Biotechnology Discoveries and Applications

Extensions to high school science curriculum

The 2014-2015 guidebook

HudsonAlpha
INSTITUTE FOR BIOTECHNOLOGY

CONTENTS

HUDSONALPHA
INSTITUTE FOR BIOTECHNOLOGY

science for life

About HudsonAlpha

HudsonAlpha is a nonprofit research institute committed to improving human health and quality of life through a unique three-fold mission of genomic research, economic development and educational outreach. A collaborative environment hastens the process from discoveries made in research laboratories into the lives of individuals, whether it be through patient care or improved agriculture.

Genomic Research

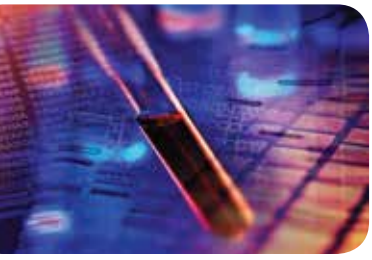
HudsonAlpha scientists are adding to the world's body of knowledge about the basis of life, health, disease and bio-diversity and seeking to enable

Earlier and/or less invasive diagnostics
Better, more customized treatments for disease
Improved food and energy sources

Current research focus areas are:

- Multiple forms of cancer, including breast, ovarian, prostate, kidney, colon and pancreatic
- Neurological and psychological disorders, including Alzheimer's, Parkinson's, ALS, bipolar disorder and autism
- Childhood genetic disorders, affecting 2 out of every 100 children born
- Immunogenomics, which is using genomics technology to understand the human immune system and related diseases such as lupus, rheumatoid arthritis, pancreatitis and psoriasis
- Agriculture and Bioenergy

Biotech Enterprises



HudsonAlpha strengthens and diversifies North Alabama's economy by attracting new and growing existing life science companies to the Tennessee Valley. Through industry recruitment, retention and expansion of the institute's Associate companies or encouraging entrepreneurship, HudsonAlpha takes a leading role in building a biotech hub in Alabama. HudsonAlpha's flagship building features 270,000 square feet of laboratory, office and collaboration areas and is the cornerstone of the 152-acre HudsonAlpha Biotechnology Campus in Cummings Research Park. Currently, 27 associate life sciences companies do business on the HudsonAlpha campus, developing new diagnostics and therapeutics, creating health-related products and offering varied services. McMillian Park, with its signature double helix walkway, runs the length of the campus and is the backbone for future expansion.

Educational Outreach

HudsonAlpha's Educational Programs

HudsonAlpha's educational outreach team inspires the next generation of researchers, while building a more biotech-literate community. The institute's dynamic educators are preparing future scientists through hands-on classroom modules, in-depth school and summer camp experiences, and digital learning opportunities. Additionally, the team builds awareness through community outreach classes and events. More than 400,000 individuals were impacted through HudsonAlpha education outreach in 2013-14.

Teacher Professional Development

Besides providing this guidebook, HudsonAlpha has several opportunities for teacher professional development, ranging from single-day workshops to a two-week academy. These increase an educator's comfort in discussing genetic concepts, terminology and associated ethical, social and legal issues. As part of all professional development activities, educators receive a genetics and biotechnology "toolkit" of laboratory activities, video clips, animations and online resources.

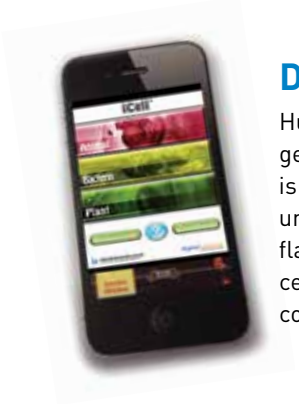
Student Experiences

Activities based on direct experience are some of the most powerful learning tools available to students. They provide a context that connects knowledge to relevancy. At HudsonAlpha, experiential learning includes field trips, classroom visits by industry leaders, summer camp sessions, in-depth internship opportunities and college-level laboratory courses. These activities engage students in biotechnology-related fields, increase exposure to career options, provide mentoring opportunities and equip students with a toolbox of content-specific skills. Communities looking to recruit science and technology occupations need to build a population of workers who can thrive in a knowledge-based economy. HudsonAlpha has crafted a pipeline of programs that blend conceptual understanding and skill acquisition to identify and engage our future workforce.

Classroom Kits and Activities

In 2007, HudsonAlpha began a partnership with the Alabama Department of Education to develop an eight-lesson module for seventh grade students matching state curriculum requirements related to DNA, how proteins are made and how genetic information is copied and segregated when cells divide. These activities have been incorporated into seventh grade classrooms across the state. Preliminary evidence on the module's impact is promising. Several teachers have shared that the percentage of their students achieving mastery on content standards addressed by the module has increased by 20 percentage points or more.

HudsonAlpha has also developed six laboratory activities for students in grades 9-12. Each activity meets state-mandated requirements for a range of courses. Activities highlight topics such as extracting DNA, exploring chromosome behavior in cells, diagnosing genetic disorders and using bioinformatics databases. Feedback has been overwhelmingly positive, with teachers expressing appreciation for the ability to expose their students to these hands-on activities.



Digital Resources

HudsonAlpha has crafted a suite of digital activities to showcase the history of genetics and biotechnology and explore the content of the human genome. iCell® is an interactive simulation that allows students and teachers to explore and understand the inner workings of a typical animal, plant or bacterial cell. Unlike flat, static images from a textbook, iCell® offers a full 3-D representation of cellular components and their dynamic interrelationships, giving students a context for learning fundamental cell structure and function.

genome cache

Build your own genome, or walk ours. **GenomeCache**® combines the challenge of a scavenger hunt with the human genome. It allows anyone to create up to 20 walkable paths that explore the human genome with over 150 challenging questions, a leaderboard and themed paths. GenomeCache combines clues, fun facts and trivia questions to create an engaging learning experience.

GenomeCache® is available on iPad®, iPhone®, through GooglePlay™ and at **genomecache.hudsonalpha.org**.

TOUCHING TRITON

Touching Triton™ is an online educational activity focused on building understanding of common complex disease risk. This serious game uses the engaging storyline of long-term space flight to highlight how risk is influenced by factors from family history, environment and an individual's genomic profile. By engaging students in an interesting storyline and graphical interface, we aim to create an activity that is both enjoyable and educational.

ARGOS

More information is available at hudsonalpha.org/education/touchingtriton

iCell®

Why use flat images from a textbook when your students can explore cell structure in 3D?

HudsonAlpha iCell®, one of Apple's featured biology apps on the iTunes® Education market, allows students to explore representative plant, animal and bacteria cells with vivid 3D models. iCell is available on multiple platforms and has been downloaded over 600,000 times by students and educators around the world.

iCell is available on Apple® and Android® **devices**, Windows 8® **tablets**, as a downloadable program for **Mac® and Windows®**, and at **icell.hudsonalpha.org**.

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**.

Genetic Resources to Empower Alabama Teachers

GREAT

Cell Division

The GREAT Workshop provides opportunities for Alabama public high school life science educators to update genetics knowledge and discover recent scientific findings that are too new for textbooks. This second round of GREAT workshops will dig into cell division by working through the following:

- How do chromosomes behave during cell division?
- What regulates the cell cycle?
- How does cancer occur?
- How is DNA sequencing changing the way cancer is treated?
- How can I teach these concepts to my students?

In two full days of small group concurrent sessions and talks by dynamic speakers, teachers will learn about recent findings and applications that relate to health, agriculture and the environment.

Coming soon to:

Huntsville SPRING 2015

Birmingham FALL 2015

Visit **www.hudsonalpha.org/GREAT** to register!

The GREAT Workshop is open to Alabama accredited, public high school life science teachers and is made possible through support from the State of Alabama.

Biotechnology Discoveries and Applications 2014-2015

HOW THIS GUIDE IS ARRANGED

Recent research findings are grouped on pages nine through twenty-one 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 teal.

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

These are identified by the  symbol in green.

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

The  symbol identifies those connections.

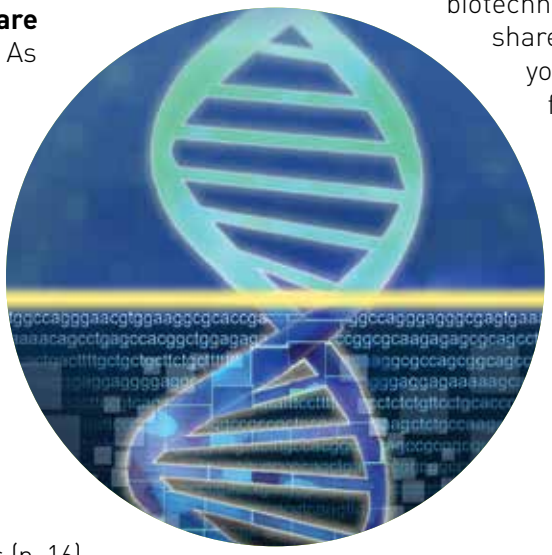
EXECUTIVE SUMMARY

I was introduced to DNA sequencing as a graduate student in 1992. At the time it was a cumbersome process. A thick sheet of plexiglass sat on the lab bench, shielding my torso from tubes of radioactively labeled DNA nucleotides that were used in the sequencing process. After the reaction was finished, the DNA fragments were separated by gel electrophoresis and the gel was exposed to X-ray film. Processing the film revealed regularly spaced smudges identifying150 bases of hard-won sequence.

Today, sequencing is radioactive-free and fully automated. The newest sequencing machines churn out hundreds of billions of bases per day. At least one sequencing company's technology produces an unannotated human genome for less than $1,000. Other emerging techniques promise the ability to sequence single strands of DNA – a mind-boggling concept.

At the HudsonAlpha Institute for Biotechnology, these technological advances allow our scientists to decipher the genetic codes of thousands of samples in less time and for less money than ever before. The data deepen our understanding of the genetic links to health, agriculture and bioenergy. This same thirst for knowledge unfolds at universities and research institutes across the globe. Collectively, our discoveries impact the clinic, farm, and power company - along with a host of other industries.

This edition of the annual *Biotechnology Discoveries and Applications* guidebook showcases both the enthusiasm and the challenge of research in genetics and biotechnology. **The guidebook provides educators with information about recent advances from around the world, allowing teachers to share those findings with their students**. As in years past, it is divided into two sections: research highlights and foundational concepts. More than 45 new discoveries are highlighted, including articles about:



- the discovery that different types of bacteria are found in wounds that do and don't heal (p. 9)
- the identification of a new organelle in plants (p.11)
- a DNA variant that increases tomato yields by 20 percent (p. 14)
- the identification of 127 genetic mutations common across cancers (p. 16)
- the potential forensic application of predicting facial shape from genetic information (p.18)

HudsonAlpha's Educational Outreach Team

Recent findings are linked to one or more foundational topics, covered in detail beginning on page 27. Each topic links to course of study objectives for science, health and relevant career technical education classes in Alabama. For quick reference, these linkages are also shown in table form on pages 22-26. Educators from states other than Alabama will find that these foundational topics easily align to their own state objectives.

Genomics is an important tool across the life science landscape, informing conversations in agriculture, health, bioinformatics and ethics. This guidebook equips teachers with the content and context that allows those conversations to occur.

On behalf of HudsonAlpha and the 27 biotechnology companies with whom we share our campus, I am excited to bring you these stories. We welcome your feedback, along with ideas for next year's edition.

Neil E. Lamb, Ph.D.
Vice President for Educational Outreach
HudsonAlpha Institute for Biotechnology
nlamb@hudsonalpha.org

SCIENCE SNAPSHOTS

a quick rundown of 10 genetics and biotech stories

1. A recently developed method rapidly identifies and tracks the microorganisms found in wounds. The microbiome of tissue samples from 44 soldiers wounded in Iraq or Afghanistan were analyzed, comparing bacterial populations present in wounds that heal to those that do not. The presence of bacteria commonly found in the gut was often associated with proper healing. *Note: This early finding does not imply that adding certain bacteria to a wound promotes healing.*

2. Two international teams have independently assembled the first drafts of the human proteome, or the entire collection of proteins produced by the genome. Each team produced a catalog of proteins identified from a collection of non-diseased human cells. The Human Proteome Map and ProteomicsDB catalogs provide a baseline for comparison with diseased tissues. Both teams uncovered new proteins in regions of the genome previously thought to be non-coding.

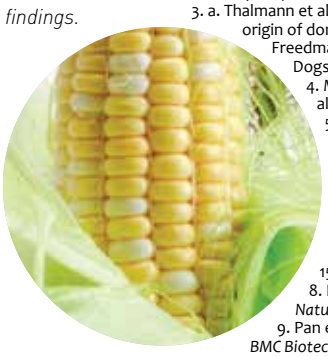
3. Recent research into dog domestication compared the genomes of grey wolves, African basenji, Australian dingo and a Boxer to see which lineage gave rise to domestic dogs. The results suggest that both modern dogs and modern wolves arose from a species of wolf that is now extinct. Similar conclusions came from a mitochondrial DNA analysis of modern wolves and dogs plus the fossilized remains of ancient dogs. In a departure from previous studies, these findings suggest dog domestication arose before the widespread use of human agriculture.

4. Scientists from The Scripps Research Institute have engineered a semi-synthetic organism that contains an artificially created base pair and can reproduce normally. This is the first time that DNA containing a synthetic base pair has been replicated successfully in a cell. This work has potential implications for information storage in DNA and the re-engineering of cells for a range of alternative applications.

5. To understand cell biology, scientists must identify the relationships between genes, RNA, proteins and metabolites - small molecules that interact with proteins to accomplish cellular functions. A multi-lab collaboration explored these relationships in yeast. More than 14,000 transcript, protein, metabolite and physical traits were measured across 22 different strains of yeast. The findings provide insight into how cellular players interact to produce trait diversity.

HudsonAlpha researcher Dr. Sara Cooper contributed to these findings.

6. The roots of ancestral varieties of corn produce a chemical called E-beta-caryophyllene when the plant is under attack from rootworms. This chemical attracts beneficial nematode worms from the surrounding soil, which kill the rootworm larva. Unfortunately, the ability to produce E-beta-caryophyllene was lost during corn domestication. Swiss scientists have introduced an oregano-based version of the responsible gene into corn. Preliminary data show the transformed corn releases the chemical consistently and suffers less rootworm root damage.



7. Recent analysis discovered that the pond-dwelling protozoan *Oxytricha trifallax* breaks its DNA into almost a quarter of a million fragments, stores them in a jumbled manner, and reassembles all the pieces when ready to mate. *Oxytricha* is a single-celled organism with two nuclei: the macronucleus carries out daily activities, and the micronucleus contains the jumbled archived version of the genome. The macronucleus has more than 16,000 nanochromosomes that average 3,200 bases in length, generally containing a single gene. They arise from a massive rearrangement of the 225,000 precursor DNA fragments found in the micronucleus.

8. Nanopore sequencing, an emerging technology that sequences single strands of DNA, recently deciphered a 4,500 nucleotide-long fragment from the genome of a bacteriophage, which is a virus that infects bacteria. Many of the current next-gen approaches generally sequence DNA in 50-250 nucleotide snippets.

9. Although the Polymerase Chain Reaction is one of the most important developments in modern biology, the process does not occur with equal efficiency across the genome. Sequences with a high percentage of either G/C or A/T nucleotide stretches do not amplify well. There are a number of reasons for bias, but the specific type of DNA polymerase used is a major contributor. A recent analysis of amplification bias from DNA polymerases collected from different organisms will help researchers account for differences in amplification frequency and increase representation of DNA from all parts of the genome.

This research was performed in the lab of HudsonAlpha researcher Dr. Jian Han.

10. An analysis of olfactory receptor genes across 13 different mammals identified more than 10,000 different genes, although only three smell receptors were common across the mammals. African elephants were the runaway winner with nearly 2,000 genes, twice as many as dogs and five times more than humans. The number of genes may not translate to better scent detection. Dogs, for example, have fewer olfactory receptor genes than mice yet still have a superior sense of smell.

REFERENCES

1. Be et al., "Microbial profiling of combat wound infection through detection microarray and next-generation sequencing," *Journal of Clinical Microbiology* (2014) 52:2583-2594 doi:10.1128/JCM.00556-14.
2. a. Wilhelm et al., "Mass-spectrometry-based draft of the human proteome," Nature (2014) 509:582-587 doi:10.1038/nature13319. b. Kim et al., "A draft map of the human proteome," *Nature* (2014) doi:10.1038/nature13302.
3. a. Thalmann et al., "Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs," *Science* (2013) 342:871-874 doi: 10.1126/science.1243650. b. Freedman A et al., "Genome Sequencing Highlights the Dynamic Early History of Dogs," *PLoS Genetics* (2014) 10(1) e1004016 doi:10.1371/journal.pgen.1004016.
4. Malyshev et al., "A semi-synthetic organism with an expanded genetic alphabet," *Nature* (2014) 509:385-388 doi:10.1038/nature13314.
5. Skelly et al., "Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast," *Genome Research* (2013) 23:1496-1504 doi:10.1101/gr.155762.113.
6. Based on a presentation by Ted Turlings at the July 2014 Society for Experimental Biology meeting: http://www.eurekalert.org/pub_releases/2014-07/sfeb-owh062714.php
7. Chen et al., "The Architecture of a Scrambled Genome Reveals Massive Levels of Genomic Rearrangement during Development," *Cell* (2014) 158:1187-98 doi:10.1016/j.cell.2014.07.034.
8. Laszlo et al., "Decoding long nanopore sequencing reads of natural DNA," *Nature Biotechnology* (2014) 32:829-833 doi:10.1038/nbt.2950.
9. Pan et al., "DNA polymerase preference determines PCR priming efficiency," *BMC Biotechnology* (2014) 14:10 doi:10.1186/1472-6750-14-10.
10. Niimura Y., Matsui A. and Touhara K., "Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals". *Genome Research* (2014) 24:1485-1496 doi:10.1101/gr.169532.113.

Sailing farmers

tracing ancient migration

Ten to twelve thousand years ago, hunter-gatherers along the eastern Mediterranean coast learned to domesticate grains. This transition to farming marks the shift between Paleolithic (Old Stone Age) and Neolithic (New Stone Age) eras. Over the next 2,000 years these Neolithic farmers expanded into Europe, intermixing with the Paleolithic populations that migrated there 30,000 to 40,000 years before.

A team of geneticists examined genetic markers from 32 modern Mediterranean and European populations to identify the migratory route these farmers traveled. The frequency of specific single nucleotide polymorphisms (SNPs) was compared across the population.

As the Neolithic people migrated and intermixed with existing populations, they introduced their particular SNPs into the native gene pool. In turn, they acquired some of the genetic information from the local population, a process repeated as the migrations continued into Europe. The signatures of these journeys are present in today's populations.

The findings suggest the Neolithic farmers first traveled over land to central Turkey. From there, they traveled by boat, sailing to Greece by way of the Dodecanese islands and the island of Crete. At that point, many continued on to Sicily and subsequently traveled across Italy into the remainder of Europe.

REFERENCE

Paschou P. et al., "Maritime route of colonization of Europe," *Proceedings of the National Academy of Science* (2014) 111:9211-9216 (2014). doi:10.1073/pnas.1320811111.

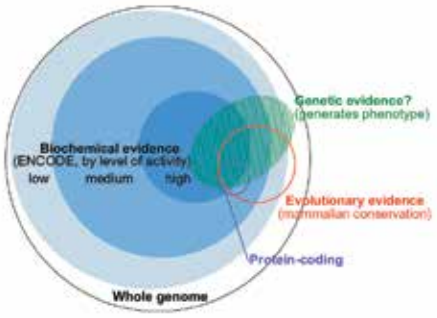
How much of the human genome is functional?

it depends on the definition of "function"

If "functional" means coding for proteins, only 1-2 percent of the genome is functional. Contrast this with the 2012 ENCODE study which suggested up to 80 percent of the genome has a "biochemical function". This definition includes regions bound by proteins that regulate transcription or folding as well as sequences that are transcribed but not translated. Many of these sequences are actually holdovers from evolutionarily ancient pathways no longer used by human cells, so 80 percent is likely an overestimation.

Recently, scientists compared the genomes of twelve mammals to identify DNA regions that remained nearly identical over 100 million years of evolution. Sequences that have undergone little change suggest the DNA has some functional purpose that requires its retention (see *Comparative Genomics* on page 33 for more details). With this definition, just over 8 percent of the human genome is functional.

However, a region of DNA can experience change over time and still have function. Many genes associated with the immune system experience rapid genetic change, but are definitely functional. Ultimately, the findings from multiple types of analyses will be necessary to determine the full set of functional sequences.



The lab of HudsonAlpha President Dr. Rick Myers participates in the ENCODE study.

REFERENCE

Rands et al.,."8.2% of the Human Genome Is Constrained: Variation in Rates of Turnover across Functional Element Classes in the Human Lineage," *PLoS Genetics* (2014) 10(7) e1004525 doi:10.1371/journal.pgen.1004525.

Kellis et al., "Defining functional DNA elements in the human genome," *Proceedings of the National Academy of Sciences* (2014) 111:6131-6138 doi: 10.1073/pnas.1318948111.

How do chromosomes condense before mitosis?

the classic model of chromatin folding is replaced with a two-phase process

During the cell cycle, chromosomes switch between two levels of condensation. Interphase chromosomes are loosely packaged and distributed across a relatively large area inside the nucleus. Different cell types organize DNA into characteristic spatial patterns, with large loops connecting genes and distant regulatory sequences and positioning needed genes into transcriptionally active conformations. In contrast, cells preparing to divide contain tightly condensed, transcriptionally silent chromosomes.

Various folding models have been suggested to explain how interphase DNA is compressed into metaphase chromosomes. A classic image in biology textbooks shows DNA coiling in a hierarchical manner to form thicker and thicker chromatin fibers. An alternate model suggests the DNA forms looping structures that attach to a protein backbone. To date, experimental results have been unable to confirm or refute either model.

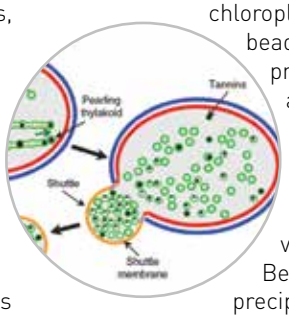
Using a combination of three-dimensional modeling, advanced computer simulation and next-generation sequencing, researchers are unraveling the process of chromosome assembly. Building on a decade of analysis, scientists found the data are inconsistent with the classical model of hierarchical DNA coiling. Instead, mitotic chromosomes are assembled in a two-phase process (see figure at right). First, the organizational scaffolds from interphase are disassembled and replaced by irregularly sized chromatin loops of 80,000-120,000 base pairs. These loops appear to form randomly along the chromosome and radiate outwards from a central DNA:protein scaffold. The second phase compresses the backbone of the loop scaffold, similar to squeezing a spring to flatten it. The DNA is

A new plant cell organelle

the tannosome originates within the chloroplast

A French research team has identified a new organelle in plant cells: the tannosome.

Originating inside chloroplasts, tannosomes produce tannins, plant-based molecules that lend a bitter, astringent taste to tea, wine and unripe fruit. They belong to the polyphenol class of organic chemicals and are thought to protect the plant from predators. The destruction or modification of tannins over time is an important part of fruit ripening and the aging of wine. Decaying vegetation along a stream or swamp can release tannins into the water, turning it a tea-like color.



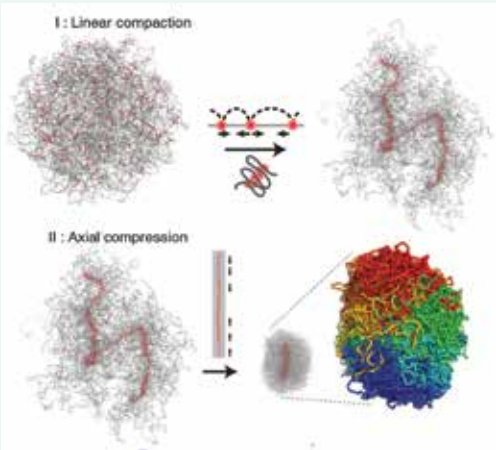
Using transmission electron microscopy, the researchers learned the pearl-shaped tannosomes form when membrane-bound thylakoids inside the chloroplast swell and form beads. The tannosomes produce tannins and are gathered into one region of the chloroplast, which buds off and shuttles the tannosomes to a vacuole for storage. Because tannins precipitate proteins and are lethal to the cell, they are kept separate from the other components of the cell and maintained at all times within membrane-bound structures.

REFERENCE

Brillouet J.M. et al., "The tannosome is an organelle forming condensed tannins in the chlorophyllous organs of Tracheophyta,"*Annals of Botany* (2014) 112:1003-1014 doi: 10.1093/aob/mct168.

condensed into a tightly folded package. Unlike the cell-specific organization found during interphase, the packaging of DNA prior to mitosis appears to be universal across all cell types.

Following mitosis, the chromosomes revert to their cell-specific spatial organization. The mechanisms that direct disassembly and reassembly of the interphase chromosomes are unknown. As most DNA binding proteins disassociate from the chromosome during mitosis, what is left behind to guide reassembly? The authors suggest histone tags, DNA methylation, or the few protein complexes that remain on DNA during mitosis may be responsible for this "cell-specific memory."



REFERENCE

Naumova N. et al, "Organization of the mitotic chromosome," *Science* (2013) 342:948-53 doi:10.1126/science.1236083.

In brief

Identifying genetic influence on disease

Functional effect of a mutation for autism

There are multiple genetic pathways for autism spectrum disorder (ASD). A population of affected individuals will contain a mixture of different genetic and environmental risk factors. A recent sequencing study of more than 6,000 children with ASD or developmental delay, identified 15 individuals with disruptive mutations in the *CHD8* gene. In addition to the behavioral characteristics associated with ASD, individuals with *CHD8* mutations displayed large head size, distinctive facial features and gastrointestinal disorders such as chronic constipation. When scientists introduced the disrupting mutations from humans into the *CHD8* gene of zebrafish, the fish also exhibited increased head size, due to neuron overgrowth, and digestive issues, caused by reduced innervation in the gut.

CHD8 belongs to a poorly understood family of genes associated with chromatin remodeling, which is the process that relaxes or tightens the coiling of DNA. This determines whether genes are accessible for transcription. The recent zebrafish findings support the idea that *CHD8* is an evolutionarily conserved master regulator of transcription for genes involved in neuron growth and development.

REFERENCE: Bernier et al., "Disruptive CHD8 mutations define a subtype of autism early in development," *Cell* (2014) 158:263-76 doi: 10.1016/j.cell.2014.06.017.

Applications of DNA sequencing

DNA barcoding identifies herbal supplements

Herbal remedies are the most rapidly growing segment of the North American alternative medicine market. No standards exist for the manufacturing and labeling of these products and some have been found with all or part of the herbal ingredients replaced by lower value fillers.

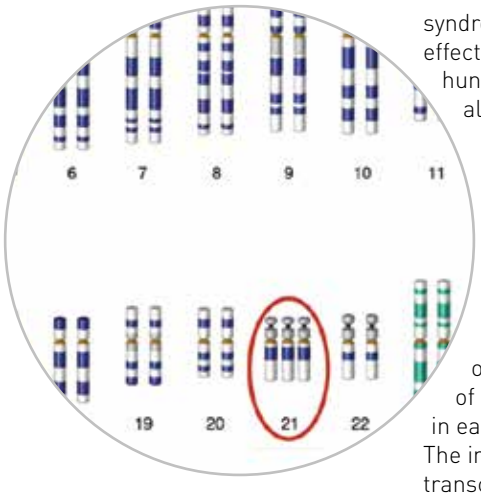
DNA barcoding is a technique that identifies plant material using standardized gene sequences. It offers a rapid and cost-effective method to identify contamination and substitution. A recent DNA barcoding study tested the authenticity of 44 herbal products from 12 companies. Nearly half of the samples did contain the herbs advertised on the packaging, but over 50 percent contained additional plant material not listed. One third of the samples included cheaper fillers, often rice, soybean and wheat. Only two of the 12 companies tested provided authentic products with no substitution, contaminants or fillers. In contrast, the supplements from three companies could not be authenticated at all . This represents a significant health risk for consumers who may be inadvertently consuming products with adverse or allergic side effects.

REFERENCE: Newmaster et al., "DNA barcoding detects contamination and substitution in North American herbal products," *BMC Medicine* (2013) 11:222 doi: 10.1186/1741-7015-11-222.

NEW FINDINGS

Silencing the effect of chromosomal trisomy

researchers exploit the process of x-inactivation to produce a chromosome 21 barr body



Typically, human cells contain two copies of each chromosome. Having more than two copies is generally not well tolerated. For example, trisomy (the presence of three copies) for chromosome 21 results in individuals with Down syndrome. The characteristics of Down syndrome are due to the effects of an extra dose of the hundreds of genes located along this chromosome.

Scientists have developed a technique to study the impact of an extra chromosome that borrows a behavior from female cells. Women inactivate one of the two copies of the X chromosome in each of their cells. The inactive X silences transcription from its genes, keeping the overall activity of these genes similar between women and men, who only have a single X. The silencing process occurs when the XIST gene on the X chromosome is transcribed to produce XIST RNA. This particular RNA molecule is not translated into protein but preferentially coats one of the X chromosomes and recruits other proteins to condense the DNA into a transcriptionally silent bundle known as a Barr body.

Scientists placed a copy of the XIST gene into a predefined place on one of the three copies of chromosome 21 in cells taken from an individual with Down syndrome and grown in a lab. As it would on the X chromosome, the transplanted XIST gene produced noncoding RNA that coated the third chromosome 21, resulting in its silencing. By shutting down the extra chromosome 21, the cells essentially returned to their typical state — two working copies of each gene.

This technique offers researchers a way to understand how extra copies of genetic information alter the biology of cells. Any type of tissue that is affected in Down syndrome could be used, comparing the function of cells before and after chromosome 21 inactivation. For example, scientists discovered that the cells grew more quickly and were better at forming neuron-making cells after chromosome 21 silencing. The approach could also shed light on the impact of other conditions caused by an altered number of other chromosomes.

While not a technologically feasible approach for therapy, this method is a promising first step towards understanding what goes awry during chromosome trisomy.

REFERENCE

Jiang et al., "Translating dosage compensation to trisomy 21," *Nature* (2013) 500:296-300 doi:10.1038/nature12394.

The Y chromosome gets a little respect

exploring the evolutionary history of the Y and its potential role in longevity

The tiny Y chromosome has been jokingly described as a genetic wasteland. Evolutionarily speaking, 200-300 million years ago it likely contained approximately 600 genes, most of which it shared in common with the ancestor of the modern X chromosome. At some point, 97 percent of those genes were lost from the Y chromosome and today, only about 19 of the shared genes remain, along with a handful that determine maleness. In the past, geneticists disagreed as to whether the Y chromosome was stable or if the remaining Y chromosome genes would ultimately be lost, their functions replaced by genes on other chromosomes during the next ten million years.

A recent paper comparing the evolution of the Y chromosome across eight mammals found that even distantly related species share most of the same functional Y chromosome genes. The Y chromosome size has remained stable over the last 25 million years. There are a dozen or so essential genes on the Y that seem to be major regulators, activated soon after conception and responsible for key cellular functions like protein synthesis or regulating the transcription of other genes. Many are present on the X chromosome as well, suggesting that two copies of these genes are required for normal development in men (XY) as well as women (XX).

Sometimes, body cells lose the entire Y chromosome through a nondisjunction event during mitosis. Although this loss of Y (LOY) has been previously associated with increasing age (older men show a greater proportion of Y-absent cells), until now a clinical effect of the loss has not been documented. A recent study of more than 1,100 men found LOY in 8.2 percent of the participants' blood samples. LOY was associated with an increased risk of cancer and a shortened life expectancy, with lifespan reduced by an average of 5.5 years.

Chromosome Y is one of the most commonly deleted chromosomes in human cancers, suggesting genes on this chromosome have a role in tumor suppression. The authors propose two hypotheses to explain their findings. The first suggests that LOY in white blood cells somehow alters the immune surveillance system that suppresses cancer formation. Alternately, LOY in blood samples may be indicative of similar chromosomal loss in other cells, which may increase cancer risk for those tissues.

HudsonAlpha researcher Dr. Devin Absher contributed to the LOY findings.

REFERENCES

Forsberg et al., "Mosaic Loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer," *Nature Genetics* (2014) 46:624-8 doi:10.1038/ng.2966.

Bellott et al., "Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators," *Nature* (2014) 508:494-99 doi:10.1038/nature13206.

A new approach to gene editing
using CRISPR to create targeted DNA changes

Although it sounds like a new-fangled refrigerator, CRISPR is a gene editing technology that creates targeted and specific DNA changes - correcting a single nucleotide mutation or inserting a DNA change with beneficial impacts. It has quickly become the go-to tool for modifying plant and animal DNA, successfully working in nearly every cell type tested.

CRISPR stands for clustered, regularly interspaced, short palindromic repeats. It is part of an ancient cellular defense system that protects bacteria and archaea from viral infection. Short sections of DNA in the bacterial genome are transcribed to produce small RNA molecules that match and bind to sequences from invading viruses. This binding activates a series of enzymes that chew up the viral genome and halt the infection.

Scientists are utilizing this process to make precise DNA changes in cells. Two components are required: the enzyme that cuts the target DNA and an RNA fragment that guides the enzyme to the target based on sequence complementarity. Both the gene for the enzyme and the guide sequence are delivered to the cell via a plasmid. After the DNA is cut, the break is repaired by an error-prone process that frequently creates a premature stop codon, inactivating the gene of interest.

Incredibly, for a couple hundred dollars and a few days work, CRISPR can inactivate almost any gene. Two research labs recently used CRISPR to knock out the function of nearly every gene in human cells.

The technique can be modified to correct an existing mutation. Cystic fibrosis causing mutations in the *CFTR gene* were recently corrected in patient stem cells.

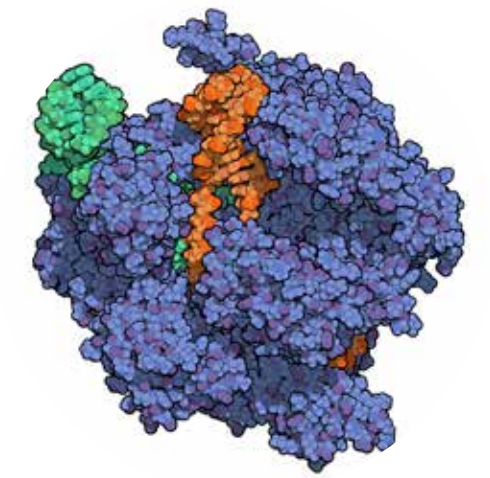
Some studies have documented "off target" effects, where DNA sequences similar but not identical to the RNA guide fragment have erroneously been cleaved. Despite its early successes, there is a long road to the clinical application of CRISPR for gene therapy.

REFERENCE

Wang et al., "Genetic Screens in Human Cells Using the CRISPR-Cas9 System," *Science* (2014) 343:80-84 doi:10.1126/science.1246981.

Shalem et al., "Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells," Science (2014) 343:84-87 doi:10.1126/science.1247005.

Schwank et al., "Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients," Cell stem cell (2013) 13:653–658 doi:10.1016/j.stem.2013.11.002.



In brief

Infectious disease

Sequencing the Ebola genome

The outbreak of Ebola virus disease that began in February 2014 is the largest known outbreak to date. A team of scientists sequenced 99 Ebola virus genomes that were isolated from the blood of 78 patients in Sierra Leone, West Africa. Like most viruses, the Ebola genome undergoes rapid mutation. Some of the mutations affect the rate of infection or severity of symptoms. Sequencing multiple individuals who have been affected allows epidemiologists and public health officials to track the spread and evolution of the disease from person to person. The Ebola virus genomes contain more than 300 genetic changes as compared to viral genomes from previous outbreaks. The functional consequences of these changes are still under analysis. Tragically, five of the co-authors of this study contracted Ebola virus disease and died before publication of the work.

REFERENCE: Gire et al., "Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak," *Science* (2014) doi:10.1126/science.1259657.

The genome of the tsetse fly

The tsetse fly transmits various species of *Trypanosoma* protozoa. Infection can lead to sleeping sickness in humans and a livestock disorder called nagana. Researchers completed a 10-year initiative to sequence the genome of a tsetse species. Analysis of the genome will provide insight into the behavior and life cycle of the tsetse fly, leading to better insecticides, traps and repellants.

REFERENCE: Internatinal Glossina Genome Initiative, "Genome Sequence of the Tsetse Fly (Glossina morsitans): Vector of African Trypanosomiasis," *Science* (2014) 344:380-386 doi:10.1126/science.1249656.

A new botulism toxin remains under wraps

Since the early days of the Human Genome Project, providing the public with access to DNA sequence has been standard. However, in the interest of public safety, the gene sequence for a newly discovered protein was omitted from the recent papers announcing its finding. A novel type of the extremely dangerous botulinum protein was identified from a child stricken with botulism. This is the eighth protein associated with botulism, all of which are produced by the bacterium *Clostridium botulinum*. Since no antitoxins have as yet been developed to counteract the effect of the protein, the Centers for Disease Control and Prevention and the Department of Homeland Security approved the publication conditional on omitting the gene sequence.

REFERENCES: Barash and Arnon, "A Novel Strain of Clostridium botulinum that Produces Type B and Type H Botulinum Toxins," *The Journal of Infectious Disease* (2014) 209:183-191 doi:10.1093/infdis/jit449.

Dover et al., "Molecular Characterization of a Novel Botulinum Neurotoxin Type H Gene," *The Journal of Infectious Disease* (2014) 209:192-202 doi:10.1093/infdis/jit450.

In brief

Genome Sequencing

An overview of recently sequenced organisms

The genomes of several organisms were sequenced recently. A sampling, along with genome size and predicted number of genes, includes:

	Genome Size (million bases)	Number of Genes
Bread wheat	17,000	100,000
African rice	316	33,150
Wild tomato	1,200	32,250
Fruit bat alphaherpesvirus 1	0.15	67
Electric eel	533	22,000
Coffee	710	25,500
Loblolly pine	22,200	50,100
Hot pepper	3,500	34,900
Migratory locust	6,500	17,300
Kiwifruit	758	39,000
Tiger	2,400	22,200

At HudsonAlpha, the Genome Sequencing Center has recently helped sequence the genomes of several plants, including:

	Genome Size (million bases)	Number of Genes
Clementine	301	25,000
Common Bean	587	27,200
Eucalyptus	640	36,350

Unlocking the power of African rice

Nine thousand years ago, hunter-gathers in Asia began domesticating wild strains of rice that evolved into what is today called Asian rice (*Oryza sativa*). Six thousands years later, early African farmers in the Niger River basin began domesticating a similar variety, producing African rice (*Oryza glaberrima*). The genome of African rice was recently sequenced, ten years after a similar analysis of Asian rice. A comparison of the two genomes shows that during the process of domestication, farmers in Africa and Asia independently selected for genetic traits that increase nutrition and make the rice easier to harvest.

Asian rice is arguably the most important food crop on earth, currently feeding half of the world's population. In contrast, African rice is cultivated in only a few places on the African continent. However, African rice is hardier and more resistant to environmental stress (drought, high salinity, flooding) than its Asian counterpart. Identifying the genes responsible for increased hardiness in African rice may lead to the production of Asian/African hybrids that produce high yields of grain, are more resistant to disease and pests, and require smaller amounts of water and fertilizer.

REFERENCE

Wang et al., "The genome sequence of African rice (Oryza glaberrima) and evidence for independent domestication.," *Nature Genetics* (2014) 46: 982-988 doi:10.1038/ng.3044.

The Arctic Apple

A GMO fruit that won't brown awaits US approval

"Arctic" apples have been genetically modified to resist browning when sliced. Developed by Australian researchers and licensed to a Canadian company, the apples are currently awaiting approval from the U.S. Department of Agriculture before they can be commercially grown and sold.

Browning occurs when enzymes called polyphenol oxidases (PPOs) react with antioxidant chemicals called phenols. Normally, phenols and PPOs are kept separate inside cells, but when the fruit is bruised, sliced or bitten, the compounds mix and browning occurs.

Arctic apples use RNAi technology to suppress the activity of PPO encoding genes, so the browning reaction does not occur. Essentially, the apple cells contain an additional fragment of DNA that is transcribed to produce an RNA molecule that binds and degrades the PPO gene transcripts. Even though the apple resists browning, it still decomposes over time, just like traditional apples.

Let there be light!

tomato DNA variant increases light tolerance & yields

Commercial growers often use artificial light to stimulate plant growth. While crops such as lettuces and peppers thrive under ever-lit conditions, tomatoes suffer a potentially fatal form of leaf damage when exposed to too much light. Consequently, growers limit tomato plants to 16 hours of light per day.

Wild tomatoes don't have overexposure problems, suggesting the domestication process inadvertently bred out light tolerance. Tomatoes were introduced to Europe by Spanish conquistadors. They brought only a few plants and seeds, meaning most of the genetic diversity of wild tomatoes remained in the New World.

To identify the genetic factors associated with light tolerance, domestic and wild tomatoes were crossed and recrossed multiple times. This produced several plants, each with a different combination of wild and domestic genetic information. The plants were tested using a series of genetic markers to identify which regions of the genome came from each parent.

All the light tolerant hybrid plants contained an identical section of chromosome 7 from the ancestral wild tomato. Additional studies of this region suggested a gene called *CAB-13* was the likely culprit. The gene encodes a component of the light harvesting complex – an array of protein and chlorophyll molecules embedded in the membrane of thylakoids found inside the chloroplast. This array captures light energy and transfers it to the photosynthetic reaction center, where it is ultimately converted into chemical energy in the form of ATP and NADPH. The light-sensitive plants

The company argues that nonbrowning apples will reduce the use of antibrowning chemicals commonly sprayed on sliced apples before packaging.

The USDA has allowed the company to grow the apples in test plots in New York and Washington. Golden Delicious and Granny Smith apples have been developed, with Gala and Fuji varieties in the research stage. USDA's preliminary review of the apple suggests it is nutritionally equivalent to traditional apples and has no elevated health or environmental risks. The review process included multiple opportunities for public comment. A final decision on whether the apples will be allowed is expected in late 2014 or early 2015. If approval is given, it will take several years before Arctic apples will appear in the produce aisle of grocery stores, although convincing stores to stock the fruit may be challenging.

For additional information about the RNAi technology used to silence the PPO genes, see *Therapeutic Approaches* on page 50.

REFERENCES

This GMO Apple Won't Brown. Will That Sour The Fruit's Image? http://www.npr.org/blogs/thesalt/2014/01/08/260782518/this-gmo-apple-wont-brown-will-that-sour-the-fruits-image.

USDA draft environmental assessment on Artic apples: http://www.aphis.usda.gov/brs/aphisdocs/10_16101p_dea.pdf.

contained a nine base deletion in a region of *CAB-13* that regulates transcription rates. Under continuous light conditions, this deletion prevents the gene from producing additional protein, disrupting the process of photosynthesis and leading to leaf damage and plant death.

Using traditional plant breeding methods, the full-length version of the *CAB-13* gene was crossed back into the domesticated varieties of tomato. The plants thrived under continuous light. Amazingly, tomato yield increased by 20 percent!

This genetic change represents a significant breakthrough for commercial growers, especially during the winter months when there is less sunlight and a greater need for artificial lighting. While continuous lighting means increased yields, it also results in higher energy costs. Continuous photosynthesis also leads to greater evaporation and increased moisture levels in the greenhouse. Large-scale trials are needed to confirm the net benefit of continuous light.

REFERENCE

Velez-Ramierez et al., "A single locus confers tolerance to continuous light and allows substantial yield increase in tomato," *Nature Communications* (2014) 5:4549 doi:10.1038/ncomms5549

Strengthening soybean strains

identifying genes that improve growing conditions

Once a commercially important crop like soybean has been sequenced, scientists and farmers can identify genetic changes responsible for desirable qualities like drought tolerance or pest resistance. This often involves comparing the genomes of domesticated plants to their wild ancestors.

Although the wild relatives of modern crops are usually less productive than their domesticated cousins, they have survived thousands of years of extreme environmental conditions, often as a result of subtle changes in their genetic code. These changes are often lost during the process of domestication.

A team of researchers working with soybeans recently identified a gene linked to salt tolerance. This gene, *GmCHX1*, contains an inactivating mutation in all salt-sensitive strains of soybean but is fully functional in wild varieties. Initial laboratory experiments suggest the gene encodes a transport protein that prevents sodium ions from accumulating in the cytoplasm of cells. Other plants increase salt tolerance by compartmentalizing sodium ions into the cell's vacuole. It remains to be seen if *GmCHX1* works through a similar mechanism.

Soil salinity is a major abiotic stress for plants. Globally, salinization affects more than 20 percent of irrigated lands. Crossing the functional version of *GmCHX1* into modern soybean varieties will hopefully increase the conditions under which this critical crop can be grown.

Temperature is also an abiotic factor that influences plant growth. Soybeans, like many plants, are often grouped into two types: determinate, which are bush-style varieties that stop growth of the main stem once they begin to flower, and indeterminate, which are often taller and bloom and set bean pods throughout the season. Determinate soybeans perform well in the longer growing season of the south, while indeterminate varieties produce better yields in northern climates. Unfortunately, the tall indeterminate soybeans are susceptible to lodging, a bending or breaking of the primary stem.

Soybean geneticists recently identified a mutation in a gene called *Dt2*, that impacts main stem growth. This dominant-acting mutation results in mid-sized, semi-determinate plants that produce as many soybean pods as their taller cousins but without the risks of lodging. This finding deepens our understanding of how growth pathways affect plant size and yields. It also opens the door to selectively breeding new high-yielding, semi-determinate strains of soybean.

HudsonAlpha researchers Jeremy Schmutz and Dr. Jane Grimwood contributed to sequencing the soybean genome.

REFERENCES

Qi et al., "Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing," *Nature Communications* (2014) 5:4340 doi:10.1038/ncomms5340.

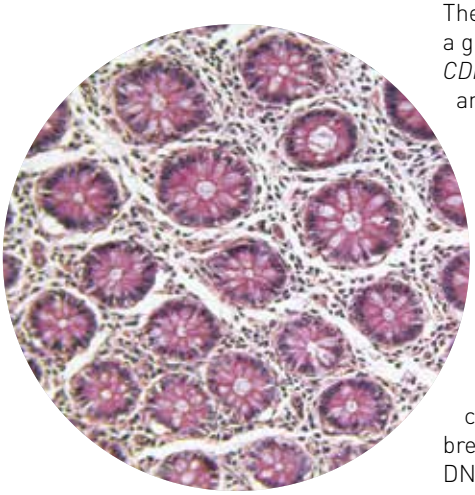
Ping et al., "Dt2 is a gain-of-function MADS-Domain factor gene that specifies semideterminancy in soybean," *Plant Cell* (2014) 26:2831-42 doi: 10.1105/tpc.114.126938.

NEW FINDINGS - GENETICS AND GENOMICS IN THE CLINIC

Understanding and treating cancer

how scientific discoveries in the lab are driving cancer diagnosis and treatment in the clinic

Fusion transcripts help drive cancer development
Two groups of researchers have implicated previously unknown gene fusions in the formation of cancer. Gene fusions can occur when two DNA strands are broken and incorrectly rejoined, connecting the beginning of one gene to the end of another.



One of the best-known gene fusions is *BCR-ABL*, which joins the *ABL* gene from chromosome 9 to the *BCR* gene on chromosome 22. Found in patients with chronic myeloid leukemia, the *BCR-ABL* fusion gene encodes a protein that is signaling for cell growth and division.

The new research identified a gene fusion between *CDKN2D* on chromosome 19 and *WDFY* on chromosome 13 in 20 percent of 60 ovarian cancer tumors. The product of this fusion appears to impact a known cancer pathway.

In a separate study, two different novel gene fusions were consistently identified in breast cancers. Instead of DNA breakage and improper rejoining, these fusions are caused by transcript read-through, where the last exon of one gene is spliced to the first exon of a neighboring gene to create one large in-frame transcript. In both instances, the fused transcript was translated. Experiments that reduced levels of the fusion transcripts also reduced breast cell proliferation, indicating the fusions play a role in cancer progression.

With further study, these fusions may become signatures for cancer detection or important targets for drug development.

The Pan-Cancer Effort
In a massive analysis, the Pan-Cancer Initiative of The Cancer Genome Atlas analyzed 5,000 cancer samples from 12 cancer types (breast, uterine, ovarian, lung, glioblastoma, head and neck, colon and rectum, bladder, kidney and leukemia). The average number of DNA mutations per million bases of DNA ranged from a low of 0.28 (leukemia) to 8.15 (lung squamous cell carcinoma). The majority of these mutations are not in genes important to cancer formation, but researchers identified 127 commonly mutated genes that are contributors to tumor development and progression. Cancers generally had mutations in two to six of these "driver" genes. Some mutations were exclusive to one tumor type, while others were present in several. The most commonly mutated gene was TP53, found in 42 percent of all cancers.

Clinical sequencing

annotating genomes and returning incidental findings

Exome sequencing continues to transition from the realm of the research lab into a clinical diagnostic tool. The approach analyzes the approximately 180,000 exons within the 22,000 human genes. Although the exome accounts for only 2 percent of the genome, more than 85 percent of known disease-causing changes occur within these regions.

Determining which variants are functionally important
Sequencing a human exome routinely identifies 200,000 to 400,000 single nucleotide variants, small insertions and deletions compared to the reference sequence. A series of analyses identify the biologically important changes. Generally, these computer-dependent processes compare variant frequency within human populations and across organisms and predict the effect of the change on protein amount, structure or function. A recently published bioinformatics program integrates information from multiple analysis methods into a single measurement. Known as combined annotation-dependent depletion (CADD), this method predicts how harmful a genetic change may be and is a valuable tool for distinguishing detrimental variants from benign DNA changes.

Incidental Findings
Exome sequencing may uncover incidental findings, DNA changes not connected to patient symptoms but with important secondary implications. These may be health-related (i.e. predisposition to adult-onset cancers, carrier status for recessive disorders) as well as societal (issues surrounding paternity). This knowledge may lead to early medical intervention but can also lead to unnecessary medical procedures and distress. There is an ongoing debate around which incidental findings should routinely be shared with patients:

- In 2013, the American College of Medical Genetics and Genomics released a list of 56 genes where incidental findings should always be reported. This year, the organization revised their stance, giving patients the choice to "opt out" of receiving the findings.

- One analysis explored the frequency of these incidental findings on 543 exomes from 159 families. Reportable variants were identified in 8.8 percent of the families.

- More than 1,500 U.S. adults responded to an online survey about a hypothetical genetic research study, with 56 percent indicating they would be willing to participate in such a project. Of those, 78 percent expressed interest in receiving incidental findings, provided there was no cost and results were only related to risk of serious or treatable diseases.

- Approximately 800 genetics experts were surveyed about reporting incidental findings. Most believed patients who undergo this type of sequencing should have the option to hear about incidental findings. Importantly, the majority felt that patient preference is important; that is, patients should have a say in what type of incidental results they want to receive.

Genetic testing on tumor samples impacts therapy choice and survival

Targeting the genetic mutations associated with cancer growth and progression has begun to impact patient care. The Lung Cancer Mutation Consortium tested more than 700 tumor samples between 2009-2012 for mutations in 10 genes known to drive cancer formation. Mutations were identified in 64 percent of patients, and more than one quarter of the patients received targeted therapies or enrolled in clinical trials based on mutation status.

Among individuals with identified driver mutations, those that received matched therapy lived an average of one year longer than those without matched therapy.

The research on breast cancer fusion genes was carried out in the lab of HudsonAlpha President Dr. Rick Myers.

For more information, see *Cancer* on page 32.

REFERENCES

Varley et al., "Recurrent read-through fusion transcripts in breast cancer," *Breast Cancer Research and Treatment* (2014) 146:298-97 doi:10.1007/s10549-014-3019-2.

Kannan et al., "CDKN2D-WDFY2 is a cancer-specific fusion gene recurrent in high-grade serous ovarian carcinoma," *PLoS Genetics* (2014) 10(3):E1004216 doi:10.1371/journal.pgen.1004216.

Kandoth et al., "Mutational landscape and significance across 12 major cancer types," *Nature* (2013) 502:333-339 doi:10.1038/nature12634

Kris et al., "Using multiplexed assays of oncogenic drivers in lung cancer to select targeted drugs," *Journal of the American Medical Association* (2014) 311:1998-2006 doi:10.1001/jama.2014.3741.

Impact on patients

Currently, exome sequencing has the greatest benefit in identifying rare disorders from mutations in a single gene. The technology provides a comprehensive approach to diagnosing unexplained disorders and saves time and costs compared with searching one gene at a time. Ongoing studies suggest exome sequencing yields a definitive (or at least plausible) biological explanation in approximately one-fourth of these patients. Because the biology of many diseases is still poorly understood, the likelihood a diagnosis will impact treatment is still frustratingly small. Additional research is needed to link changes in DNA to disease biology and improve patient care.

HudsonAlpha researcher Dr. Greg Cooper helped develop the CADD annotation tool.

REFERENCES

Kircher et al., "A general framework for estimating the relative pathogenicity of human genetic variants," *Nature Genetics* (2014) 46:301-315. http://dx.doi.org/10.1038/ng.2892.

ACMG press release on the update to reporting incidental findings: https://www.acmg.net/docs/Release_ACMGUpdatesRecommendations_final.pdf

Lawrence et al., "The implications of familial incidental findings from exome sequencing: the NIH Undiagnosed Diseases Program experience," *Genetics in Medicine* (2014) doi:10.1038/gim.2014.29.

Bollinger et al., "Public preferences for the return of research results in genetic research: a conjoint analysis," *Genetics in Medicine* (2014) doi:10.1038/gim.2014.50.

Yu et al., "Attitudes of genetics professionals toward the return of incidental results from exome and whole genome sequencing," *American Journal of Human Genetics* (2014) 95:77-84 doi:10.1016/j.ajhg.2014.06.004.

In brief

Identifying genetic influence on disease

CFTR mutations - 1 gene; 2 diseases?

The cystic fibrosis transmembrane conductance regulator (*CFTR*) gene encodes a channel protein that spans the membrane of epithelial cells. When the channel is open, anions (particles with a negative charge) move across the membrane. This regulates the movement of water across cell membranes producing thin, freely flowing mucus. Mutations in this gene can result in cystic fibrosis (CF), an autosomal recessive disorder. These CF mutations eliminate the movement of chloride ions into and out of cells. Consequently, CF patients have thick and sticky mucus that can damage organs, clog airways, and form extensive scar tissue in lungs.

The CFTR channel also moves bicarbonate, a chemical the pancreas produces to neutralize stomach acid. Bicarbonate is also used to control mucus consistency in the sinuses and plays a role in male reproductive health. New research has identified several *CFTR* mutations that preserve chloride secretion but impair bicarbonate transport. Individuals with these mutations do not have CF but are instead affected by painful pancreatitis. They cannot neutralize highly acidic digestive enzymes from the pancreas, resulting in inflammation and pancreatic scarring. Additionally, these mutations are associated with recurrent sinus infections and male infertility.

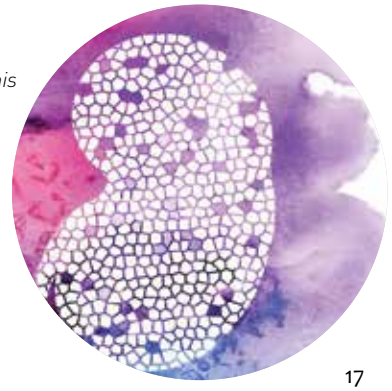
This research suggests that depending on the functional impact of the mutation, CF could be considered as two diseases, one that affects multiple organs, including the lungs, and one that doesn't impact the lungs at all.

REFERENCE: LaRusch et al., "Mechanism of CFTR functional variants that impair bicarbonate permeation and increase risk for pancreatitis but not for cystic fibrosis," *PLoS Genetics* (2014) 10(7): e1004376 doi:10.1371/journal.pgen.1004376.

Distinguishing heterogeneity and phenocopy

Some diseases show *locus heterogeneity*, where mutations in different genes lead to identical disease symptoms. This is different from *clinical phenocopy*, where mutations lead to diseases with similar-looking symptoms but may require different treatments. Heterogeneity and phenocopy make diagnosis a challenge. Correctly identifying a disorder can be critical for determining treatment and recurrence risk. As an example, patient exomes were sequenced across eight families with chronic kidney disorders. Based on symptoms, all the families had been given a diagnosis of nephronophthisis (NPHP), a recessive disorder that results in end-stage kidney failure. Six families actually harbored mutations in one of the 15 genes responsible for NPHP, but two families had been misdiagnosed and had genetic mutations that caused different kidney disorders. This points to the strength of genetic testing for providing definitive diagnoses.

HudsonAlpha researcher Dr. Shawn Levy participated in this work.



REFERENCE: Gee et al., "Whole-exome resequencing distinguishes cystic kidney diseases from phenocopies in renal ciliopathies," *Kidney International* (2014) 85:880-887 doi:10.1038/ki.2013.450.

In brief

Understanding microbial communities

Malnutrition and the human gut microbiome

The populations of bacteria living in children's intestines change dramatically as they are weaned from breast milk or formula to solid foods. By the time they are toddlers, children's gut microbiomes are pretty similar to those of adults. Researchers recently developed a model that tracks the microbial transition towards maturity by measuring the prevalence of 24 key gut bacteria. These microbial species were analyzed among 64 severely malnourished infants and toddlers in Bangladesh, both before and after diet therapy.

When compared to healthy controls, the microbiomes of the underfed children were immature – they looked more like the microbiome from much younger children. After a month of treatment, the children gained weight and their microbiomes matured dramatically. However, this maturation was short-lived, and by four months after treatment had ended and the children were on standard diets, the bacterial populations regressed to a less mature state.

It is well known that even after treatment, many malnourished children remain below normal weight and suffer from stunted growth. Poorly developed communities of gut microbes may contribute to this frustrating outcome. Different forms of diet therapy may need to be developed, including those that incorporate probiotics (useful bacterial populations) and/or prebiotics (nutrient sources that encourage the growth of useful bacteria).

REFERENCE: Subramanian et al., "Persistent gut microbiota immaturity in malnourished Bangladeshi children," *Nature* (2014) 510:417-421 doi:10.1038/nature13421.

Don't put your money where your mouth is

Researchers in New York City recently set out to examine the microbiome of money, specifically, the bacteria living on U.S. one dollar bills. Analyses were carried out during February and July of 2013, allowing a comparison between Winter and Summer. DNA from the bacteria was extracted, sequenced and analyzed for 80 circulating bills.

More than 3,000 different types of bacteria were identified, about half of which are naturally found living in and on humans. Most were harmless, although antibiotic resistant strains such as methicillin-resistant *Staphylococcus aureus* (MRSA) were identified. There was a seasonal effect as well, with increasing concentration of pneumonia microbes present during winter. This suggests that money may serve as a method of transmission for illness.

The number of bacteria found on money is perhaps to be expected. Cash carried in a back-pocket wallet or a purse held close to the body provides a warm space for microbes to thrive. This is another reminder of the ever-present nature of bacteria and nothing to overreact about. Still, it's probably best to avoid licking your fingers while counting out cash.

REFERENCE: Douiey, Caren Professor's study finds microbes on dollars. Washington Square News April 29, 2014. Accessed at http://www.nyunews.com/2014/04/29/dirty-2/ on September 15, 2014.

Trying to detect Alzheimer's

looking for biomarkers that predict progression

Alzheimer's affect one in nine Americans over the age of 65. In 2011, the National Institute on Aging and the Alzheimer's Association proposed three stages to define the progression of the disorder: preclinical Alzheimer's disease, mild cognitive impairment (MCI) due to Alzheimer's disease, and dementia due to Alzheimer's disease.

Accurately diagnosing the disorder is a challenge. A major focus of Alzheimer's research is around identifying clinical biomarkers for diagnosis. Biomarkers are proteins and other small molecules in the blood, brain or cerebrospinal fluid that accurately diagnose individuals at each phase and predict the likelihood of progression.

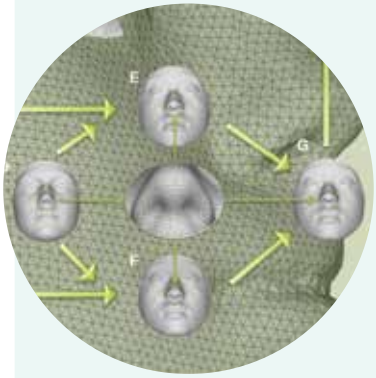
After testing thousands of potential blood-based biomarkers, researchers recently identified 10 lipids (fats, oils, and certain hormones) that may help identify individuals who will develop MCI. The panel has a reported accuracy of 90 percent. The researchers calculated a 90 percent sensitivity (meaning it will miss one in 10 people who will advance to MCI) and a 90

Using DNA to predict facial shape

emerging technique may support forensic efforts

Only a handful of genes contribute to eye, hair and skin pigmentation. In contrast, the number of genes that contribute to facial shape is significantly greater. A recent attempt was made to identify many of these genes.

High-resolution photographs were taken of the faces of nearly 600 individuals of mixed European and West African ancestry. These images were converted to 3D models, and more than 7,000 data points were placed across the digital face. These points measured how much specific facial features differed from the average – for example, if the upper lip was fuller or the cheekbones were higher.



With this information in hand, researchers examined the participants' DNA at 76 single nucleotide polymorphisms (SNPs) in candidate facial genes. Keeping gender and ancestry in mind, the likelihood that any given single nucleotide polymorphism SNP helped shape a facial feature was calculated. Two dozen SNPs from 20 genes were significantly associated

percent specificity (it will give a false positive result for one in 10 people who will remain mentally sharp). Separately, a different panel of 10 proteins was identified that predicts the progression from MCI to the dementia phase with 87 percent accuracy.

The false positive and false negative frequencies of both tests are much too high to be of use in a clinical setting. For example, imagine a population of 1000 seniors who take the lipid-based test. Previous research has found that the prevalence for converting to MCI is about 5 percent, meaning 50 people will develop MCI and 950 will not.

If the lipid test has a 90 percent sensitivity, 45 of the 50 will have a positive result (five will have a false negative). Similarly, with a 90 percent specificity rate, 855 of the 950 will correctly test negative, but 95 will receive a false positive. Out of the 1,000 people tested, 45+95 = 140 will receive a positive result, but only 45 of these will be truly positive. In other words, two out of every three positive tests are actually false positives.

In spite of these limitations, the findings represent an important early step towards more accurate and earlier diagnosis - especially for a disorder sorely lacking in predictive biomarkers.

REFERENCES

Mapstone et al., "Plasma phospholipids identify antecedent memory impairment in older adults," *Nature Medicine* (2014) 20:415-418 doi:10.1038/nm.3466.

Hye et al., "Plasma proteins predict conversion to dementia from prodromal disease," *Alzheimers & Dementia* (2014) early online access DOI: 10.1016/j.jalz.2014.05.1749 (2014).

with facial shape. These were incorporated into a computer program that converts genetic information into facial shape.

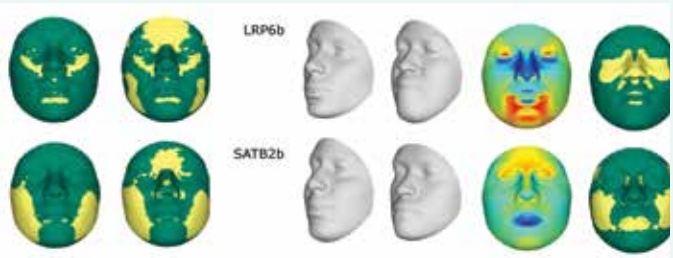
A similar study recently connected facial features from more than 1,000 Han Chinese to five SNPs in regions different from the European/West African analysis. Taken together, these represent early, but important steps towards "forensic molecular photo fitting," or the ability to reconstruct the identifying features of an individual based on genetic evidence alone. Although they currently produce crude reconstructions, the modeling will improve as more genetic predictors are identified.

HudsonAlpha researchers Dr. Greg Barsh and Dr. Devin Absher contributed to this work.

REFERENCES

Claes et al., "Modeling 3D Facial Shape from DNA," PLoS Genetics (2014) doi:10:e1004224 doi:10.1371/journal.pgen.1004224.

Peng at al., "Detecting Genetic Association of Common Human Facial Morphological Variation Using High Density 3D Image Registration," PLoS Computational Biology (2013) 9:e1003375 doi:10.1371/journal.pcbi.1003375.



In brief

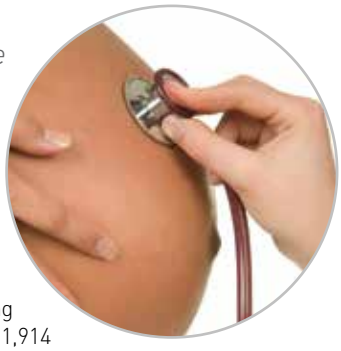
Gene therapy

Gene therapy for Parkinson's disease

Clinical scientists recently reported early results from a gene therapy trial to treat Parkinson's disease. In a twist on the classic gene therapy approach, the trial delivered not one but three genes simultaneously. The genes control the production of a neurotransmitter called dopamine, which helps nerve cells control muscle movement. The cells that usually produce dopamine slowly die in the brains of people affected with Parkinson's disease. The genes were carried by a lentivirus that was injected into a region in the center of the brain called the striatum. The virus infected the cells, activated the three genes it carried, and turned those cells into dopamine-producing factories. During the year following treatment, most of the side effects associated with the therapy were mild. Improvements in motor function were noted, although the improvements were mild and could be due to the placebo effect. This highlights the ongoing challenges associated with identifying the correct gene and dosing level to produce a truly therapeutic patient response. Regardless, the data demonstrate that as the field matures, the safety challenges associated with this type of gene therapy are being overcome.

REFERENCE: Palfi et al., "Long-term safety and tolerability of ProSavin, a lentiviral vector-based gene therapy for Parkinson's disease: a dose escalation, open-label, phase 1/2 trial," *Lancet* (2014) 383:1138-46 doi:10.1016/S0140-6736(13)61939-X.

Prenatal testing



The effectiveness of noninvasive prenatal testing

Previous versions of the Guidebook have highlighted the development of noninvasive prenatal tests that screen for chromosome aneuploidy by rapidly sequencing fragments of fetal DNA released into the maternal bloodstream from the placenta. A 2014 study finds this technique trumps standard screening methods. The multi-center trial involved 1,914 women from the general population. Trisomy for chromosomes 21 and 18 were detected either by sequencing fetal cell-free DNA (cfDNA) or through an analysis of maternal blood markers with or without an associated ultrasound. Diagnostic testing or data from newborn examination confirmed the screening results.

The false positive rates (a positive screening test even though no trisomy is present) were much lower with cfDNA testing: 0.3 percent vs 3.6 percent for trisomy 21 and 0.2 percent vs 0.6 percent for trisomy 18. These findings are impressive, although there are important limitations. Aside from18 and 21, analysis rates are not consistent among other chromosomes; there are several technological and ethical hurdles to be addressed before sequencing fetal genomes in this manner. In addition, because the cfDNA is shed from the placenta and not directly from the fetus, there is always a small risk that chromosome nondisjunction could have occurred after placental and fetal cells differentiated. Follow up diagnostic testing (amniocentesis) is still recommended to make a prenatal diagnosis.

For more details, see *Noninvasive Prenatal Diagnosis* on page 42.

REFERENCE: Bianchi et al., "DNA sequencing versus standard prenatal aneuploidy screening," *New England Journal of Medicine* (2014) 370, 799–808 doi: 10.1056/NEJMc1405486.

Neil Lamb, Ph.D.
Vice President for Educational Outreach

Genomics-driven oncology

Nearly all cancers are caused by genetic changes that alter important biological pathways controlling cell growth and survival. Specific genetic changes influence the rate of cell growth, determine how aggressively the cancer will spread and control whether one drug will be more effective than another at killing the cancer cells.

Over the past decade, advances in genomic technologies, tumor analysis and drug development have changed the landscape of cancer diagnosis and treatment.

In the laboratory, genomic information obtained from cancer cells has reshaped understanding of how cancer forms. In the clinic, this same information is beginning to guide therapeutic decisions, improving outcomes for patients with cancer.

Glioblastoma

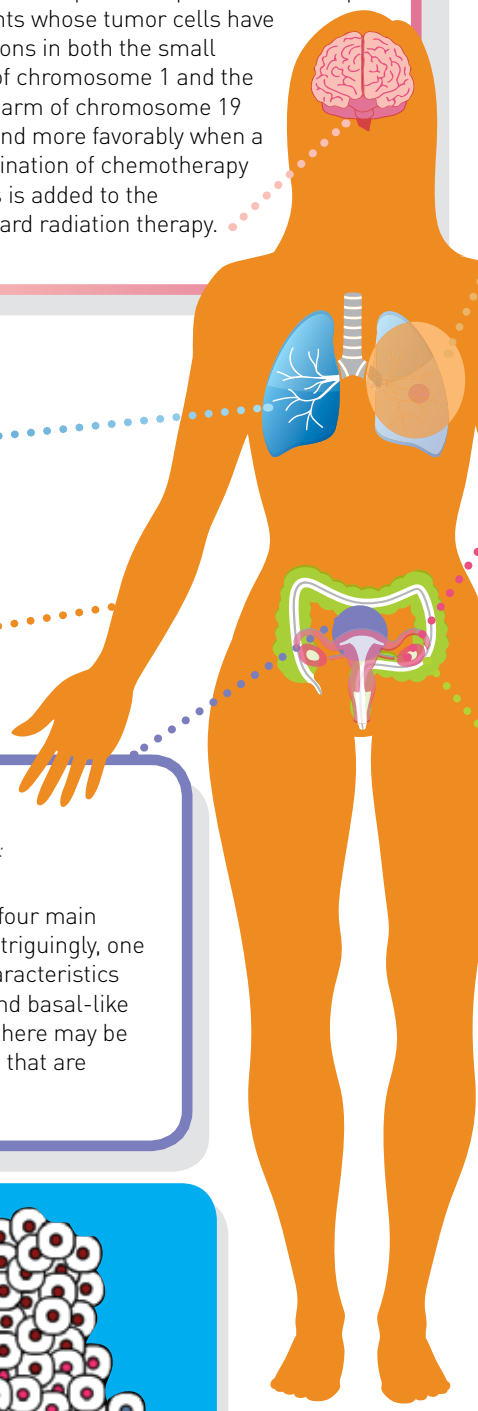
Estimated U.S. annual incidence:
9,500 new cases

A series of genetic changes has been identified that classifies glioblastoma into subtypes. Some subgroups preferentially respond to certain medications, meaning these genetic markers can be used to predict therapeutic response. For example, patients whose tumor cells have deletions in both the small arm of chromosome 1 and the large arm of chromosome 19 respond more favorably when a combination of chemotherapy drugs is added to the standard radiation therapy.

Lung adenocarcinoma

Estimated U.S. annual incidence: **77,585 new cases**

Lung adenocarcinoma is the most common form of lung cancer. At least 60 percent of patients has identifiable genetic mutations that impact the rate of cell division. Approved or experimental anti-cancer drugs target more than half of these mutations. For example, tumors with activating mutations in the EGFR gene can be successfully treated with the drugs getfitinib and erlotinib, which bind to and silence the mutated EGFR protein. However, this therapy is completely ineffective if mutations are also present in a separate gene known as KRAS — a striking example of the complex genetic nature of cancer.



Melanoma

Estimated U.S. annual incidence: **76,690 new cases**

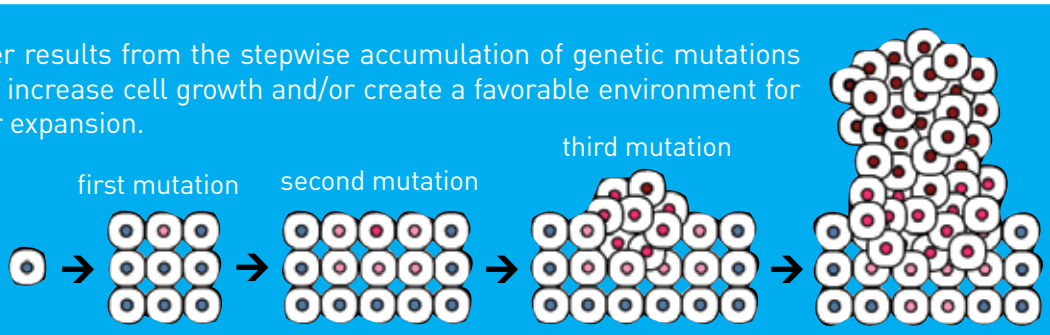
Nearly 50 percent of melanomas have mutations in a gene called BRAF and the U.S. Food and Drug Administration has approved two drugs that target BRAF as part of a treatment plan. Melanoma has also been linked to mutations in the TERT gene, which encodes a component of telomerase. This protein regulates the length of telomeres – those repeating DNA sequences found at the ends of chromosomes. The cancer-associated mutations are believed to increase the level of telomerase, which allows cells to divide for a longer period of time. Found in over 70% of analyzed melanomas, this may be one of the most common drivers of cancer growth.

Uterine

Estimated U.S. annual incidence:
49,560 new cases

Genetic analysis has identified four main subgroups of uterine cancer. Intriguingly, one type shares several genetic characteristics with both high-grade ovarian and basal-like breast cancers. This suggests there may be common drug-based therapies that are effective for all three cancers.

Cancer results from the stepwise accumulation of genetic mutations which increase cell growth and/or create a favorable environment for tumor expansion.



Breast

Estimated U.S. annual incidence:
234,580 new cases

A majority of breast tumors contains mutations in a category of genes that regulate when cells divide. This includes the genes CCND1, ERBB2, FGFR1 and PIK3CA. The mutations often result in proteins that continually signal for cell growth and division. Fortunately, approved drugs now exist that target many of these genetic mutations.

Ovarian

Estimated U.S. annual incidence:
22,240 new cases

A two-tier classification system was recently introduced for ovarian cancer. Low grade tumors are generally slow-growing and have a more favorable outcome. Approximately two-thirds have mutations in the BRAF, ERBB2 or KRAS genes. In contrast, high-grade ovarian cancer develops rapidly and nearly all cases have mutations not only in the TP53 gene, but show gains and losses in large chunks of genetic material throughout the genome.

Colorectal

Estimated U.S. annual incidence:
142,820 new cases

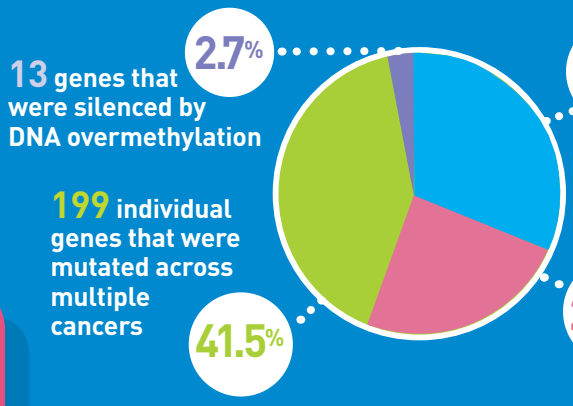
Most colorectal cancers arise through a stepwise accumulation of genetic mutations that occur over the span of many years. Commonly, mutations arise in genes such as AKT1, BRAF, KRAS, PIK3CA, PTEN and SMAD4. Many of these are targets for small molecule drugs. A significant fraction of colorectal cancers have mutations in the system that monitors and repairs DNA damage. Not surprisingly, these cancer cells have an unusually high frequency of mutation across their genome.

References:
Ciriello G, et al. *Nature Genetics* 45:1127-1135 (2013).
Garraway L.A., *Journal of Clinical Oncology* 15:1806-1814 (2013).
Lim, D. and Oliva, E. *Pathology* 45:229-242 (2013).
Nana-Sinkam, S.P. and Powell, C.A. Chest 143(supplement): e30S-e39S (2013).

Annual incidences based on estimates from Cancer: facts and figures, 2013, American Cancer Society, and Ostrom, Q.T. et al. Neuro-Oncology 15 (supplement 2): ii1-ii56 (2013).

Comparing mutation patterns across cancer

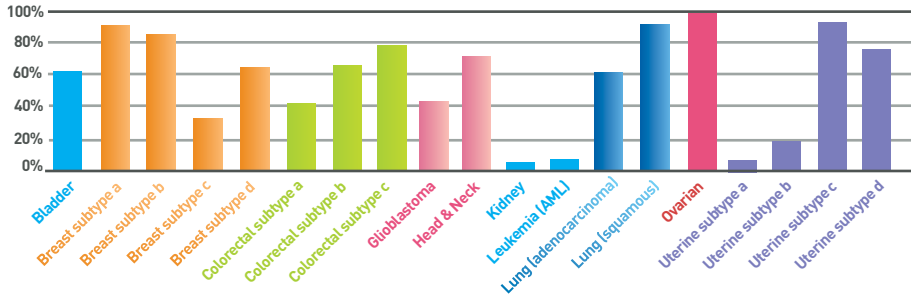
A recent study by The Cancer Genome Atlas analyzed the genetic changes present in over 3,000 tumors from 12 different cancer types. Alterations were consistently identified in over **479 regions** of the genome.



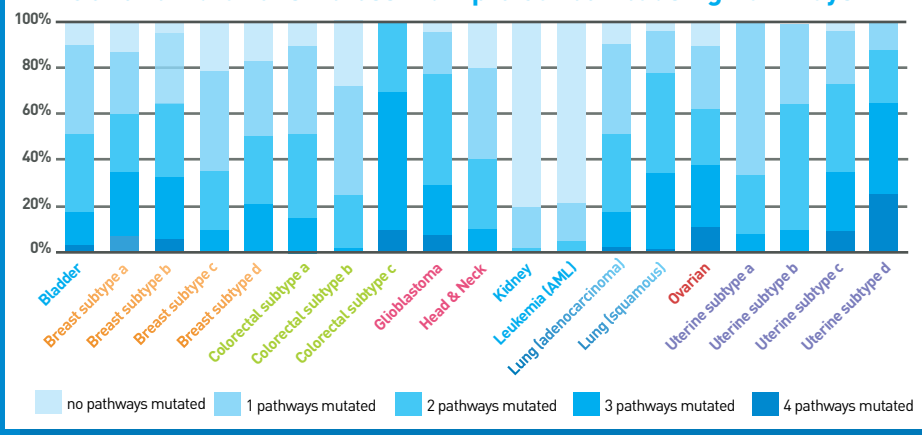
Classification of genetic alterations

The various mutations and alterations can be loosely grouped into one of four major biological pathways: two involved in receiving and transmitting "grow" signals from outside the cell, one that oversees DNA replication and cell division, and one that searches for and repairs DNA damage. Mutations within the same pathway are common to many tumor types. Additionally, most cancers have a combination of mutations that impacted multiple pathways.

Percentage of Tumors with Mutations in the DNA Repair Pathway



Genetic Mutations Across Multiple Cancer-causing Pathways



Tumors with mutations in the four pathways

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	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,
		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	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
		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 APPLICATIONS

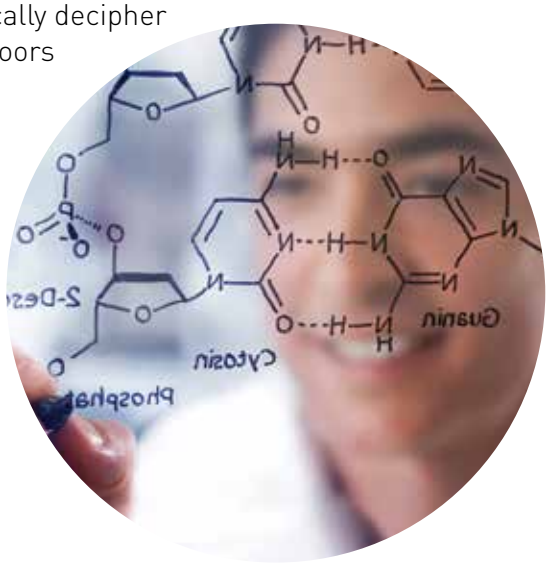
DNA Sequencing

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 something that is approaching $1,000..

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.



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 has been incorporated into the AMSTI high school program across Alabama and is available for purchase to out of state teachers through a partnership with Carolina Biological Supply.

RNA and Protein Analyses

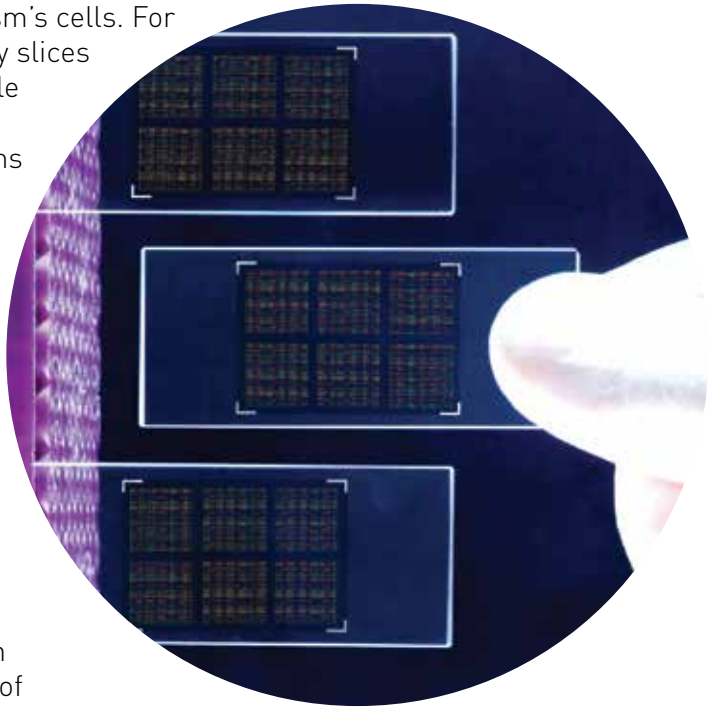
As sequencing techniques identify the genetic recipes of an organism, understanding the function of those genes becomes increasingly important. Many of the fundamental measurements used by molecular biologists seek to determine the presence, absence or relative amounts of RNA produced by a gene. Initially, 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 continuing towards 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

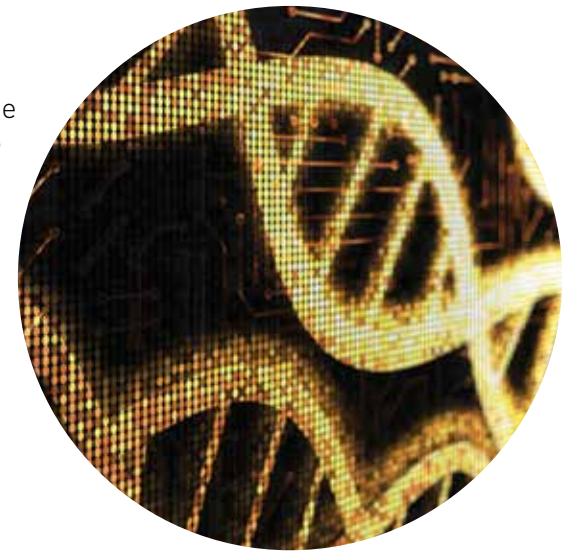
Acquiring DNA sequence has now become routine and new technologies can sequence a bacterial genome in a single day. Similarly, microarray experiments shed light on the RNA levels produced by tens of thousands of genes. Current analysis platforms are capable of generating terabytes of data in a single 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.

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. In Alabama, the lab is incorporated into the AMSTI Science in Motion program. It is available for purchase to out of state teachers through a partnership with Carolina Biological Supply.

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.



APPLICATION

Agriculture

The demand for crop production is rising due to increased human population, greater worldwide meat and dairy consumption, and the expanding role of biofuels. Studies suggest that agricultural production must double between 2005 and 2050 to meet this growing need. Increasing crop yields, rather than clearing additional farmland, is believed to be the more sustainable path. However, crop yields are not increasing fast enough to keep up with projected demands. The additional challenges of drought, temperature change and poor soil quality further strain the productivity of agricultural systems.

Developing new high-yield seeds adapted for our present and future environmental conditions is a cornerstone of increased food production. This begins with the ability to locate and characterize agriculturally important versions of specific genes. These discoveries can then be shared with the farmers and commercial plant breeders who are developing new varieties of crops. Such a collaborative approach blends the emerging field of genomics with the ancient practice of agriculture, increasing yields and ensuring global food security.

Sequencing Plant Genomes for Food and Bioenergy Needs

Over the last decade, genome sequencing projects have been completed for a number of plants, including rice, corn, soybean, canola, and orange. These efforts provide a better understanding of the 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, and switchgrass. For example, soybean not only accounts for 70 percent 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 allows researchers to select for specific type high-yield strain as well as develop soybean plants that are more resistant to drought or disease.

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).

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 are in commercial production and 15 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 crops reduce the need for plowing to control weeds, leading to better conservation of soil and water and 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.

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.

Cancer

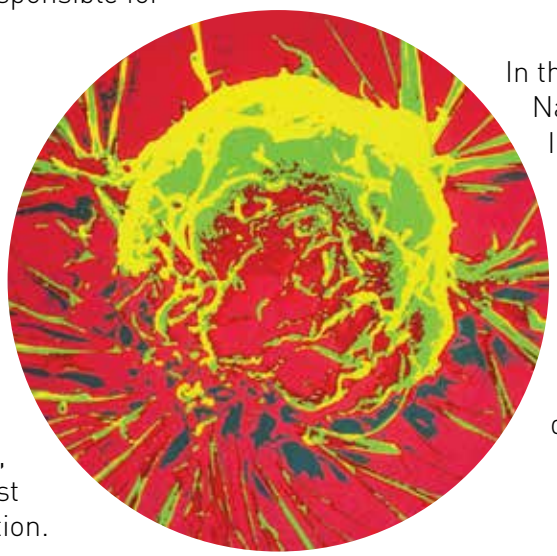
Cancer is a collection of diseases that are characterized by uncontrolled growth of cells and their spread to surrounding tissues. All cancers are genetic diseases, because changes in the genes that control cell growth and division are involved. However, only about 5 percent of cancers are strongly hereditary – 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 for high school life science teachers across Alabama. It is available for purchase to out of state teachers through a partnership with Carolina Biological Supply.

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.

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.



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.

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.

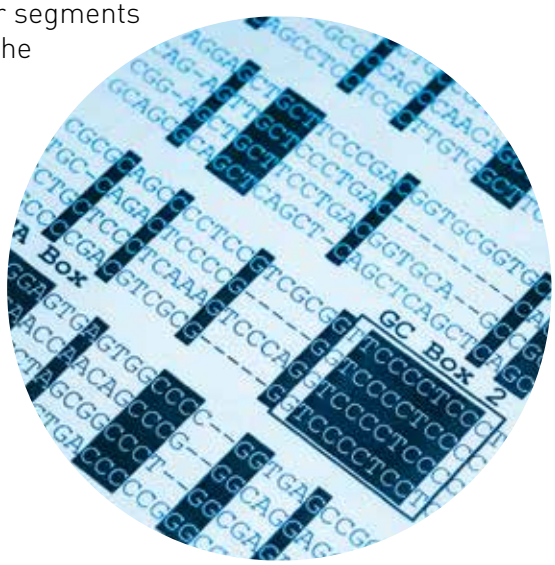
Microarray experiments are currently too cumbersome to perform in a clinic, so it is not likely they will be used routinely to diagnosis patients.

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

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 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).

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 percent 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.

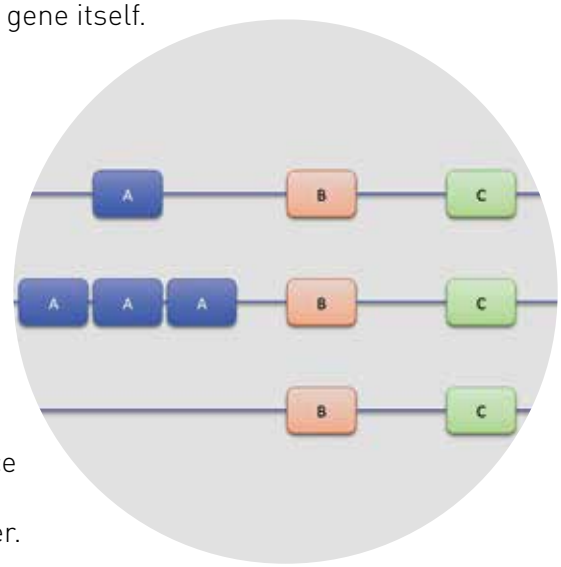
Relating genetic variation to human disease and inheritance is identified in the Biology COS under objective 8 and is described in detail in the Genetics COS objectives 2 and 5, particularly as it connects with genetic patterns of inheritance and multiple alleles. Genetic variation as it relates to human 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 percent. CNVs revise that estimate: the two genomes differ by at least 1.0 percent. 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 percent 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 'Disorder Detectives', where 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 across Alabama. Disorder Detectives is available for purchase to out of state teachers through a partnership with Carolina Biological Supply.

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 49).

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.

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".

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.

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 three-fourths 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 percent of individuals with two copies of the blue eyed OCA2 allele actually had blue eyes. Blue eye color was also found among 7.5 percent 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.

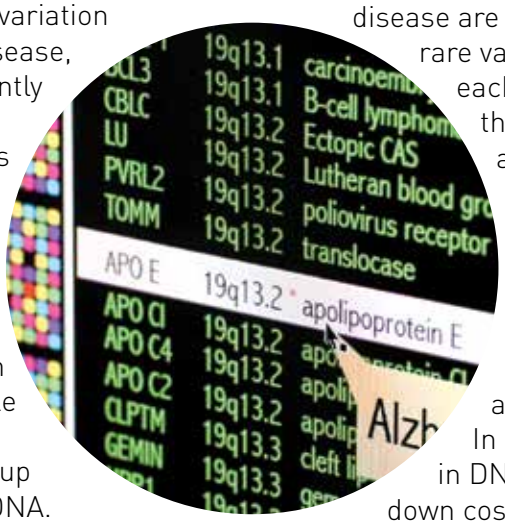
Identifying Genetic Influence on Disease

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 percent 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 percent. 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. Basically, 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 percent 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 36) - 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 are being 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. Several companies offer an analysis 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. In 2013, the FDA ordered the health-related versions of these tests to be halted, restricting companies to offering genetic information about ancestry alone.

Two additional companies (Knome and Illumina) offer to sequence the entire 3 billion base pairs of an individual's genome.

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

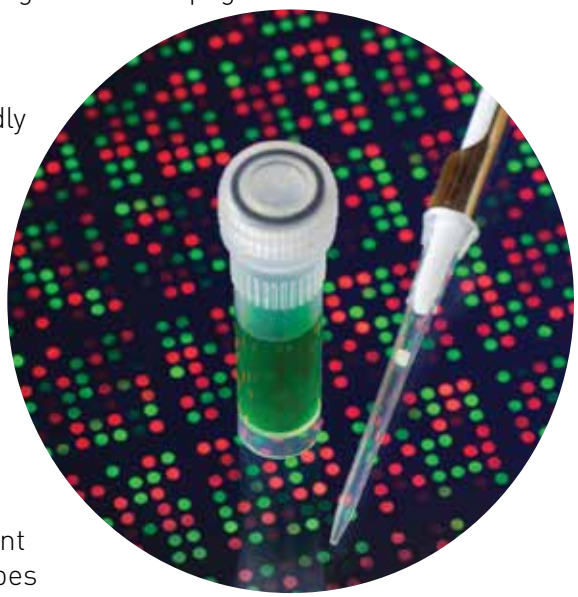
At its core, personalized medicine uses information about a person's genetic background to tailor strategies for the detection, treatment or prevention of disease. This may include genetic screening tests to identify susceptibility to disease or more precisely pinpoint existing conditions. It may also 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).

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 have moved 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?



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. 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 genetic profile.

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. 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 percent 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.

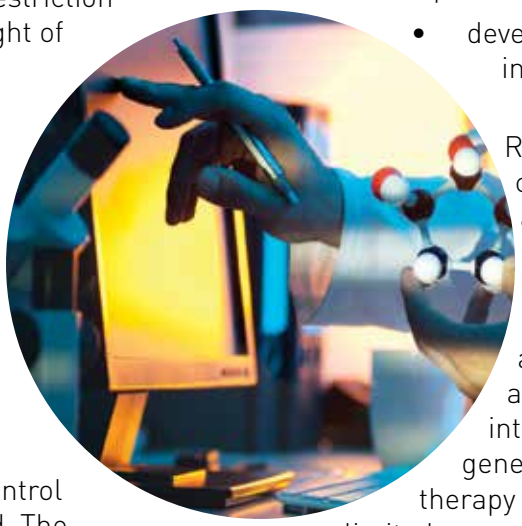
There are over 200 pharmaceutical products that either recommend genetic testing or point to the influence of genetic variability on the drug's response.

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 100 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 or soybean, adding pest or herbicide resistance, or increasing nutrient content (see Agricultural Applications, page 31)
- 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.

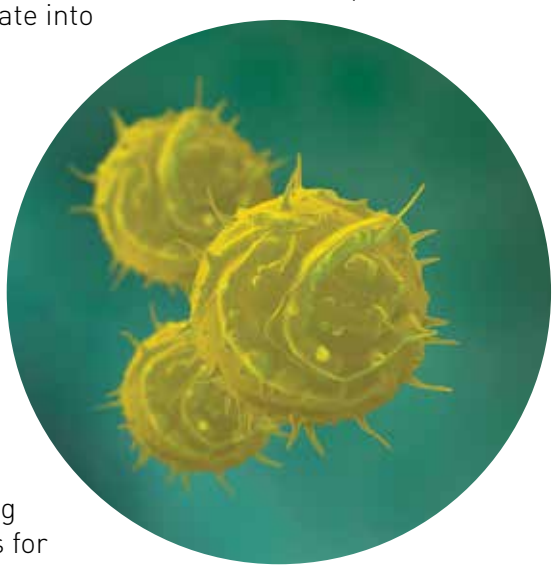
Stem cells

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.



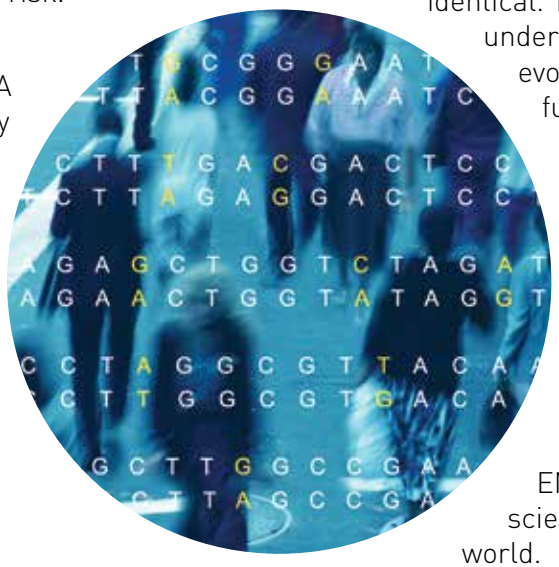
The concept of stem cells connects to several components of the standard Biology Course. It can be highlighted during explanation of the cell cycle (COS objective 6), although some biology curriculum models include discussions of stem cells during instruction on the Cell Theory instead (COS objective 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. Data released in 2012, suggest as much as 80% of the genome is involved in some sort of "biochemical function". This includes sequences for noncoding RNA (which is transcribed, but not translated), as well as DNA regions bound by proteins to regulate processes of transcription or DNA folding. Many of these sequences likely include evolutionarily ancient mechanisms and pathways not used by human cells, over-estimating the true functionality of the human genome.

A recent study adds to the conversation by compared the genomes of twelve mammals – representing approximately 100 million years of evolution - to identify regions of the DNA that have remained nearly identical. Those sequences that have undergone very little change throughout evolution suggest that DNA has some functional purpose that requires its retention. Researchers estimate that just over 8% of the human genome, approximately 253 million bases of sequence, meets this definition for function.

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

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.

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.

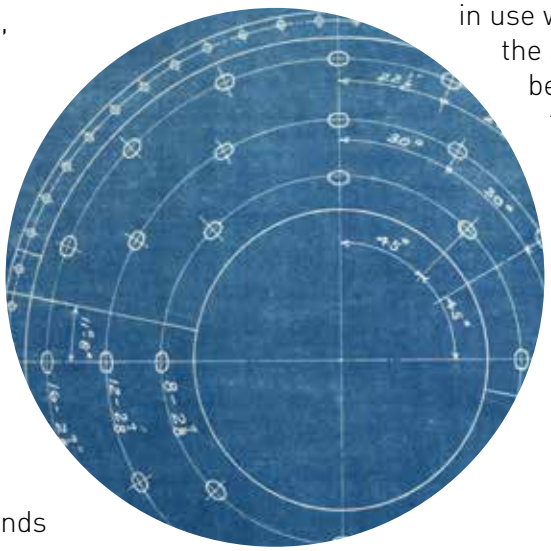
Powerful computers offer in-depth modeling and simulation to predict the behavior of the part, device or system before it is assembled. The relevant DNA instructions are then artificially synthesized and inserted into a biological cell, such as bacteria. The bacterial cell is the "chassis" or vehicle 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 link 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.



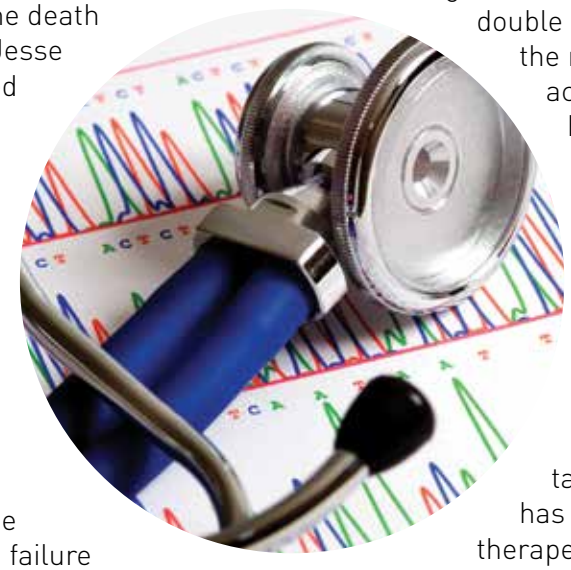
THERAPEUTIC APPROACHES

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.

Image Credits:

Page 8
DNA helix - NHGRI Digital Media Database, Jonathan Bailey http://www.genome.gov

Page 9
Oxytricha trifallax - by Unknown - Licensed under Public domain via Wikimedia Commons -http://commons.wikimedia.org/wiki/File:Oxytricha_trifallax.jpg#mediaviewer/File:Oxytricha_trifallax.jpg

Page 10
Functional percentages of the human genome - Kellis et al., "Defining functional DNA elements in the human genome," *Proceedings of the National Academy of Sciences* (2014) 111:6131-6138 doi: 10.1073/pnas.1318948111.

Human Chromosomes – NHGRI Digital Media Database, Jane Ades http://www.genome.gov

Page 11
Model of Chromosome Folding - From Naumova et al., "Organization of the mitotic chromosome," *Science* (2013) 342:948-53. Reprinted with permission from AAAS.

Tannosome formation - Brillouet J.M. et al., "The tannosome is an organelle forming condensed tannins in the chlorophyllous organs of Tracheophyta,"*Annals of Botany* (2014) 112:1003-1014 doi: 10.1093/aob/mct168, by permission of Oxford University Press.

Page 12
Trisomy 21 karyotype - NHGRI Digital Media Database, Darryl Leja http://www.genome.gov

Page 17
Chromophobe renal cell carcinoma - NHGRI Digital Media Database, Sang Eun Lee, Samsung Medical Center http://www.genome.gov

Pages 18-19
Three Dimensional facial recognition - Claes et al., "Modeling 3D Facial Shape from DNA," PLoS Genetics (2014) 10:e1004224 doi:10.1371/ journal.pgen.1004224.

Contact

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