nonsense-mediated decay can play a role in the disease process. The first occurs when the machinery is functioning correctly: if mutant mRNAs are removed, then there will be a reduction in the amount of mRNA and protein available in the cell. The second is when a mutation occurs in the nonsense-mediated decay process itself, such as a mutation in RENT1, a human homolog of Upf1p, resulting in a population of truncated proteins, which could be harmful when targeted to their site of function.

Use BLAST to search for relatives of yeast Upf1p.

Created: October 13, 1999

Click on the link below to start an html tutorial.

Find relatives of yeast Upf1p



Nonsense-mediated decay (NMD) in yeast, as a model for NMD in humans.
Ribonuclear proteins that bind to mRNAs in the nucleus remain associated with the mRNA as it becomes attached to the ribosome. When a premature stop codon is present, one of these proteins could be Upf3p. If Upf3p, or another as yet unidentified factor, is recognized by the surveillance complex (represented here by the eye), then the NMD mechanism is triggered. In yeast, this trigger may be assisted by the binding of Upf2p to Upf3p, after which the Upf1p helicase unwinds the mRNA, leaving it open for degradation by a decapping enzyme and exonuclease. Should the premature stop codon not be recognized, translation of the mRNA proceeds and results in the production of a truncated protein.

The compound eye of flies divulges evolutionary secrets

analysis of fly eye development may shed light on human eye disease

Created: September 29, 1999.

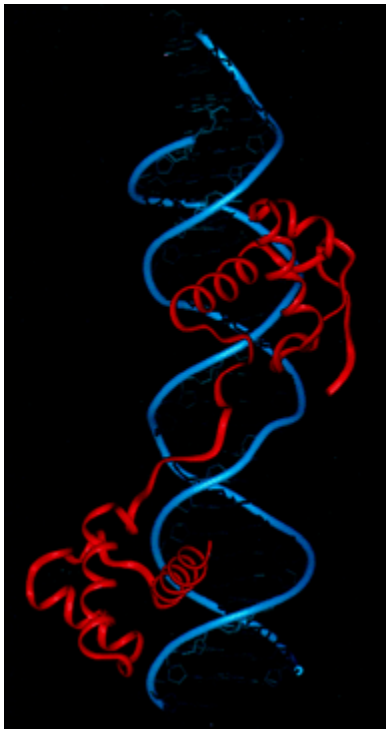
The *Pax6* group of genes belongs to a larger class of homeobox-containing genes, found in organisms from yeast to humans. They code for transcription factors and are distinguished by the presence of a specific DNA-binding motif (a homeodomain) that serves to regulate gene expression. The "helix-turn-helix" 3D structure of the homeodomain is the same structure that is seen in bacterial gene regulatory proteins, suggesting that this is an ancient conformation that has been conserved throughout billions of years of evolution.

In mice, sea squirts, and squid, *Pax6* has been shown to activate the program that leads to eye formation during the development of the organism. In mouse, where the *Pax6* gene is expressed in the developing eye and brain, a mutation called *Small eye (Sey)* results from defects in *Pax6.* This makes it a good model for studying aniridia, a condition caused by a mutation in human *Pax6* in which an incomplete iris can lead to poor vision, light sensitivity, and a tendency to develop progressive glaucoma.

Although the eyes of vertebrates have a single lens, the compound eye of *Drosophila* consists of about 750 units, each unit containing a lens, retina, and photoreceptor cells. Even so, parallels exist between these two types of eye. In flies, the eye precursor cells differentiate into these distinct units at a distinct step of development, when *ey*, a Pax6 homolog, can be detected.

Recently, a second *Drosophila Pax6* gene was reported and named *twin of eyeless (toy).* Perhaps surprisingly, *toy* in some ways is more similar to the evolutionarily distant vertebrate Pax6 proteins than to Ey. In particular, fly Toy and mouse Pax6 have a similar DNA-binding pattern; they have a much higher affinity for DNA than Ey. This can be attributed to the mutation of a single residue (Asn?Gly) in a highly conserved part of Ey, known as the paired domain.

The existence of two Pax6 genes in flies, but not in vertebrates, suggests that a gene duplication event occurred sometime during fly evolution. That Toy is more closely related to vertebrate Pax6 suggests that Toy is the more ancient form, which gave rise to Ey. The key point in the evolution in these two genes was probably when the Asn?Gly mutation occurred, radically altering the DNA-binding function of Ey. At this point Ey could have become dependent on Toy for its activation, triggering a divergence in function of the two proteins. Today, genetic evidence suggests that Toy is found upstream of Ey in the regulatory pathway of eye development, and they each regulate distinct developmental events.



The structure of the paired domain found in human *Pax6*.
Two distinct domains are found in *Pax6* — a homeodomain and a paired domain. Recent interest has focused on the paired domain, mutations in which cause several human disorders, including aniridia. This figure depicts the structure as ribbons drawn through the main carbon backbone of the protein (red) and through the phosphate atoms of the DNA backbone (blue). Mutations in the paired domain interfere with the DNA-binding properties of Pax6, altering its function, and causing a detrimental effect on the health of an individual.

(Reproduced from Xu, H.E., Rould, M.A., Xu W., Epstein, J.A., Maas, R.L. and Pabo, C.O. (1999) 'Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding' *Genes Dev.* 13, 1263-1275, with permission.)

Search PubMed for *Pax6* mutations.

Created: September 29, 1999

Click on the link below to start an html tutorial.

Find information on *Pax6* mutations

Search GeneMap99 for *Pax6*.

Created: September 29, 1999

Click on the link below to start an html tutorial.

Where is *Pax6* found in the human genome?

Use BLAST to search for proteins similar to Pax6.

Created: September 29, 1999

Click on the link below to start an html tutorial.

See how Drosophila Ey compares to other Pax6 proteins

The neighborhood of Alzheimer's amyloid precursor protein

new clues on Alzheimer's pathology from other proteins linked to amyloid plaque formation

Created: September 15, 1999.

The identification of amyloid-rich plaques has long been a diagnostic tool for pathologists investigating Alzheimer's disease (AD). The plaques are formed through the accumulation and aggregation of beta-amyloid peptides derived from the amyloid precursor protein (APP; see figure) and are characteristically found in the brain parenchyma and around blood vessels. Although it is clear that APP plays a role in AD pathology, the normal function of APP is currently unknown.

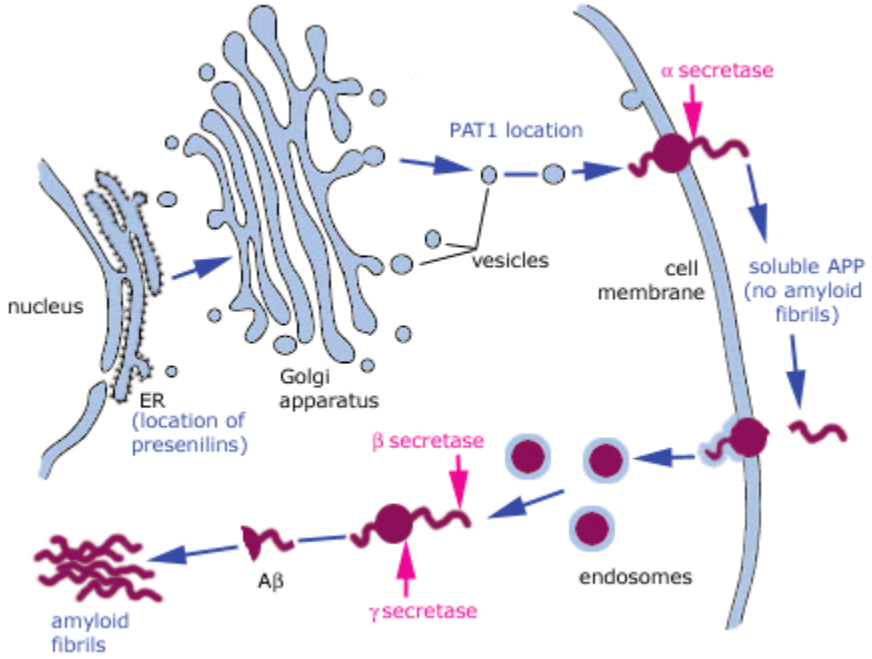
Mutations in APP itself are linked to only a small fraction of familial AD cases. However, more recently, mutations in two more genes have been linked to a much larger subset of familial AD cases, presenilin-1 (PS-1) and presenilin-2 (PS-2).

Similar to APP, the presenilins are also membrane-spanning proteins but are found mostly in the endoplasmic reticulum, whereas APP is found distinctly on the basolateral surface of cells. Mutations in the presenilins increase the production of beta-amyloid, suggesting that they influence the metabolism of APP in some way. This could be through a direct interaction but could also be a result of an indirect effect, perhaps on the trafficking of APP as it travels from its site of synthesis on the ribosome, via the endoplasmic reticulum and Golgi apparatus, to the cell surface, or as it is internalized from the cell surface for break down or recycling (see figure).

Several APP-interacting proteins have been reported recently, including the presenilins themselves and adaptor proteins, which could act as a scaffold for intracellular signalling molecules. Another protein, called PAT1, has also been found to bind to the cytoplasmic tail of APP. PAT1 appears to recognize and bind to a specific signal sequence, called the basolateral sorting sequence, which ensures that APP is transported to the basolateral membrane.

As it turns out, PAT1 bears more than a passing resemblence to kinesin light chain, found by searching the sequence database. Kinesins are motor proteins and help transport molecules around cells by connecting the cargo molecule (in this case, APP) to microtubules, which provide a network of "railroad tracks" on which to shuttle molecules around the cell. Combining this result with other experimental evidence, it seems likely that PAT1 plays a role in sorting and delivering APP to the right place in the cell.

Although the roles of APP, the presenilins, and other molecules implicated in AD still require significant investigation, the characterization of genes and proteins that are linked to amyloid plaque formation may help build a picture of the events that lead to AD. Not least, the identity of the enzymes that cleave APP and how the presenilins exert their effect on APP would provide insight into the most common neurodegenerative disease in the world.



The trafficking and metabolism of amyloid precursor protein (APP).
After synthesis on the ribosome, APP enters the endoplasmic reticulum (ER) and is transported via the Golgi apparatus to the cell membrane on the basolateral surface of the cell. The presenilins, which are thought to play a role in APP metabolism, reside mainly in the ER. PAT1, the subject of this article, may attach APP to the microtubules — the 'railroad tracks' of the cell — to assist in transporting APP to the membrane. The generation of the toxic beta-amyloid peptides requires that APP on the cell surface enters a cellular recycling program, after which two cleavages occur, one in the extracellular domain and one in the transmembrane domain. However, which enzymes carry out these cleavages, where in the cell they occur, and how the events are regulated, are not known. An alternative pathway to beta-amyloid formation involves another cleavage by an enzyme termed 'alpha-secretase', which is thought to be located near the plasma membrane. If this second pathway is triggered, beta-amyloid can not be formed, so amyloid plaques do not accumulate.

Search PubMed for APP and the presenilins.

Created: September 15, 1999

Click on the link below to start an html tutorial.

Find articles that discuss both amyloid precursor protein and the presenilins

Search PubMed for kinesin-related links to Alzheimer's disease.

Created: September 15, 1999

Click on the link below to start an html tutorial.

Are there other kinesin-related links to Alzheimer's disease?

Use BLAST to search for PAT1.

Created: September 15, 1999

Click on the link below to start an html tutorial.

What can PAT1 do for amyloid precursor protein?

What do Lyme disease and syphilis have in common?

two pathogenic spirochetes unexpectedly share an ATP synthase

Created: September 1, 1999.

Diseases spread by ticks are not recent phenomena to the New World. However, one tick-borne disease, known in Europe since the beginning of the century, has only recently emerged in the United States and is now the most common tick-borne disease on this side of the Atlantic.

In the mid-1970s, a number of residents of the small town of Lyme in Connecticut were afflicted with an unusual arthritis-like condition. Most new cases were observed in the summer and early fall of each year, and the incidence was soon linked to recent tick bites in a high percentage of patients.

The mouthparts of a deer tick, as viewed down an electron microscope. The tick is responsible for transmitting the bacterium that causes Lyme disease.

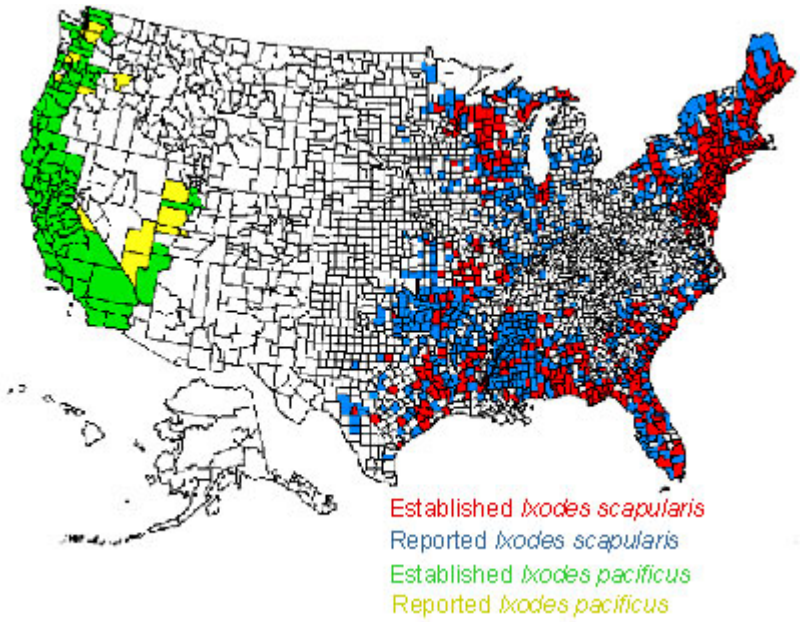
Reproduced from: The American Lyme Disease Foundation Inc.

The organism that causes Lyme disease is a bacterium, *Borrelia burgdorferi*, which was cultured from the midgut of *Ixodes* ticks in the mid-1980s. It is shaped like a wave or helix and belongs to a whole class of bacteria that can be identified by this distinctive shape, the spirochetes, which include some other nasty human pathogens that can cause syphilis, tick-borne relapsing fever, dysentery, and leptospirosis.

We now have the full sequence of the *B. burgdorferi* genome. It has a total of 853 genes on one linear chromosome and an additional 430 genes on 11 plasmids. The chromosomal genes are required for cell growth, although there are no genes present that are involved in the synthesis of amino acids, fatty acids, enzyme cofactors, or nucleotides. The genes on the plasmids may be involved in infectivity and virulence. One of the chromosomal genes codes for "outer surface protein A" (OspA), which has been the focus of a search for a vaccine against Lyme disease. Recent clinical trials using a recombinant form of OspA as a vaccine look extremely promising.

But what of the relatives of *B. burgdorferi*? Perhaps one of the most infamous of these is *Treponema pallidum*, the bacterium that causes syphilis. in contrast to Lyme disease, syphilis is not a modern affliction; it was first recognized in the 15th century in Europe. It is similar to *B. burgdorferi* in that it is a spirochete with a relatively small genome and requires a host to survive; however, at the genomic level, the two organisms are not very closely related to each other at all.

Of the 1041 protein-coding regions in *Treponema pallidum*, 476 are shared with *B. burgdorferi*, but nearly half of these are common to other bacteria too, and most have a predicted biological function. Of the genes with unknown function, there are about 50 that are conserved only in the spirochetes and are not found in any other



Distribution of Lyme disease vectors in the US.
The map shows the established and reported distribution of the Lyme disease vectors in the United States, by county, 1907-1996.

'Established' is defined as at least six ticks or two life stages (larvae, nymphs or adults) being identified; 'reported' is defined as at least one tick being identified.

bacteria. Some of these are likely to represent genes that code for spirochete-specific traits, such as their helical shape and perhaps elements of their pathogenicity.

The comparison of whole genomes may help pinpoint genes that make an organism a successful pathogen or suited to living in a particular environment. Furthermore, the identification of related genes in different species gives a new twist to the classification of different organsims, deepening our understanding of their relationship to each other across the whole evolutionary spectrum.

Search PubMed for gene sequencing of Lyme disease.

Created: September 1, 1999

Click on the link below to start an html tutorial.

How is gene sequencing helping us understand Lyme disease?

Search PubMed for Lyme disease and syphilis.

Created: September 1, 1999

Click on the link below to start an html tutorial.

Where do Lyme disease and syphilis converge?

Use BLAST to search for an ATPase found in spirochetes.

Created: September 1, 1999

Click on the link below to start an html tutorial.

An unusual way to generate energy in the spirochetes?

Dissecting the mechanism of our internal clock

how living organisms tune in to the time of day

Created: August 18, 1999.

As any jet-setter knows, it takes time to adapt to the shifted day-night cycle of a foreign time zone. We have an internal circadian clock that times many physiological and behavioral events on a 24-hour cycle, according to day length. The clock can also reset itself, so we can cope with the seasonal variation in day light hours and the trappings of 20th century living such as shift work and air travel.

Not only humans have circadian rhythms. The eyes of marine molluscs, for example, show a correlation between perception of light and a circadian rhythm, as do the pineal glands of lizards and birds. The underlying clock that gives rise to these rhythms is dependent on feedback loops that regulate the expression of certain genes. Two animals in particular have given insight into the molecular mechanisms of internal clocks: the fungus *Neurospora crassa* and the fruit fly *Drosophila melanogaster*.

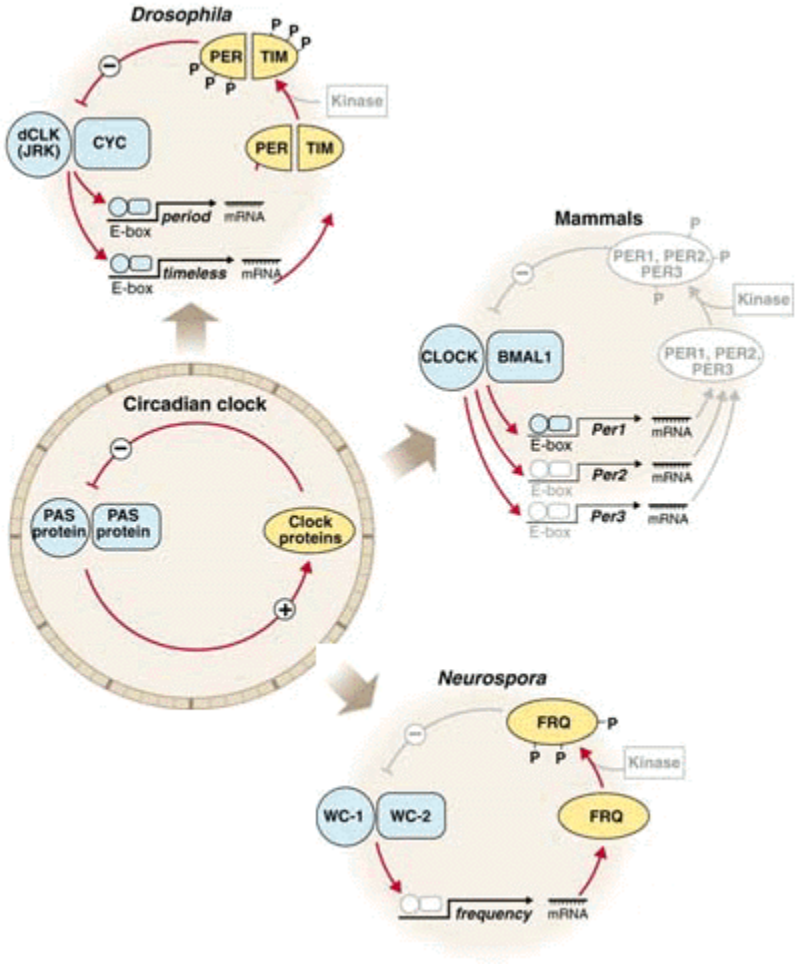
Several components of molecular clocks have now been cloned and sequenced. In *Neurospora*, the *frq* gene was the first found to be associated with period length; then two more genes, *wc-1* and *wc-2*, were discovered in a strain of *Neurospora* that was blind to light. Both *wc-1* and *wc-2* are transcription factors that contain zinc fingers and transcriptional activation domains. Furthermore, these two proteins have PAS domains.

PAS domains were first identified in the *Drosophila* period clock protein PER, the vertebrate aryl hydrocarbon receptor nuclear translocator (ARNT), which is involved in a cell's response to lowered oxygen levels, and the *Drosophila* single-minded protein (SIM1), involved in the regulation of development. Many proteins have since been found to have PAS domains, which have now been shown to mediate protein-protein interactions.

A series of recent papers have confirmed that there is a common pattern to molecular clocks that has been conserved across evolution, from fungi to mammals. Part of the pattern is that PAS domains glue proteins such as wc-1 and wc-2 together to form a complex that switches on other clock components, such as frq, as a part of the organism's response to light. The frq protein then feeds back to inhibit the action of wc-1 and wc-2, thereby ultimately effecting its own expression. Signals from the environment, such as different light levels or temperature, could impact upon the loop to add more layers of regulation.

There are certain to be more feedback loops that are linked to this core component, because several observations have been made that do not quite fit this model, and it is not yet clear whether the clocks of plants or cyanobacteria will work in the same way. Perhaps these other cogs will be specific to different organisms, with only the "master clock", outlined here, being conserved across species.

Time will tell.



Control of circadian rhythms at the molecular level.
We are just beginning to unravel the secrets of how circadian rhythms are controlled at the molecular level.

Several components of what might be the 'core' molecular clock have now been cloned and sequenced in a number of different organisms. In particular, efforts have focussed on the fungus *Neurospora crassa*, the fruit fly *Drosophila melanogaster*, and mammals such as mice and humans. Many of these components share a PAS dimerization domain and a basic helix-loop-helix DNA-binding motif.

Recent elegant work carried out in several laboratories has led to the current model, illustrated above. This proposes that proteins with PAS domains form heterodimers that bind DNA at specific sites, called E-boxes. E boxes are located in the promotor region of oscillator genes such as Drosphila per and tim. Binding of the heterodimer to the E box leads to the transcription of the oscillator genes. The proteins produced then feedback to ultimately inhibit their own production. Faded parts of the figure are speculatory

Figure reproduced, with permission, from Dunlap, J. (1998) *Science* 280, 1548-1549.

Search PubMed for day length and melatonin levels.

Created: August 18, 1999

Click on the link below to start an html tutorial.

How might day length effect melatonin levels?

Search PubMed for new clock components.

Created: August 18, 1999

Click on the link below to start an html tutorial.

What's the recent progress in identifying clock components?

Use BLAST to search for CLOCK protein.

Created: August 18, 1999

Click on the link below to start an html tutorial.

CLOCK protein from the fruit fly is one of the most recently discovered clock components.

Ubiquitin links Parkinson's disease genes

a tantalizing link between two new genes

Created: August 4, 1999.

Although we have known about Parkinson's disease for almost 200 years (it was discovered by the English physician James Parkinson in 1817), the role of genetics in the pathology of the disease has only recently emerged.

Two new genes associated with Parkinson's disease have been reported. The first, called α-synuclein, is mutated in an autosomal-dominant type of Parkinson's disease. The second, uncovered in April 1998, codes for a protein called parkin, which is associated with a juvenile autosomal-recessive form of Parkinson's disease (AR-JP).

One of the pathological features of most types of Parkinson's disease is the appearance of an inclusion body, known as a Lewy body, in many regions of the brain. The Lewy body is associated with neuronal degeneration and is also seen in Alzheimer's disease. It may be no coincidence that α-synuclein is found in Lewy bodies. In AR-JP, however, there is no Lewy body formation. In spite of these apparent differences between these two types of Parkinson's disease, there are some tantalizing links at the molecular level.

The best way to visualize Lewy bodies in brain tissue is to use antibodies against ubiquitin, because they contain high levels of ubiquitinated proteins. In cells, ubiquitin first binds to proteins displaying degradation signals, then recruits additional ubiquitin molecules to form a polyubiquitin chain. This 'flags' the protein for destruction by the 26S proteasome. In neurodegenerative disease, incompletely degraded, ubiquitinated proteins accumulate in Lewy bodies. It is intriguing that α-synuclein, which is responsible for a subset of cases of Parkinson's disease, should be found in the same place.

Enter parkin. After cloning and sequencing the suspected Parkinson's disease gene, the researchers compared their sequence to those in the database. They found that parkin contains an ubiquitin-like domain and a RING finger motif, implicated in interactions with DNA (and more recently, with other proteins). Could parkin function in a similar way to ubiquitin proteins, and could its defect in AR-JP interfere with the ubiquitin-mediated protein degradation pathway?

Although all roads (for the present) lead to a defect in ubiquitin-mediated pathways as a cause of Parkinson's and perhaps other neurodegenerative diseases, the two types of Parkinson's disease described here account for only a small fraction of the total number of Parkinson's disease cases. However, these early forays will prove extremely valuable for tracking down other genetic factors that could be involved in neurodegenerative disease.

Search PubMed for the role of alpha-synuclein.

Created: August 4, 1999

Click on the link below to start an html tutorial

What role do alpha-synuclein and parkin play in Parkinson's disease?

Use BLAST to search for parkin.

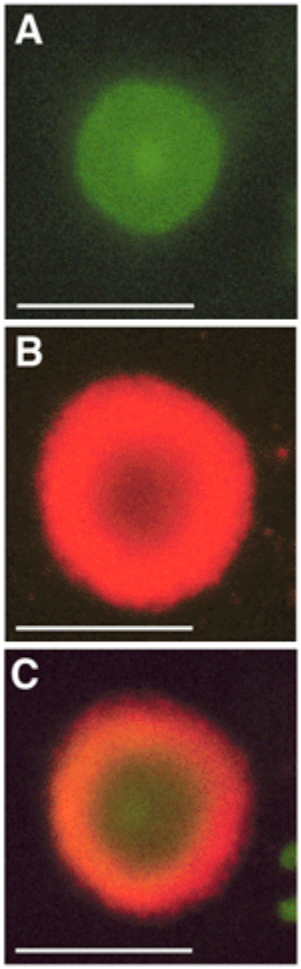
Created: August 4, 1999

Click on the link below to start an html tutorial

continues on next page...

continued from previous page.

Protein that match parkin give clues to its function



Lewy bodies visualized by staining for ubiquitin.
Lewy bodies, found in Parkinson's disease brain and other related neurodegenerative disorders, can be visualized by staining for ubiquitin.

(A) An antibody against ubiquitin stains Lewy bodies green.

(B) A red stain for alpha-synuclein.

(C) If both ubiquitin and alpha-synuclein are found in the same place, the green and red stains combine to give a yellow/orange color. This demonstrates that alpha-synuclein and ubiquitin cluster in the same place - a tantalizing parallel with parkin, which has been found to be similar to ubiquitin proteins.

The ubiquitin family of proteins are known to be involved in the pathogenesis of a number of neurodegenerative disease, including Alzheimer's disease, where they are a component of paired helical filaments.

Plant genes contribute to a sexually transmitted disease?

how plant genes found their way into a human parasite

Created: July 15, 1999.

The most frequently reported sexually transmitted disease in the United States is caused by the parasitic bacterium *Chlamydia trachomatis*. Although about half a million cases of infection are reported annually, a more realistic incidence is around 4 million cases per year. This is because there is a large pool of asymptomatic individuals within the population. If left untreated, chlamydial infections can develop into pelvic inflammatory disease (PID) and can also cause severe but curable eye disease (trachoma).

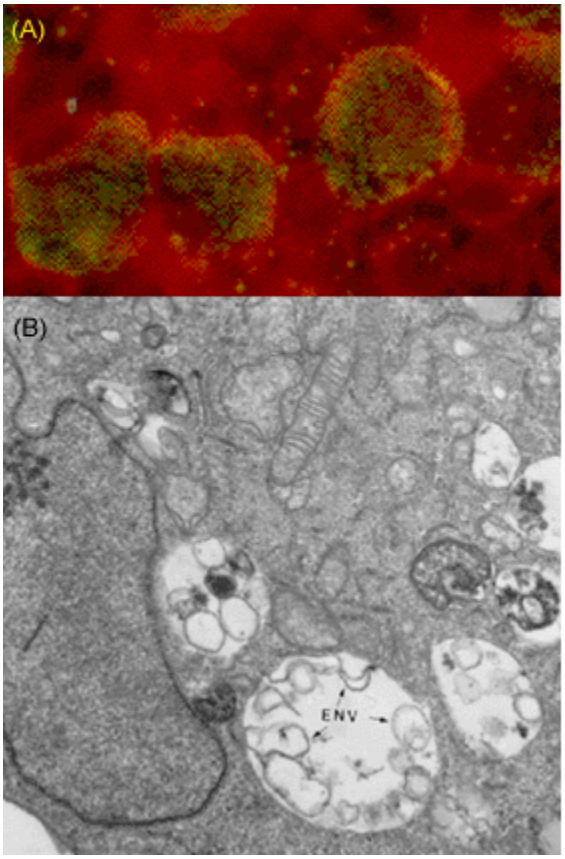
The Chlamydia Genome Project consortium has recently sequenced the *C. trachomatis* genome. It has a circular chromosome of about 1,045,000 base pairs, about one-quarter the size of *Escherichia coli*. Analysis of the sequence has identified 888 protein-coding genes. Among these, some proteins appear to have an unconventional natural history.

The genome appears to have undergone an unusually high number of horizontal gene transfer events, suggesting that the parasitic nature of *C. trachomatis* provides greater opportunity for gene transfer to occur. More bizarre, though, is that some of the *C. trachomatis* proteins are more related to green plants than to other bacteria or their human hosts.

FabI and FabF, two chlamydial proteins involved in the fatty acid synthesis component of membrane biogenesis, seem to be most closely related to their plant counterparts, whereas plsB, involved in lipopolysaccharide biosynthesis, has only plant orthologs.

How this came to be relies on linking two pieces of information. First, a *Chlamydia*-like parasite has been found recently in *Acanthamoeba*, a free-living protozoan usually found in fresh water or soil but which may occur as a human pathogen. Perhaps *Acanthamoeba* represents the original host for *Chlamydia*, and served as a vector to transfer its *Chlamydia* parasite to humans. Second, inferences made from 16S-like RNA provide evidence that *Acanthamoeba* is phylogenetically related to green plants. One would therefore expect some *Acanthamoeba* and green plant genes to be highly related. Horizontal transfer between *Acanthamoeba* (host) and *Chlamydia* (parasite) could therefore give plant-like genes to *Chlamydia*. If this horizontal transfer occurred before *Chlamydia* was passed to humans, then it is possible that a human parasite has plant-like genes.

Analysis of the genome of *Chlamydia* will provide a starting point for a deeper understanding of other eukaryotic parasites, including those responsible for human disease.



***Chlamydia trachomatis* inclusions in infected cells.**
Chlamydia trachomatis inclusions in infected human genital epithelial cells.

(A) The inclusions are seen as green or yellow spots by fluorescence photomicroscopy. (Photograph by Stephen T. Knight.)

(B) Inside a *Chlamydia trachomatis*-infected human genital epithelial cell. If the cells are exposed to the antibiotic azithromycin, the viable *Chlamydia* are killed, leaving residual chlamydial envelope material ("ENV"; one of these vesicles is equivalent to one of the yellow/green spots in (A) above, at this much higher magnification). (Photograph by Jane E. Raulston.)

Both photographs are from studies at the *Chlamydia* laboratories of Drs P. B. Wyrick and J. E. Raulson, Dept of Microbiology and Immunology, University of North Carolina School of Medicine, Chapel Hill, NC 7290, USA.

Search PubMed for chlamydia adapting to live in human cells.

Created: July 15, 1999

Click the link below to start an html tutorial.

How has chlamydia adapted to live in human cells?

Search PubMed for horizontal gene transfer.

Created: July 15, 1999

Click the link below to start an html tutorial.

Horizontal gene transfer in pathogenesis

Use BLAST to search for chlamydia fatty acid synthesis enzymes.

Created: July 15, 1999

Click the link below to start an html tutorial.

Chlamydia fatty acid synthesis enzymes match plant enzyme sequences

Tutorial list

The following list of Coffee Breaks is organized according to the NCBI bioinformatics resource featured in the tutorial that accompanies the story.

BLAST (Basic Local Alignment Search Tool)

A gut feeling:
bugs are critical for good health
July 24, 2013

ClinVar

Turning the white fat brown:
a new approach to obesity?
March 3, 2014

Genetic Testing Registry

DNAs of our lives:
the role of pharmacogenomics in modern medicine
April 25, 2013

GEO (Gene Expression Omnibus)

Do brains have a freshness date?
the effect of aging on the human brain
January 12, 2005

NIH Clinical Research Trials

How low can you go?
the promise of a new class of cholesterol lowering drugs
March 25, 2014

Structure and Cn3D

Yoda and the fountain of youth?
the many hats of IGF-1
July 21, 2014

Variations on a gene:
investigating the causes of iron overload
August 25, 2003

dbSNP (Database of Single Nucleotide Polymorphisms)

Roses, noses, and underarms:
how one variation in our DNA influences underarm perspiration (and ear wax)

March 25, 2013

PopSet (Population Sets)

What you see is not what you get!
DNA barcoding is helping scientists unveil nature's most hidden diversity
August 06, 2005