Biotechnology Discoveries and Applications

Extensions to high school science curriculum

The 2012 guidebook

HUDSONALPHA
INSTITUTE FOR BIOTECHNOLOGY

Genetic Technologies for Alabama Classrooms (GTAC)

a two week teacher academy

GTAC is an intensive two-week professional development academy for high school biology teachers held at the HudsonAlpha Institute for Biotechnology in Huntsville, AL. The academy is designed to help Alabama educators more effectively teach genetics by updating content knowledge, identifying common student misconceptions and gaining familiarity with hands on genetic activities and classroom tools.

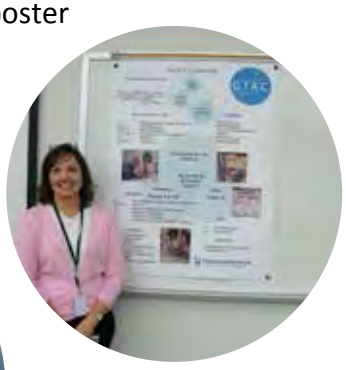
June 16 - 28, 2013

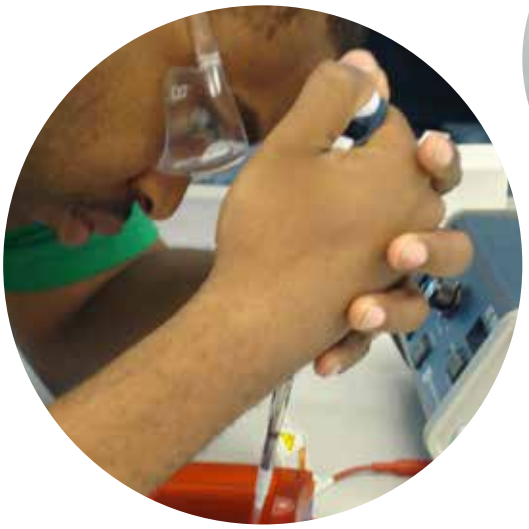
Applications will be accepted December 5, 2012 – February 20, 2013

www.hudsonalpha.org/education

What can I expect?

- Practice using hands on activities
- Hear from scientists involved in cutting edge biotechnology research
- See and use modern biotechnology equipment and laboratories
- Implement learning through individual and group projects
- Create and present a professional poster showcasing GTAC concepts





Walk Away with:

- 80+ Professional Learning units
- Stipend
- Toolkit of equipment and resources
- Updated genetic content knowledge
- Modern applications in biotechnology
- Network of teachers from across the state

CONTENTS

iCell®

HudsonAlpha is proud to announce the redesign of iCell. The updates have been made in response to feedback from educators and users. Additions include an all new interface, more accurate graphics, and availability on iPad, iPhone and online at icell.hudsonalpha.org. With over 100,000 downloads from Apple's App store, iCell is becoming a standard tool for biology education.

The Progress of Science

The Progress of Science is an online timeline that details over 200 major accomplishments and milestones in genetics and biotechnology during the past 10,000 years. The digital timeline is an interactive navigation tool that offers details on each major event and links out to other online resources where available. The timeline is frequently updated, keeping the content current for classroom discovery. The Progress of Science can be accessed at timeline.hudsonalpha.org.

genome cache

Build you own genome or walk ours. The newest digital education project from HudsonAlpha combines the challenge of a scavenger hunt with the human genome. GenomeCache and its associated website, genomecache.hudsonalpha.org allow anyone to create up to 20 walkable paths that explore the human genome. GenomeCache allows you to experience and learn more about the human genome through clues, fun facts and trivia questions. GenomeCache is available on iPad and iPhone and features over 150 challenging questions, a leaderboard, and themed paths.

Biotechnology Discoveries and Applications
2012

How this guide is arranged

Recent research findings are grouped on pages seven through nineteen and provide a quick update on the genetics/genomics/biotechnology field. **This section represents discoveries, treatments or applications that have been announced during the past year.** Some are described in only a few sentences while others get a more thorough explanation.

Each new finding connects to one of twenty-three key technologies or concepts described in detail on subsequent pages. Language and concepts are intentionally geared to a high school or public audience.

Within each overview, linking course of study objectives are identified for Alabama High School Courses:

Look for the  symbol in blue.

Where relevant, the experiments and activities developed by HudsonAlpha are also described:

These are identified by the  symbol in orange.

Where appropriate, an acknowledgement of research occurring at HudsonAlpha is given:

The  symbol identifies those connections.

EXECUTIVE SUMMARY

Thanks to a revolution in DNA sequencing technology, scientists are able to accurately and efficiently sequence a human genome at a reasonable cost - currently somewhere in the neighborhood of $5,000 to $7,500. This is more than a one million-fold reduction in cost from the early days of the Human Genome Project. Simultaneously, sophisticated computer algorithms comb through the 3 billion bases of sequence information, helping identify DNA variants with potential clinical significance.

Sequencing the entire genome or only the exome (that two percent of the genome that is protein-coding) is rapidly becoming an integral part of many research plans. When the 2010 edition of the *Biotechnology Guidebook* was written, fewer than 50 human genomes had been sequenced. By the end of 2012, the National Institutes of Health estimates they will have funded exome or genome sequencing of samples for approximately 70,000 research subjects. Clearly, the era of genomics is upon us.

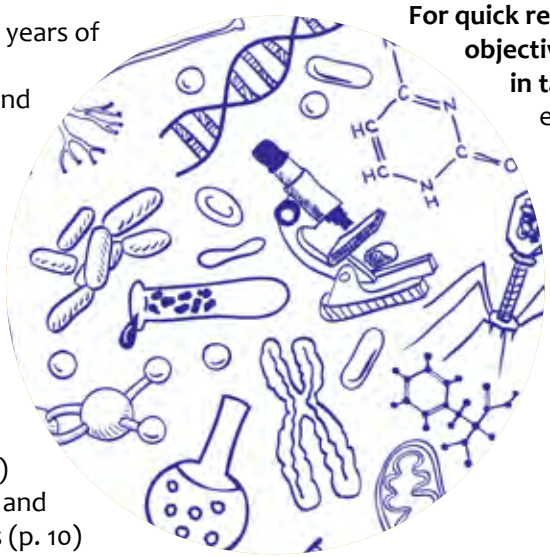
Against this dizzying backdrop, how does a science, health or career technical education teacher keep his or her students current regarding the impact of genomics? This is the aim of our annual *Biotechnology Discoveries and Applications* educator guidebook: **to provide educators with an overview of the recent advances in genetics and biotechnology, allowing them to share these findings with their students.**

The guidebook is divided into two sections: research highlights and foundational concepts. This year's research section highlights 41 new discoveries, including articles on:

- processes for controlling gene activity across the human genome (p. 8)
- successes in gene therapy after years of disappointment (p. 11)
- collisions between replication and transcription machinery (p. 8)
- the genetic pathway for hair color patterns in cats (p. 7)
- insight into the 100 trillion microbes that live in and on the human body (p. 12)
- carrot cell cultures that produce medication to treat disease (p. 14)
- genome sequencing to track a deadly bacterial outbreak (p. 15)
- cancer-causing genetic changes and targeted treatment possibilities (p. 10)

A not-for-profit research organization, HudsonAlpha is located in a 270,000 square foot building in Huntsville, Alabama, the cornerstone of a planned 150-acre biotechnology campus. Opened in 2008, the institute is a joint venture between private philanthropy and support from the state of Alabama. HudsonAlpha aims to harness the power of biotechnology to improve human health, stimulate economic growth and inspire youth to seek careers in the field of science through educational outreach.

These new findings are linked to 23 foundational topics, covered in detail beginning on page 26. Each includes the basic concepts underlying each topic and connections to relevant course of study objectives for Alabama's science, health and career technical education classes. Genomics has become an important tool across the life science landscape, entering classroom discussions ranging from agriculture and human health to bioinformatics and ethics.



For quick reference, the course of study objectives and foundational topics are listed in table format on pages 20-24. For educators who are not in Alabama, many of these course of study objectives will align across states fairly easily.

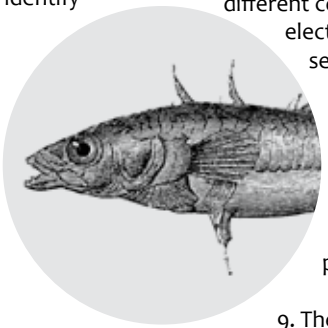
A list of suggested readings for additional exploration of many of the foundational topics can be found on page 52.

SCIENCE SNAPSHOTS

a quick rundown of 10 genetic and biotech stories

1. The Genetic Testing Registry, created by the National Center for Biotechnology Information, is designed to help physicians, patients and researchers make sense of the many genetic tests that are currently available. The database includes tests for over 2,500 disorders, submitted voluntarily by the testing organizations. Details about a test's purpose and its limitations are also included.

2. At the end of the last ice age, marine stickleback fish separated into different populations across salty oceans and freshwater lakes and streams, evolving different traits such as body length, eye size, skeletal structure and behavior. These differences allowed sticklebacks to thrive in their specific habitat. Sequencing and comparing the genomes across multiple forms of stickleback allow scientists to identify DNA changes related to fish appearance and behavior. Many changes involve large chromosomal inversions. Single nucleotide changes predominantly appear in regulatory rather than protein-coding areas of the gene, suggesting the evolutionary process altered how and when genes were active, rather than directly changing the gene product.
HudsonAlpha researchers Dr. Rick Myers, Dr. Jane Grimwood and Jeremy Schmutz participated in this work.



3. Patients undergoing heart surgeries like angioplasty or stenting are often given the drug clopidogrel (also known as Plavix™) to inhibit platelet activity and reduce the risk of post-surgery blood clots. The *CYP2C19* gene codes for an enzyme that activates this drug in the body. A common loss-of-function variant in this gene means that some patients poorly activate the drug and should be given a different medication. In a Canadian study, patients underwent *CYP2C19* genetic testing just after surgery and the results determined medication choice. When compared to a non-tested population, the tested patients had fewer cases of high platelet activity (a risk for post-surgery heart attack and stroke), suggesting this test could be useful at the patient's bedside.

4. Researchers at Harvard and Johns Hopkins universities encoded a 53,000-word book into DNA fragments as a method of digital data storage. The text was converted into binary language, with a series of 0s and 1s representing each character. The binary string was then synthesized as DNA fragments, with nucleotide bases A and C representing 0s, and G and T as 1s. To read the data, the DNA is sequenced and the nucleotide message converted back into binary. The fragments are stable at room temperature for very long periods of time. Using this approach, one gram of DNA can store 700 terabytes of data – the equivalent of 14,000 Blu-ray 50 gigabyte discs.

5. The sequences of 126 proteins, including collagen and albumin, were identified in a 43,000-year old woolly mammoth bone. Modifications in sample preparation maximized the ability to detect trace amounts of protein, which is more stable than DNA. As this technique is refined, it will provide an alternative for analyzing ancient samples where the DNA has degraded.

6. Prized for centuries, Thoroughbred racehorses traced their ancestry to three Arabian stallions and 74 primarily British mares. The breed can be subdivided into sprinters and distance runners. Thoroughbred speed is strongly influenced by a single nucleotide polymorphism in the first intron of the myostatin gene - which encodes a protein important to muscle mass development. Homozygous C/C horses excel in sprinting, C/T heterozygotes compete well in middle distances and T/T homozygotes have long distance stamina. Genetic and pedigree analysis from 593 horses (including bone samples from 18th century stallions) traced the origin of the C allele to a British mare that lived roughly 300 years ago.

7. The race for cheaper and faster DNA sequencing continues as genomic companies introduce new sequencing technologies. Various systems for reading individual DNA bases include different colors of light, tiny changes in pH or disruptions in electrical current. Manufacturers claim some systems can sequence a human genome in less than a day.

8. The FarGen project aims to sequence all 50,000 individuals living on the Faroe Islands, incorporating the information into the country's healthcare system. A pilot study to create the infrastructure and discuss the scientific, ethical, legal and social aspects of the project is currently underway.

9. The HIrisPlex test is an example of forensic DNA phenotyping - predicting visible traits from forensic samples to build a more complete suspect or victim profile. Highlighted in the *2011 Biotechnology Guidebook*, a panel of 24 DNA variants now predicts hair as well as eye color. The technology accurately predicts brown versus blue eye color 94 percent of the time and average prediction accuracy for hair color ranges from 69.5 percent for blond to 87.5 percent for black.

10. Researchers have identified the genetic pathways that explain the molecular basis of hair color patterns in felines from house cats to snow leopards. The studies point to an interaction between two genes called *Taqpep* and *Edn3*. *Taqpep* appears to establish the coloration pattern during skin development. This influences the expression of *Edn3* in the hair follicles, ultimately leading to yellow or black pigment production.
HudsonAlpha researcher Dr. Greg Barsh contributed to these findings.

REFERENCES

1. U.S. Department of Health and Human Services, (2012) "Confused by genetic tests? NIH's new online tool may help [Press Release]," <http://www.nih.gov/news/health/feb2012/od-29.htm>.
2. Jones et al., "The genomic basis of adaptive evolution in threespine sticklebacks," *Nature* 484:55 (2012), doi:10.1038/nature10944.
3. Roberts et al., "Point-of-care genetic testing for personalization of antiplatelet treatment: a prospective, randomized, proof-of-concept trial," *The Lancet* published online March 29, 2012, doi:10.1016/S0140-6736(12)60161-5.
4. Church et al., "Next-Generation Digital Information Storage in DNA," *Science*, published online August 16, 2012, doi:10.1126/science.1226355.
5. Cappellini et al., "Proteomic Analysis of a Pleistocene Mammoth Femur Reveals More than One Hundred Ancient Bone Proteins," *Journal of Proteome Research* 11:917 (2012), doi:10.1021/pr200721u.
6. Bower et al., "The genetic origin and history of speed in the Thoroughbred racehorse," *Nature Communications* (2012) 3:643 doi:10.1038/ncomms1644.
7. Baker, M., "Benchtop sequencers ship off," Nature News Blog, Posted September 14, 2012. < http://blogs.nature.com/news/2012/09/benchtop-sequencers-ship-off.html>.
8. Eliasen, B., "Sequencing the entire nation of the Faroe Islands," <http://cphx.org/2012/08/bogi-eliasen-sequencing-the-entire-nation-of-the-faroe-islands/>.
9. Walsh et al., "The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA," *Forensic Science International: Genetics*, published online August 22, 2012, doi:10.1016/j.fsigen.2012.07.005.
10. Kaelin et al., "Specifying and Sustaining Pigmentation Patterns in Domestic and Wild Cats," *Science*: 337:1536, (2012) doi:10.1126/science.1220893.

Wanted: gamers

citizen-scientists at play

Thin-slice volunteering refers to the collectively useful tasks that volunteers can accomplish in micro-slices (while waiting at the doctor's office, sitting in an airport, waiting to pick up kids at the soccer fields, etc.). The idea is to develop engaging tasks that can be done with a smartphone or laptop in just a few minutes, turning the chore of waiting into something fun for the collective good. Scientific discovery games represent thin-slice challenges that take advantage of humanity's skill in pattern recognition to optimize predictive computer programs. Three of the best known that link to genetics and cellular biology are listed below.

1. **Foldit** (fold.it) is a protein folding game where users play to competitively fold the best proteins. This was one of the first online science games and the best scoring players have contributed to scientific understandings of cancer and Alzheimer's disease.

2. **Phylo** (phylo.cs.mcgill.ca) asks users to align colored blocks that represent DNA sequences to identify disease-causing regions of the genome.

3. **EteRNA** (eternal.cmu.edufold.it) challenges players to design RNA molecules that fold into a target shape. EteRNA allows the public to give scientists suggestions for better molecular design – the highest scoring model of the week is synthesized in a lab and experimentally tested to compare predicted and actual folding patterns.

Cell cycle traffic jam

when the processes of transcription and replication collide in cells

DNA replication and transcription are critical molecular processes required for cell growth and division. Both involve large multi-protein complexes traveling across long chromosomal distances at high speeds. When possible, eukaryotic cells coordinate these processes at different times in the cell cycle. Most short genes are transcribed in G1, before DNA replication in S phase. Surprisingly, new research suggests that genes longer than 800kb begin transcription after replication, at G2 or M. For especially long genes, transcription may continue through G1 of the next cell cycle and into S. At this point, the RNA machinery risks colliding with the replication fork, stalling both systems. If this occurs and the impass is not resolved, a persistent loop forms between the DNA of the transcribed strand and the newly-made RNA. The DNA:RNA hybrid loop increases the likelihood of DNA breakage at the collision site.

Throughout the genome, common fragile sites (CFS) are associated with cancer-causing translocations, deletions and other rearrangements. Many of these sites apparently map within the late-transcribed regions of large genes, suggesting that collisions between transcription and replication can be an early step in tumor formation.

REFERENCES

Helmrich et al., "Collisions between Replication and Transcription Complexes Cause Common Fragile Site Instability at the Longest Human Genes," *Molecular Cell* 44:966 (2011) doi:10.1016/j.molec.2011.10.013.

Interpreting the book of life

the ENCODE Project offers new clues in understanding how our genome functions

ENCODE stands for the **ENC**yclopedia **O**f **D**NA **E**lements. It is something like a catalog of the regions in the human genome that control how genes function under certain conditions.

As an analogy, think of a passage from a favorite book. Imagine the text is missing punctuation, spacing and those visual cues that make sense of the language. All that is present are the individual letters that make up the words. Adding in other parts of language provides meaning to the string of text – things like capital letters, commas, quotation marks and periods. From a string of seeming nonsense, great literature appears.

In a similar way, the ENCODE Project helps scientists make sense of the human genome – by understanding the biological language contained in the chemical bases of the DNA. The Human Genome Project determined the location of more than 21,000 genes in the 3.2 billion DNA nucleotides. However, these genes only account for 1-2 percent of the entire genome. This means the functional significance of much of the genome has been a mystery. The ENCODE Project is the genetic equivalent of providing spacing and punctuation – a set of experiments to determine which pieces of DNA regulate the action and storyline of the genome.

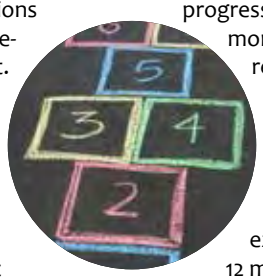
Some analyses identified regions where transcription factors bind DNA and control gene transcription. Other experiments searched for DNA methylation – small molecules that attach to the DNA sequence. Transcription factor binding and DNA methylation can greatly increase or completely silence the activity of a corresponding gene - dramatically altering the level of protein it produces. This may have important consequences for how the cell functions or interacts with neighboring cells.

The decade-long ENCODE project involved 442

Exon-skipping

bypassing mutations during RNA splicing

Duchenne Muscular Dystrophy (DMD) is one of the most common forms of muscle wasting. Treatments for this X-linked recessive disorder have been elusive, but promising medications are moving into late-stage development. In a follow-up from the *2011 Biotechnology Guidebook*, two exon-skipping drugs have yielded clinically significant phase 2 results. About 13 percent of patients with DMD have mutations in exon 51 of the dystrophin gene that lead to a nonfunctional protein. The drugs, both more than six years in development, are targeted antisense oligonucleotides that bind the RNA from this gene, causing the cell to splice out (skip) exon 51. This produces a slightly shortened, but hopefully functional dystrophin protein. Clinical benefit is measured by the distance patients are able to walk in six minutes. In one study, patients experienced a significant slowing of disease progression over an eight-month timeframe. Early results from a different drug showed a slight increase on walking test distances for patients. Additional clinical information is expected over the next 12 months, which may lead to a formal request for FDA approval.



REFERENCES

References: Goemans, N.M., Systemic Administration of PRO051 in Duchenne's Muscular Dystrophy. *New England Journal of Medicine* (2011) 364:1513. doi: 10.1056/NEJMoa1011367. Opar, A., Exon-skipping drugs pulls ahead in muscular dystrophy field. *Nature Medicine* 18:1314 (2012) doi: 10.1038/nm0912-1314.

researchers from 32 institutions around the world. All told, 1,649 experiments were performed. From these, scientists identified the location of millions of functional elements in our genome. The results were simultaneously published in 30 different papers (www.nature.com/encode).

As with many large-scale genomic projects, the data is publicly available (www.encodeproject.org), allowing scientists in all fields to better understand how human genomes are organized and regulated, providing important insight into health and disease.

The lab of HudsonAlpha researcher Dr. Richard Myers is part of the ENCODE Project.

For additional information about this topic, see *Studying the Genome to Understand the Sequence* on page 48. The biological process of DNA methylation overlaps with the growing field of epigenetics, which is explored in detail on the infographic on pages 16-17 and in the article *Epigenetics* on page 35.

REFERENCE

The ENCODE Project Consortium, "An integrated encyclopedia of DNA elements in the human genome," *Nature* 489:57 (2012) doi:10.1038/nature11247.

In brief

Studying the genome to understand the sequence

long range enhancers influence pigmentation levels

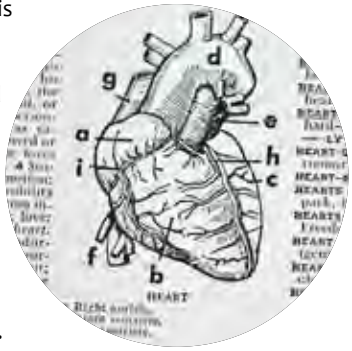
Pigmentation of the eye, skin and hair is one of the most variable common human traits. Several genes are involved in this process and DNA changes are linked to variation in pigmentation. One of the strongest links is a single nucleotide polymorphism (SNP) called rs12913832. Surprisingly, this SNP is in the intron of a non-pigment gene - 21,000 bases away from *OCA2*, the nearest pigment gene. The C allele of the SNP is associated with blue eyes and light pigmentation, while the T allele links to darkly pigmented skin and brown eyes.

Researchers believed the SNP functions as an enhancer, regulating the transcription of that distant *OCA2* gene in pigment-producing cells. Scientists in the Netherlands have confirmed this view and identified the underlying enhancer mechanism. Human melanocytes containing the T allele do a more effective job than their C allele counterparts at recruiting specific transcription factors - proteins that bind DNA and help control the rate of gene transcription. Once the transcription factors are present at the SNP's genomic location, a large loop of chromatin forms that brings them in contact with the OCA2 gene. This leads to greater *OCA2* transcription and ultimately, darker pigmentation in the eye and skin melanocytes. This is an excellent example of how a noncoding SNP can impact gene activity, even from thousands of bases away.

REFERENCE: Visser et al., "HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter," *Genome Research* **22:446 (2012) doi:10.1101/gr.128652.111.**

genetic switches in heart development

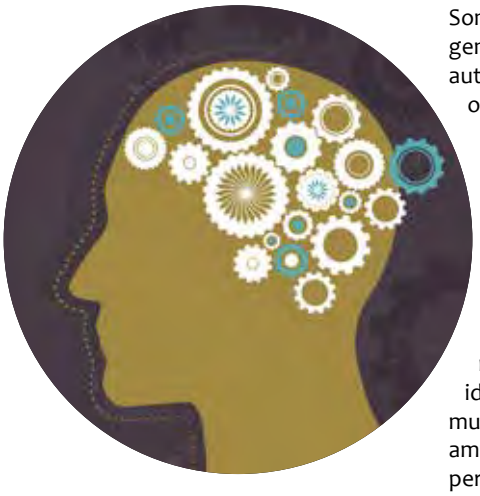
During embryonic development, stem cells differentiate into the hundreds of specialized cell types that make up an organism. This differentiation process requires extreme coordination among thousands of genes that are activated at specific times and in precise patterns. If this regulation goes awry, organ function can be disrupted, leading to disorders like congenital heart defects. As a first step to understanding the developmental pathway for the heart, mouse embryonic stem cells were studied as they underwent cardiac differentiation. Using an ENCODE-like approach, transcription factors and other regulatory marks were identified at various developmental time points. Groups of genes were found that function in a coordinated manner – switching on and off collectively at specific times. Some of these genes had never been associated with heart formation. Other analyses identified chemical tags involved in gene "pre-activation" – a process that sets the stage for future transcription. As this developmental landscape becomes better defined, genetic and environmental disruptions that lead to congenital heart disease can be identified and explored.



REFERENCE: Wamstad et al., "Dynamic and Coordinated Epigenetic Regulation of Developmental Transitions in the Cardiac Lineage," *Cell*, **published online September 12, 2012, doi:10.1016/j.cell.2012.07.035.**

Alzheimer's disease

genetic pathways for amyloid formation, clues to disease timing and challenges with clinical trials



An estimated 5.4 million Americans are affected with Alzheimer's disease (AD), a neurodegenerative disorder characterized by twisted fibers inside cells called tangles and extracellular deposits called plaques. The plaques are made of fragments from a protein called beta-amyloid that is encoded by the *APP* gene on chromosome 21.

Some mutations in the *APP* gene have been linked to autosomal dominant early-onset forms of AD. These mutations alter the pattern of beta-amyloid processing, leading to a greater percentage of the plaque-forming fragment.

A sequencing study of nearly 1800 Icelanders identified a protective mutation in *APP* that reduces amyloid formation by 40 percent. Individuals in the study aged 85 years or older who had this DNA change were 7.5 times less likely to have AD than those without the variant.

The variant is rare, occurring in only 1 out of 5,000 Americans. Still, the finding reinforces the central role of amyloid processing in the development of AD and shows that if toxic fragments can be reduced, the disease may be prevented.

This concept is at the heart of recent therapeutic attempts to target beta-amyloid using an antibody that would attach to and clear the amyloid from the brain. Disappointingly, several recent, large-scale phase three clinical trials failed to identify any cognitive benefit from these types of medications in patients with mild to moderate AD.

The cause behind these failures may be the population they are trying to treat – patients who already have symptoms. A new analysis of presymptomatic patients suggests amyloid formation may begin more than 20 years before symptoms appear. This implies that prevention may need to be started decades earlier than previously thought. If true, the recent failure of amyloid-busting medications is not surprising – the treatments were begun too late to make a difference. Consequently another round of clinical trials is set to begin, this time in high-risk presymptomatic individuals.

For more information, see *Therapeutic Approaches* on page 50.

REFERENCES

Jonsson et al., "A mutation in APP protects against Alzheimer's disease and age-related cognitive decline," *Nature* 488:96 (2012) doi:10.1038/nature11283. Gandy S., "Lifelong Management of Amyloid-Beta Metabolism to Prevent Alzheimer's Disease," *New England Journal of Medicine* 367:864 (2012) doi:10.1056/NEJMe1207995. Mullard A., "Sting of Alzheimer's failures offset by upcoming prevention trials," *Nature Reviews Drug Discovery* 11:657 (2012) doi:10.1038/nrd3842.

Cancer update

key research discoveries shed light on three of the most common forms

Colorectal Cancer

The Cancer Genome Atlas (TCGA), a large collaborative research initiative, released findings for colorectal cancer showing the pattern of genetic change is nearly identical whether the cancer originates in colon or rectal tissue. Twenty-four genes were consistently mutated across the over 200 tumors studied, including a number of genes that may serve as potential therapeutic drug targets.

Lung Cancer

TCGA also released data related to lung squamous cell cancer, which often develop in the center of the lungs. Historically difficult to treat, researchers identified a set of promising therapeutic targets, including three classes of tyrosine kinase genes that function as on/off switches for many cell growth functions. A majority of tumors also contained mutations in the tumor suppressor genes *TP53* and *CDKN2A*, leading to unregulated cell growth. Intriguingly, mutations were also found in the immune system gene *HLA-A*, which may help explain how tumors escape the body's surveillance system.

A separate study explored the genetic background and smoking history of patients with non-small cell lung cancer - primarily adenocarcinoma. Patients with a history of smoking had 10 times more genetic mutations in their tumors. In an encouraging finding, many of the patients had at least one mutated gene that could be targeted with a drug currently in clinical trials or already available, reinforcing the value of genetic data in therapeutic decision-making.

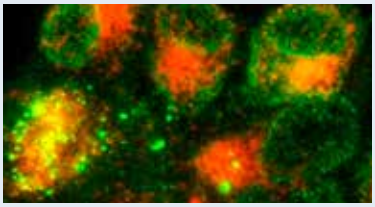
Breast Cancer

By analyzing genetic sequence and the RNA expression levels from almost 2,000 breast cancers, researchers have developed a system that classifies the cancers into 10 distinct types, each associated with a different prognosis. Each accounts for 4-17 percent of breast cancer and 10-year survival rates vary from 40-90 percent depending on the category. Full validation of this model, along with therapeutic correlations will require thousands of additional patients, followed for 10 or more years.

Refer to *Cancer* on page 30 for additional information.

REFERENCES

The Cancer Genome Atlas Research Network, 1."Comprehensive molecular characterization of human colon and rectal cancer," *Nature* 487:330 (2012) doi:10.1038/nature1252. 2. "Comprehensive genomic characterization of squamous cell lung cancers," *Nature*. Published online Sept 9, 2012. doi:10.1038/nature11404. Imielinski et al., "Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing," *Cell* 15:1107 (2012) doi: 10.1016/j.cell.2012.09.29. Curtis et al., "The genomic and transcriptomic architecture of 2000 breast tumors reveals novel subgroups," *Nature* 486:346. (2012) doi:10.1038/nature10983.



Antisense therapy

using mouse models for myotonic dystrophy

Studies of disease research in animal models contribute greatly to our understanding of how diseases develop. These studies can also identify new therapeutic approaches, as is the case in a recent finding of a mouse model for myotonic dystrophy.

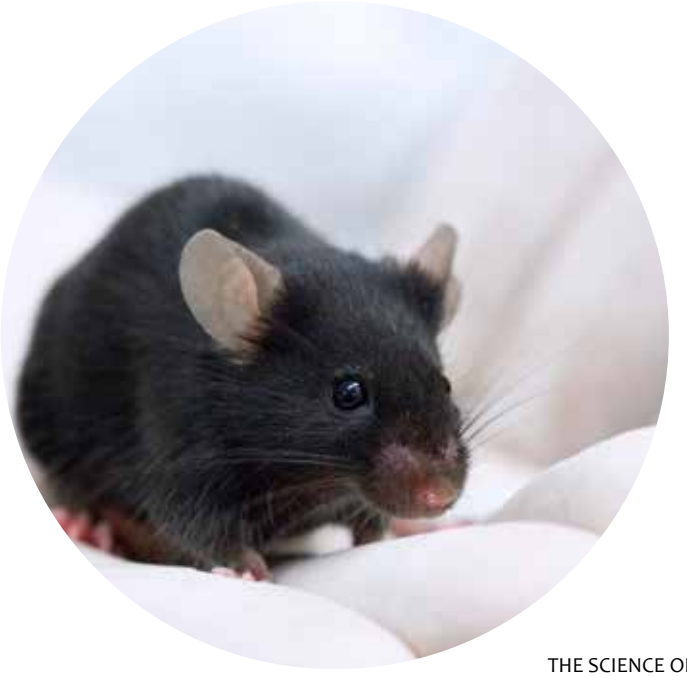
In humans, myotonic dystrophy is an inherited disorder that leads to progressive muscle weakness and stiffness. The causative mutation produces an abnormal messenger RNA molecule that remains trapped in the nucleus. Over many decades these toxic fragments accumulate to such high levels that they interfere with the processing of RNA transcripts from other genes. It appears that genes encoding proteins critical for muscle integrity and relaxation are especially sensitive to this disruption, leading to the clinical symptoms of myotonic dystrophy. A therapeutic "holy grail" would be some process that specifically destroys the mutated mRNAs that perturb nuclear function.

Along these lines, antisense oligonucleotides have been created that selectively bind the mutated mRNA and target them for degradation. After decades of work, antisense technology is yielding promising results. Eight doses across four weeks reduced the level of toxic mRNA by 80 percent, markedly eased muscle stiffness and improved the microscopic structure of the muscle cells. The effects lasted more than a year.

While the findings are encouraging, caution is warranted. There are numerous examples of promising studies in animal research that are not replicated when brought to human clinical trials. An important unknown is whether the drug will improve the muscle-wasting associated with myotonic dystrophy in humans. That symptom has been hard to mimic in mouse models, so the interaction with antisense oligonucleotide treatment has not been examined. Even so, the mouse model has provided key pre-clinical data in the search for an effective future cure.

REFERENCE

Wheeler et al., "Targeting nuclear RNA for in vivo correction of myotonic dystrophy," Nature 488:112 (2012) doi:10.1038/nature11362.



In brief

Therapeutics

gene therapy success stories

After several years of false starts, disappointments and challenges, gene therapy scored several recent successes.

1. An adeno-associated viral vector successfully produced FIX ('factor 9') in six patients with hemophilia B. Four patients produced enough of the clotting factor to remain free of bleeding episodes over the 16 month timecourse of the study.

2. Leber's congenital amaurosis is a group of inherited diseases where genetic mutations impair an enzyme needed in the retina, ultimately leading to blindness. In 2009, researchers injected one eye in each of 12 patients with a modified adeno-associated virus carrying a normal version of the *RPE65* gene. Six saw enough improvement in vision to no longer be classified as legally blind. Recently, three of these patients received gene therapy for their other eye, with significant improvements in light sensitivity and the ability to navigate in dim light. There were no safety problems or significant immune responses.

3. ADA-deficient severe combined immunodeficiency (SCID) has been a target for gene therapy since the earliest days of the field. In a recent clinical trial, physicians collected stem cells from patient bone marrow into which they inserted a working copy of the defective gene using a retroviral vector. Before reintroducing the genetically-enhanced stem cells, patients received low doses of a chemotherapy drug to destroy many of the faulty bone marrow cells. Although there were complications with infection, half of the patients now produce enough of the missing enzyme that they no longer need other forms of treatment, even 3-5 years after the procedure.

REFERENCES: Nathwani et al., "Adenovirus-Associated Virus Vector-Mediated Gene Transfer in Hemophilia B," *New England Journal of Medicine* **365:2357 (2011) doi:10.1056/NEJMoa1108046. Bennett et al., "AAV2 Gene Therapy Readministration in Three Adults with Congenital Blindness,"** *Science Translational Medicine*, **4:120ra15 (2012) doi:10.1126/scitranslmed.3002865. Candotti et al., "Gene therapy for adenosine deaminase-deficient severe combined immune deficiency: Clinical comparison of retroviral vectors and treatment plans,"** *Blood* **published online September 11, 2012, doi:10.1182/blood-2012-02-400937.**

mutation-dependent cystic fibrosis treatment

The U.S. Food and Drug Administration approved the first drug to treat the genetic cause of a subset of patients with cystic fibrosis. The approval process for Kalydeco™ took only three months - one of the fastest ever for the FDA. The drug is approved for patients with at least one G551D mutation in the *CFTR* gene. This mutation encodes a defective protein that moves to the correct place on the surface of the cell but does not function properly. Kalydeco™ restores partial function to the protein, allowing a proper flow of salt and fluid in and out of the cell and thinning the sticky mucus that collects in the lungs of patients with CF. G551D represents only a small percentage of total CF mutations. Other drugs currently in the clinical pipeline are targeted for patients with Delta F508 mutation, which is much more common.

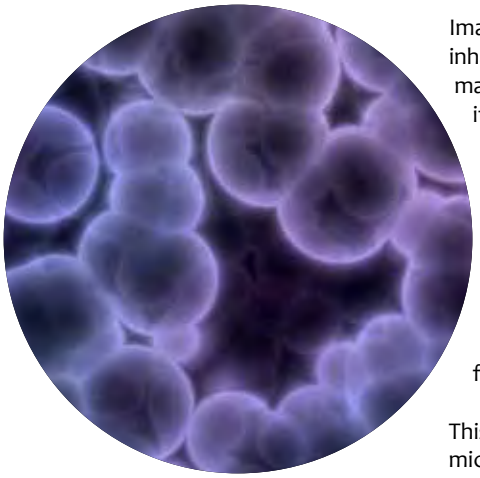
REFERENCE: Ramsey et al., "A CFTR Potentiator in Patients with Cystic Fibrosis and the G551D Mutation," *New England Journal of Medicine* **(2011) 365:1663 doi: 10.1056/NEJMoa1105185.**

For more details, refer to *Therapeutic Approaches* on page 50.

NEW FINDINGS - THE IMPACT OF LARGE-SCALE SEQUENCING

You and your microbiome

the trillions of bacteria, viruses and fungi that call you home



"An estimated 100 trillion microbes live on and inside every individual - 250x as many stars in the Milky Way Galaxy."

Imagine a thriving metropolis, inhabited by millions. Like many large communities, it is grouped into distinct neighborhoods – the Upper East Side, Old Town, the manufacturing district, Little Italy. Using existing transportation systems, the citizens travel from one region to another, carrying out daily functions.

This is the story of the human microbiome: the collection of microscopic organisms (primarily bacteria, but also viruses and fungi) found on skin, throughout the digestive system and inside various body cavities like the mouth and nose. Humans serve as the infrastructure for these thriving communities, providing nutrients and a favorable climate for growth. Although each microbe is one-tenth to one-hundredth the size of a human cell, in total, microbes outnumber human cells by a factor of 10. Collectively, the microbiome accounts for two to five pounds of an average adult's bodyweight.

The vast majority of these microbial inhabitants are helpful to the body. In the digestive system, bacteria break down plant material into digestible pieces. Other species make vitamins humans cannot produce. The bacteria on skin secrete a moisturizing film that keeps the skin supple, preventing cracks that serve as entry points for harmful pathogens.

Over the last five years, the Human Microbiome Project (HMP) has cataloged the microbiome of nearly 250 healthy Americans. Historically, bacteria and other microbes have been studied by growing them on petri dishes but most bacteria do not thrive under these conditions. Thanks to the development of relatively inexpensive, high-throughput DNA sequencing, the HMP directly identified microbes by their genetic sequence. This gave researchers a snapshot of bacterial populations and their relative abundance at any given site. Over 10,000 different species were identified. There is a great deal of variation from person to person, likely influenced by factors such as diet, environmental exposures, genetic contributions and early life experiences. Researchers observed microbiomes changing over time. If a study subject became sick and took an antibiotic, specific bacterial species were often wiped out. The antibiotic not only impacted the illness-

Genome sequencing finds its stride

searching for rare variants that influence traits and disease

Technological advances are driving down the cost to sequence large stretches of DNA and sophisticated bioinformatics pipelines are more accurately distinguishing meaningful DNA change from benign genetic variation. Consequently, research projects are chosing to sequence the exome (the protein-coding regions) or entire genome of subjects to identify genetic contributions in traits and disorders. A sampling of recent findings is included below:

Broken genes in healthy individuals: Whole-genome sequences from 185 individuals were screened for loss-of-function variants – genetic changes that substantially disrupt the gene's product. The analysis suggests every human has around 100 of these changes. Most are present in only a single copy of the gene, but as many as 20 may be present in both copies. Only a small number are associated with severe diseases: the majority appear in genes that are less functionally critical or have similar (redundant) genes elsewhere in the genome.

New genes in cognitive disorders: Historically, research into intellectual disability (ID) has focused on the X chromosome, in part because of the large number of X-linked disorders with cognitive impairment. An analysis of 136 families with a child affected by ID, where the parents were distantly related, looked instead at autosomal recessive models. Say two parents share a common ancestor: both could have inherited the identical copy of a disease-causing allele and then passed it to their child, leading to ID. Along these lines, researchers scanned the genome of the affected child to identify long stretches of DNA that were homozygous (where both alleles were inherited from the same ancestor). These regions were then extensively sequenced to identify rare ID-causing mutations. Disease-causing variants were identified in 78 families, uncovering 50 genes not previously associated with ID.

DNA mutation rate and age
DNA sequence changes were studied between groups of identical twins of various ages. Structural changes in the DNA occur more commonly in older individuals. More than three percent of healthy individuals age 60 and older contained genetic rearrangements greater than 500,000 nucleotides long in their white blood cells. These mutations were not inherited from a parent, but likely arose due to errors in the DNA replication process as cells grew and divided over the lifespan of the individual.

HudsonAlpha researcher Dr. Devin Absher contributed to the analysis of this finding.

De novo genes in autism
Exome sequencing was used in a number of studies to find mutations linked to autism spectrum disorders (ASD). Collectively, these projects sequenced nearly 1,000 families with a child affected by ASD. Researchers were looking for *de novo* mutations – those that arise spontaneously in egg, sperm or the newly developing embryo. Several genes were identified, six of which had *de novo* mutations in more than

causing bacteria but also related species in the normal microbial community. The microbiome eventually re-established its equilibrium, although at times the affected bacteria was replaced with a different species.

Surprisingly, nearly every individual carried so-called opportunistic pathogens – microorganisms that peacefully coexist with the rest of the microbiome but under unusual circumstances can cause illness. These included low levels of destructive strains of *Escherichia coli* and the ulcer-causing *Helicobacter pylori*. Scientists seek to understand why and under what conditions these bacteria become harmful.

Several new research questions have sprung from the HMP's initial results, including how microbes initially colonize the body and why the immune system – programmed to identify and destroy foreign organisms – permits the microbiome to thrive. The findings also allow researchers to search for microbial changes associated with disorders like Crohn's disease, ulcerative colitis, acne, psoriasis and immunodeficiency.

For additional information about this topic, see *Infectious Disease* on page 39.

REFERENCE

The Human Microbiome Project Consortium, "Structure, function and diversity of the healthy human microbiome," *Nature* 486:207 (2012), doi:10.1038/nature11234.

one individual with ASD. These genes provide important insight into the biology of ASD. Taken together, the findings suggest *de novo* mutations may be involved in 10 percent of cases, but reinforce the idea that no single genetic event is sufficient to cause autism.

HudsonAlpha researcher Dr. Shawn Levy contributed to the analysis of this finding.

For more details, see *Identifying Genetic Influences on Disease* on page 38.

REFERENCES

Phimister E.G., "Realizing Genomic Medicine," *New England Journal of Medicine*, 366:8 (2012) doi:10.1056/NEJMe1200749.
MacArthur et al., "A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes," *Science*, 335:823 (2012) doi: 10.1126/science.1215040.
Najambadi et al., "Deep sequencing reveals 50 novel genes for recessive cognitive disorders," *Nature* 478:57 (2011) doi:10.1038/nature10423.
Forsberg et al., "Age-Related Somatic Structural Changes in the Nuclear Genome of Human Blood Cells," *American Journal of Human Genetics* 90:217 (2012) doi:10.1016/j.ajhg.2011.12.009.
Iossifov et al., "De novo gene disruptions in children on the autistic spectrum," *Neuron* 74:285 (2012) doi:10.1016/j.neuron.2012.04.009.
Sanders et al., "De novo mutations revealed by whole-exome sequencing are strongly associated with autism," *Nature* 485:237 (2012) doi: 10.1038/nature10945.
O'Roak et al., "Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations," *Nature* 485:246 doi: 10.1038/nature10989.
Neale et al., "Patterns and rates of exonic de novo mutations in autism spectrum disorders," *Nature* 485:242 doi:10.1038/nature11011.

In brief

Identifying genetic influence on disease

insight into irritable bowel syndrome

Details about rare inherited forms of a disorder often provide insight into the biological mechanisms shared with more common versions of the disease. For example, the recent genetic characterization of a large family with chronic diarrhea may offer insight to the 15-20 percent of adults who suffer from irritable bowel syndrome. Extensive analysis uncovered a heterozygous missense mutation (a point mutation that results in one amino acid substituted for another) in the *GUCY2C* gene in all affected family members. This gene encodes a protein that spans the membrane of intestinal cells and helps coordinate a cell's response to changes in the intestinal environment. The mutation creates a protein that responds more strongly to external stimuli, initiating a cellular chain of events that ultimately pumps large amounts of chloride and water out of the cell and into the intestines, leading to the diarrhea. The symptoms of the family overlap with both irritable bowel syndrome and Crohn's disease, suggesting that other, more subtle variation in this gene may be important in the development of those disorders.

HudsonAlpha researcher Dr. Shawn Levy contributed to the analysis of this finding.

For details, see page 38 - *Identifying Genetic Influences on Disease*.

REFERENCE: Fiskerstrand et al., "Familial Diarrhea Syndrome Caused by an Activating GUCY2C Mutation," *New England Journal of Medicine* **366:1586 (2012) doi:10.1056/NEJMoa1110132.**

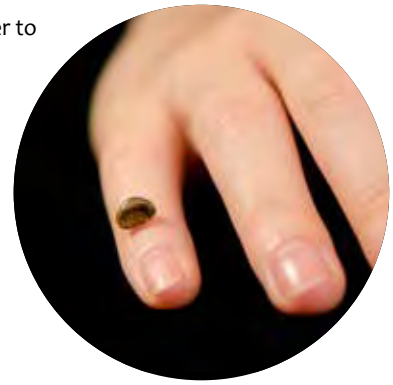
Comparative Genomics

bringing an ancient hominin genome into focus

In 2008, a pinky bone was discovered in Denisova Cave in southern Siberia. DNA sequencing produced a draft genome that suggested the individual, a female who lived over 50,000 years ago, was neither Neandertal nor modern human, but a member of a distinct lineage that came to be known as Denisovans. Additional details were scarce - the ancient DNA was heavily fragmented, resulting in a low quality genome analysis. A new technological approach allows sequencing from single strands of DNA, rather than the previously required double strands. Consequently, additional genetic information can be deciphered from fragmented DNA. A more complete version of the Denisovan genome reveals the woman likely had dark skin, brown eyes and hair. It also identifies more than 100,000 single nucleotide differences between the Denisovan genome and modern humans. Only a handful alter the amino acid sequence in a protein. Intriguingly, several are involved in brain function and development.

For more information, refer to *Comparative Genomics* on page 31.

REFERENCE: Meyer et al., "A High-Coverage Genome Sequence from an Archaic Denisovan Individual," *Science*, **published online August 30, 2012. doi: 10.1126/science.1224344.**



In brief

Plant Genomes

genomic details about agriculturally-important crops

	Genome Size (Million bases)	Number of Genes
Cucumber	200	21,000
Peach	230	28,000
Strawberry	240	35,000
Orange	325	25,000
Foxtail millet	410	35,000
Cacao	430	29,000
Rice	450	41,000
Grape	490	30,000
Sorghum	730	28,000
Potato	844	39,000
Tomato	1,000	35,000
Soybean	1,200	46,000
Maize	2,300	33,000
Bread wheat	17,100	unknown
for comparison		
human	*3,200*	*23,000*

REFERENCE: Morrell et al., "Crop genomics: advances and applications," *Nature Reviews Genetics* **13:85 (2012) doi: 10.1038/hrg3097.**

Fungal Genomics

white rot and worldwide coal deposits

Coal is the fossilized remains of plants that lived during the Carboniferous period 360 million - 290 million years ago. During this window, trees with bark made from a carbon-containing molecule called lignin grew in vast swamp forests. When the trees died, the lignin was absorbed into the soil, ultimately giving rise to vast deposits of coal in modern-day Europe, North America and Asia.

A new study suggests the evolution of a type of fungi known as white rot may have ended this period of coal formation. This form of fungi is capable of breaking down lignin. By sequencing the genomes of several fungi species, the specific genes involved in wood decay could be compared. A technique called molecular clock analysis tracked the evolution of these genes back through fungal lineages. The findings suggest that the oldest fungal ancestor with the ability to degrade lignin first appears near the end of the Carboniferous period. According to the theory, as this white rot spread, it destroyed much of the decaying plant material that would have otherwise been fossilized as coal.

In an interesting twist, these same fungi are being studied as potential players in the development of new biofuels. White rot could be used to break down lignin in plants, releasing cellulose from cell walls. This can be further degraded into sugars and fermented into the alcohols that form the basis of biofuel.

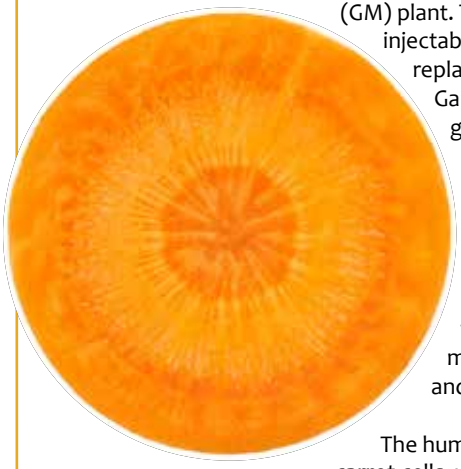
The HudsonAlpha Genome Sequencing Center, led by Jeremy Schmutz and Dr. Jane Grimwood, was part of the white rot sequencing study.

REFERENCE: Floudas et al., "The Paleozoic Origin of Enzymatic Lignin Decomposition Reconstructed from 31 Fungal Genomes," *Science* **336:1715 (2012) doi: 10.1126/science.1221748.**

Biopharming for disease treatment

using plant cell cultures to produce medication

The U.S. Food and Drug Administration has approved the first human medication produced by a genetically modified (GM) plant. The drug, Elelyso™ is an injectable form of an enzyme replacement used to treat Gaucher disease, a rare genetic disorder where patients lack the enzyme glucocerebrosidase, which typically is produced in the lysosome of cells. Without the enzyme, patients experience a toxic accumulation of fatty material in the liver, spleen and other cells.



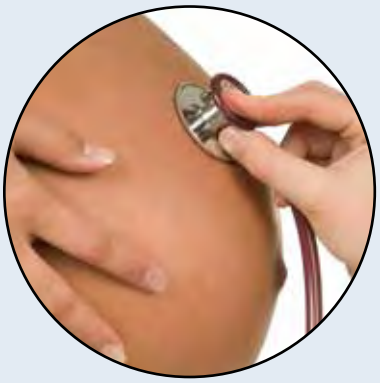
The human gene was inserted into carrot cells, which are grown as cultures of cells, rather than fully formed carrots. The cells are maintained in a water-based growth broth inside bioreactors made from large sterile plastic containers. The protein is then purified from the cells.

Similar approaches have been used to produce proteins from bacterial, bird or mammalian cells for decades. The first human-like enzyme produced in plants was from tobacco in 1992, although it has taken 20 years to develop an approved pharmaceutical. Proponents argue that plant-based systems are

Low-risk prenatal DNA tests available

studying fetal genomes from within the mother's blood

This year witnessed the launch of non-invasive prenatal testing using cell-free fetal DNA from maternal blood. During pregnancy, a small fraction of the cells from the placenta undergo a programmed cell death, continuously releasing fetal DNA into the maternal bloodstream. Using approximately a teaspoon of maternal blood, the genome of the fetus can be sequenced.



In a follow-up to an article from the *2011 Biotechnology Guidebook*, at least three noninvasive tests are now commercially available to detect Down syndrome and other common fetal trisomies using this approach. The tests detect chromosomal anomalies as early as ten weeks into the pregnancy and have been developed by companies such as Sequenom, Ariosa Diagnostics, Verinata Health and Natera. Current pricing ranges from $800 to over $2500, depending on the test.

It is not clear how these tests will be integrated into the current prenatal screening and testing pipeline. Non-invasive testing approaches reduce the risk of miscarriage compared to more invasive tests such as amniocentesis or chorionic villus sampling.

easier to maintain and carry lower production costs - two key challenges with traditional pharmaceutical manufacturing. For example, the two other manufacturers of the Gaucher enzyme use mammalian cells, which require more complicated culture conditions and additional protein purification steps to remove infectious viruses that may be present. Manufacturing challenges related to these factors led to temporary shortages in supply of the replacement enzyme in 2011.

Even with the challenges posed by mammalian-produced proteins, the path to regulatory approval of Elelyso™ was complex. Concerns over the risk of allergic reactions due to differences in how plants process proteins held up approval for months. A number of similar GM plant medications are in development or clinical trials, including flu vaccines produced in the leaves of tobacco plants and a hepatitis C therapy from duckweed. Now that the first plant-based GM drug has been approved, these companies hope the precedent has been set.

For more information, refer to *Recombinant DNA and Genetic Engineering* on page 46.

REFERENCE

Maxmen, A., "First plant-made drug on the market," *Nature News Blog*. Posted May 2, 2012. <http://blogs.nature.com/news/2012/05/first-plant-made-drug-on-the-market.html>.

At the same time, bioethicists note that the availability of these procedures may ultimately lead to parental testing and decision-making on the basis of less serious conditions. How much genetic information will expectant couples want to know? Is that different from the amount of information they should be allowed to know? As these tests expand beyond trisomic conditions and into deciphering the entire fetal genome, genetic susceptibilities for late-onset disorders will be identified. Does a pregnant woman have the right to obtain all of this information? At what point does the future child's "right not to know" override a parent's desire to have the genetic data? Last, but certainly not least, the time and resources required to interpret and provide counseling on a genome's worth of noteworthy genetic information will be substantial. Consequently, there must be a major shift in the way genetic education is offered to healthcare professionals and to the public.

For more information and a deeper discussion of the ethical challenges associated with this topic, refer to *Non-invasive Prenatal Diagnosis* on page 40.

REFERENCES

Bianchi, D.W., "Fetal genes in mother's blood," *Nature* 487:304, doi:10.1038/487304a.
Netzer et al., "To know or not to know the genomic sequence of a fetus" *Nature Reviews Genetics* published online September 4, 2012, doi:10.1038/nrg3333.

Antibiotic resistance in bacteria

genomic tools trace its evolution and track an outbreak

How resistance occurs: To study the process by which bacteria develop antibiotic resistance, researchers grew *Escherichia coli* in the presence of increasing antibiotic concentrations. At the end of each test, whole genome sequencing identified mutations associated with resistance pathways. Resistance to doxycycline and chloramphenicol, which inhibit protein synthesis, occurred through mutations in genes that encode membrane proteins or proteins integral to the machinery of transcription or translation. Multiple pathways were identified, with each requiring a small number of mutations across a diverse set of genes.

In contrast, resistance to trimethoprim arose almost solely from mutations in the gene for dihydrofolate reductase, an enzyme needed to produce the DNA nucleotide thymidine. By competitively binding to this enzyme, the antibiotic reduces thymidine production and prevents bacterial replication. The resistance-causing mutations changed a specific set of amino acids, altering the enzyme's shape so it no longer bound the antibiotic.

Mapping an outbreak: Bacterial infections from *Klebsiella pneumoniae* are often found in hospital intensive care units. In 2011, a patient known to have a form of carbapenem-resistant *K. pneumonia* (KPC) was admitted to the U.S. National Institute of Health Clinical Center. The patient was isolated and appropriate infection-control precautions were taken. Three weeks after this patient left the hospital, another individual tested positive for KPC. Over a 10-month window, nineteen patients were infected. Seven subsequently died from KPC-related complications.

Whole-genome sequencing compared KPC sequences across patients to track the outbreak. Because bacteria rapidly grow and divide, genetic changes can be introduced across many rounds of DNA replication. These changes can differentiate strains. The initial patient had been infected for several months and three genetically distinct strains had developed at various places on her body. Analyses of the genetic sequence from other patients gave KPC signatures that matched two of these three strains, meaning the infection was transmitted from the first patient on more than one occasion. As the outbreak spread, additional genetic changes rendered KPC resistant to other antibiotics, leaving some patients with no effective treatment options. The KPC strain was also found in six sink drains and a ventilator that had been used by one of the patients. The ventilator had been cleaned three separate times, yet the pathogen had survived the cleaning procedures.

For more information, refer to *Infectious Disease* on page 39.

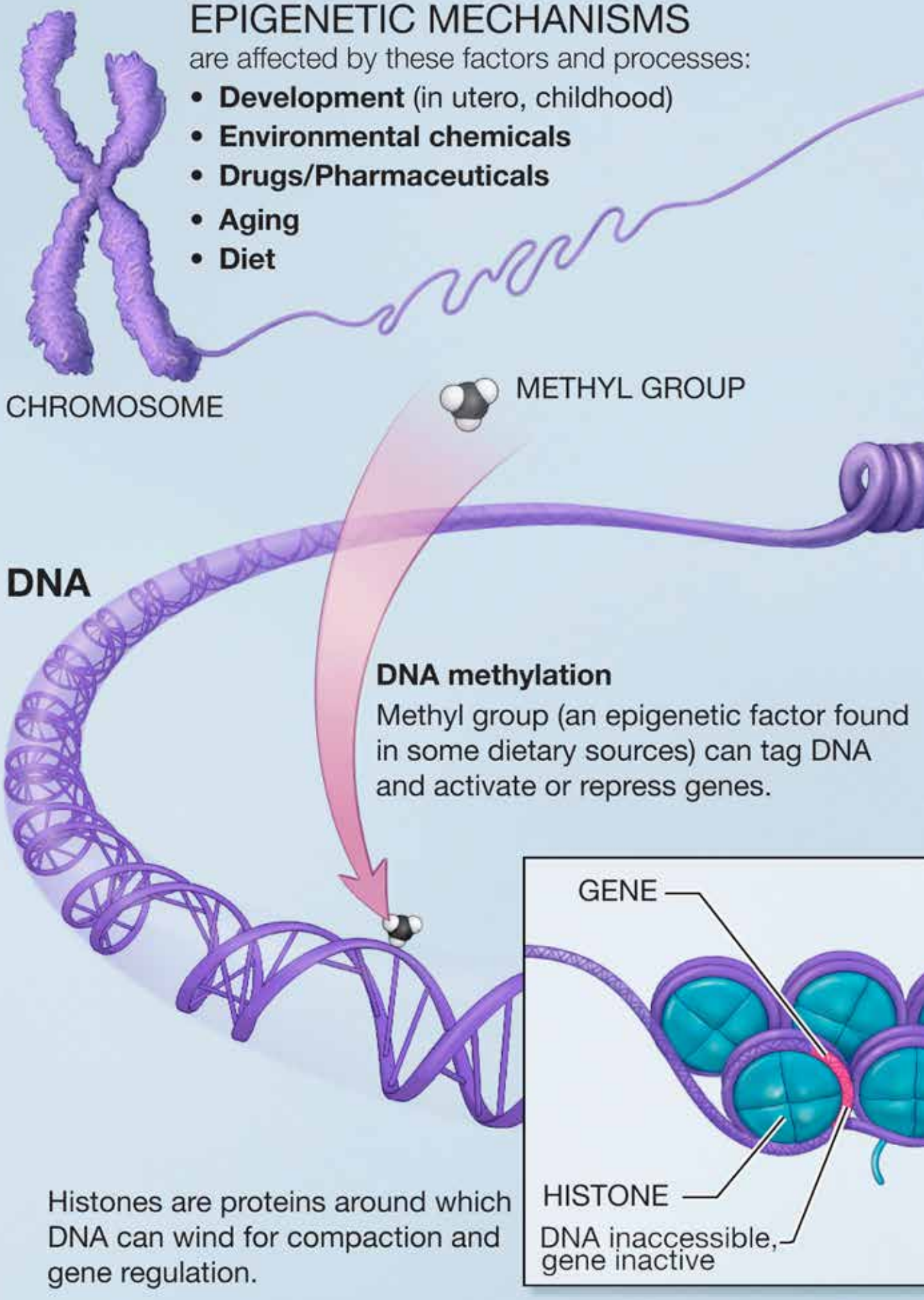
REFERENCES

Toprak et al., "Evolutionary paths to antibiotic resistance under dynamically sustained drug selection," *Nature Genetics* 44:101 (2012) doi: 10.1038/ng.1034.
Snitkin et al., "Tracking a Hospital Outbreak of Carbapenem-Resistant Klebsiella pneumonia with Whole-Genome Sequencing," *Science Translational Medicine* 4:148ra116 (2012), doi: 10.1126/scitranslmed3004129.

EPIGENETIC MECHANISMS

are affected by these factors and processes:

- **Development** (in utero, childhood)
- **Environmental chemicals**
- **Drugs/Pharmaceuticals**
- **Aging**
- **Diet**



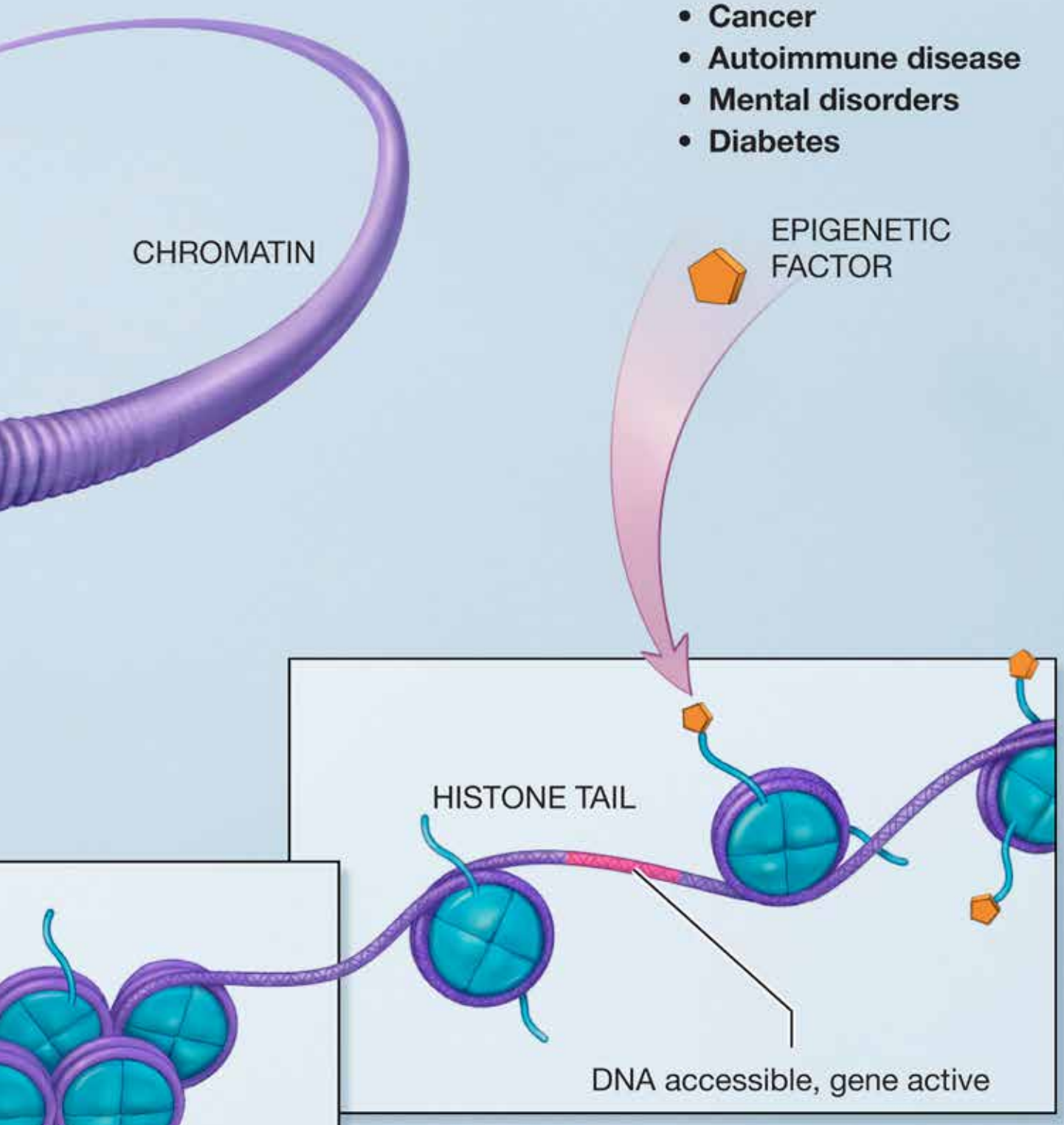
DNA methylation

Methyl group (an epigenetic factor found in some dietary sources) can tag DNA and activate or repress genes.

Histones are proteins around which DNA can wind for compaction and gene regulation.

HEALTH ENDPOINTS

- **Cancer**
- **Autoimmune disease**
- **Mental disorders**
- **Diabetes**



Histone modification

The binding of epigenetic factors to histone "tails" alters the extent to which DNA is wrapped around histones and the availability of genes in the DNA to be activated.

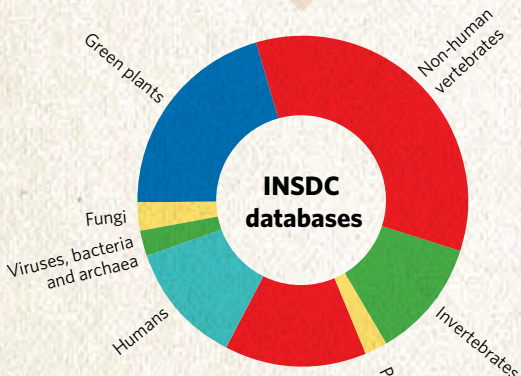
THE SEQUENCE EXPLOSION

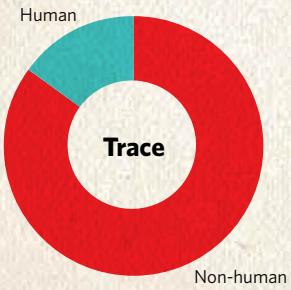
At the time of the announcement of the first drafts of the human genome in 2000, there were 8 billion base pairs of sequence in the three main databases for 'finished' sequence: GenBank, run by the US National Center for Biotechnology Information; the DNA Databank of Japan; and the European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database. The databases share their data regularly as part of the International Nucleotide Sequence Database Collaboration (INSDC). In the subsequent first post-genome decade, they have added another 270 billion bases to the collection of finished sequence, doubling the size of the database roughly every 18 months. But this number is dwarfed by the amount of raw sequence that has been created and stored by researchers around the world in the Trace archive and Sequence Read Archive (SRA).

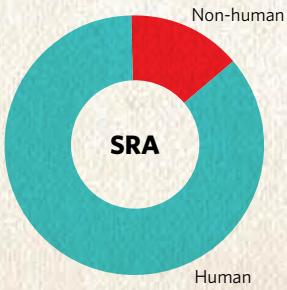
See Editorial, page 649, and human genome special at www.nature.com/humangenome

DNA SEQUENCES BY TAXONOMY

International Nucleotide Sequence Database Collaboration: The main repositories of 'finished' sequence span a wide range of organisms, representing the many priorities of scientists worldwide.

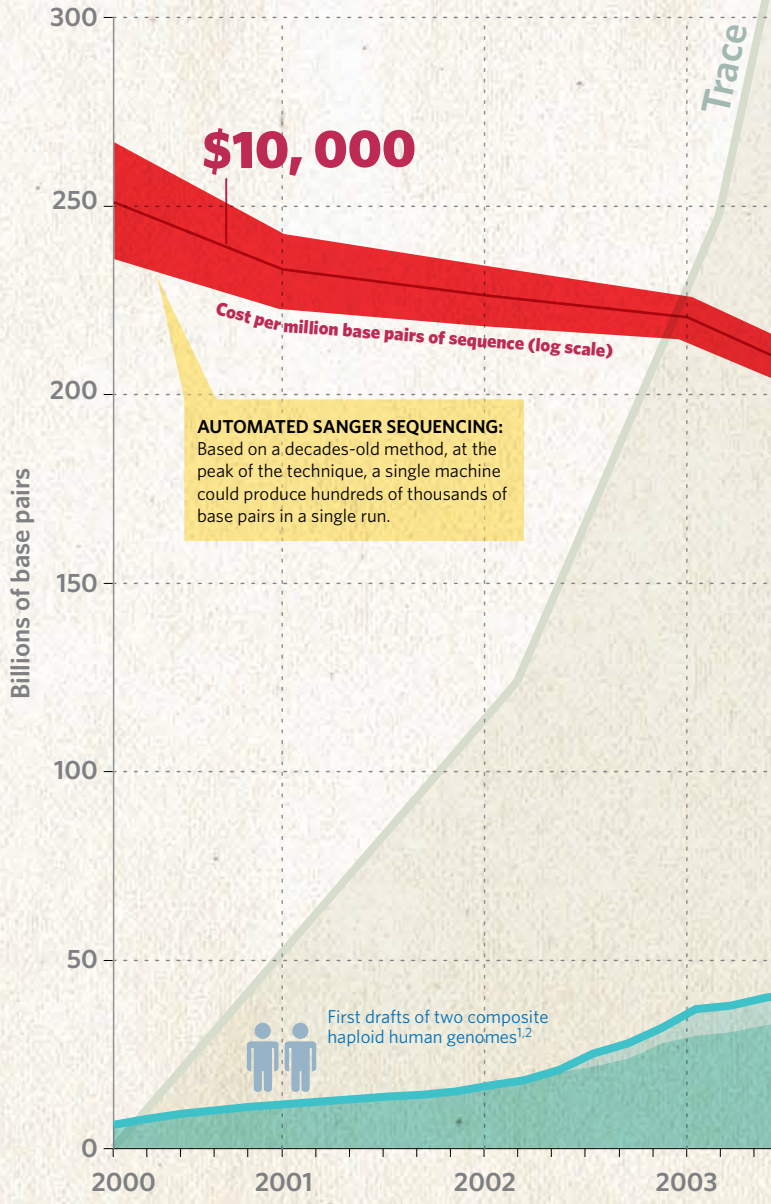






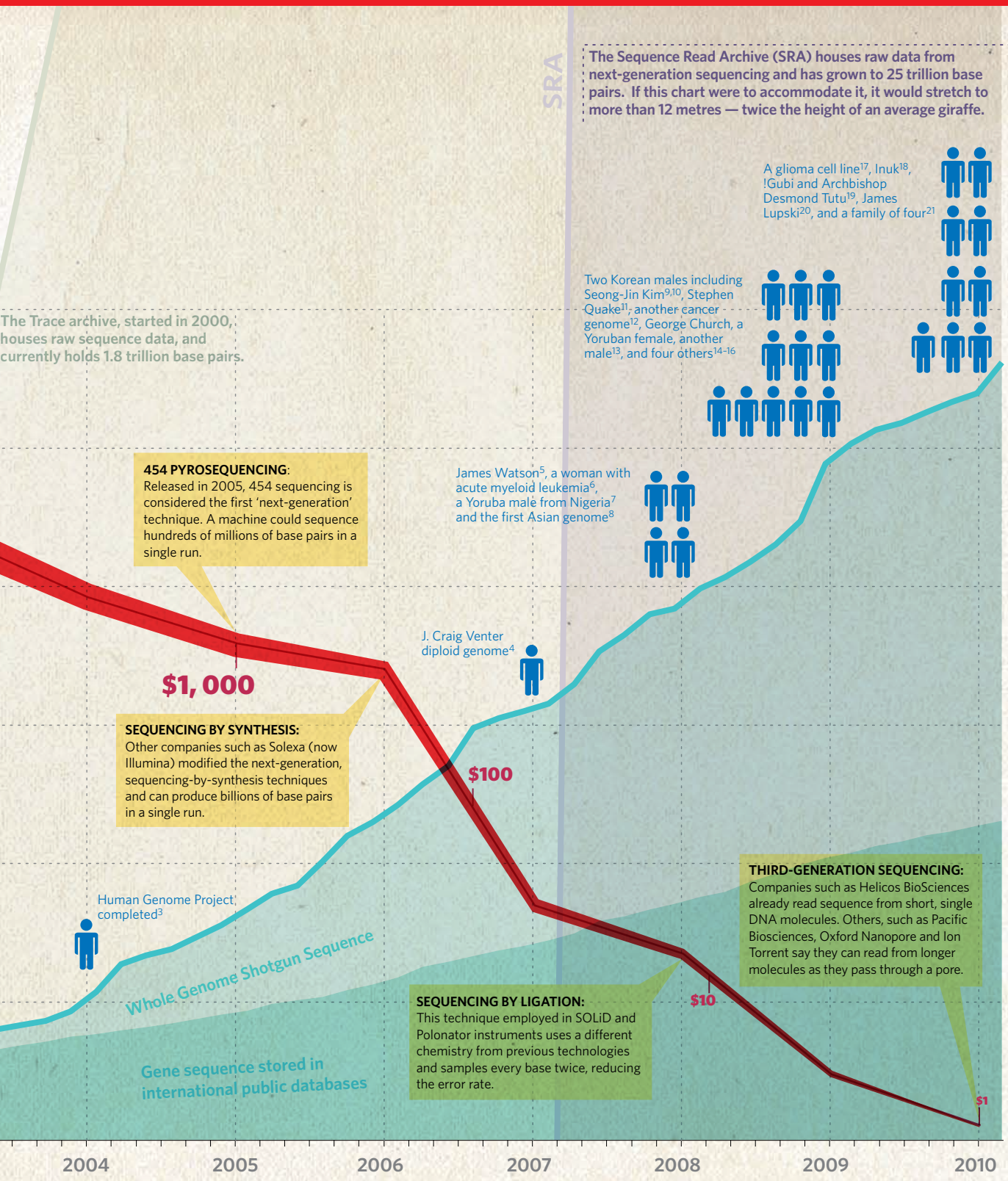
Trace Archive: Developed to house the raw output of high-throughput sequencers built in the late 1990s, the trace archive spans a wide range of taxa.

Sequence Read Archive: Houses raw data from next-generation sequencers. Dominated by human sequence, including multiple coverage for more than 170 people.



HOW MANY HUMAN GENOMES?

The graphic shows all published, fully sequenced human genomes since 2000, including nine from the first quarter of 2010. Some are resequencing efforts on the same person and the list does not include unpublished completed genomes.



SOURCE: NCBI; GRAPHICS BY N. SPENCER & W. FERNANDES

1. Venter, J. C. *et al. Science* **291,** 1304–1351 (2001).
2. International Human Genome Sequencing Consortium *Nature* **409,** 860–921 (2001).
3. International Human Genome Sequencing Consortium *Nature* **431,** 931–945 (2004).
4. Levy, S. *et al. PLoS Biol.* **5,** e254 (2007).
5. Wheeler, D. A. *et al. Nature* **452,** 872–876 (2008).
6. Ley, T. J. *et al. Nature* **456,** 66–72 (2008).
7. Bentley, D. R. *et al. Nature* **456,** 53–59 (2008).
8. Wang, J. *et al. Nature* **456,** 60–65 (2008).
9. Ahn, S.-M. *et al. Genome Res.* **19,** 1622–1629 (2009).
10. Kim, J.-I. *et al. Nature* **460,** 1011–1015 (2009).
11. Pushkarev, D., Neff, N. F. & Quake, S. R. *Nature Biotechnol.* **27,** 847–850 (2009).
12. Mardis, E. R. *et al. N. Engl. J. Med.* **10,** 1058–1066 (2009).
13. Drmanac, R. *et al. Science* **327,** 78–81 (2009).
14. McKernan, K. J. *et al. Genome Res.* **19,** 1527–1541 (2009).
15. Pleasance, E. D. *et al. Nature* **463,** 191–196 (2010).
16. Pleasance, E. D. *et al. Nature* **463,** 184–190 (2010).
17. Clark, M. J. *et al. PLoS Genet.* **6,** e1000832 (2010).
18. Rasmussen, M. *et al. Nature* **463,** 757–762 (2010).
19. Schuster, S. C. *et al. Nature* **463,** 943–947 (2010).
20. Lupski, J. R. *et al. N. Engl. J. Med.* doi:10.1056/NEJMoa0908094 (2010).
21. Roach, J. C. *et al. Science* doi:10.1126/science.1186802 (2010).

Page size by comparison

COURSE OF STUDY CONNECTED TO GUIDEBOOK TOPICS

Course		Objective and Applicable Subheading	Linking Scientific Concept
Biology	2	Describe cell processes necessary for achieving homeostasis, including active and passive transport, osmosis, diffusion, exocytosis, and endocytosis.	
		Identifying functions of carbohydrates, lipids, proteins, and nucleic acids in cellular activities	RNA and Protein Analysis
	4	Describe similarities and differences of cell organelles, using diagrams and tables.	See HudsonAlpha iCell (pg 4)
		Identifying scientists who contributed to cell theory	Stem Cells, See also Biotechnology Timeline (pg 4)
	5	Identifying cells, tissues, organs, organ systems, organisms, populations, communities, and ecosystem as levels of organization in the biosphere.	
		Recognizing that cells differentiate to perform specific functions	Comparative Genomics, RNA and Protein Analysis, Stem Cells
	6	Describe the roles of mitotic and meiotic divisions during reproduction, growth, and repair cells.	Cancer, Stem Cells
		Comparing sperm and egg formation in terms of ploidy	Diagnosing Chromosomal Disorders, Noninvasive Prenatal Diagnosis
	7	Apply Mendel's law to determine phenotypic and genotypic probabilities of offspring.	Genetics of Eye Color
		Defining important genetic terms, including dihybrid cross, monohybrid cross, phenotype, genotype, homozygous, heterozygous, dominant trait, recessive trait, incomplete dominance, codominance, and allele	Epigenetics
		Interpreting inheritance patterns shown in graphs and charts	Cancer
	8	Identify the structure and function of DNA, RNA and Protein.	RNA and Protein Analysis, Recombinant DNA and Genetic Engineering, Therapeutic Approaches
		Explaining relationships among DNA, genes and chromosomes	Diagnosing Chromosome Disorders, Noninvasive Prenatal Diagnosis,Recombinant DNA and Genetic Engineering, Studying the Genome to Understand the Sequence
		Listing significant contributions of biotechnology to society, including agricultural and medical practices	Agricultural Applications, Cancer, DNA sequencing, Genetic Information Nondiscrimination Act, Noninvasive Prenatal Diagnosis, Personal Genomic Analysis, Personalized Medicine, Pharmacogenomics, Recombinant DNA and Genetic Engineering,RNA and Protein Analysis, Stem Cells, Synthetic Biology, See also Biotechnology Timeline (pg 4)
		Relating normal patterns of genetic inheritance to genetic variation	Cancer, Comparative Genomics, Copy Number Variation, Identifying the Genetic Influences on Disease, Personalized Medicine
		Relating ways chance, mutagens and genetic engineering increase diversity	Agricultural Applications, Cancer, Diagnosing Chromosomal Disorders, Epigenetics, Personal Genomic Analysis, Studying the Genome to Understand the Sequence

Course		Objective and Applicable Subheading	Linking Scientific Concept
Biology	8	Relating genetic disorders and disease to patterns of genetic inheritance.	Identifying Genetic Influence on Disease
	9	Differentiate between the previous five kingdom and current six kingdom classification system.	
		Identifying ways in which organisms from the Monera, Protista, and Fungi Kingdoms are beneficial and harmful	Infectious Disease
		Justifying the grouping of viruses in a category separate from living things	Infectious Disease
	12	Describe protective adaptations of animals, including mimicry, camouflage, beak type, migration, and hibernation.	
		Identifying ways in which the theory of evolution explains the nature and diversity of organisms	Comparative Genomics
		Describing natural selection, survival of the fittest, geographic isolation, and fossil record	Comparative Genomics
Environmental Science	9	Describe land-use practices that promote sustainability and economic growth.	Agricultural Applications
Forensic Science	4	Describe presumptive and confirmatory tests.	Criminal Justice and Forensics, DNA Sequencing
	5	Describe the importance of genetic information to forensics.	Criminal Justice and Forensics, DNA Sequencing
Genetics	2	Describe factors such as radiation, chemicals, and chance that cause mutations in populations.	Cancer, Comparative Genetics, Identifying Genetic Influence on Disease, Infectious Disease, Studying the Genome to Understand the Sequence
		Describing effects of genetic variability on adaptations	Agricultural Applications, Comparative Genomics, Copy Number Variation, Criminal Justice and Forensics, RNA and Protein Analysis
	4	Describe the process of meiosis and the cell cycle, including the hereditary significance of each.	Cancer, Diagnosing Chromosomal Disorders, Noninvasive Prenatal Diagnosis, Stem Cells
	5	Describe inheritance patterns based on gene interactions.	Diagnosing Chromosomal Disorders, Epigenetic, Genetics of Eye Color, Identifying Genetic Influence on Disease
		Identifying incomplete dominance, codominance, and multiple allelism	Copy Number Variation, Epigenetics
	6	Describe occurrences and effects of sex linkage, autosomal linkage, crossover, multiple alleles, and polygenes.	Epigenetics, Identifying Genetic Influence on Disease, RNA and protein analysis
	7	Describe the structure and function of DNA, including replication, translation, and transcription.	DNA Sequencing, Recombinant DNA and Genetic Engineering, RNA and Protein Analysis
		Describing methods cells use to regulate gene expression	Comparative Genomics, Epigenetics, Recombinant DNA and Genetic Engineering, RNA and Protein analyses, Therapeutic Approaches
		Defining the role of RNA in protein synthesis	Recombinant DNA and Genetic Engineering, RNA and Protein analyses, Therapeutic Approaches

Course		Objective and Applicable Subheading	Linking Scientific Concept
Genetics	8	Explain the structure of eukaryotic chromosomes, including transposons, introns, and exons.	Bioinformatics, Diagnosing Chromosomal Disorders, Studying the Genome to Understand the Sequence
	9	Differentiate among major areas in modern biotechnology, including plant, animal, microbial, forensic, and marine.	Agricultural Applications, Bioinformatics, Cancer, Criminal Justice and Forensics, DNA Sequencing, Personalized Medicine, Pharmacogenomics, Recombinant DNA and Genetic Engineering, RNA and Protein Analysis
		Describing techniques used with recombinant DNA	Agricultural Applications, Recombinant DNA and Genetic Engineering, RNA and Protein Analyses
	10	Explain the development and purpose of the Human Genome Project.	Bioinformatics, Criminal Justice and Forensics, DNA Sequencing, Identifying Genetic Influence on Disease, Studying the Genome to Understand the Sequence; See also Biotechnology Timeline (pg 4)
		Analyzing results of the Human Genome Project to predict ethical, social, and legal implications.	Cancer, Copy Number Variation, Criminal Justice and Forensics, Genetic Information Nondiscrimination Act, Personal Genomic Analysis, Personalized Medicine, Pharmacogenomics, Therapeutic Approaches
		Describing medical uses of gene therapy, including vaccines and tissue and antibody engineering.	DNA Sequencing, Infectious Disease, RNA and Protein Analysis, Therapeutic Approaches
AP Biology	II	Evolution	Agricultural Applications, Comparative Genomics
	IV	Continuity and Change	Agricultural Applications, Bioinformatics, Cancer, Comparative Genmoics, Copy Number Variation, Criminal Justice and Forensics, DNA Sequencing, Genetics of Eye Color, Identifying Genetic Influence on Disease, Stem Cells, Studying the Genome to Understand the Sequence
	V	Relationship of Structure to Function	Epigenetics, RNA and Protein Analysis, Recombinant DNA and Genetic Engineering, Studying the Genome to Understand the Sequence
	VI	Regulation	Cancer, Copy Number Variation, Epigenetics, RNA and Protein Analyses
	VIII	Science, Technology and Society	Agricultural Applications, Cancer, Comparative Genomics, DNA Sequencing, Genetic Information Nondiscrimination Act, Identifying Genetic Influence on Disease, Noninvasive Prenatal Diagnosis, Personalized Medicine, Personal Genomic Analysis, Pharmacogenomics, Recombinant DNA and Genetic Engineering, Therapeutic Approaches, Synthetic Biology
Health	5	Evaluate negative and positive impacts of technology on health.	Agricultural Applications, Cancer, Identifying Genetic Influence on Disease, Noninvasive Prenatal Diagnosis, Personalized medicine, Pharmacogenomics, Recombinant DNA and Genetic Engineering, Stem Cells, Synthetic Biology
	6	Discuss valid and essential information for the safe use of consumer goods and health products.	Agricultural Applications, Cancer, Noninvasive Prenatal Diagnosis, Personal Genomic Analysis, Pharmacogenomics
	10	Determine the causes of disability and premature loss of life across life stages.	Cancer, Identifying Genetic Influence on Disease

Course		Objective and Applicable Subheading	Linking Scientific Concept
Technology Education	26	Explain uses and advantages of databases.	Bioinformatics
	27	Apply appropriate techniques for producing databases.	Bioinformatics
Agriscience	10	Determine characteristics and functions of plants.	
		Explain how agricultural crops can be utilized as alternative fuel sources	Agricultural applications
Forensic and Criminal Investigations	7	Describe presumptive and confirmatory forensic tests. Examples: blood type comparison, DNA testing	Criminal Justice and Forensics
	8	Describe the importance of genetic information to forensics Using the process of gel electrophoresis for deoxyribonucleic acid (DNA) fingerprinting.	Bioinformatics, Criminal Justice and Forensics
Foundations of Health Sciences	10	Recognize legal responsibilities, limitations, and implications within the health care delivery setting. Examples: Patients' Bill of Rights, legal documentation requirements, Health Insurance Portability and Accountabilitiy Act (HIPPA)	Genetic Information Nondiscrimination Act, Personal Genome Analysis
Health Informatics	5	Describe legal and ethical regulations as they relate to health informatics. Examples: Patients' Bill of Rights, legal documentation requirements, Health Insurance Portability and Accountabilitiy Act (HIPPA)	Genetic Information Nondiscrimination Act, Personal Genome Analysis
Introduction to Agriscience	16	Analyze biotechnology to determine benefits to the agriculture industry. Example: Improved productivity, medical advancements, environmental benefits	Agricultural Applications, Bioinformatics, Recombinant DNA and Genetic Engineering
Introduction to Pharmacy	9	Identify classifications of selected drugs. Examples: analgesic, antibiotic, antiemetic	Personalized Medicine, Pharmacogenomics
	11	Differentiate among drug interactions, drug reactions, and side effects.	Personalized Medicine, Pharmacogenomics
Introduction to Biotechnology	1	Trace the history of biotechnology.	See also Biotechnology Timeline (pg 4)
		Describing both scientific and non-scientific careers, roles, and responsibilities of individuals working in biotechnology.	Agricultural Applications, Bioinformatics, Criminal Justice and Forensics, Diagnosing Chromosome Disorders, DNA Sequencing, Pharmacogenomics, See also Biotechnology Timeline (pg 4)
	4	Correlate key cellular components to function.	See HudsonAlpha iCell (pg 4)
	5	Describe the process of meiosis and the cell cycle, including the hereditary significance of each.	Cancer, Diagnosing Chromosome Disorders, Noninvasive Prenatal Diagnosis, Stem Cells,
	8	Describe occurrences and effects of sex linkage, autosomal linkage, crossover, multiple alleles, and polygenes.	Cancer, Copy Number Variation, Genetics of Eye Color, Identifying Genetic Influence on Disease
	9	Describe the structure and function of deoxyribonucleic acid (DNA), including replication, translation, and transcription.	Recombinant DNA and Genetic Engineering, Studying the Genome to Understand the Sequence
		Applying the genetic code to predict amino acid sequence	Bioinformatics

Course		Objective and Applicable Subheading	Linking Scientific Concept
Introduction to Biotechnology	9	Describe methods cells use to regulate gene expression.	Cancer, Comparative Genomics, Epigenetics, RNA and Protein Analysis, Therapeutic Approaches
		Defining the role of ribonucleic acid (RNA) in protein synthesis	Recombinant DNA and Genetic Engineering, RNA and Protein Analysis, Therapeutic Approaches
	11	Describe factors such as radiation, chemicals and chance that cause mutations.	Cancer, Infectious Disease
	13	Differentiate among major areas in modern biotechnology, including plant, animal, microbial, forensic, and marine.	Agricultural Applications, Bioinformatics, Criminal Justice and Forensics, DNA Sequencing, Infectious Disease
		Describing techniques used with recombinant DNA	Agricultural Applications, DNA Sequencing, Synthetic Biology
	14	Explain the development, purpose, findings, and applications of the Human Genome Project.	Comparative Genomics, Copy Number Variation, DNA Sequencing, Identifying Genetic Influence in Disease, Personalized Medicine, Pharmacogenomics, Studying the Genome to Understand the Sequence
		Analyzing results of the Human Genome project to predict ethical, social and legal implications	Criminal Justice and Forensics, Genetic Information Nondiscrimination Act, Personalized Genomic Analysis
		Describing medical uses of gene therapy, including vaccines and tissue and antibody engineering.	Cancer, DNA Sequencing, Infectious Disease, Recombinant DNA and Genetic Engineering, RNA and Protein Analysis
		Using computer bioinformatics resources to provide information regarding DNA, protein, and human genetic diseases	Bioinformatics, Cancer, Comparative Genomics, Copy Number Variation
	15	Describe the replication of DNA and RNA viruses, including lytic and lysogenic cycle.	Infectious Disease
Plant Biotechnology	1	Identify career opportunities associated with plant biotechnology.	Agricultural Applications
	14	Describe the ecological and economic importance of plants.	Agricultural Applications
	16	Explain the historical development of plant biotechnology.	Agricultural Applications
		Identify medical advancements in plant biotechnology	Agricultural Applications, Comparative Genomics
		Describing environmental advancements in plant biotechnology	Agricultural Applications; See also Biotechnology Timeline (pg 4)
	17	Describe methods of genetic engineering.	Agricultural Applications

FOUNDATIONAL CONCEPTS AND THEIR APPLICATIONS

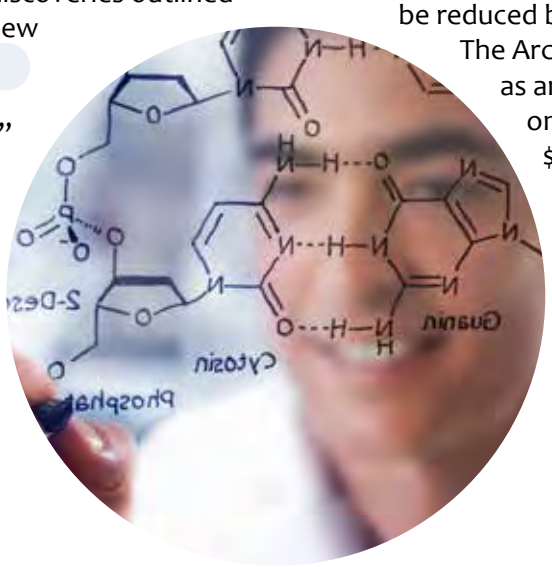
DNA Sequencing

In 1977 Fred Sanger and Alan Coulson published a method to rapidly determine the specific order of the adenine, thymine, cytosine and guanine nucleotides in any DNA sequence. This technology ultimately transformed biology by providing a tool for deciphering complete genes and later entire genomes. Improvements in process parallelization (running hundreds or thousands of samples simultaneously), automation and analysis led to the establishment of factory-like enterprises, called sequencing centers. These facilities spearheaded the effort to sequence the genomes of many organisms, including humans.

Today, the need for even greater sequencing capability at a more economical price has led to the development of new technologies based on different chemistries and refined for accuracy and speed. These "second generation" approaches reduce the necessary volume of reagents while dramatically increasing the number of simultaneous sequencing reactions in a single experiment. They are capable of producing nearly 150 times more sequence than the first generation systems, at 1/150th the cost. For example, the cost of sequencing all 3 billion letters in the human genome has dropped from $15,000,000 to less than $10,000.

The ability to quickly and economically decipher large swaths of DNA has opened doors to research previously deemed out of reach. Many of the discoveries outlined in this guide are in part due to this new technology.

The first so-called "third generation" sequencing system debuted in 2009, producing an entire human sequence. Based on the analysis of a single molecule of DNA, a major technological improvement, it is believed that these systems will become widespread within the next 2-3 years, further decreasing sequencing costs.



Looking towards the third (and fourth) generation sequencing systems, there remains a long list of necessary improvements. Chief among them is cost reduction: in order to deliver on the goal of sequencing a human genome for $1,000, sequencing costs must be reduced by an order of 1-2 magnitudes. The Archon Genomics X-Prize serves as an incentive to groups working on sequencing technologies: a $10M prize to the first group who sequences 100 human genomes in 30 days or less at a per genome cost of no more than $1,000. The competition will officially begin in September 2013 and 100 centenarians will donate the DNA to be sequenced.

Second- and third-generation sequencing technologies should be briefly discussed in Biology courses as part of course of study (COS) objective 8, particularly as it relates to significant contributions of biotechnology to society. These topics should be more thoroughly explored in Genetics classes, relating to COS objectives 7, 9 and 10, especially with respect to the impact such technologies have on identifying genetic risks, personalized medicine and pharmacogenomics. They may also be incorporated in the Forensic Science class in preparation for a discussion about DNA phenotyping (see page 8) as part of COS objective 4 and 5 or in an AP Biology course as part of the "Science, Technology and Society" and "Continuity and Change" general themes. This topic would also be appropriate for discussion in the Career/Tech Intro to Biotechnology course as part of objectives 1, 13 and 14.

HudsonAlpha educators have developed a high school lab activity, "Genes & ConSEQUENCES", that connects the information produced by a DNA sequencing system to genes, mutations and disease. The activity incorporates biological databases used by genetic researchers on a daily basis and links changes in DNA sequence to common genetic disorders (see "Bioinformatics" on page 26 for more details). The lab was incorporated into the AMSTI high school program (Science in Motion) statewide during the 2010-2011 academic year.

RNA and Protein Analyses

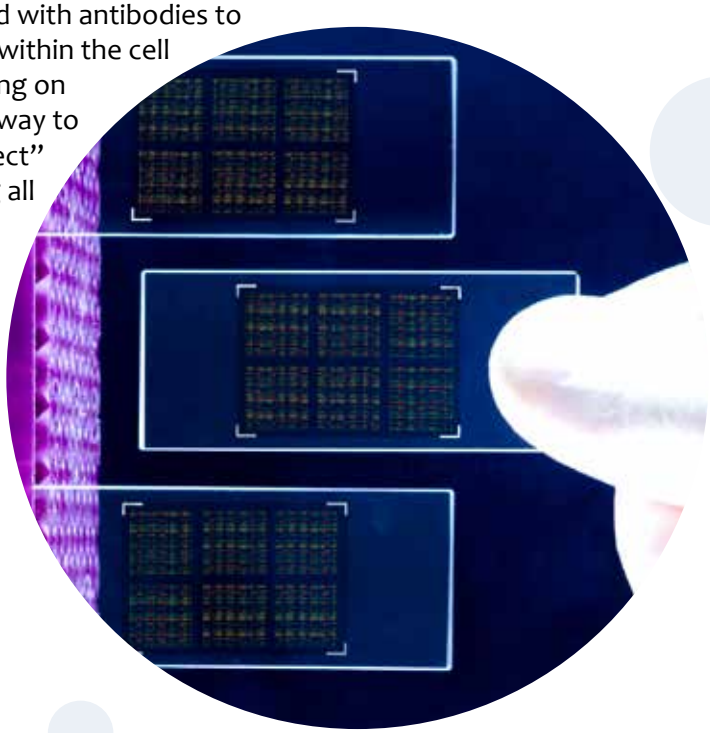
As sequencing techniques identify the genetic recipes of an organism, understanding the function of those genes becomes increasingly important. Many of the fundamental measurements used by molecular biologists seek to determine the presence, absence or relative amounts of RNA produced by a gene. Initially, these approaches examined one or only a handful of RNA sequences at a time. During the last decade, researchers developed techniques to study tens of thousands of RNA fragments simultaneously arrayed on a glass slide. Called "microarrays", these could be used to identify which genes are active or silent in a given cell type, classifying, for example, the genes that distinguish a liver cell from a neuron or the set of genes activated or silenced across different types of cancer.

RNA- and Protein-based technologies should be noted in a Biology course, as it relates to both COS objectives 2, 5 and 8 as they strive to identify the function of proteins and nucleic acids in cellular activities. These technologies can be examined in greater detail for either an AP biology course (under the "Relationship of Structure to Function" and "Regulation" themes) or a Genetics course, where they can be incorporated into activities that describe the occurrence and effects of genetic variability on populations (COS 2 and 6), methods used to regulate gene expression (COS objective 7), techniques using recombinant DNA and antibody engineering (COS objectives 9 and 10). These are also useful technologies to cover in the Career/Tech Intro to Biotechnology course, linking to COS objectives 9 and 14.

Second-generation sequencing technology has recently been extended to also identify RNA expression across cells. Scientists have shown that this approach, known as RNA-seq, yields more precise results than microarray analysis. It is expected that RNA-seq will become the standard tool for measuring genome-wide gene expression.

Many of the fundamental measurements used by molecular biologists seek to determine the presence, absence or relative amounts of RNA produced by a gene.

Large-scale, high-throughput technologies have also been developed to identify protein activity and interactions. This represents part of the emerging field of proteomics, which seeks to understand the entire protein complement (amounts, locations, interactions, and even activities) of an organism's cells. For example, "tissue microarrays", tiny slices of tissue from a single or multiple samples, can be tested with antibodies to identify the locations of proteins within the cell and their relative amounts. Building on these methods, efforts are underway to initiate a "Human Proteome Project" that would systematically catalog all the proteins manufactured in the body. The scale and complexity of this project is much greater than the Human Genome Project as a single gene can direct the production of multiple different versions of a protein and each protein can in turn be modified in a number of different ways.

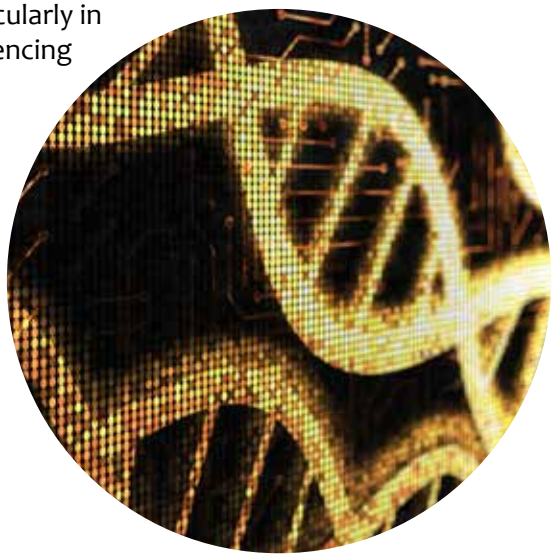


Bioinformatics

Acquiring DNA sequence has now become routine and new technologies can sequence a bacterial genome in a single day. Similarly, microarray experiments shed light on the RNA levels produced by tens of thousands of genes. Current analysis platforms are capable of generating terabytes of data in a single run. For reference, 1 terabyte is equal to 1,000 gigabytes - enough storage space to hold 500 copies of your favorite box office movie or the music libraries from nearly 125 iPod nanos.

Understanding the meaning of all that information is a daunting challenge. Deciphering the data requires a biological knowledge of what to look for, algorithms (computer programs) capable of detecting interesting features, and computers powerful enough to perform complex analyses efficiently and rapidly. Fortunately, advances in all three areas have kept pace and the resulting field of bioinformatics seeks to characterize functional sequences in genes and genomes through computational models. In addition, the data must be managed – stored in a form that is useful to the researcher and readily accessible. This has led to the development of many databases that store and provide data and analytical tools for researchers. The primary mission of all these databases is to provide unlimited free access to anyone, including Alabama students, interested in studying genomic sequences. It is no exaggeration to say that these databases and the immediate access to them through the Internet have changed the way that nearly all biological research is done.

Many bioinformatics experts, particularly in the early days of the genome sequencing efforts, were computer scientists who formed partnerships with biologists. With the growth of the field of genomics, it is not unusual today for a student to be trained in a truly interdisciplinary way by developing deep expertise in both biology and computational science.



The concept of bioinformatics is a critical component to understanding modern genomic discoveries. It provides tools capable of exploring the structure of chromosomes and predicting the likelihood of a genetic match in a forensics case. Bioinformatics databases also manage, search and store the data produced by the human genome project and more recent large-scale studies (Genetics COS objectives 8, 9 and 10). This topic should be incorporated in an AP Biology class under the general theme "Continuity and Change", as well as Career/Tech courses in Forensic and Criminal Investigations (COS objective 8), Introduction to Agriscience (COS objective 16) and Intro to Biotechnology (COS objectives 1, 9, 13 and 14). Lastly, the creation, management and utilization of bioinformatics databases can be incorporated into the Technology Education course (COS objectives 26 and 27).

HudsonAlpha educators have developed a high school lab activity, "Genes & ConSEQUENCES", that connects the information produced by a DNA sequencing system to genes, mutations and disease. The activity incorporates several biological databases used by genetic researchers on a daily basis. Students access a portion of the NCBI (National Center for Biotechnology Information) database known as BLAST. This program compares sequence data entered by the student to known sequences from a number of organisms, including human, and identifies genetic matches. Students then explore their matches on another NCBI database called Genes & Diseases. This dataset allows students to determine the chromosomal location of the gene and its role in disease. The lab has been incorporated into the AMSTI Science in Motion program statewide during the 2009-2010 academic year.

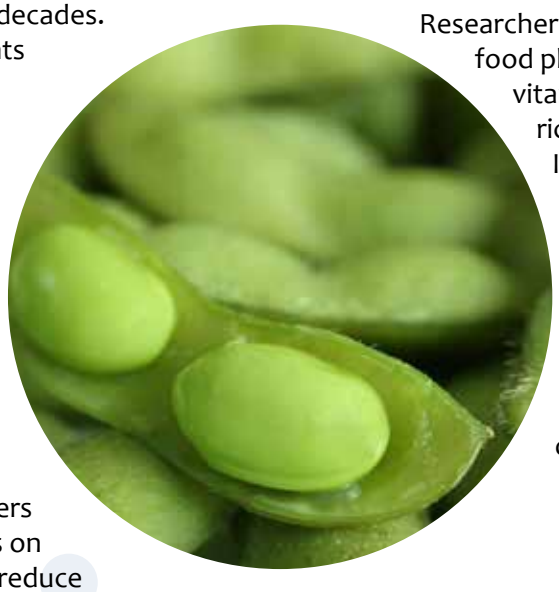
Agriculture

Sequencing Plant Genomes for Food and Bioenergy Needs

Over the last decade, genome sequencing projects have begun for a number of plants, including rice, corn, soybean, canola, and orange. The goal of these sequencing efforts is a better understanding of the underlying genes that contribute to growth rate, seed and fruit characteristics and susceptibility to climate change or infectious agents. In addition, a number of plants have been or are being sequenced for their potential contribution to bioenergy. These include corn, soybean, loblolly pine, poplar and switchgrass. For example, soybean not only accounts for 70% of the world's edible protein, but soybean oil is the principle source of biodiesel. Detailed knowledge of the soybean genome, published in December 2008, allows for crop improvements and better applications of this plant to the generation of clean energy. Knowing which genes control specific traits, researchers could potentially change the type and quantity of oil produced by the crop as well as develop soybean plants that are more resistant to drought or disease.

Genetically Modified (GM) Crops

More than 13 million farmers across 25 countries currently plant biotech crops (also known as genetically modified organisms or GMOs). To date, over two billion acres of biotech crops have been harvested globally. At least 57 different plants have been the focus of biotech research over the last two decades. Of this number, eight different plants are in commercial production, and 15 different plants have received regulatory approval in the United States. Currently, biotech soybean is the principal genetically modified crop worldwide, followed by corn, cotton and canola. Herbicide tolerance has consistently been the primary trait introduced into the crops, followed by insect resistance and the combination of both traits. Biotechnology has enabled producers worldwide to produce higher yields on existing land. Biotechnology crops reduce the need for plowing to control weeds, leading to better conservation of soil and water and a decrease in soil erosion and soil compaction. A reduction in plowing also allows farmers to significantly reduce the consumption of fuel and decrease greenhouse gas emissions. Plants modified to contain genetic variants that confer a degree of drought tolerance, particularly important for developing countries, will become available within the next five years.



Researchers are also developing biofortified food plants to boost the levels of nutrient, vitamins and minerals in foods such as rice, cassava, carrots and tomatoes. It is hoped that these fortified foods will reduce the incidence of global hunger and micronutrient malnutrition (taking in adequate calories, but lacking appropriate vitamins and minerals) which, according to a 2004 United Nations report, impacts up to half of the world's population.

The application of genetic information and Genetically Modified Organisms to increase agricultural yields, improve nutritional content, craft insect resistance or increase bioenergy yields has a direct connection to COS objective 8 for Biology and COS objective 9 for the Environmental Science class. It can also be discussed in a Genetics course (COS objectives 2 and 9) and AP Biology as part of general themes "Evolution", "Continuity and Change" and "Science, Technology and Society". It also has a direct connection to Career/Tech courses in Agriscience (COS objective 10), Intro to Agriscience (COS objective 16) , Intro to Biotechnology (COS objectives 1 and 13) and Plant Biotechnology (COS objectives 1, 14, 16 and 17).

HudsonAlpha has created a lab that allows students to test foods available at their local grocery store (such as chips and cookies) for the presence of genetically modified crops. The test primarily identifies various forms of herbicide resistant corn and soybeans and exposes students to DNA extraction, DNA amplification by the polymerase chain reaction (PCR) and DNA electrophoresis to separate fragments of varying length. Students test foods for both the genetic modification and a control gene from plant cells. This activity is an excellent link between key food safety techniques, foundational biotechnology methods and a subject of interest to all high school students – food. The ethical challenges associated with biotech crops and the varying global viewpoints are also presented for discussion, and the lab links to careers in food safety and inspection. The G-Mod lab is available to all Alabama high school students through the AMSTI Science in Motion program.

Cancer

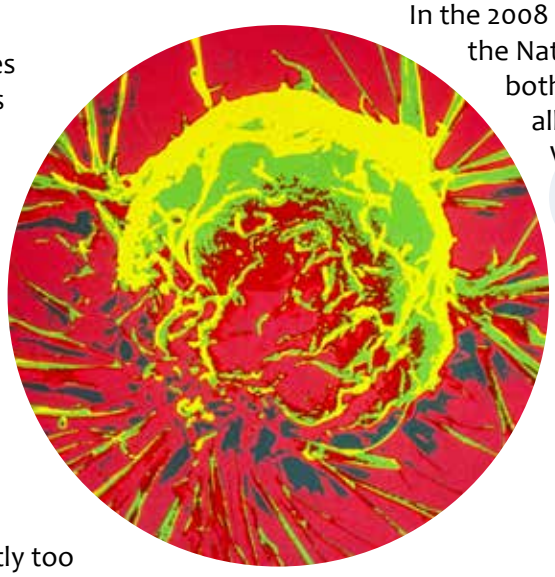
Cancer is a collection of diseases that are characterized by uncontrolled growth of cells and their spread to surrounding tissues. All cancers are genetic diseases, because changes in the genes that control cell growth and division are involved. However, only about 5% of cancers are strongly hereditary – primarily caused by mutations that are inherited from parent to child. Therefore, most cancers do not result from inherited mutations, but instead develop from an accumulation of DNA damage acquired during our lifetime. These cancers begin with a single normal cell that becomes genetically damaged. The transformation from that initial cell into a tumor is a stepwise progression. The number of genetic mutations that are required to convert a genetically normal cell into an invasive tumor is not known but most likely varies among cancer types. These genetic changes may involve single "letter" or base substitutions, large deletions or duplications, or chromosomal rearrangements impacting vast sections of the genome. Most cancer cells have a number of both large-scale chromosome abnormalities as well as single letter mutations.

The idea that all cancers are genetic in nature and occur as a stepwise addition of mutations, many of which are initiated by environmental factors, is a useful addition to a discussion on common causes of disability and premature loss of life in a Health class (COS objective 10). These concepts should also be incorporated into Biology (COS objectives 6, 7 and 8), Genetics (COS objectives 2, 4, 9 and 10), and AP Biology (general themes "Continuity and Change", "Regulation" and "Science, Technology and Society"). There are also several points of linkage with the Career/Tech Intro to Biotechnology course (COS objectives 5, 11 and 14). In all cases, the distinction should be made between a relatively small number of cancer types with strong inherited risks and most forms of cancer that are primarily due to mutations acquired throughout the life of the individual.

HudsonAlpha has developed a high school lab that focuses on various forms of cancer and methods for their detection. This lab gives students experience in drawing a family pedigree (a genetic family tree) and interpreting the pedigree with respect to a specific form of inherited colon cancer. The students will then complete and analyze a DNA-based diagnostic test to identify which family members have inherited the cancer-causing mutation. The lab activity also introduces students to a genetic counselor and laboratory technician for career exploration. The HNPCC lab has been incorporated into the AMSTI Science in Motion program and is currently available to high school life science teachers across Alabama.

Historically, the diagnosis and staging of cancers has been based on the appearance of the cancer cells under a microscope, and the spread to surrounding or distant tissues. Treatment decisions and options are often based upon this information. However, in many cases, individuals with similar-appearing tumors will show markedly different responses to treatment. We now know that differences at the molecular level, not visible under a microscope, are responsible for the varying outcomes.

Microarray-based expression studies can be used to identify which genes are activated or silenced in the formation of cancer. Expression patterns can classify patients into groups that correlate with cancer subtypes and responses to a specific drug or clinical outcome. If validated, these differences can be used to predict outcomes for new patients, helping physicians identify the most optimal treatment or course of action.



Microarray experiments are currently too cumbersome to perform in a clinic, so it is not likely they will be used routinely to diagnosis patients. However, once a small subset of the genes most relevant to predicting disease or treatment outcome is discovered, it becomes possible to detect the corresponding protein levels in the cancer cells using specially labeled antibodies. For example, some of these proteins have been identified for breast cancer. Detecting whether each protein is present and at what level is useful in determining which therapy will be most effective for treatment. See the table on page 42-43 for specific genetic tests used in this manner.

In the 2008 "Annual Report to the Nation", the National Cancer Institute noted that both the incidence and death rate for all cancers combined is decreasing. While cancer death rates have been declining for several years, this marks the first decline in cancer incidence, the rate at which new cancers are diagnosed.

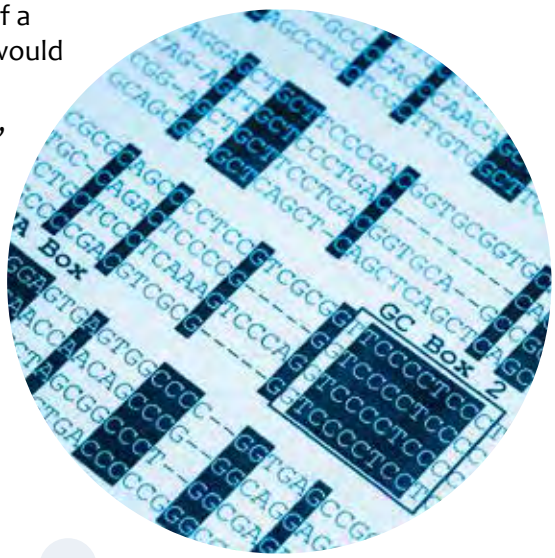
Comparative Genomics

Although the human genome is perhaps the most famous sequencing project, scientists have assembled a genomic library of over 200 different organisms. Knowing the genome of each species provides insight into the function of its DNA; however, there is additional information gained by comparing genomes across organisms. This field of comparative genomics helps discover previously undetected genes, identify the regulatory regions that control gene activity and determine gene function as it relates to health and disease

Comparative genomics provides evidence for the molecular process that underlies evolutionary theory and explains the nature and diversity of organisms, as outlined in the Biology COS objectives 5, 8 and 12 as well as in the Genetics COS objectives 2 and 7. Comparative genomics and its relationship to evolution intersects AP Biology, particularly with respect to general themes "Evolution", "Continuity and Change" and "Science Technology and Society". Career/Tech courses will also benefit from a discussion of comparative genomics, including Veterinary Science (COS objective 3) and Intro to Biotechnology (COS objective 9, 11 and 14).

While humans may seem to have little in common with organisms such as fruit flies, roundworms or mice, they are all composed of cells that must take in nutrients and remove waste, interact with neighboring cells and the outside environment, and grow and divide in response to specific signals. To varying degrees, each of these organisms contains a digestive, circulatory, nervous and reproductive system and is impacted by disorders that impair these systems. During the evolutionary process, as organisms diverged and gave rise to new species, many key proteins such as enzymes, underwent little change. In general, the nucleotide and amino acid sequences of these key proteins have similarly been conserved across the species.

Scientists directly compare the DNA sequence of these organisms, using sophisticated computer programs that line up multiple genome sequences and look for regions of similarity. These similar segments or conserved sequences suggest the DNA sequence has an important functional role – for example, a gene or a regulatory element that controls the activity of a gene. Less critical DNA segments would accept sequence changes without clinical consequence: subsequently, these segments would vary among species. Genes that have relatively high sequence similarity are referred to as homologous genes or homologues. Comparative genomics provides a powerful tool for studying evolutionary changes among organisms, identifying genes that are conserved among species as well as gene and genetic changes that give each organism its unique characteristics.



Genomic comparison also extends to genes involved in disease. If we examine the current list of human disease genes, approximately 20% have a homolog in yeast and nearly two-thirds have one in flies and worms. Initial studies suggest these counterparts may function in nearly identical ways, meaning these organisms can serve as models for understanding human disease and potential treatment. For example, studying genes involved in DNA repair in yeast or bacteria has offered valuable insight into this process in humans and the role that mutations of these genes play in the development of some cancers.

Copy Number Variation

For years single nucleotide polymorphisms (SNPs) were thought to be responsible for the majority of human variation. Until recently, larger scale changes (1000+ nucleotides in length), known as Copy Number Variants (CNV), were thought to be relatively rare. However, scientists have discovered that CNVs occur much more frequently than was suspected. These structural changes alter the number of copies of a specific DNA segment.

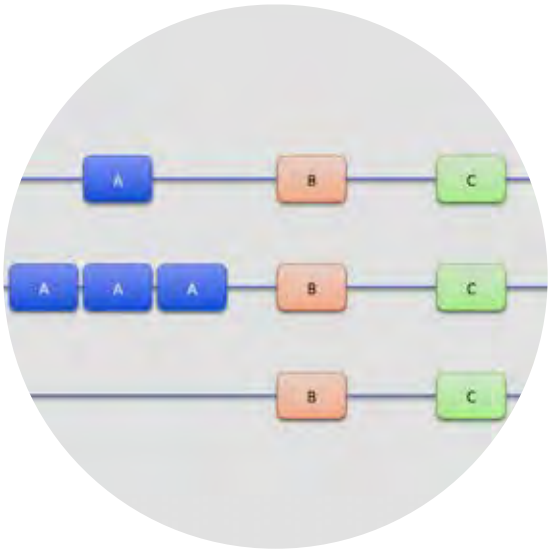
Relating genetic variation to human disease and inheritance is identified in the Biology COS under objective 8 and is described in detail in the Genetics COS objectives 2 and 5, particularly as it connects with genetic patterns of inheritance and multiple alleles. Genetic variation as it relates to human disease also is highlighted under objective 10, which explores the ongoing impacts from the Human Genome Project. AP Biology themes "Continuity and Change" and "Regulation" also intersect the topic of copy number variation, as does Career/Tech course Intro to Biotechnology (COS objective 8).

It came as a surprise to many scientists just how much DNA variation is due to copy number changes. Previous studies based primarily on SNPs suggested that any two randomly selected human genomes would differ by 0.1%. CNVs revise that estimate: the two genomes differ by at least 1.0%. While this may not seem like a major increase, remember that the human genome is composed of approximately 3 billion nucleotides, so the estimated number of nucleotides that vary between two random individuals has increased from 3 million to 30 million. Humans are still nearly 99% identical at the DNA sequence level, but the CNV research has broadened our understanding of how and where we differ.

Preliminary studies have linked copy number variation to lupus, Crohn's disease, autism spectrum disorders, Alzheimer disease, HIV-1/AIDS susceptibility, rheumatoid arthritis and Parkinson's disease.

It has been suggested that CNV regions influence gene activity by directly increasing or decreasing the number of copies of that gene, leading to a concurrent change in the amount of protein. Alternately, CNVs may alter the performance of nearby regulatory signals that activate or silence genes without directly impacting the copy number of the gene itself.

Preliminary studies have linked CNVs to lupus, Crohn's disease, autism spectrum disorders, Alzheimer disease, HIV-1/AIDS susceptibility, rheumatoid arthritis and Parkinson's disease. In some cases the associated CNV is rare, but in other diseases, the identified risk variant is quite common. It is also likely that CNVs may influence individual drug response and susceptibility to infection or cancer.



Criminal Justice and Forensics

DNA profiling, popularly known as DNA fingerprinting, has transformed personal identification, whether in forensic cases, missing persons, mass disasters or paternity disputes. It has become ubiquitous in law enforcement. It is used to exclude individuals suspected of crimes, help convince a jury of an individual's guilt and in some cases, set free individuals wrongly convicted of crimes.

DNA analysis is also used to suggest ancestral origins; there are several companies offering Y-chromosome and mitochondrial DNA studies to determine, for example, to which of the ancient tribes of Britain a man belongs or whether a man or woman has African, Native American or Celtic DNA markers. It is possible to use forensic DNA profiling in the same way to determine the ethnic or geographical origin of the individual from whom the DNA sample came, providing additional information that could be used to narrow the number of potential suspects. For example, in 2007, a DNA test based on genetic biomarkers indicated that one of the suspects associated with a bombing in Madrid was of North African origin. Using other evidence, police confirmed the suspect was an Algerian, confirming the test result.

It has been suggested that this testing could be extended to identify external and behavioral features as well. Scientists have recently identified the genetic variants related to hair, skin and eye color and are exploring other genes that influence traits such as facial height and width as well as nose and lip shape. This "forensic molecular photo fitting" may one day serve as a genetically-based police sketch. Today this approach is still primarily theoretical and currently has little concrete value. As noted throughout this guide, it will take years before the genetic markers associated with all physical and behavioral traits are known.

DNA profiling is a critical component of the Forensics science elective, as part of COS objectives 4 and 5, as well as the Career/Tech course Forensic and Criminal Investigation (COS objectives 7 and 8). It can also be explored in AP Biology as part of the general theme "Continuity and Change", in Genetics as part of COS objectives 9 and 10 and in the Career/Tech course Intro to Biotechnology linked to COS objectives 1, 13 and 14. DNA phenotyping should be an extension of the discussion in all three of these classes, highlighting the concepts and technological challenges still facing the field. The ethical complications of phenotyping should also be incorporated into the discussion.

Legislatively, forensic phenotyping is allowed on a limited basis in some countries (such as the UK) and forbidden in others (Germany). However, for most of the world, legislation that addresses DNA forensic methods is silent about the ability to infer ethnicity or physical traits.

Diagnosing Chromosome Disorders

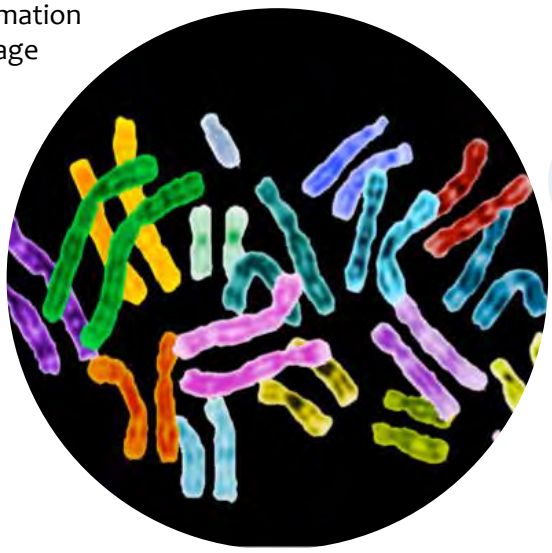
Although scientists have been able to microscopically observe chromosomes since the mid-1800's, a century passed before staining techniques were developed to examine them on a specific and individual basis. The chromosomes could then be arranged according to size and banding pattern for detailed examination - a display called a karyotype. Once it became possible to accurately identify individual chromosomes, abnormalities in chromosome number (such as trisomy 21, also known as Down syndrome) were discovered. Karyotypes can also identify deletions, duplications, and inversions of chromosomal segments.

Although abnormalities on the order of millions of base pairs can be detected using the basic chromosomal banding techniques, smaller alterations cannot be discerned. More recent technologies, such as fluorescence in situ hybridization (FISH) and array comparative genome hybridization (array CGH), allow a finer level of resolution, with the ability to identify submicroscopic chromosome changes.

Although array CHG is still relatively new, it appears to hold great promise for detecting chromosome disorders both large and small. Over the next 3-5 years, this technology will likely become the standard chromosome diagnostic tool to detect abnormalities in chromosome number, microdeletions and other chromosome imbalances. In 2009, clinicians in the UK developed a screening method based on array CGH to identify the most viable eggs obtained from older women undergoing in vitro fertilization (IVF). Array CGH was used to examine the chromosomes from the polar body, a by-product of egg formation that generally serves as a mirror image of the chromosomes found in the egg itself.

Chromosome studies, their behavior in cell division, the formation of egg and sperm and the concept of karyotyping are regularly discussed in Biology classes under the requirements of COS objectives 6 and 8. Karyotypes and their ability to diagnose chromosomal disorders are examined in Genetics classes as part of COS objectives 4,5 and 8, as well as in the Career/Tech course Intro to Biotechnology (COS objectives 1 and 5). The techniques of FISH and aCGH should also be discussed with students in these classes, although many of the technical details need not be described. It is important for students to realize that there are a number of genetic disorders that cannot be identified at the karyotype level, but the newer technologies bridge the gap between studies of stained chromosomes and DNA sequencing.

The HudsonAlpha education team has crafted a karyotype lab as a modification to an existing AMSTI Science in Motion chromosome lab for high school biology and genetics classes. In 'Disorder Detectives', students take on the role of a cytogeneticist working in a hospital or clinic and are given a case study and a set of human chromosomes. They arrange the chromosomes on a prepared board into a completed karyotype, analyze the karyotype and diagnose their patient. Many types of normal and abnormal chromosomal cases are presented. Students also explore the more recent techniques of FISH and aCGH to learn how these technologies provide the ability to diagnose increasingly small genetic imbalances. Geneticists, genetic counselors, and laboratory technicians are highlighted as careers that utilize these types of technologies. The module has been incorporated into AMSTI training at all 11 sites across Alabama and is currently in use by students.



Epigenetics

While identical twins (twins who share the same genetic information) generally look alike when young, obvious differences often emerge as they age. The differences may be due to the varied environment of each twin – for example, one may lift weights and become very muscular while the other never exercises and gains weight. Recent advances in the relatively new field of epigenetics suggest an additional role for the environment in health and disease by altering the activity of particular genes. Activating genes to begin the protein-making process is a key area of study. By identifying the signals that turn genes "on" and "off", investigators hope to understand not only gene function under normal conditions, but also how improper on/off signaling may lead to disorders such as cancer, diabetes, heart disease and obesity.

Epigenetics encompasses modification to DNA, including the addition of small chemical tags called methyl groups. These modifications alter the patterns of gene activity, but do not change the actual DNA sequence. The modifications are not permanent, but can be remembered across thousands of cell divisions and at times from parent to child. This field includes some of the most fascinating biological phenomena, including X-chromosome inactivation, imprinting (when the DNA copy inherited from a particular parent is silenced, while the other copy remains active) and cellular differentiation (see the article on stem cells, page 45).

Studies of identical twins suggest that at birth, twins share similar patterns of epigenetic modification. As they age and are exposed to different diets and environments, the twin's patterns become markedly different, leading to altered activation and silencing patterns.

Current research suggests environment alterations to these epigenetic patterns can change an individual's risk for disease. For many mammals (humans included), differences in diet and level of stress during fetal development and shortly after birth alter the pattern of on/off gene activity, leading to higher risk of obesity, type 2 diabetes and cardiovascular problems. These observations have a number of clinical and public health implications.

Epigenetic changes in DNA often lead to unusual patterns of inheritance for specific disorders. This could be discussed as part of a lesson on exceptions to standard Mendelian inheritance for Biology COS objectives 7 and 8, Genetics COS objectives 5-7,and Intro to Biotechnology COS objective 9. The relationship between the methyl modifications on the DNA and the gene silencing links epigenetics to AP Biology through general themes "Relationship of Structure to Function" and "Regulation".

Epigenetics involves DNA modifications that alter the patterns of gene activity, but do not change the actual DNA sequence. This field includes some of the most fascinating biological phenomena, including X-chromosome inactivation, imprinting and cellular differentiation.

Genetic Information Nondiscrimination Act

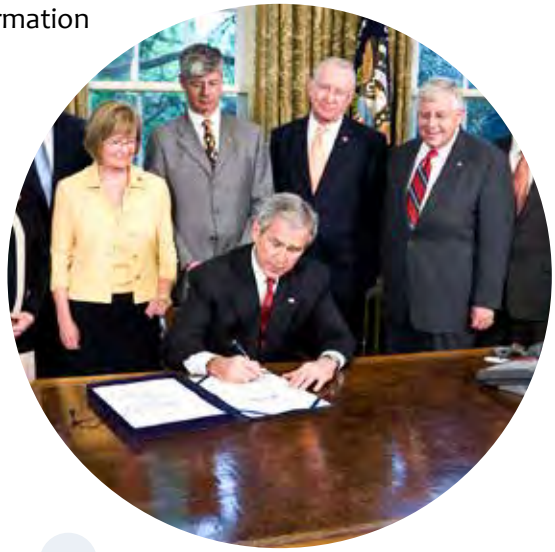
While most Americans are optimistic about the use of genetic information to improve health, many have been concerned that genetic information may be used by insurers to deny, limit or cancel health insurance and by employers to discriminate in the workplace. There has also been concern that some insurers may choose to not insure healthy individuals who are genetically pre-disposed to future disease onset: such people incur more health-related costs for the insurance company than individuals who are not predisposed. A similar fear is that some employers might only employ or retain individuals who are not pre-disposed to future disease onset, since healthy individuals are more productive. Consequently, for many years lawmakers, scientists and health advocacy groups have argued for federal legislation to prevent genetic discrimination.

In 2009, the Genetic Information Nondiscrimination Act (GINA) took effect across America, paving the way for people to take full advantage of the promise of personalized medicine without fear of discrimination. The act had been debated in Congress for 13 years and was signed into law in 2008. GINA protects Americans against discrimination based on their genetic information when it comes to health insurance and employment. The law, together with existing nondiscrimination provisions from other laws, prohibits health insurers or health plan administrators from requesting or requiring genetic information of an individual or the individual's family members, or using it for decisions regarding coverage, rates, or preexisting conditions. The law also prohibits most employers from using genetic information for hiring, firing or promotion decisions.

GINA's protection does not extend to life, disability, or long-term care insurance. In addition, GINA does not prohibit a health insurer from determining eligibility or premium rates for an individual who is already exhibiting clinical symptoms of a disease or disorder.

Genetic discrimination should be briefly discussed in Biology courses as part of COS objective 8, particularly as it relates to significant contributions of biotechnology to society. It could be explored in AP Biology courses under "Science, Technology and Society" general theme and in Genetics classes in light of the ethical, social and legal implications of the Human Genome Project (COS objective 10). There are additional linkages to the Career/Tech courses Foundations of Health Science (COS objective 10), Health Informatics (COS Objective 5) and Intro to Biotechnology (COS objective 14).

In 2009, the Genetic Information Nondiscrimination Act (GINA) took effect across America, paving the way for people to take full advantage of the promise of personalized medicine without fear of discrimination.



Genetics of Eye Color

In 1907, Charles and Gertrude Davenport developed a model for the genetics of eye color. They suggested that brown eye color is dominant over blue eye color. This would mean that two blue-eyed parents would always produce blue-eyed children but never ones with brown eyes. For most of the past 100 years, this version of eye color genetics has been taught in classrooms around the world. It is one of the few genetic concepts that adults often recall from their high school or college biology classes. Unfortunately, this model is overly simplistic and incorrect – eye color is actually controlled by several genes.

In humans, eye color depends on the level of a pigment called melanin present in the iris. Melanin is produced and stored inside specialized cells known as melanocytes. Blue eyes contain minimal amounts of melanin. Irises from green–hazel eyes show moderate pigment levels, while brown eyes are the result of high melanin concentrations.

To date, eight genes that impact eye color have been identified. The *OCA2* gene, located on chromosome 15, appears to play the major role in controlling the brown/blue color spectrum. *OCA2* produces a protein called P-protein that is involved in the formation and processing of melanin. *OCA2* alleles (versions of the gene) related to eye color alter P-protein levels by controlling the amount of *OCA2* RNA that is generated. The allele that results in high levels of P-protein is linked to brown eyes. Another allele, associated with blue eye color, dramatically reduces the P-protein concentration.

While studies suggest that about ¾ of the eye color variation can be explained by genetic changes in and around *OCA2*, it is not the only genetic influence on color. A recent study that compared eye color to *OCA2* status showed that only 62% of individuals with two copies of the "blue eyed " *OCA2* allele actually had blue eyes. Blue eye color was also found among 7.5% of the individuals with the brown-eyed *OCA2* alleles. A number of other genes (such as *TYRP1*, *ASIP*, and *SLC45A2*) also function in the melanin pathway and shift the total amount of melanin present in the iris. The combined efforts of these genes may boost melanin levels to produce hazel or brown eyes or reduce total melanin resulting in blue eyes. This explains how two parents with blue eyes can have green or brown eyed children (an impossible situation under the Davenport single gene model) – the combination of color alleles received by the child resulted in a greater amount of melanin than either parent individually possessed.

The multifactorial genetics of eye color should be discussed in Biology courses as part of COS objective 7, and in Genetics courses under COS objective 5, especially since most textbooks still explain this trait in terms of a single gene effect. It could also be explored in AP Biology courses under "Continuity and Change" general theme. In the Career/Tech Intro to Biotechnology courses, eye color genetics could be explored under COS objectives 8 and 11.

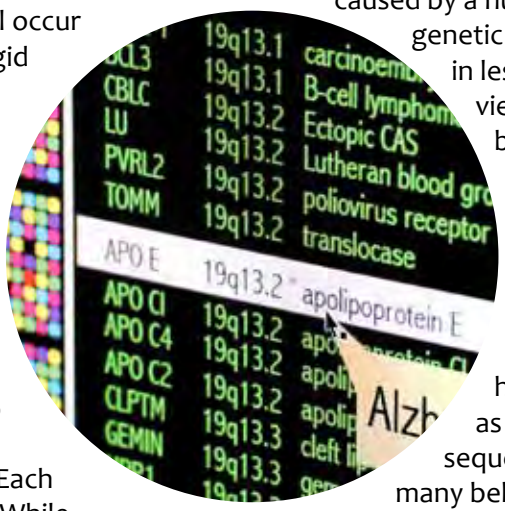
Identifying Genetic Influence on Disease

Much progress has been made in identifying the genetic causes of single gene diseases such as cystic fibrosis, phenylketonuria and Huntington disease. This has led to more accurate risk analysis, better testing approaches and, in some instances, more effective methods of treatment. Even though there are thousands of single gene disorders, they are rare, affecting less than 3% of the population.

In contrast, other diseases, including cleft lip, cardiovascular disease, psychiatric disorders, and cancer, affect much of the world's population. While these diseases have a strong genetic component, they arise from a combination of genetic risk factors that are also influenced by the environment. Few of the contributing genes are believed to make more than a modest contribution to overall risk, perhaps increasing it by 5 or 10%. It is the specific combination of multiple predisposing alleles (DNA changes) and environments that leads to physical symptoms. For this reason, they are often called complex or multifactorial disorders. Identifying the factors that influence disease is a major goal for biomedical research.

Traditional methods of determining the genes responsible for single-gene disorders do not work well for complex diseases. Fortunately, thanks to the advent of second-generation technology to cheaply analyze DNA changes, scientists have used a process known as genome-wide association (GWA) to identify the genetic factors involved in complex disease.

The basic premise behind GWA studies is straightforward: if a specific genetic variation increases the risk of developing a disease, that variation will occur more frequently - and hold up under rigid tests for statistical significance - in individuals who have the disease compared to those not affected. In other words, there is an association between the specific allele and the incidence of disease.



Successful genome-wide association studies test large numbers of variable DNA sites, using DNA microarrays (also called "gene chips") that contain up to one million microscopic spots of DNA. Each spot corresponds to a genetic change. While many of these changes occur with genes, others are in DNA sequences that may be important in regulation or expression of genes.

Relating genetic variation to human disease and inheritance is identified in the Biology COS under objective 8 and is described in detail in the Genetics COS objectives 2 and 5, particularly as it connects with genetic patterns of inheritance and multiple alleles. Genetic variation as it relates to human disease also is highlighted under Genetics objectives 6 and 10, which explore influence of multiple alleles as well as the ongoing impacts from the Human Genome Project. This would also be an appropriate discussion for an AP Biology course ("Continuity and Change and "Science, Technology and Society"), Health (COS objectives 5 and 10) and the Career/Tech Intro to Biotechnology course (COS objective 14).

This technology allows a researcher to simultaneously examine hundreds of thousands of genetic variants that span the human genome – a previously unfathomable accomplishment.

Until recently, researchers knew of almost no genetic variants involved in complex diseases. As of 2010, over 800 genetic single nucleotide polymorphisms have been associated with more than 150 complex diseases or traits. Most of the newly associated genes have not previously been linked to the disease of interest. Intriguingly, some genetic regions have been associated with multiple disorders, suggesting common chemical pathways that influence a number of different processes.

Even with these successes, the majority of the genetic risk for common disease remains undiscovered and the contribution by a single genetic variant to the overall clinical picture is often small. As a result, scientists believe that many of the genetic risks for disease are caused by a number of so-called "rare variants", genetic changes that are each present in less than 1% of the population. This view represents a shift from previous beliefs that complex diseases were caused by variants that were much more common. Projects aimed at sequencing the genomes of a larger number of individuals will hopefully identify many of these "rare variants", allowing this hypothesis to be tested. In addition, as emerging technologies in DNA sequencing continue to drive down costs, many believe GWA studies will shift from examining specific sites of known genetic variation towards full sequencing of the entire genome. At that point, identifying even the rarest of variation becomes feasible.

Infectious Disease

The impact of infectious disease is a major healthcare challenge. Antibiotic resistant strains of pneumonia and staph infections are surfacing in hospitals, nursing homes and locker rooms. The 2009 H1N1 virus confirms long-held concerns about a pandemic influenza virus spreading unchecked across the globe. In both cases, the infectious agents seem to evolve with speed, evading treatment methods. What are we facing and how do these organisms change so quickly?

Similarities and differences between bacteria and viruses connects with the Biology course as part of COS objective 9. Discussions about mutation in both organisms and how it leads to diversity useful for both detection and treatment could be explored in a Genetics course under COS objectives 2 and 10. In the Career/Tech Intro to Biotechnology courses, infectious disease could be explored under COS objectives 11, 13, 14 and 15.

Infectious disease can be classified into two broad categories based on the infectious agent: bacterial or viral. Bacteria are single-celled organisms that live in nearly every environment on the planet including in and on the human body. Most bacteria associated with humans are beneficial and help with daily functions like digestion and protection. Other versions (strains) of bacteria are pathogenic, meaning they can cause illness or harm. If pathogenic bacteria enter the body, they may temporarily escape the body's immune system. Once recognized, the body's immune response attacks invading bacterial cells. Most healthy individuals will be able to fight off a bacterial infection, often with the help of an antibiotic. Antibiotics weaken the bacteria by interfering with its ability to carry out functions like protein synthesis and cell division.

In recent years there has been an increase in bacteria that are resistant to the effects of antibiotics, such as the antibiotic-resistant form of *Staphylococcus aureus*, better known as MRSA. Bacteria reproduce quickly, copying their DNA before each cell division. In some cases, the copying process introduces small DNA changes. By chance, these changes may make the bacteria more resistant to a particular antibiotic. If these bacteria spread to other individuals, then a strain with antibiotic resistance has formed. As additional changes occur, the bacteria may become resistant to a wide range of antibiotics (a "super-bug"), becoming difficult to effectively treat.

In contrast to bacteria, viruses are small packages of genetic material that infect and take-over a cell, converting it to a virus-producing factory. The take-over may occur immediately after the individual is exposed, as happens with the flu, leading quickly to symptoms. Other viruses (e.g. the herpes simplex virus 1 that leads to cold sores) cause a delayed infection with symptoms appearing weeks, months or even years after exposure. Delayed infection viruses hide their genetic material in the cell until conditions are optimal for the virus to reproduce itself. Unlike bacteria, viral infections cannot be treated with antibiotics, although antiviral medications, such as Tamiflu, may be helpful in certain instances.

Viruses reproduce very quickly once activated and like bacteria randomly change their genetic material, often leading to new strains. In addition, if two viruses simultaneously infect the same organism, their genetic information may mix, leading to a completely new strain. This is what occurred with the 2009 novel H1N1 influenza virus. Studies have shown that 2009 H1N1 contains genetic material from pig- bird- and human-based flu viruses.

Understanding the genetic and molecular basis of these organisms allows scientists to develop better diagnostic test, treatments and preventatives. Although the genomes of pathogens have the capability to change rapidly, the genomes are small and often change in semi-predicable ways. Scientists may never be able to cure the flu or common cold, but through genetics and biotechnology more accurate and faster diagnostics can be made.

Non-invasive Prenatal Diagnosis

Prenatal diagnosis involves the use of tests during pregnancy to determine whether a fetus is affected with a particular disorder. These tests have been a part of prenatal medicine for over 30 years. Testing methods vary both in level of invasiveness to the fetus as well as the degree of accuracy. Generally, a set of non-invasive screening methods - such as maternal serum analysis or ultrasound - are initially performed. Suspicious results are followed up with more invasive diagnostic testing e.g. amniocentesis or chorionic villus sampling (CVS). These invasive approaches obtain amniotic fluid and/or fetal cells that are then biochemically or genetically analyzed. Genetic tests may be genome wide - such as karyotyping or array comparative genome hybridization (see page 32) - or more narrow in scope, e.g. testing a single gene. Both amniocentesis and CVS carry a small but significant risk of miscarriage.

Scientists have recently developed a testing method that is both non-invasive and diagnostic. In the 1990s it was discovered that fetal DNA crosses the placenta into the maternal bloodstream. Relatively straightforward techniques have been developed to isolate and analyze this DNA, beginning as early as seven weeks gestation. This test can be performed several weeks earlier than conventional techniques and carries no risk to the health of the fetus. As a result, a larger number of pregnant women may chose to undergo prenatal diagnosis. In 2012, three companies introduced this form of non-invasive prenatal diagnosis into the clinic. Initially only the most common trisomies will be diagnosed, although as the technology matures it will likely be applied to other genetic disorders.

Whether this method ultimately replaces CVS and amniocentesis will depend upon the sensitivity and specificity of the testing. However a number of significant ethical issues are associated with safer, earlier prenatal diagnosis. For example, by offering early non-invasive diagnosis, will there be increased social pressure to have the test and terminate an "abnormal" pregnancy? What or who decides the definition of "abnormal"? As the genetic components of many disorders become better understood, would non-invasive diagnostic testing allow parents, with only a blood test to identify mild, adult-onset disorders, as well as nonmedical traits such as eye color?

Prenatal diagnosis is a standard part of discussions around egg and sperm formation and the abnormalities that can occur during meiosis. The advent of non-invasive techniques is an exciting addition for Biology (COS objectives 6 and 8), Genetics (COS objective 4) and the Career/Tech Introduction to Biotechnology (COS objective 5). The application of this new technology to health and society links to classroom conversations in AP Biology ("Science, Technology and Society") and Health (COS objectives 5 and 6). Clearly, there are a number of ethical concerns related to non-invasive prenatal testing, Depending on the context of the conversation and the maturity of the class, these questions may be appropriate for exploration and detailed discussion.

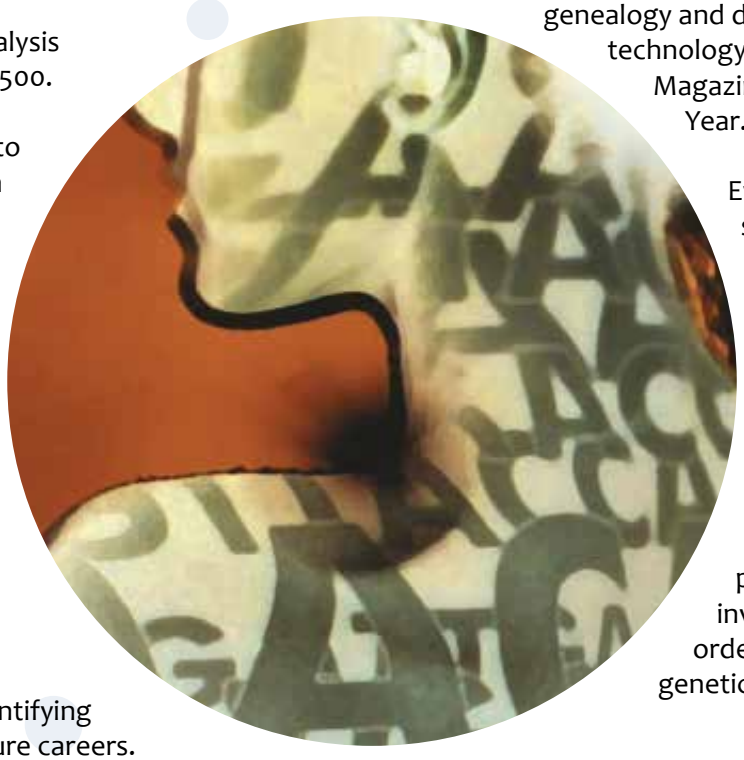
Personal Genome Analysis

The past few years have seen the rise of genomics research aimed towards sequencing groups of individuals, such as the "PGP-10", ten individuals who have volunteered to share their DNA sequences, medical records and other personal information as part of the personal genomes project (PGP). The public profiles of the PGP-10 are freely available online at http://www.personalgenomes.org/. An additional large-scale genome sequencing project is the 1000 Genomes Project, an international research collaboration that hopes to sequence the genome of approximately 1200 individuals from across the globe. Sequencing such a large number of individuals will create an index of genetic variation including previously unidentified "rare variants", genetic changes which scientists increasingly believe are responsible for much of the genetic influence on disease.

As an initial step in the direction of personalized, commercially available genomic sequencing, several companies have begun offering consumer genomics testing. Four companies (Navigenics, deCODEme, Pathway Genomics and 23andme) offer a similar product, namely a read-out of between 500,000 and 1,000,000 variable regions from across the genome. A small but increasing proportion of these variable regions has identified connections to ancestry, physical traits or disease risk, although the predictive value for medical decisions of many of these traits remains marginal or unclear.

The cost of this personal analysis varies between $100 and $2,500. Two additional companies (Knome and Illumina) offer to sequence the entire 3 billion base pairs of an individual's genome for between $48,000 and $100,000.

In addition to genome-wide analysis, consumer genomics testing is available for individual genes, such as the *ACTN3* genetic variant involved in muscle strength and sprint ability. A number of companies offer parents genetic testing on their children, in the hopes of identifying characteristics linked to future careers.



The first wave of personal genome studies offered direct-to-consumer should be a component of a Genetics course as part of COS objective 10 regarding ethical, social and legal implications from the Human Genome Project. The availability of personal information from the PGP-10 is also fertile ground for a discussion on the implications of genetic information. These topics can also be incorporated into a Biology course under COS objective 8 - significant contributions of biotechnology to society, the Career/Tech Intro to Biotechnology (COS objective 14) and an AP Biology course as part of the general theme "Science, Technology and Society". Outside the traditional science classroom, this could form the basis of an excellent conversation with students in Health (COS objective 6), and the Career/Tech electives Foundations of Health Sciences (COS objective 10) and Health Informatics (COS objective 5) outlining valid and essential information for the safe use of consumer goods and health products.

Such programs are poor predictors of athletic aptitude, intelligence or musical or artistic talent. Much of the genetic and environmental influences on these traits are still unknown.

There is little data regarding the response of people who have received information about their genetic risk factors from one of these consumer genomic companies. At the same time, there is a growing recognition among personal genomic stake-holders that consumer genomics may provide a positive impact on an individual's life and actions even if its direct health benefit is uncertain or marginal.

Regardless, there appears to be a strong consumer appetite for genetic information related to both genealogy and disease risk - the underlying technology was named Time Magazine's 2008 "Invention of the Year."

Even so, a number of scientists and health care providers have argued that these services are akin to "practicing medicine without a license". The American College of Medical Genetics has issued a statement recommending "a knowledgeable health professional should be involved in the process of ordering and interpreting a genetic test."

Personalized Medicine

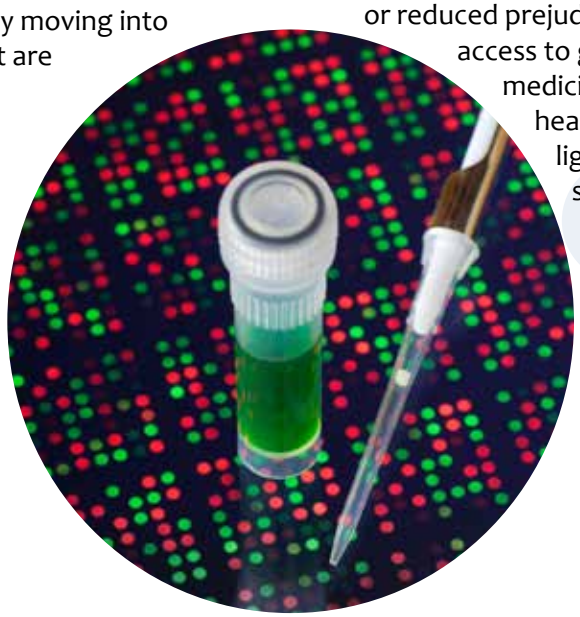
At its core, personalized medicine uses information about a person's genetic background to tailor strategies for the detection, treatment or prevention of disease. This may include genetic screening tests to identify susceptibility to disease or more precisely pinpoint existing conditions. It may also be used to guide pharmaceutical choices, highlighting the brand and dose of medication best suited for a patient. The goal of personalized medicine is to help physicians and their patients identify the best course of action to prevent or manage a disease based upon the patient's genetic and environmental profile.

The implications of personalized medicine impacts biology-based science courses, Health Education and pre-healthcare options at the high school level. Biology COS objective 8 and AP Biology theme "Science, Technology and Society" discuss significant contributions of biotechnology to society. Diagnosing genetic variants that increase the risk of human disease is a key focus of the Genetics COS objectives 9 and 10, particularly as it explores the ongoing impacts from the Human Genome Project and their application to disease. At the Health level, COS objective 5 asks students to evaluate negative and positive impacts of technology on health. Personalized medicine is an excellent candidate for this discussion, as well as showing application to the Career/Tech courses Introducation to Pharmacy (COS objectives 9 and 11) and Intro to Biotechnology (COS objectives 11 and 14).

Drawing an analogy from the world of fashion, personalized medicine is the equivalent of a custom-made suit or outfit, designed with an individual's unique body measurements. This type of tailored approach provides a much better fit than purchasing something "off the rack."

As has already been noted in this guide, people vary from one another in many ways – what they eat, their lifestyle, the environmental factors to which they are exposed, and variations in their DNA. Some portion of this genetic variation influences our risk of getting or avoiding specific diseases. Certain changes in the DNA code influence the course of disease, impacting the age of onset for symptoms or the speed of progression. Genetic variation also contributes to differences in how drugs are absorbed and used by the body (see the section on pharmacogenomics on page 43).

This newfound knowledge is rapidly moving into the clinical setting. At the forefront are a series of drugs such as Gleevac™, Herceptin™ and Iressa™ known to be most effective in people with a specific genetic profile (set of genetic variants). Straightforward genetic tests are performed to identify who will benefit from these medications. At the same time, more precise diagnostic tests are in development that better classify disease subtypes or progression. The information identified in our genome will help develop a lifelong plan of health maintenance tailored to our unique genetic profile. For an overview of current medical approaches based on genetic information, see the table "Selected Personalized Medicine Drugs, Treatments and Diagnostics as of March 2009" on pages 44-45.



One of the "holy grails" in personalized medicine is the so-called $1,000 genome – the ability to sequence a human's genetic information at an economically feasible price. Recent advances in sequencing technology (highlighted in "Genome Sequencing in the Clinic" on page 8 and discussed in detail on page 24) are steadily moving the field closer to this figure. In addition to issues of cost, there are other challenges to personalized medicine, including concerns about patient privacy, confidentiality and insurability after taking a genetic test. Will the knowledge that specific genetic variation increases disease risk lead to greater or reduced prejudice or discrimination? How will access to genetic testing and personalized medicine be equitable? Does our current healthcare system need to change in light of this genetic approach and if so, which new model will be best?

Pharmacogenomics

Pharmacogenomics deals with how a patient's specific genetic variation affects the response to certain drugs. In part, the genetic variation among individuals helps explain why one drug may work spectacularly in one person, not at all for another and produce harmful side effects in a third. For example, variation in the *CYP2C9* and *VKORC1* genes impact whether someone is likely to develop a dangerous reaction to warfarin, a blood-thinning medication often prescribed for people at risk for blood clots or heart attacks.

A genetic test that identifies those susceptible to that reaction has now been developed to help doctors adjust warfarin doses based on each patient's genetic profile. For an overview of current pharmacological approaches based on genetic information, see the table "Selected Personalized Medicine Drugs, Treatments and Diagnostics as of March 2009" on pages 44-45. In addition, there are over 200 pharmaceutical products that either recommend genetic testing or point to the influence of genetic variability on the drug's response.

Pharmacogenomics has most rapidly developed in the field of cancer. For example, the *HER2* receptor, often found on the surface of a cell, helps regulate when the cell divides and grows. In many instances of breast cancer, the *HER2* receptor is present at very high levels, leading to increased cell growth and tumor formation. In these cases, the anti-cancer drug Herceptin™ is added to the patient's treatment plan where it increases the efficacy of chemotherapy.

Molecular testing is needed because only 25% of breast cancer patients will see any benefit from Herceptin™ -- the rest should be given another treatment. In a similar manner, Gleevac™ and Erbitux™ may be respectively prescribed for specific forms of chronic myeloid leukemia and colorectal cancer. Both medications prevent tumor cells from continuing growth but each operates in a very pathway-specific process that is unique to a subset of each cancer type. This type of therapy based on molecular targets is slowly but surely gaining in success as additional genetic pathways for disease are identified.

The implications of pharmacogenomics as a part of personalized medicine impact Health Education as well as Biology-based courses. Biology COS objective 8 and AP Biology general theme "Science, Technology and Society" discusses significant contributions of biotechnology to society. Diagnosing genetic variants that lead to specific drug recommendations is also a part of the Genetics COS objectives 9 and 10, particularly as it explores the ongoing impacts from the Human Genome Project and their application to disease. At the Health level, COS objectives 5 and 6 address negative and positive impacts of technology on health the safety of health products and like personalized medicine, pharmacogenomics is an ideal discussion topic. Classroom discussions concerning pharmacogenomics would clearly also be appropriate in the Career/Tech Intro to Pharmacy (COS objectives 9 and 11) and Intro to Biotechnology (COS objectives 1, 11 and 14) courses offered to Alabama students.

There are over 200 pharmaceutical products that either recommend genetic testing or point to the influence of genetic variability on the drug's response.

Table 1: Selected Personalized Medicine Drugs, Treatments, and Diagnostics as of March 2009*

Therapeutic product label contains pharmacogenomic information as:
- Information only
- Recommended
- Required

THERAPY	BIOMARKER/TEST	INDICATION
Herceptin® (trastuzumab) **Tykerb® (lapatinib)**	HER-2/neu receptor	**Breast cancer**: "…for the treatment of patients with metastatic breast cancer whose tumors over-express the HER2 protein and who have received one or more chemotherapy regimens for their metastatic disease."
Pharmaceutical and surgical prevention options and surveillance	*BRCA 1,2*	**Breast cancer**: Guides surveillance and preventive treatment based on susceptibility risk for breast and ovarian cancer.
Tamoxifen	Aviara Breast Cancer Index℠ (*HOXB13, IL17BR*)	**Breast cancer**: Calculates a combined risk analysis for recurrence after tamoxifen treatment for ER-positive, node-negative breast cancer.
Chemotherapy	Mammostrat®	**Breast cancer**: Prognostic immunohistochemistry (IHC) test used for postmenopausal, node negative, estrogen receptor expressing breast cancer patients who will receive hormonal therapy and are considering adjuvant chemotherapy.
Chemotherapy	MammaPrint®	**Breast cancer**: Assesses risk of distant metastasis in a 70 gene expression profile.
Coumadin® (warfarin)	*CYP2C9*	**Cardiovascular disease**: "an increased bleeding risk for patients carrying either the *CYP2C9*2* or *CYP2C9*3* alleles."
Coumadin® (warfarin)	*VKORC1*	**Cardiovascular disease**: "Certain single nucleotide polymorphisms in the *VKORC1* gene (especially the -1639G>A allele) have been associated with lower dose requirements for warfarin."
Coumadin® (warfarin)	PGx Predict™: Warfarin	**Cardiovascular disease**: Determines *CYP2C9* and *VKORC1* genotypes to predict likelihood of adverse events with warfarin therapy.
Coumadin® (warfarin)	Protein C deficiencies	**Cardiovascular disease**: Hereditary or acquired deficiencies of protein C or its cofactor, protein S, has been associated with tissue necrosis following warfarin administration.
Pharmaceutical and lifestyle prevention options	Familion® 5-gene profile	**Cardiovascular disease**: Guides prevention and drug selection for patients with inherited cardiac channelopathies such as Long QT Syndrome (LQTS), which can lead to cardiac rhythm abnormalities.
Statins	PhyzioType SINM	**Cardiovascular disease**: Predicts risk of statin-induced neuro-myopathy, based on a patient's combinatorial genotype for 50 genes.
Atorvastatin	*LDLR*	**Cardiovascular disease**: "Doses should be individualized according to the recommended goal of therapy. Homozygous Familial Hypercholestremia (10-80mg/day)and heterozygous (10-20mg/day)."
Camptosar® (irinotecan)	*UGTIA1*	**Colon cancer**: "Variations in the *UGT1A1* gene can influence a patient's ability to break down irinotecan, which can lead to increased blood levels of the drug and a higher risk of side effects."
Erbitux® (cetuximab) **Gefitinib** **Vectibix® (panitumab)**	*EGFR* expression	**Colon cancer**: "Patients enrolled in the clinical studies were required to have…evidence of positive *EGFR* expression using the DakoCytomation *EGFR* pharmDx™ test kit." *EGFR* positive individuals are more likely to respond to the drug than those with reduced *EGFR* expression.
Erbitux® (cetuximab) **Gefitinib** **Vectibix® (panitumab)**	*KRAS*	**Colon cancer**: Certain *KRAS* mutations lead to unresponsiveness to the drug.
Erbitux® (cetuximab) and Vectibix® (panitumab) **Fluorouracil** **Camptosar® (irinotecan)**	Target GI™	**Colon cancer**: Provides information of the expression of key molecular targets—*KRAS*, *TS*, and *TOPO1*—to guide therapy.
Tagretol (carbamazepine)	*HLA-B*1502*	**Epilepsy and bipolar disorder**: Serious dermatologic reactions are associated with the *HLA-B*1502* allele in patients treated with carbamazepine. "Prior to initiating Tegretol therapy, testing for *HLA-B*1502* should be performed in patients with ancestry in populations in which *HLA-B*1502* may be present."
Immunosuppressive drugs	AlloMap® gene profile	**Heart transplantation**: Monitors patient's immune response to heart transplant to guide immunosuppressive therapy.
Ziagen® (abacavir)	*HLA-B*5701*	**HIV**: "Patients who carry the *HLA-B*5701* allele are at high risk for experiencing a hypersensitivity reaction to abacavir. Prior to initiating therapy with abacavir, screening for the *HLA-B*5701* allele is recommended."
Selzentry® (maraviroc)	CCR5 receptor (1)	**HIV**: "Selzentry, in combination with other antiretroviral agents, is indicated for treatment experienced adult patients infected with only CCR5-tropic HIV-1 detectable..."

Budesonide	IBD Serology 7	**Inflammatory bowel disease**: Identifies subset of patients who will benefit from budesonide.
Gleevec® (imatinib mesylate)	*BCR-ABL*	**Leukemia**: "Gleevec® (imatinib mesylate) is indicated for the treatment of newly diagnosed adult and pediatric patients with Philadelphia chromosome positive [indicated by presence of *BCR-ABL*] chronic myeloid leukemia (CML) in chronic phase."
Dasatinib	Philadelphia Chromosome	**Leukemia**: "Dasatinib is indicated for the treatment of adults with Philadelphia chromosome-positive acute lymphoblastic leukemia (Ph+ ALL) with resistance or intolerance to prior therapy"
Busulfan	Philadelphia Chromosome	**Leukemia**: "Busulfan is clearly less effective in patients with chronic myelogenous leukemia who lack the Philadelphia (Ph1) chromosome."
Purinethol® (mercaptopurine) Thiaguanine Azathioprine	TPMT	**Leukemia**: Guides adjustment of dose in treatment of acute lymphoblastic leukemia: "Patients with inherited little or no thiopurine S-methyltransferase (TPMT) activity are at increased risk for severe Purinethol toxicity from conventional doses…"
Tarceva® (erlotinib)	*EGFR* expression	**Lung cancer**: The test determines patients most likely to respond.
Capecitabine	DPD	**Multiple cancers**: "Rarely, unexpected severe toxicity (e.g., stomatitis, diarrhea, neutropenia and neurotoxicity) associated with 5-fluorouracil has been attributed to a deficiency of dihydropyrimidine dehydrogenase (DPD) activity."
Pharmaceutical and surgical treatment options and surveillance	*MLH1, MSH2, MSH6*	**Multiple cancers**: Guides surveillance and preventive treatment based on susceptibility risk for colon and other cancers.
Chemotherapy	CupPrint™	**Multiple cancers**: Determines cancer classification for tumors of unknown primary origin.
Chemotherapy	Aviara CancerTYPE ID®	**Multiple cancers**: Classifies 39 tumor types from tumors of unknown primary origin, using a gene expression profile.
Elitek® (rasburicase)	G6PD deficiency	**Multiple cancers**: "Rasburicase administered to patients with glucose- phosphate dehydrogenase (G6PD) deficiency can cause severe hemolysis. … It is recommended that patients at higher risk for G6PD deficiency … be screened prior to starting ELITEK therapy."
Drugs metabolized by CYP P450 **2C19**: Celecoxib, Codeine, Diazepam, Esomeprazole, Nelfinavir, Omeprazole, Pantoprazole, Rabeprazole, Voriconazole **2D6**: Acetaminophen, Aripiprazole, Atomoxetine, Carvedilol, Cevimeline hydrochloride, Clozapine, Fluoxetine HCl, Fluoxetine HCL and Olanzapine, Metoprolol, Propranolol, Propafenone, Protriptyline HCl, Risperidone, Tamoxifen, Terbinafine, Thioridazine, Timolol maleate, Tiotropium bromide inhalation, Tolterodine, Tramadol, Venlafaxine	Amplichip® *CYP2D6/CYP2C19*	**Multiple diseases**: FDA classification 21 CFR 862.3360: "This device is used as an aid in determining treatment choice and individualizing treatment dose for therapeutics that are metabolized primarily by the specific enzyme about which the system provides genotypic information."
Rifampin Isoniazid Pyrazinamide	NAT	**Multiple diseases**: N-acetyltransferase slow and fast acetylators and toxicity- "slow acetylation may lead to higher blood levels of the drug, and thus, an increase in toxic reactions."
Rituximab	PGx Predict™: Rituximab	**Non-Hodgkin's lymphoma**: Detects CD-20 variant (polymorphism in the IgG Fc receptor gene *FcgRIIIa*) to predict response to cancer drug rituximab.
Celebrex® (celecoxib)	*CYP2C9*	**Pain**: "Patients who are known or suspected to be P450 2C9 poor metabolizers based on a previous history should be administered celecoxib with caution as they may have abnormally high plasma levels due to reduced metabolic clearance."
Risperdal® (resperidone) Zyprexa® (olanzapine)	PhyzioType PIMS	**Psychiatric disorders**: Predicts risk of psychotropic-induced metabolic syndrome, based on a patient's combinatorial genotype for 50 genes.
Gleevec® (imatinib mesylate)	*c-KIT*	**Stomach cancer**: "Gleevec® is also indicated for the treatment of patients with Kit (*CD117*) positive unresectable and/or metastatic malignant gastrointestinal stromal tumors (GIST)."

*This list is not intended to be comprehensive but reflects commonly used or available products as of March 2009. Some products, for which the FDA recommends or requires pharmacogenomic testing or which have pharmacogenomic information in their label, are listed at the FDA's Web site (http://www.fda.gov/cder/genomics/genomic_biomarkers_table.htm). Other listed products that are novel, and/or that address large populations, have been identified via websites and public announcements.

Indications in quotes are taken from the therapeutic product label.

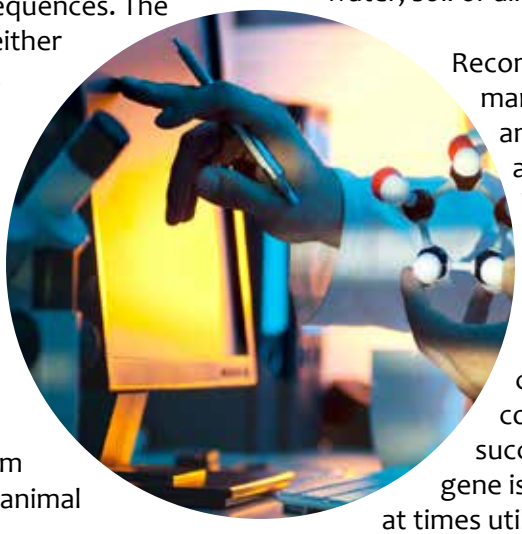
BCR-ABL = breakpoint cluster region – Abelson
BRCA 1,2 = breast cancer susceptibility gene 1 or 2
c-KIT = tyrosine kinase receptor
CYP = cytochrome P450 enzyme
DPD = dihydropyrimidine dehydrogenase
G6PD = glucose 6 phosphate dehydrogenase
HER2 = human epidermal growth factor receptor 2
NAT = N-acetyltransferase
TOPO1 = topoisomerase 1
TPMT = thiopurine S-methyltransferase
TS = thymidylate synthase
UGT1A1 = UDP-glucuronosyltransferase 1A1

Recombinant DNA and Genetic Engineering

For centuries, humans have used selective breeding techniques to modify the characteristics of both plants and animals. Typically, organisms with desired traits like a high grain count, specific petal color or fragrance, consistent milk production or ability to herd livestock have been chosen to pass those traits to the next generation. These breeding practices, while very successful, require a large number of generations to yield the desired results. In addition, only traits that are naturally expressed in a species can be selected. For example, traditional breeding methods do not allow characteristics to be transferred from a plant to an animal.

Research during the last one hundred years has identified the relationship that exists between physically observed traits and the genetic information that codes for those traits. This understanding has been coupled with modern molecular laboratory techniques to transfer certain traits expressed in one species into a different (and maybe very distant) species. Scientists can modify the DNA of bacteria, plants and animals to add genetic information (and the associated characteristics) from a different organism. This process has historically been called genetic engineering but more recently is referred to as recombinant DNA technology or genetic modification.

To make a recombinant organism, the gene of interest must first be isolated from the initial donor organism. To isolate the gene, scientists use restriction enzymes, proteins that can be thought of as molecular scissors that cut DNA at specific nucleotide sequences. The restriction enzymes cut the DNA on either side of the gene of interest. The DNA fragment containing the gene is then ligated (fused) into a different piece of DNA called a vector. The vector serves as a mechanism to carry the gene of interest into the host. It often includes additional genetic information such as selectable markers and genetic signals that control when and where it will be expressed. The vector is then introduced into a single host cell. From this cell, an entire organism, plant or animal is grown.



Recombinant DNA offers an excellent way to re-emphasize central dogma (the information in DNA is transcribed into RNA and then translated into protein) in the context of key molecular biology techniques, e.g. restriction enzyme digestion and DNA transformation. This approach of combining concept with application can be successfully incorporated into a number of life science as well as career/tech courses, many of which mention genetic engineering by name. This includes Biology (COS objective 8), Genetics (COS objectives 7 and 9), AP Biology (general themes "Relationship of Structure to Function" and "Science, Technology and Society"), Health (COS objective 5), Introduction to Agriscience (COS objective 16), and Introduction to Biotechnology (COS objectives 9, 13 and 14).

The organism must be tested to make sure the gene is functioning correctly and the organism is exhibiting the desired trait. Multiple generations are grown and tested before the crop, therapeutic drug or sensor is made commercially available.

Since the first recombinant DNA molecule was created in 1973, the technology has been used across a wide variety of fields:

- amending crops such as corn, soybean and rice, adding pest or herbicide resistance, or increasing nutrient content (see Agricultural Applications, page 29)
- modifying bacteria by adding genes that produce enzymes used in industry (Chymosin™ - used for making cheese)
- producing therapeutic products such as human insulin (Humulin™), blood clotting factors (rFVII™) and components of the immune system (Enbrel™)
- developing biosensors to identify toxins in the water, soil or air

Recombinant DNA forms the core of many key biotechnology applications and continues to result in new approaches that impact agriculture, healthcare and the environment. The technology is also at the core of gene therapy, a series of techniques aimed at introducing the correct version of a gene into the cells of a patient. Gene therapy is a complicated process, with only limited success to date. Silencing an overactive gene is a related form of therapy that at times utilizes recombinant DNA. More information about this approach, known as RNAi, can be found on page 50.

Stem cells

Stem cells can be thought of as "master cells", the raw materials from which a complete individual is crafted. The power of a stem cell lies in its "pluriopotency" - the ability to divide and develop (differentiate) into any one of the 220 various types of cells found in the body. As cells differentiate, they lose this ability; a liver cell for example, can only renew itself to form more liver cells - it cannot become lung or brain.

Because of this pluripotency, stem cells have great medical potential. They could be used to recreate insulin-producing cells in the pancreas to treat type I Diabetes, to repopulate neurons destroyed due to Parkinson's disease or to replace cells lost in spinal cord injuries. In the laboratory, stem cells have been used to successfully treat animals affected with paralysis, muscular dystrophy, Parkinson's disease and sickle cell anemia.

Multiple types of stem cells have been identified or developed. Embryonic stem cells (ES cells) were the first category discovered. These cells are fully pluripotent, but only found in young embryos (the stage of human development from conception to eight weeks gestation). Because the process to collect ES cells destroys the embryo, some religious groups are opposed to their use.

In the tissues of many developed organs, scientists have identified so called "adult stem cells" that retain a portion of the ability to differentiate into other cell types. The primary role of adult stem cells is to maintain and repair the tissue in which it is found. For example, bone marrow contains adult stem cells, which can give rise to all the types of blood cells. This is why a bone marrow transplant can repopulate the blood and immune cells in a patient. It appears that adult stem cells may not have the full range of pluripotency found in ES cells, although researchers are exploring techniques to use adult stem cells for certain forms of therapy.

The concept of stem cells connects to several components of the standard Biology Course. It can be highlighted during explanation of the cell cycle (COS objective 6), although some biology curriculum models include discussions of stem cells during instruction on the Cell Theory instead (COS objective 4). In addition, exploring the similarities and differences between stem cells and differentiated cells would reinforce concepts about structure and function of cell and how specific functions are performed (COS objective 5) as well as the role of biotechnology in developing iPS cells (COS objective 8). Discussion of stem cells in relation to cell cycle is also connected to Genetics (COS objective 4) and Introduction to Biotechnology (COS objective 5). Highlighting the pros and cons of each stem cell type provides links to AP Biology (general theme "Continuity and Change") and Health courses (COS objective 5).

Recent genetic discoveries have identified key genes that are active only in ES cells. Working in the laboratory, scientists have used this information to modify differentiated cells to "reactivate" these genes, in effect regressing the cells into pluripotent stem cells. These cells are known as induced pluripotent stem (iPS) cells and early research suggests they behave in much the same way as ES cells. Because iPS cells could be created by reprogramming a patient's own tissues, they lack the ethical concerns posed by ES cells. In addition, because they are a genetic match, therapies using iPS cells would not be rejected by the patient's immune system. While there are a number of technical hurdles that must be overcome before iPS cells are ready for clinical applications, several companies are beginning to explore treatment possibilities.

Studying the Genome to Understand the Sequence

In 2001 the "completion" of the Human Genome Project (HGP) was announced with much fanfare. The published DNA sequence was akin to an operations manual or book of recipes, identifying the genetic instructions for how cells build, operate, maintain and reproduce themselves, all the while responding to varying conditions from the surrounding environment. While the completion of the HGP may have felt like the end of an era, in reality it was only the beginning. Scientists had very little knowledge of how cells utilized the information found in each genetic recipe to function and interact. Nor was there a clear understanding of how genes keep humans healthy or predispose them to disease. A representative genome had been sequenced, but how many differences would be found if peoples from around the world were compared? How did the human sequence compare to those of other organisms? Sequencing the human genome raised more questions than it answered.

Two large-scale projects aimed at expanding our understanding of the human genome have begun to answer many of these questions. The International HapMap Project was created to compare the genetic sequences of different individuals. The HapMap identifies DNA variants across the genome and examines how the variants are distributed within and across world populations. The project does not connect the variation to a specific illness, but rather provides the raw information that researchers can use to link genetic variation to disease risk.

ENCODE, the Encyclopedia Of DNA Elements, was launched to identify and classify the functional elements in the human genome that activate or silence regions of DNA. Based on data released in 2012, the majority of DNA in the human genome appears to have some sort of functional role.

The history of and findings from the Human Genome Project are addressed in the Genetics COS objective 10. The subsequent HapMap and ENCODE studies shed light on the effects of genetic variability on adaptation (Genetics COS objective 2 and AP Biology general themes "Continuity and Change" and "Relationship of Structure to Function) and the structure of eukaryotic chromosomes (Genetics COS objective 8). The influence of genetic change and mutation on increasing diversity is also a key concept in the HapMap study that is identified in the Biology COS under objective 8. These findings also have merit for discussion in the Career/Tech Veterinary Science (COS objective 3) and Intro to Biotechnology (COS objectives 9 and 14) courses.

HudsonAlpha has modified an existing AMSTI Science in Motion lab dealing with extracting DNA. This is a foundational activity that a Biology class would perform before exploring DNA or the findings of studies such as HapMap or ENCODE. The original lab followed a very simple protocol and left no room for inquiry or student input. The expanded lab provides students an opportunity to learn about the composition and structure of cells and their DNA. Students chose from a variety of plant and animal samples (fruits, fish, liver etc). Then, using a hands-on, inquiry based approach, the students design and make the necessary buffers to break open cell membranes and extract DNA, using everyday household materials.

This firmly puts to rest any view that the human genome consists of a relatively small set of functional elements (the genes) along with a vast amount of so-called "junk" DNA that is not biologically active.

Just like the HGP, information generated from HapMap and ENCODE is freely accessible by scientists and the public around the world.

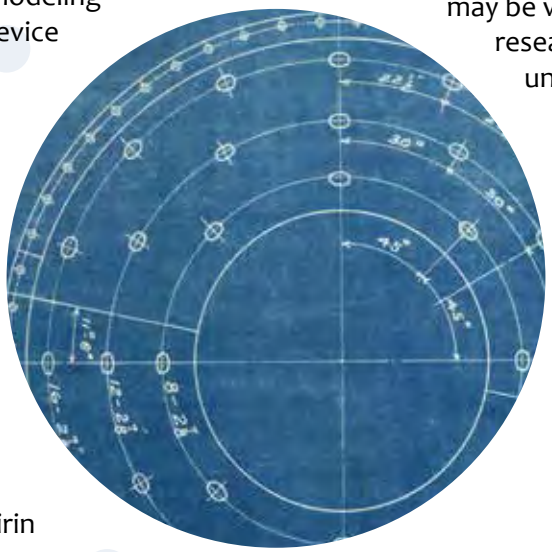
Synthetic Biology

Synthetic biology seeks to apply engineering principles to biology. It has an ultimate goal of designing and building biological systems for specified tasks (e.g. drug development, material fabrication and energy production). The field is a collaborative effort between not only engineers and biologists, but also chemists and physicists.

Synthetic biology aims to use engineering methods to build novel and artificial biological tools. This is done using an established engineering approach - defining the specification for a device or system and then using a set of standard parts to create a model that meets that specification. The basic building block is a biopart - a fragment of DNA with a specific function such as producing a protein or activating a "start/stop" switch. Bioparts are combined into devices that carry out a desired activity, like producing fluorescent protein under a given condition. Multiple devices can be connected into a system, which performs more complex, higher-level tasks.

Powerful computers offer in-depth modeling and simulation to predict the behavior of the part, device or system before it is assembled. The relevant DNA instructions are then artificially synthesized and inserted into a biological cell, such as bacteria. The bacterial cell is the "chassis" or vehicle that interprets the DNA instructions. If the synthesized information is read and processed correctly, then the specification and design were appropriately crafted. If not, the original design is modified, continuing the design-modeling-testing cycle. Once complete, the device or system becomes a component created from standard bioparts, rather than constructed each time from scratch.

The rise of synthetic biology has been compared to that of synthetic chemistry, a field that developed and matured during the past century as chemists learned how to synthesize compounds that previously only existed in nature. Early examples such as dyes and medicines like aspirin gave way to the creation of plastics, semiconductors and complex pharmaceuticals.



The concepts behind synthetic biology links to the COS objective 8 for a standard Biology course, particularly as it relates to significant contributions to biotechnology. Discussion of synthetic biology also connects to the AP Biology general theme "Science Technology and Society." Lastly, the Genetics COS objective 9 and CTE Introduction to Biotechnology COS objective 13 highlight areas of biotechnology that deal with recombinant DNA. This is a natural connection to synthetic biology, which uses recombinant DNA techniques as the cornerstone to creating the artificial bioparts, systems and devices.

Many supporters believe that synthetic biology has the potential to achieve equally important results such as producing inexpensive new drugs, developing environmental biosensors and more efficiently producing biofuels from biomass.

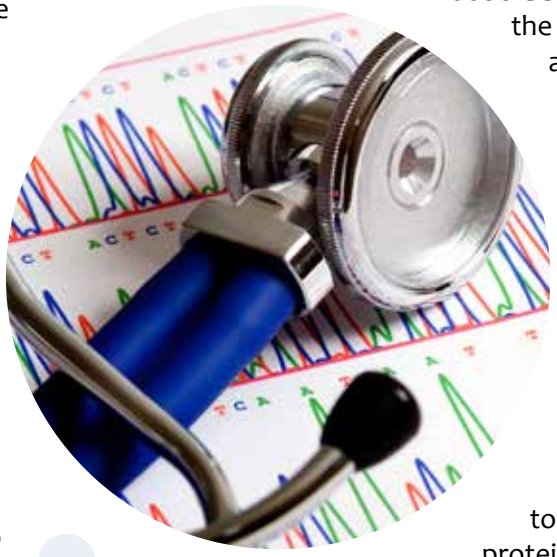
Given that synthetic biology involves creating novel living organisms, it isn't surprising that security, safety and ethical concerns have been raised. Like many other "dual use technologies," synthetic biology offers the potential for great good, but also for harm. There are concerns that the increasing accessibility of this technology may spawn a new era of "biohackers" leading to the accidental or deliberate creation of pathogenic biological components. Safety measures taken by the research community include incorporating genetic signals that prevent uncontrolled spreading outside the lab environment. It is worth noting that in many ways, these mechanisms are already in place as part of the guidelines developed for recombinant DNA techniques that are currently in use worldwide. From this perspective, the advances in synthetic biology may be viewed as a natural extension of this research, rather than a great leap into unchartered scientific territory.

Therapeutic Approaches

Gene Therapy

Gene therapy is defined as the correction of a nonfunctioning gene responsible for causing a disease. For example, a normal (functioning) copy of the gene could be inserted into a cell to replace a nonfunctioning gene. As genes will not enter cells on their own, there must be a mechanism in place to carry the corrected gene into the body's cells. The most common mechanism (vector) is an altered form of a virus. Viruses have the capability of infecting and inserting their genetic information into cells. Researchers are able to exploit this capability of viruses while removing the viral genes responsible for causing illness.

Although the concept of gene therapy is simple in theory there are several technical roadblocks that have to be overcome for these treatments to become a reality. For gene therapy to cure a disorder, the inserted gene must remain active in the body's cells long-term. Currently it is difficult to retain the added gene through multiple rounds of cell division, making it hard to achieve successful gene therapy in actively dividing cells. In addition, it is difficult to ensure that the vector containing the therapeutic gene reaches the organs and body tissues where symptoms occur. Some of the recent successes in gene therapy research have been in ocular (eye) diseases in which the targeted body area is easily accessible.
One of the major setbacks in the gene therapy research occurred in 1999 with the death of 18-year-old Jesse Gelsinger. Jesse had a rare genetic condition called ornithine transcarboxylase deficiency (OTCD) in which a gene mutation causes an enzyme, important for the removal of nitrogen from the body, to be absent. Jesse enrolled in a clinical trial for gene therapy of OTCD aimed at determining a safe dose for treatment and documenting potential side effects. Four days after starting the treatment, Jesse passed away from multiple organ failure thought to have been triggered by an immune response to the viral vector.



Gene Therapy, RNAi and their role in altering/silencing protein synthesis should be discussed in the Genetics course as a part of COS objective 7. The potential as treatment for disease, is described under Genetics COS objective 10 and AP Biology under the general theme "Science, Technology and Society." It could also be incorporated into a discussion about the relationship between DNA, RNA and proteins (COS objective 8) for a Biology class or Introduction to Biotechnology course (COS objective 9).

Researchers are working to overcome many of the roadblocks described above and are making promising strides in developing safe and effective methods for introducing functional genes into the body.

RNAi

Another type of gene therapy currently being researched is RNAi. Much like turning off a light switch, RNA interference (RNAi) offers the ability to selectively silence or "turn off" the activity of a single gene. This technology has the potential to revolutionize our understanding of how genes work and offers new promise in therapy and treatment.

In addition to mRNA and tRNA found in cells, researchers in the 1990s noted an additional form of RNA composed of small double-stranded molecules. These fragments could effectively stop protein production by coordinating the destruction of the single stranded mRNA. In other words, the double stranded RNA "interfered" with the mRNA, effectively silencing the activity of the gene. Researchers have utilized the RNAi pathway to explore the effects of systematically silencing genes. Short synthetic double-stranded RNA molecules can be created in the laboratory and delivered into cells, leading to partial or complete cessation of protein production for specific targeted genes. The ability to target and deplete specific proteins has identified RNAi as a potential therapeutic pathway.

STATUTES AND SESSION LAW

Title 40 REVENUE AND TAXATION.
Chapter 9 EXEMPTIONS FROM TAXATION AND LICENSES.
40-9-34 HudsonAlpha Institute for Biotechnology.

(a) The following is hereby found and declared by the Legislature of Alabama:

(1) The lack of content in natural and bio-science education offered to students in kindergarten through high school is a nationwide problem.

(2) Such lack in curricular offerings to students will be detrimental in the long-term to the economy of the state and the welfare of the citizens during the scientific revolution now engulfing the world.

(3) The biotechnology institute can provide to education leaders of the distance learning program of the state cutting edge biotechnology curriculum recommendations and content for Alabama high schools, by providing information about cutting edge biotechnology curriculum and content to students in kindergarten through high school pursuant to the distance learning program of the state, the state course of study, and state textbooks.

(4) By educating Alabama high school students in the field of biotechnology, such students are more likely to pursue careers in the biological sciences, thereby providing the state with a better educated workforce able to support the growing biotechnology industry, in turn attracting and encouraging biotechnology companies to locate in the state and create additional challenging and rewarding job opportunities for the citizens of the state.

(5) The reputation, economic status, and educational system of the state will be further enhanced by the addition of an internationally renowned biotechnology institute that will support internationally recognized scientists and researchers, with a focus on scientific discoveries that are intended, when possible, to be proven in the state and provided by companies in the state to patients suffering from diseases.

(6) By establishing a biotechnology campus, the biotechnology institute will be in a better position to join with the economic development leaders of the state to attract biotechnology companies to the campus and to the state, thereby creating additional job opportunities for the citizens of the state.

(b) The HudsonAlpha Institute for Biotechnology, a nonprofit corporation, and any real and personal property owned by the corporation, shall be exempt from the payment of any and all state, county, and municipal taxes, licenses, fees, and charges of any nature whatsoever, including any privilege or excise tax heretofore or hereafter levied by the State of Alabama or any county or municipality thereof.

(c)(1) In exchange for the tax exemption granted in subsection (b), beginning October 1, 2008, and for each year thereafter, **the HudsonAlpha Institute for Biotechnology shall make a report to the State Board of Education detailing the curricular content in biotechnology which could enhance the state distance learning program.** This subdivision shall not apply in the event that the distance learning program is discontinued, or is no longer in existence. **Further, the HudsonAlpha Institute for Biotechnology shall report annually to the State Board of Education, the State Course of Study Committee, and the State Textbook Committee all new developments in the field of biotechnology which could be integrated into the curriculum for high school courses in science and health.**

RECOMMENDED READING FOR MORE DETAILS

DNA sequencing

The impact of next-generation sequencing technology on genetics. Mardis E.R.,*Trends in Genetics*, Volume 24, Issue 3, March 2008, Pages 133-141

Rise of the machines. Gresham D. Kruglyak L., *PLoS Genetics*. 4(8):e1000134, 2008.

RNA analysis

Mapping and quantifying mammalian transcriptomes by RNA-Seq. Mortazavi A. Williams BA. McCue K. Schaeffer L. Wold B. *Nature Methods*. 5(7):621-8, 2008 Jul.

A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Sultan M. Schulz MH. Richard H. Magen A. Klingenhoff A. Scherf M. Seifert M. Borodina T. Soldatov A. Parkhomchuk D. Schmidt D. O'Keeffe S. Haas S. Vingron M. Lehrach H. Yaspo ML. *Science*. 321(5891):956-60, 2008 Aug 15.

Protein analysis

Biologists initiate plan to map human proteome. Pearson H., *Nature* vol 492, 920 (April 2008)

Bioinformatics

Steady progress and recent breakthroughs in the accuracy of automated genome annotation. Brent M., *Nature Reviews Genetics* 9, 62-73 (January 2008).

Agricultural applications

Where's the Super Food? Grant B. *The Scientist*. 23(9):30, September 2009
GM Crops: The First Ten Years - Global Socio-economic and Environmental Impacts. Brookes G, Barfoot P. *Economics*. 2006

Global Status of Commercialized Biotech/GM Crops: 2007. James C. *International Service for the Acquisition of Agri-Biotech Applications* (ISAAA). February 2008.

Cancer

http://www.cancerquest.org/ **CancerQuest is an excellent online resource that details both normal and cancer biology. It was developed as an educational outreach program by Emory University.**

Comparative genomics

Approaches to comparative sequence analysis: towards a functional view of vertebrate genomes. Margulies, E. H. & Birney, E. *Nature Reviews Genetics* 9, 303–313 (2008).

Mammalian karyotype evolution. Ferguson-Smith MA & Trifonov V. *Nature Reviews Genetics* 8, 950-962 (December 2007).

Copy number variation

Structural variation in the human genome. Feuk L, Carson AR, Scherer SW. *Nature Reviews Genetics* 7, 85-97 (February 2006).

Criminal Justice

Forensic DNA Phenotyping: Regulatory Issues, Koops BJ, & Schellekens M. *Columbia Science and Technology Law Review* 158 (2008)

To Sketch a Thief: Genes Draw Likeness of Suspects, Naik G, *The Wall Street Journal*, March 27, 2009 – accessed online at http://online.wsj.com/article/SB123810863649052551.html

Diagnosing chromosomal disorders

Pre- and postnatal genetic testing by array-comparative genomic hybridization: genetic counseling perspectives. Darilek S. Ward P. Pursley A. Plunkett K. Furman P. Magoulas P. Patel A. Cheung SW. Eng CM. *Genetics in Medicine*. 10(1):13-8, Jan, 2008.

Epigenetics

Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. Gluckman P, Hanson M, Buklijas T, Lowe F, Beedle A. *Nature Reviews Endocrinology* 5, 401-408 (July 2009)

Environmental epigenomics and disease susceptibility. Jirtle R, Skinner M. *Nature Reviews Genetics* 8, 253-262 (April 2007).

DNA methylation landscapes: provocative insights from epigenomics. Suzuki M, Bird A. *Nature Reviews Genetics* 9, 465-476 (June 2008).

Genetic information nondiscrimination act

A fact sheet and the text of the Legislative Act can be found at
http://www.genome.gov/24519851

Keeping pace with the times--the Genetic Information Nondiscrimination Act of 2008. Hudson KL. Holohan MK. Collins FS. *New England Journal of Medicine*. 358(25):2661-3, June 19, 2008.

Genetics of Eye Color

Molecular genetics of human pigment diversity. Strum RA. *Human Molecular Genetics*. 18(Review issue 1):R9-R17, April, 2009.

Identifying genetic influence on disease

http://www.nature.com/scitable/topicpage/Complex-Diseases-Research-and-Application-748

http://www.nature.com/scitable/topicpage/Genome-Wide-Association-Studies-GWAS-and-Obesity-752

These two webpages are from an educational website known as Scitable. Developed by Nature Publishing, Scitable is a free resource for educators, students and the public that is linked to the scientific reports published by the Nature publishing group.

http://www.genome.gov/20019523 **Fact Sheet on Genome Wide Association Studies – developed as an educational resource by the National Institutes of Health National Human Genome Research Institute**

Infectious disease

http://www.nlm.nih.gov/medlineplus/infectiousdiseases.html **This website provides an overview of bacteria, viruses and the body's response to infectious agents. It is developed in partnership with the National Library of Medicine and the National Institutes of Health.**

Non-invasive prenatal genetics

An offer you can't refuse? Ethical implications of non-invasive prenatal diagnosis. Schmitz D, Netzer C, Henn W. *Nature Reviews Genetics*. 10:515, August, 2009.

Practical and Ethical Considerations of Noninvasive Prenatal Diagnosis. Benn PA, Chapman AR. *Journal of the American Medical Association* 301, 2154-5, May 27, 2009.

Personal genome analysis

Research ethics and the challenge of whole-genome sequencing. McGuire A, Caulfield T, Cho MK *Nature Reviews Genetics* 9, 152-156 (February 2008).

American College of Medical Genetics Statement on Direct-to-Consumer Genetic Testing (2008) - http://www.acmg.net/StaticContent/StaticPages/DTC_Statement.pdf

Pharmacogenomics

http://www.ornl.gov/sci/techresources/Human_Genome/medicine/pharma.shtml **Information on Pharmacogenomics, including links for additional information. Developed by the US Department of Energy as part of their Human Genome Project overview and application pages.**

RNA interference

Exploring the Uses of RNAi — Gene Knockdown and the Nobel Prize. Bernards R *New England Journal of Medicine*. 355:2391, December 7, 2006.

Stem cells

http://stemcells.nih.gov/info/basics/ **A primer on stem cells developed by the National Institutes of Health.**

Studying the Genome to Understand the Sequence

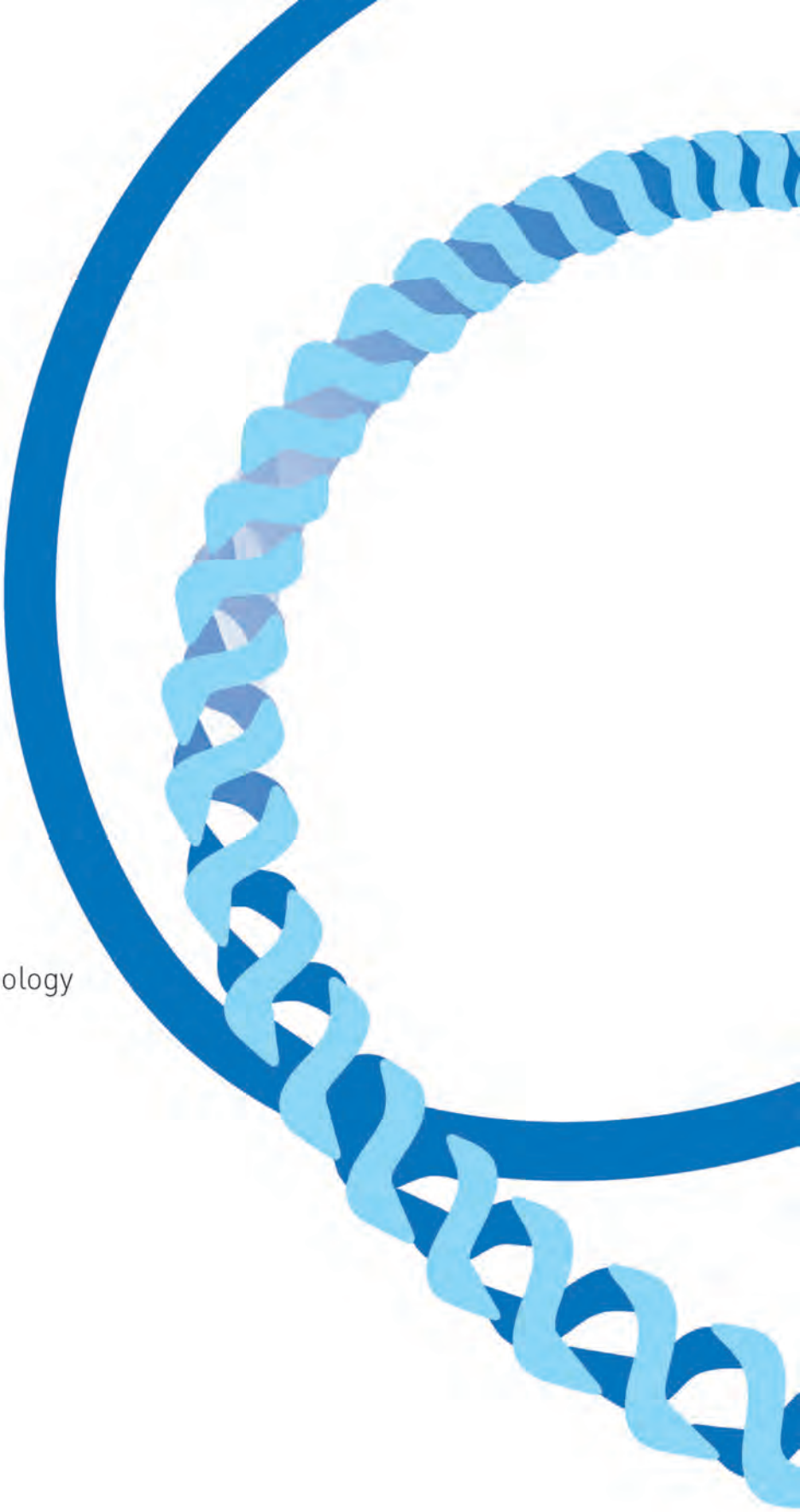
Comparing whole genomes using DNA microarrays. Gresham D, Dunham M., Botstein D., *Nature Reviews Genetics* 9, 291-302 (April, 2009).

Image Credits:

Page 13
Pinky bone – Max Planck Institute for Evolutionary Anthropology - http://www.eva.mpg.de/

Pages 16-17
Epigenetics – NHGRI
http://commonfund.nih.gov/epigenomics/epigeneticmechanisms.aspx

NOTES



Contact

Neil E. Lamb Ph.D.
Director of Educational Outreach
HudsonAlpha Institute for Biotechnology
601 Genome Way
Huntsville, AL 35806
nlamb@hudsonalpha.org
hudsonalpha.org

THE SCIENCE OF PROGRESS