Biotechnology Discoveries and Applications

Extensions to high school science curriculum

The 2015-2016 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.

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 700,000 individuals were impacted through HudsonAlpha education outreach during the 2014-15 academic year.

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.

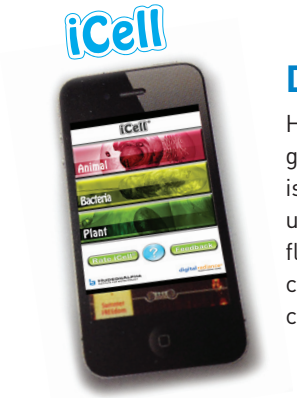
Student Experiences

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 triton.hudsonalpha.org

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 1 million 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 Standards

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 third round of GREAT workshops will dig into **the newly adopted Alabama Science Course of Study** by working through the following:

- Using models to help students predict cell membrane behavior
- Using genetic variant analysis to argue from evidence
- Addressing large scale genetics projects such as ENCODE and the 1000 Genomes Project
- Analyzing data from environmental and genetic factors to inform lifetime risk for complex disease
- Evaluating and modifying classroom resources in light of the new course of study

In two full days of small group concurrent sessions and talks by dynamic speakers, teachers will learn about recent findings in genetics and genomics and methodologies to address this content with their students.

Teachers who complete both days of the workshop will return to the classroom with lesson plans and hands-on materials that are student-tested, informative and that link to state course of study objectives.

For more information, visit **www.hudsonalpha.org/GREAT**

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.

HUDSONALPHA
INSTITUTE FOR BIOTECHNOLOGY

601 Genome Way Huntsville, AL 35806 | 256.327.0400 | www.hudsonalpha.org

Biotechnology Discoveries and Applications 2015-2016

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

The past several years have witnessed an unprecedented growth in genetics, genomics and biotechnology research. Deciphering the instructions encoded in an organism's DNA has rewritten our understanding of biology. The applications of this research shape our society on a global scale. It is an exciting time to study genomics and the number of fields that intersect with our work continues to expand.

This edition of the annual *Biotechnology Discoveries and Applications* guidebook shares the enthusiasm and challenges of the research and commercial applications associated with these fields. **The guidebook provides educators information about recent advances in genetics and biotechnology, allowing them to share those findings with their students.** As in years past, it is divided into two sections: research highlights and foundational concepts. More than 40 new discoveries are highlighted, including articles about:

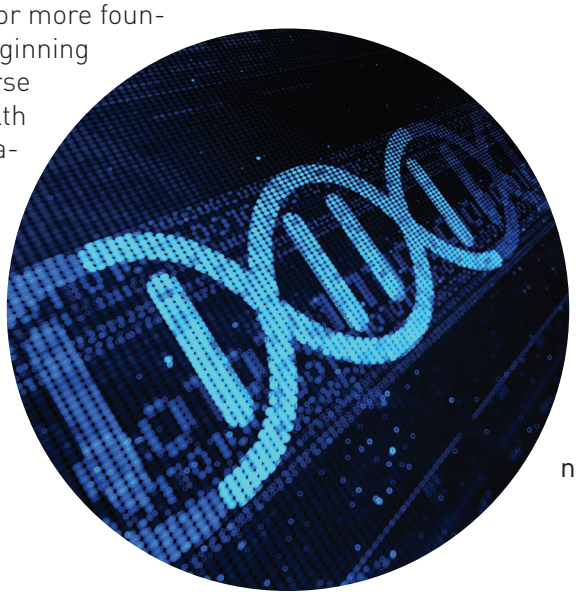
- an endangered shark-like fish that appears to be temporarily boosting its declining numbers through asexual reproduction (p.9)
- the first city-scale map of microbial populations, collected from the subway stations of New York City (p.10)
- an attempt to calculate the total amount of DNA present on earth (p.12)
- the ethical challenges associated with "3 parent babies" and editing the genome of embryos (both featured on p.14)
- a sampling of results from gene therapy trials and follow-up studies (p.17)
- the recent approval of genetically engineered apples and potatoes, designed to resist browning and bruising (p.18)

The new findings are linked to one or more foundational topics, covered in detail beginning on page 28. 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 align to their own state objectives fairly easily.

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

It's said that science builds upon itself - yesterday's research findings often serve as the foundation for new experiments and subsequent discoveries. Accordingly, many of the stories in this year's guidebook grew from research initially presented in earlier editions. PDF versions of all prior guidebooks can be freely explored on the HudsonAlpha website at http://hudsonalpha.org/education/teachers/classroom-resources.

Today's students step into a world increasingly shaped by genomic technologies associated with food production, energy generation and the delivery of healthcare. To make informed decisions, students will need to possess at least a rudimentary understanding of the underlying scientific and engineering concepts. The guidbook offers educators a resource to talk with their students about the current research in genetics, genomics and biotechnology, as well as the applications of those discoveries.



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. More than 100 genetic regions have been associated with susceptibility for autism spectrum disorders (ASD). A recent study sequenced the genomes of 85 families, each with two children affected by autism. Only 42% of the families carried known ASD-associated mutations. In addition, the study challenged the assumption that siblings with the disorder were inheriting the same autism-predisposing genes. Only 31% of siblings shared the same ASD-relevant allele, suggesting that many of the genetic influences on autism remain undiscovered.

2. In an early win for exome sequencing to identify complex disease genes, researchers identified a new gene associated with amyotrophic lateral sclerosis (ALS). A comparison of rare DNA variants between more than 2800 ALS cases and 6400 controls identified TBK1 as an ALS associated gene. The protein produced by TBK1 binds to and phosphorylates several proteins previously linked to ALS and plays an important role in cellular destruction.

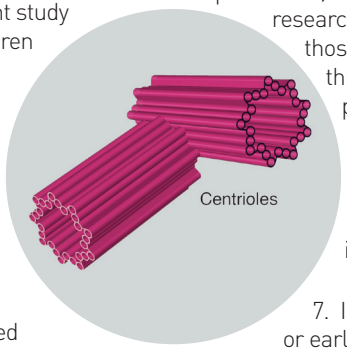
HudsonAlpha reserchers Shawn Levy, Ph.D., Greg Cooper, Ph.D., and Rick Myers, Ph.D. contributed to this work

3. The.tomato originated in the South American Andes mountains as a pea-sized fruit one-hundredth the size of the juicy, red, varieties we know today. To discover how manmade selection shaped the genetics of the modern tomato, a team of researchers sequenced the genomes of 333 different red tomatoes, 10 different wild tomatoes, and 17 different hybrid tomatoes created for commercial use. The resulting "variation map" will help breeders identify genetic changes associated with disease resistance, as well as fruit size and taste.

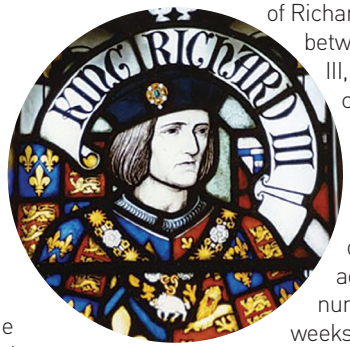
4. Over the last 10 years, The Cancer Genome Atlas (TCGA), has created the largest tumor collection ever analyzed from a molecular perspective - 11,000 patients from 33 tumor types. TCGA results have been published in over 2,700 research articles and this past year's findings highlighted papillary thyroid, head and neck tumors, lower-grade gliomas and cutaneous melanoma. For each case, researchers identified key driver genes that, when mutated, influence cancer development. This knowledge opens the door for developing treatment approaches that target the action of these driver genes.

5. Family health history (FHx) is a written record of the diseases and health conditions present in a family. FHx captures shared genetic, behavioral and environmental risk factors and can be used to identify individuals at increased disease risk who may benefit from preventive interventions. Not everyone has access to personal FHx – for example, both adoptees and orphans often lack such critical information. Given the advances in technology, groups have begun to ask whether genome-wide testing may provide useful information where FHx is missing. Testing could begin with a targeted set of well-known genes where mutations confer relatively high levels of disease risk and advanced knowledge allows for surveillance or preventative action.

This research was performed by HudsonAlpha researcher Kim Strong, Ph.D.

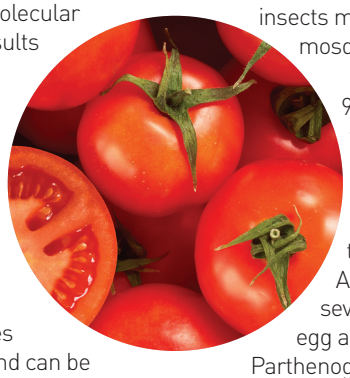


6. Centrioles are barrel-shaped cellular organelles that ensure replicated chromosomes are appropriately distributed during mitosis. Mutations in the genes that encode centriolar components have been implicated in respiratory conditions, male sterility and cancer. During reproduction, only the sperm contributes centrioles. A Swiss team of researchers has discovered that in the roundworm *C. elegans,* those original centrioles are remarkably stable, persisting through as many as ten cell divisions. This raises the possibility that centrioles may carry information across multiple cell cycles. Damaged paternal centrioles could impair mitosis and contribute to developmental anomalies. It remains to be seen if centriole persistence is found in other organisms, including humans.

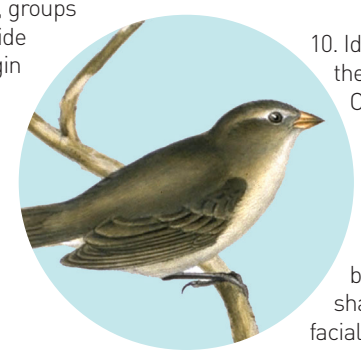


7. In 2012, the skeletal remains of a male in his late 20s or early 30s were discovered beneath a parking lot in Leicester, U.K. The grave was on the former site of the Greyfriars Abbey, where King Richard III was buried after dying in battle in 1485. The exact site of the burial had been lost, but the archeological, skeletal and radiocarbon dating evidence were all consistent with the remains being those of Richard III. Mitochondrial DNA analysis was compared between the remains and two living relatives of Richard III, identified through a female-only line. The results confirmed the skeleton is Richard III, portrayed by Shakespeare as one of history's most famous villains.

8. A targeted release of genetically modified mosquitoes designed to produce offspring that die before adulthood has dramatically reduced overall mosquito numbers in a Brazilian field trial. Over the course of six weeks, more than 185,000 modified male mosquitoes were released across a suburb in Itaberaba, Brazil. The mosquito population dropped by 80-95%, depending on measuring techniques, consistent with previous releases in the Cayman Islands (profiled in the 2011 Guidebook.) Developed by the British company Oxitec, the modified insects may become a useful weapon in the battle to control mosquito-borne diseases like dengue.



9. A recent study of the endangered smalltooth sawfish suggests some of the females are reproducing without a mate – a process called facultative parthenogenesis. Previously observed in birds, sharks and reptiles kept in captivity, this marks the first time the process has been found in the wild. A catch and release study of 190 sawtooths identified seven that were likely created by a fusion between the egg and one of the polar bodies produced during meiosis. Parthenogenesis may occur in populations where females have trouble finding mates. It may be a short term method to stave off extinction, but since only females are produced, it is not a long-term solution.

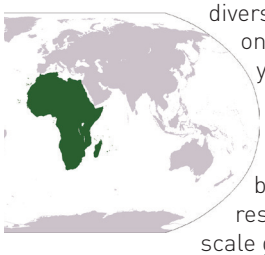


10. Identifying variation in beak shape and size among the species of Galapagos finches helped inspire Charles Darwin's theory of evolution. Recently, the genomes of 120 finches were sequenced to identify the genes behind this variation. Fifteen genomic regions were linked to blunt or pointed beak shape. The strongest finding was with the ALX1 gene, which has an ancestral version associated with pointed beaks and a more recent version correlated with blunt shape. This gene plays an important role in skull and facial development.

NEW FINDINGS

Africa's variants
tracking genetic diversity

Around the world, human populations exhibit genetic diversity as a consequence of geographical separation, selection for evolutionarily useful traits and genetic drift.



Africa is the most genetically diverse continent on the planet, yet Sub-Saharan African populations have generally been under represented in large-scale genomics projects. Consequently, many databases lack ancestrally and medically relevant DNA variations primarily found in individuals of African heritage.

In the first large-scale effort to address this, the African Genome Variation Project (AGVP) has cataloged genetic variation from whole genome and genotyping data of 1,481 individuals across sub-Saharan African. More than 30 million variants were identified, with nearly one-fifth previously uncataloged.

Several variants exhibit selective pressure in association with risk for infectious agents. In addition to the relationship between sickle cell disease and malaria, gene variants also influence the risk of contracting Lassa fever, trypanosomiasis and trachoma - parasitic infections endemic in certain regions of Africa.

The AGVP also uncovered clues into historical patterns of human migration. Many Africans have some Eurasian genetic markers within their genomes. This suggests that Eurasian populations migrated back into Africa thousands of years after first leaving the continent.

REFERENCE: Gurdasani, D. et al. The African Genome Variation Project shapes medical genetics in Africa. *Nature* (2014) 517:327-32 doi:10.1038/nature13997.

Evolve or die
strong selective pressures repurpose gene function

Through a stroke of luck, researchers identified how evolutionary pressures can repurpose one gene to replace and carry out the function of a different gene.

While studying how the soil bacteria *Pseudomonas fluorescens* colonizes plant roots, scientists intentionally deleted the gene that encodes the FleQ transcription factor. This protein binds to and activates the transcription of genes required for flagellum production. Lacking a flagellum, these bacteria are unable to move.

A graduate student inadvertently incubated the mutant bacteria much longer than intended. Expecting to find the bacteria had consumed the nutrients imediately around them and then died, he was shocked to find a handful of colonies had reacquired the ability to form flagella, allowing them to find additional nutrients. The experiment was repeated - within four days bacteria had regained flagella.

Comparing the bacterial genomes, each revertant had undergone the same two-step process to

repurpose a nitrogen regulatory protein called NtrC that shares ~30% amino acid similarity with FleQ. The "re-evolution" process began with mutations in genes that overexpressed NtrC production, which led to a minor activation of flagella synthesis genes. The bacteria then acquired mutations in the *NtrC* gene itself, allowing the encoded protein to more effectively bind and transcribe the genes normally regulated by FleQ.

Under strong selective pressure (evolve or die), the bacteria rapidly and consistently repurposed the *NtrC* gene, putting its product to a new use and allowing the bacteria to survive.

REFERNCE: Taylor T. et al. Evolutionary resurrection of flagellar motility via rewiring of the nitrogen regulation system. *Proceedings of the National Academy of Sciences* 111:6131-6138 (2014).

Mapping New York's microscopic residents
the first large-scale portrait of microbial diversity

PathoMap is a census of the microscopic organisms living throughout the New York City subway system. The study began in 2012, when researchers swabbed over 500 locations in the New York City subway and parks systems. Over 1,400 samples were obtained from kiosks, benches, turnstiles, railings,

seats, poles, and doors. DNA isolated from each sample was sequenced and compared to databases of reference microbes.

This massive dataset is the first city-scale molecular portrait of microbial diversity. Nearly half of the genetic sequences did not match to known organisms, confirming there are many bacteria, viruses and fungi that still await our discovery. Over 1,600 organisms were identified, not surprising given that 5.5 million people ride the subway each weekday. Bacteria accounted for half of the species that were identifiable.

At Penn station, samples taken every hour across a single workday demonstrated that while some bacterial populations remained relatively constant, others fluctuated markedly with time.

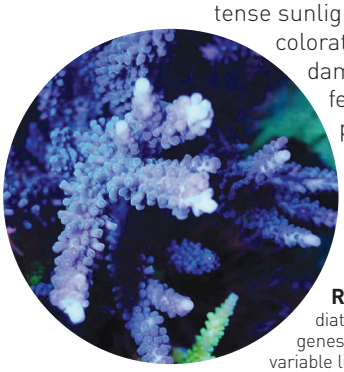
Samples were taken from a different subway station that had been permanently closed after being flooded in 2012 by Hurricane Sandy. The microbial community existing at that station was different from the rest of the subway system - researchers identified DNA fragments on the floors and walls associated with cold-water marine-related bacteria.

Across the city, the vast majority of identified bacteria

Increasing gene expression
copy number variation drives coral coloration

Scientists have long wondered how members of the same species of corals could exist side by side, yet display different colors. By studying the staghorn coral, researchers at the University of Southampton discovered that color intensity is determined by copy number variation within the coral's pigment-encoding genes.

Multiple identical copies of each gene encode fluorescent pigments that act as a form of sunscreen for the algae residing inside the coral. In a symbiotic relationship, corals offer algae protection and nourishment in return for sugars the algae create during photosynthesis. High levels of sunlight passing through the waters can kill the algae, ultimately killing the corals as well. Consequently, corals adjust to differing light conditions by controlling the number of pigment genes being actively transcribed. Those in areas of intense sunlight generate more pigment, have darker coloration and protect their algae from sun damage. Corals in shadier areas activate fewer copies of pigment genes and remain paler. Because pigment production is an energy-intensive process, corals must balance how much of their resources are allocated to algae protection versus growth and reproduction.



REFERENCE: Gittins J.R. Fluorescent protein-mediated colour polymorphisms in reef corals: multicopy genes extend the adaptation/acclimatization potential to variable light environments. *Molecular Ecology* (2015) 23:453 doi:10.1111/mec.13041

were innocuous, although some DNA sequences did match organisms responsible for gastrointestinal distress, and other infectious diseases. In addition, many bacteria contained genes that conferred antibiotic resistance.

It is important to point out that these large-scale metagenomics surveys, their computational analyses and the reference databases they rely on are still in their infancy. Using different methods, the same dataset can yield dramatically different answers. Additionally, even if DNA sequences from potentially harmful bacteria were present, they were at extremely low levels. None of these findings should be interpreted as a public health concern.

Ultimately, once the databases and computational methods reach maturity, these types of results will establish a reference picture of a "typical" microbial ecosystem. In turn, that reference can be used to detect, track and prevent outbreaks of disease.

REFERENCES: Afshinnekoo E. et al. Geospatial resoution of human and bacterial diversity with city-scale metagenomics. *Cell Systems* (2015) 1:1-15 doi:10.1016/j.cels.2015.01.001

Pathomap website: http://www.pathomap.org

HudsoAlpha researcher Shawn Levy, Ph.D. was involved in this work.

In brief

Understanding microbial communities

The impact of kissing on the oral microbiome

Although not common in other animal or primate groups, intimate kissing (involving tongue contact) occurs in over 90% of human cultures. Previous research has indicated that people living together have much more similar microbiomes than non-related people. Scientists sought to characterize the impact of kissing on the oral microbial community of couples. DNA sequencing was used to identify the bacterial strains present in the mouths of couples before and after kissing. Not surprisingly, couples had a more similar oral microbiotia than non-couples and those partners who reported frequent kissing had a higher similarity index than infrequent kissers.

To measure the actual amount of bacteria exchanged in a kiss, researchers took swabs from the mouths of both members of a couple before a kiss. One partner was then given a probiotic yogurt drink (to introduce novel bacteria). Both partners were resampled after a 10 second kiss. An average of 80 million bacteria were swapped from the yogurt-drinking partner to the nondrinker. While many of these kiss-introduced bacteria were transient, some took up permanent residence in the recipient partner's mouth.

REFERENCE: Kort, R. et al. Shaping the oral microbiota through intimate kissing. *Microbiome* (2014) 2:41 doi:10.1186/2049-2618-2-31.

Curating the microscopic biodiversity of the oceans

Although it represents the largest ecosystem on earth, scientists understand very little about the various organisms that call the ocean their home. New findings shed light on the tiniest members of those saltwater ecosystems – oceanic plankton.

Between 2009-2013, more than 250 biologists, oceanographers, sailors, journalists, writers and artists took turns on board the 110-foot schooner Tara, sampling 200+ sites across all major oceanic regions. Most sampling was from the sunlit-upper region of the ocean (the epipelagic layer), although the deeper twilight mesopelagic zone was also analyzed. DNA and RNA sequencing generated a census of plankton diversity incorporating viruses, prokaryotes, phytoplankton and zooplankton.

The Tara Ocean project also utilized a number of techniques to study the interactions between the plankton and the influence of surrounding environmental conditions on their place within the larger ecosystem. Peter Bork, one of the scientists associated with the project noted that the results increased our previous knowledge of the ocean's microbial diversity "one thousand-fold ... and yet, this is still the tip of the iceberg."

REFERENCES: Bork P. et al. Tara Oceans studies plankton at a planetary scale. *Science* (2015) 348:873 doi:10.1126/science.aac5605.

The Tara Ocean project: http://oceans.taraexpeditions.org/en/.

NEW FINDINGS

In brief

Identifying genetic influence on disease

A link between genetics, obesity and year of birth

Obesity is a complex trait, resulting from the interaction of multiple genetic and environmental factors. Some findings connect the dramatic rise in obesity rate to altered environmental risks – more people consuming a high calorie diet but getting less overall exercise. At the same time, a significant link has been found between body mass index (BMI) and a specific variant in the fat mass and obesity associated *(FTO)* gene.

Looking for connections between genetic and environmental risks, researchers examined changes in BMI over time using data from over 5,000 individuals whose health has been tracked regularly from 1971. All subjects were categorized by birth year and the number of copies of the *FTO* "risk" variant.

Surprisingly, the *FTO* variant was not associated with higher BMI in individuals born before 1942. In contrast, for those born after 1942, mean BMI was significantly influenced by the presence of the *FTO* genetic risk. The results suggest that as the environment changes, it may increase the penetrance of genetic variants, resulting in higher rates of obesity

REFERENCE: Rosenquist H.N. et al. Cohort of birth modifies the association between FTO genotype and BMI. *Proceedings of the National Academy of Sciences* (2015) 112:354-9 doi:10.1073/pnas.1311893111.

Mitochondrial mutations and aging

Because of their role in generating ATP during cellular respiration, mitochondria are often described as the "cell's powerhouse." They are thought to have originated from an ancient symbiotic relationship in which a nucleated cell engulfed a smaller prokaryote. Mitochondria have retained a small subset of genes encoded on a separate non-nuclear genome present in multiple copies inside the mitochondria. Mitochondria are inherited only through the mother and mitochondrial DNA (mtDNA) has a 10 fold higher mutation rate than nuclear DNA.

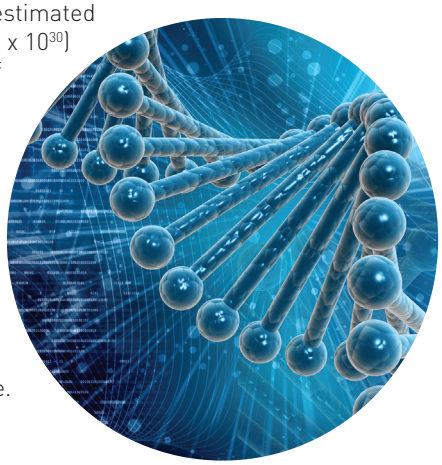
Damaged mtDNA has been implicated in the aging process and recent work in mice suggests that inheriting even low levels of mtDNA damage may have lifelong consequences. The health and lifespan were compared between two sets of mice with genetically identical nuclear DNA. One set of mice were born to mothers whose eggs contained mtDNA mutations; the other set were born to typical mothers with no mtDNA mutations. Mice with higher levels of mtDNA mutations experienced premature aging (hair loss, reduced body size and lower weight) and on average lived 30% shorter lifespans than their less-damaged counterparts (100.2 versus 141.1 weeks.)

REFERENCE: Ross J.M. at al. Maternally transmitted mitochondrial DNA mutations can reduce lifespan. *Scientific Reports* (2014) 4:6569 doi:10.1038. srep06569.

How much DNA?

estimating the total number of nucleotides on earth

Researchers recently attempted to tally the total amount of DNA present in Earth's biosphere, quantified across the five major subgroups of life (prokaryotes, protists, fungi, plants and animals) plus viruses. Different methods were used to determine the total number of cells for each subgroup, including the estimated number of individuals, relative densities, and total biomass weight. For example, the average prokaryotic genome size (3.2 million bases) was multiplied by the estimated number of prokaryotic cells (5×10^{30}) to yield a total DNA amount of 1.6×10^{31} million bases.



Life on earth is estimated to contain at least 5.3×10^{31} million bases of DNA (see table at right), which corresponds to ~ 5×10^{10} tons of nucleic acid. Hypothetically, the earth's DNA would require one billion standard shipping containers of storage.

Astronomical? Try "genomical"

Will genomics become the king of big data?

"Big Data" broadly describes data sets so large that traditional methods for capturing, searching, sharing and storing are inadequate. Examples of big data include astronomy, particle physics, YouTube and Twitter.

Genomics is increasingly counted as a big data producer, with the total amount of DNA sequence data doubling approximately every seven months.

Scientist recently estimated the current data usage and future growth for genomics, suggesting it may surpass other big data players in terms of storage needs.

To date, it's estimated that over 250,000 human genomes have been sequenced, most as part of disease-related research projects. The 20 largest genomics institutions are currently using more than 100 petabytes of data to store these genomes (see the box for a crash course in big data). The authors estimate within the next 10 years, scientists will sequence the genomes of nearly all of the 1.2 million currently known plants and animals, along with several million microbes. They also predict 100 million to two billion human genomes will

The authors note that this estimate is based on a number of uncertainties. There is a lack of confidence regarding the actual number of organisms present in various environmental biomes. The figures don't include DNA from chloroplasts, mitochondria, plasmids, fossils and leaf litter. Aside from animals, most organisms were counted as having only a single copy of the genome, which is often not the case. Accordingly, the final calculation should be treated as an underestimate of the total number of nucleotides.

DNA content in the biosphere

DNA amount (Million bases)

Prokaryotes	1.6×10^{31}
Protists	1.3×10^{29}
Fungi	1.7×10^{27}
Animals	4.2×10^{29}
Plants	3.6×10^{31}
Viruses	4.0×10^{29}

REFERENCE:
Landenmark H.K.E., Forgan D.H., and Cockell C.S. An estimate of the total DNA in the biosphere. *PLoS Biology* (2015) 13:e1002168 doi: 10.1371/journal.pbio.1002168.

be sequenced within this same timeframe. Storing these results would require 2-40 exabytes of space, potentially surpassing the data requirements of all other big data generators.

REFERENCE: Stephens Z.D. at al Big Data: Astronomical or Genomical. *PLoS Biology* (2015) 13:e1002195 doi:10.1371/journal.pbio.1002195

Digital information is measured as bytes. One thousand bytes is approximately the amount of space required to store this paragraph of text. An average MP3 song is about 3.5 million bytes (megabyte) of digital data, while streaming a movie utilizes 1 billion bytes (gigabyte) per hour. The Microsoft Xbox One game console ships with 1 trillion bytes (1 terabyte) of storage, which could hold 1,000 copies of the Encyclopedia Britannica. A quadrillion (10^{15}) bytes of digital information (1 petabyte) would fill approximately 20 million 4-drawer filing cabinets with text. An exabyte is approximately 1,000 petabytes - it's been suggested that 5 exabytes would equal all the words ever spoken by mankind.

In brief

Identifying genetic influence on disease

A copy number variation map of the human genome

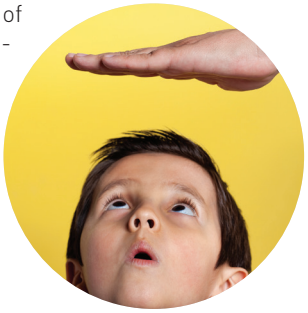
Copy number variation (CNV) has been recognized as a major driver of genetic diversity in humans. CNVs are regions of the genome 50 nucleotides or larger that vary in the number of copies present - they increase or decrease the content of DNA at a specific site. Two groups of researchers recently assessed the frequency and type of CNVs in humans. Zarrei and colleagues assembled a CNV map with over 24,000 variants. These CNVs were unevenly distributed across the genome and DNA deletions outnumbered duplications nearly 8:1. Intriguingly, they identified over 100 genes where both copies can be deleted without any observable consequences. These seemingly non-essential genes may have nearly-identical counterparts elsewhere in the genome or may contribute to unidentified late-onset disorders.

In a separate study, Sudmant et al. analyzed 236 human genomes from 125 diverse global populations. Showing the power of working with historically underrepresented world populations, they identified over 15,000 previously unidentified CNVs. These data suggest human genomes are much more likely to differ due to CNVs than changes in a single base: 7% of the human genome is variable due to CNVs vs. 1% from single nucleotide variants.

REFERENCES: Zarrei M. et al. A copy number variation map of the human genome. *Nature Reviews Genetics* (2015). 16:171-83 doi:10.1038/nrg3871.

Sudmant P.H.et al. Global diversity, population stratification, and selection of human copy number variation. *Science* (2015). epublication ahead of print aab3761 doi:10.1126/science.aab3761.

Genetics of human height



Human height is a textbook example of a multigenic trait, shaped by a combination of genes that each contribute to the overall rate of growth. A recent meta-analysis combined data from 79 prior genome-wide studies into one massive study with over a quarter of a million participants. The larger sample size increases the ability to detect relevant DNA variants. Nearly 700 DNA changes were identified, clustered in 423 regions of the genome. Not surprisingly, genes in these clustered regions are highly expressed in cartilage, joint, spine and other musculoskeletal tissues. These variants explain 20% of total variation in height that is caused by genetics. The results suggest several thousand genes likely contribute to human height. The impact of changes in the sequence of these genes range from fractions of an inch (most genes) to a foot or more (in disorders like achondroplasia).

HudsonAlpha researcher Devin Absher, Ph.D. participated in this research.

REFERENCE: Wood, A.R. et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* (2014) 46:1173086 doi:10.1038/ng.3097.

NEW FINDINGS

In brief

Precision medicine

Identifying exceptional responders to therapy
The National Cancer Institute has established the Exceptional Responders Initiative to collect and sequence both normal and tumor cells from up to 100 exceptional responders - rare patients who respond to treatments that are not effective for most other people. These unusual cases will be studied in an effort to identify patterns linking genetic mutation and drug response. By learning why some individuals respond so favorably, scientists hope to identify others who might respond equally well.

For example, a 70 year old male had metastatic bladder cancer that in spite of standard treatments, continued to grow. After being given a combination of two chemotherapy drugs, his tumors disappeared and he was tumor free for 14 months – an exceptional response. Sequencing the tumor identified a specific combination of genetic mutations that, in essence, was the Achilles' heel of the cancer. Going forward, looking for the identical combination of mutations in the other cancer patients could identify those who would respond to the same type of therapy.

REFERENCES: Wagle N. et al. Activating mTOR mutations in a patient with an extraordinary response on a phase 1 trial of everolimus and pazopanib. *Cancer Discovery* (2014) 4:546-53 doi:10.1158/2159-8290.CD-13-0353.

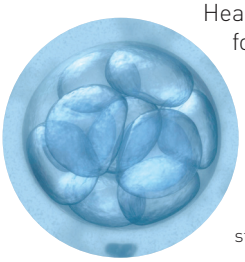
http://www.cancer.gov/news-events/press-releases/2014/exceptionalrespondersinitiativesept2014

Ethical, social and legal issues in research

Gene editing performed on human embryos
CRISPR is a molecular technology that allows scientists to create targeted changes in an organism's DNA. The genetic equivalent of "Find and Replace", CRISPR has rapidly become a general-purpose tool to edit the genome in living cells. It holds promise as a means to restore function to a mutated gene, or to silence an overactive one.

Human CRISPR studies had been limited to cultured cells until Chinese scientists announced they had attempted to edit human embryos to correct mutations in the gene that causes beta thalassemia. The researchers used discarded embryos from a fertility clinic that had been fertilized simultaneously by two sperm. The embryos were nonviable, because the extra set of chromosomes prevented development beyond a few hundred cells. The CRISPR system didn't target the gene effectively and repair attempts were often incomplete. Additionally, the system went "off target", cutting DNA in places it wasn't supposed to edit.

This study demonstrates CRISPR is not ready for clinical application. It also raises a number of ethical red flags, including whether changing the germline of an embryo should be allowed. Regulatory groups have called for a complete moratorium on gene editing of human embryos. Citing the ethical ambiguities, the National Institutes of Health has prohibited the use of its funds for this type of work.



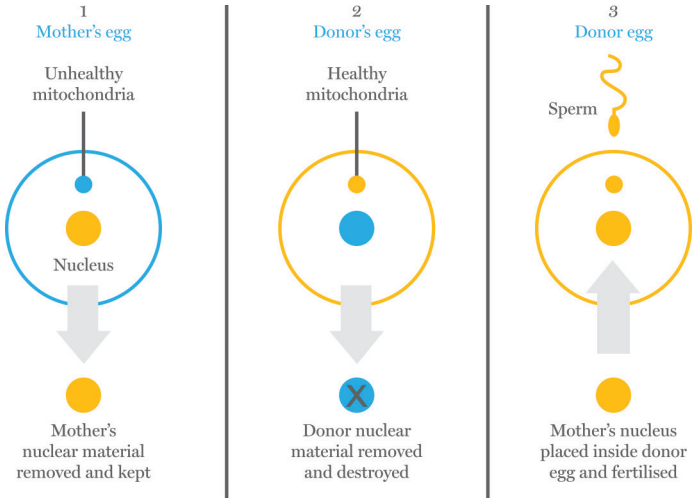
REFERENCES: Puping L. et al. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein Cell* (2015) 6:363-72 doi:10.1007/s13238-015-1053-5.

NIH statement regarding embryonic gene editing: http://www.nih.gov/about/director/04292015_statement_gene_editing_technologies.htm

"3-parent" IVF babies

digging beneath the surface of this hot-button topic

So-called "three parent babies" have been big news the last couple of years. The underlying technology, known as in vitro fertilization with mitochondrial donation, was approved in the UK in February 2015. Mitochondria contain their own tiny genomes of 37 genes; the other ~2,000 genes required to produce a functioning mitochondria reside in the nuclear genome. Sperm lack mitochondria,



Genome discoveries

highlights from the cat, monarch butterfly and octopus

Cat

A domestic Abyssinian female provided the first published genome of the cat (*Felis catus*). Relative to dogs, cats have a smaller repertoire of olfactory genes but a larger collection of genes for pheromone detection. This is not surprising, as cats don't rely on their sense of smell for hunting, but do depend on pheromones to find a mate.

Cats are only semi-domesticated - as many cat owners can readily attest – and most breeds focus on physical rather than behavioral traits. To explore the impact of domestication, the genomes of wildcats and domestic breeds were compared. Domestic cats showed significant changes in genes associated with memory, fear and motivation. These behaviors are part of the reward seeking process. Offering wild cats food as a reward for controlling rodents may have enticed normally solitary animals to stay with humans. Over time, humans selected the most docile ones as pets.

Monarch Butterfly

The monarch butterfly (*Danaus plexippus*) is well known for its annual migration across North America. However, not all monarchs migrate, most notably populations found throughout Central and South America, Europe and across the Pacific. To search for genetic factors associated with migration, the genomes of 101 monarch specimens from across the globe were sequenced. DNA variants were found in genes that affect flight muscle. These changes allow migrating

therefore children receive all their mitochondria from the maternal egg. Mutations in mitochondrial DNA can disrupt the metabolic processes that lead to ATP production, resulting in severe and often life-threatening disorders. Symptoms generally begin in childhood and steadily worsen, often leading to death before adulthood. During in vitro fertilization with mitochondrial donation, the nucleus is removed from a donor oocyte and replaced with a nucleus from one of the mother's eggs. The donor cytoplasm contains healthy mitochondria, each with the donor's DNA. The oocyte is then fertilized with sperm from the father. The technology prevents the transmission of certain forms of mitochondrial disease, resulting in infants with 0.2% of their total DNA from the donor woman.

The process is technically difficult and there are significant ethical and regulatory questions about mitochondrial donation. The new combination of nuclear and mitochondrial DNA is present in all of an offspring's cells. If that offspring is female, this novel blend will be passed to future generations through her oocytes - a form of germline modification that many individuals are hesitant to permit. Some worry that allowing this form of genetic modification implicitly authorizes genetic engineering to enhance desired traits. Proponents argue that preventing serious, life-threatening disorders is a far cry from selecting for vanity traits. It is possible that the first baby could be born using this technique as early as 2016.

REFERENCE: UK approves three-person babies. BBC News Website, accessed September 10, 2016. http://www.bbc.com/news/health-31594856

monarchs to more efficiently utilize energy resources. Active flight is exceptionally energy demanding, requiring 25 times more energy than resting. Genetic changes that minimize energy usage during flight provide clear selective advantages for migrating monarchs, but are less important for stationary populations.

Octopus

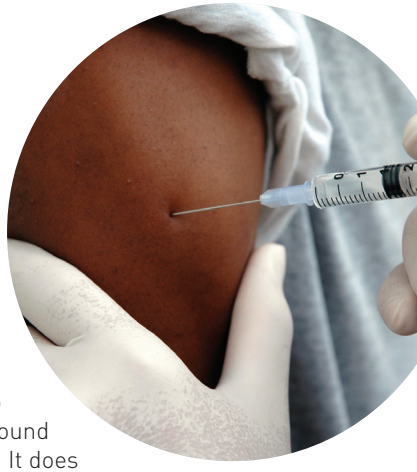
The genome of the California two-spot octopus (*Octopus bimaculoides*) is 2.7 billion base pairs in size and predicted to contain 33,000 protein-coding genes. A member of the mollusk phylum, the octopus is dramatically different from its snail and bivalve relatives. Its large brain and well-developed nervous system contain six times as many neurons as a mouse brain and it is generally regarded as the most intelligent of the invertebrates. The octopus genome yields a tantalizing explanation: it has a huge number of genes from the protocadherin family. These genes establish and maintain the nervous system and are abundant in animals with well-developed brains. Several octopus-specific genes were also identified that are primarily expressed in the suckers and skin. These may provide insight into how the octopus is able to taste what it touches, and control its rapid system of camouflage.

REFERENCES: Montague M.J. et al. Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc Natl Acad Sci USA* (2014) 111:17230-5 doi:10.1073/pnas.1410083111.
Zhan S. et al. The genetics of monarch butterfly migration and warning colouration. *Nature* (2014) 514:317-21 doi:10.1038/nature13812.
Albertin C.B. et al. The octopus genome and the evolution of edphalopod neural and morphological novelties. *Nature* (2015) 524:220-4 doi:10.1038/nature14668

In brief

Infectious disease

Testing Ebola vaccines



In the face of the ongoing Ebola virus disease (EVD) outbreak in West Africa, clinical trials have been testing the effectiveness of candidate vaccines. Encouragingly, the results from one phase 3 trial show a nearly 100% effective rate with a single dose.

The rVSV-ZEBOV vaccine is based on a vesicular stomatitis virus, genetically modified to express a portion of the protein found on the surface of the Ebola virus. It does not contain a live Ebola virus. When injected, the vaccine causes mild flu-like symptoms but produces a rapid immune response to the Ebola virus surface protein. If an individual who has been vaccinated comes into contact with the actual Ebola virus, their immune system is primed to identify and eliminate the virus without the individual getting sick.

Conducted in Guinea, the trial used a ring vaccination approach - the same strategy that eradicated smallpox in the 1970s. Upon detection and confirmation of a new case of EVD, the people who have close contact with the patient are identified. These individuals plus their own circle of contacts compose the "ring". Everyone within the ring is vaccinated. The vaccine doesn't protect individuals who have already infected by the original patient, but it can prevent the virus from being transmitted to the next level of contacts. This helps creates a buffer of immune people around each new case, preventing the spread of EVD.

In a traditional case/control study, the control group is offered a placebo or a vaccine against a different disease. Leaving vulnerable individuals completely unprotected against EVD was viewed as unacceptable, so the study leaders proposed a three week vaccination delay for the controls. Nearly all cases of EVD arise within 21 days of infection, meaning the control group wouldn't be protected from the initial contact with the virus but would receive benefit against later interactions. Rings were randomly placed into immediate or delayed vaccination cohorts and observed for 84 days after vaccination.

Of the 2,014 adults that received immediate vaccinations, no new infections were identified after 10 days (to account for a pre-vaccine infection). When compared to 16 reported infections from the 2,380 delayed vaccinations, the findings indicate the vaccine may provide complete protection. The data also suggest the ring vaccine strategy may be the most effective way to deliver the vaccine to a broad population. The trial is ongoing in Guinea and has been extended to Sierra Leone. The delayed vaccination arm has been dropped and individuals within all rings will be offered the vaccination immediately.

REFERENCE: Henao-Restrepo A.M. et al. Efficacy and effectiveness of an rVSV-vectored vaccine expressing Ebola surface glycoprotein: interim results from the Guinea ring vaccination cluster-randomised trial. *Lancet* (2014) 386:857-66 doi:http://dx.doi.org/10.1016/S0140-6736(15)61117-5.

NEW FINDINGS - GENETICS AND GENOMICS IN THE CLINIC

Exome and genome sequencing

grappling with incidental findings

Whether focused on the entire genome or just the exome (the protein-coding regions), large-scale sequencing is an integral part of many research studies and is gaining a foothold in clinical practice. As well as detecting genetic variants that answer a specific clinical question, genome sequencing may also identify additional DNA changes with important secondary implications-- referred to as incidental findings (IFs). Many are health-related, i.e. a predisposition to adult-onset cancer or identification of carrier status for recessive disorders. Others, such as nonpaternity detection, are societal in nature. IFs can lead to early medical intervention but also can result in unnecessary additional testing or emotional distress. They are a much-debated topic among researchers, ethicists and clinicians. Given the speed at which genome-wide sequencing is being integrated into both research and clinical settings, the question of whether to return IFs appears to be rapidly transitioning to how to return, who should return and when to return.

Estimating IF prevalence

A collaborative team of scientists and clinicians sought to estimate the frequency of IFs that would be observed in clinical exome sequencing studies. Single nucleotide variants in 112 genes with medically actionable results were examined in more than 6,500 research participants from the National Heart, Lung and Blood Institute's Exome Sequencing Project. Under this criteria, 2.0% of European and 1.1% of African ancestry participants would be given results with highly-penetrant disease-causing or likely disease-causing DNA variants.

Stakeholder attitudes towards IFs

An online survey explored the attitudes around returning IFs identified as part of a research study. Nearly 7,000 participants were drawn from four different stakeholder groups: members of the public, genetic health professionals, non-genetic health professionals and genomic scientists. Of those surveyed, 98% were personally interested in learning about IFs associated with preventable life-threatening conditions. The stakeholders did not expect researchers to actively search for IFs that weren't relevant to their research. The biggest factor influencing attitudes towards returning IFs was membership in a particular stakeholder group: members of the public (who form the bulk of research participants) want access to more information than genetic health professionals (who handle the clinical interpretation of IFs) consider appropriate to disclose.

Attitudes regarding the return of genomic IFs were also assessed from 258 pediatric and primary care health professionals attending a "Best Practices in Pediatrics" conference in Wisconsin. As in other studies, there was strong support for the

Cancer formation and detection

scientific discoveries about how cancer forms and potential early-stage biomarkers

Impact of stem cell divisions on cancer risk

Cancer arises from a stepwise accumulation of mutations in genes that encode factors associated with cell growth, the normal process of cellular development or the ability to identify and repair DNA damage. What causes these mutations? Generally, researchers point to environmental factors (i.e. smoking, alcohol use and ultraviolet light) as well as inherited genetic changes (mutations in *BRCA1* and *BRCA2* genes linked to heritable breast and ovarian cancer).

Scientists also recognize a third cause – replicative mutations that spontaneously arise within stem cells that give rise to each tissue – but historically have not been able to measure this factor's relative importance. A recent analysis calculates this replicative mutation risk by correlating the lifetime risk of different cancer types with the total number of normal, self-renewing stem cells that maintain tissue homeostasis. For example, colorectal and basal cell tissues are among the most frequently observed cancer types and have the highest number of stem cell divisions. In contrast, bone tissue of the pelvis, head and arms have the lowest lifetime risks of cancer and the fewest numbers of stem cell divisions. A striking association emerged, suggesting that as much as 2/3 of cancer risk among tissues is due to unavoidable random mutations arising in stem cells (described as "bad luck.")

The authors stress these findings should not be taken to mean all types of cancer are unpreventable and note the importance of lifestyle changes and early detection to reduce the risks of many cancers.

DNA methylation as an early-detection biomarker

As with many forms of cancer, kidney tumors are often lethal when metastatic, but can be successfully treated if identified early. As early stage tumors are rarely symptomatic, there is a strong push to identify cancer-related biomarkers – measurable indicators that signify the presence of cancer. Researchers compared patterns of DNA methylation between kidney tumor and adjacent normal tissue from 96 patients, noting over 9,000 significantly different regions. Further analysis identified a panel of 20 markers that best differentiated normal from malignant tissue, which was then validated using an independent set of samples. Many of these markers are located within promoter regions or introns of genes associated with immune function or cancer aggression. In the future, these methylated sequences could potentially compose a non-invasive test from patient blood or urine that could be used for early detection or to monitor patient response to therapy.

The analysis of DNA methylation took place in the labs of HudsonAlpha researchers Devin Absher, Ph.D., and Rick Myers, Ph.D.

return of IFs that are clinically actionable in children as well as adults. There was substantially less support for returning non-actionable IFs and findings of uncertain clinical significance. Perhaps not surprisingly, these views were correlated with the provider's desire for genome sequencing in his/herself.

Views on returning IFs

Approaches for returning genome-based results were compared across the International Federation of Human Genetics Societies – a 62 member group representing organizations from around the world. Guidelines generally grouped into four broad categories: 1. utilizing panels of specific genes or targeted sequences, which automatically reduces the potential for incidental findings; 2. only returning results that meet a strict set of criteria for analytical validity, clinical significance and actionability (the most prevalent option identified); 3. returning results on a case-by-case context basis; and 4. no incidental findings are returned.

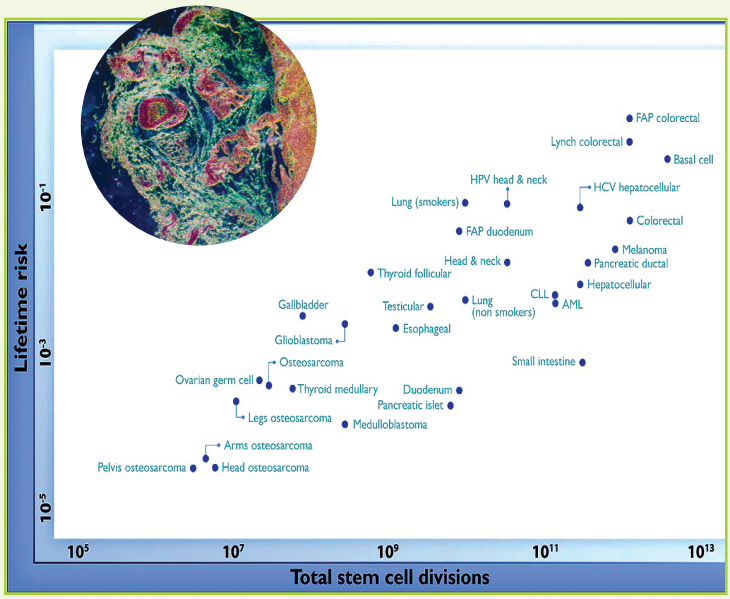
REFERENCES: Amendola L.M. et al. Actionable exomic incidental findings in 6504 participants: challenges of variant classification. Genome Research (2015) 25:305-15 doi:10.1101/gr.183483.114.

Middleton A. et al. Attitudes of nearly 7000 health professionals, genomic researchers and publics toward the return of incidental results from sequencing research. European Journal of Human Genetics (2015) advance online publication April 29, 2015. doi:10.1038/ejhg.2015.58.

Strong K.A. et. al. Views of primary care providers regarding the return of genome sequencing incidental findings. Clinical Genetics (2014) 86:461-8 doi:10.1111/cge.12390.

Knoppers B.M., Zawati M.H. and Senecal K. Return of genetic testing results in the era of whole-genome sequencing. Nature Reviews Genetics (2015) advance online publication, August 4, 2015. doi: 10.1038/nrg3960.

HudsonAlpha researcher Greg Cooper, Ph.D., contributed to the study estimating the frequency of IFs in a clinical population; Kim Strong, Ph.D., and David Bick, M.D. led the study examining attitudes of Wisconsin primary care providers.



(Figure used with permission from Tomasetti et. al)

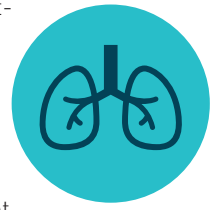
REFERENCES: Tomasetti C. and Vogelstein B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* (2014) 347:78-81 doi:10.1126/science.1260825.

Lasseigne B.N. et al. DNA methylation profiling reveals novel diagnostic biomarkers in renal cell carcinoma. *BMC Medicine* (2014) 12:235-45 doi:10.1186/s12916-014-0235-x.

In brief

Gene therapy

Inhaled therapy for cystic fibrosis



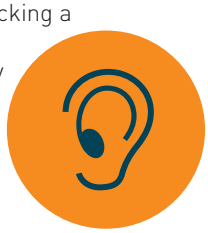
Cystic fibrosis (CF) has long been a "poster child" for the promise of gene therapy, in part because it was assumed that inhaling an aerosol-based treatment would provide a straightforward delivery route to lung cells. Like many things associated with gene therapy, the reality was much more difficult than the concept. A recent clinical trial in the UK offers hope that airway delivery may be a viable option after all. Encapsulated within fat globules, unmutated copies of the *CFTR* gene were administered monthly by nebulizer to 78 patients with CF. After one year of treatment, the lung function of those receiving the therapy remained stable, compared to a decline in the group receiving a placebo. The effect, while statistically significant, was modest - further improvements in efficacy are needed before the therapy has widespread clinical value. Still, the results offer hope that non-viral gene therapy may find a place in the CF treatment toolkit.

Therapy-restored sight fades for patients with LCA

In 2007, researchers showed that gene therapy could restore sight among patients with Leber's congenital amaurosis (LCA), an inherited blindness caused by mutations in the *RPE65* gene. A normal copy of the gene, carried on a viral vector, was injected into the eye. Two recent follow-up studies have determined that the improvement in sight peaked 6 to 12 months after injection but then declined, as retinal photoreceptors continued to die as part of the natural course of the disease. The findings point to the need for a more efficient gene delivery system. Along those lines, patients treated using a newer gene therapy approach have improvements in light sensitivity lasting beyond the timeframe experienced by the initial cohort.

Treating a mouse model of deafness

Over 100 genes have been associated with hearing loss in humans. In a first step towards future treatments, gene therapy has rescued hearing to near normal levels in a mouse model of monogenic deafness. Mice lacking a functional copy of either the *Tmc1* or *Tmc2* gene received a nonmutated copy injected into their inner ear. Within a month of treatment, there was a partial, although temporary, restoration of hearing.



REFERENCES: Alton W.F.W. et al. Repeated nebulization of non-viral CFTR gene therapy in patients with cystic fibrosis: a randomized, double-blind, placebo-controlled, phase 2b trial. Lancet Respiratory Medicine (2015) 3: 684-91 doi:10.1016/s2213-2600(15)00245-3.

Jacobson S.G. et al. Improvement and decline in vision with gene therapy in childhood blindness. New England Journal of Medicine (2015) 372:1920-6 doi:10.1056/NEHMoa1412965.

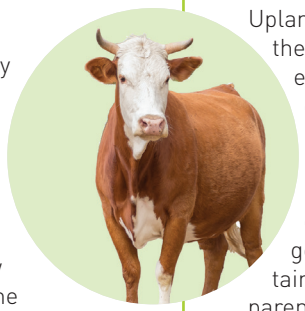
Askey C. et al. Tmc gene therapy restores auditory function in deaf mice. Science Translational Medicine (2015) 7:295ra108 doi:10.1126/scitranslmed.aab1996.

NEW FINDINGS - AGRISCIENCE

In brief

Identifying agriculturally important traits

Coat color in cattle



Breeders studied coat color inheritance in dairy cattle for generations prior to the advent of DNA sequencing. The dominant black/white spotted coat and the recessive red/white spotted coats are controlled by four alleles of the mc1r gene, which controls the production of black/brown eumelanin pigment. Disruption of the eumelanin pathway results in red/yellow pheomelanin pigment production, producing the red coat color.

In 1980 a female calf was born that displayed a novel coat color, similar to the recessive red, but from parents who should only have produced black offspring. Despite having the genotype for black coat, 50% of her offspring with black-coated fathers showed the new red coat, now termed dominant red.

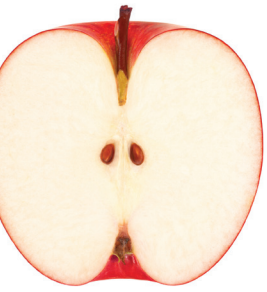
Researchers determined the dominant red phenotype is due to a missense mutation that changes an arginine to cysteine in a gene involved in vesicle transport. This is interesting for several reasons. The common assumption is that highly conerved regions of the genome are critical for life and changes in those areas result in disease. This mutation occurs in a highly conserved region, yet results in healthy animals. This new phenotype provides novel directions for research into the biology of mammalian pigmentation.

HudsonAlpha researcher Greg Barsh, M.D., Ph.D. contributed to these findings.

REFERENCE: Dorshorst B. et al. Dominant red coat color in Holstein cattle is associated with a missense mutation in the coatomer protein complex, subunit alpha (COPA) gene. *PLoS One* (2015) 10 :e0128969 doi:10.1371/journal.pone.0128969

Genetically modified foods

Apples and potatoes receive approval



The Food and Drug Administration has determined that both Arctic Apples (profiled in the 2014-15 Guidebook) and Innate potatoes are as safe and nutritious as their conventional counterparts. These plants are genetically engineered using RNA interference to reduce the transcription of genes that encode browning-associated enzymes. Consequently, the apples and potatoes do not brown when cut or bruised. The genetic sequences to silence the browning genes are directly derived from the apple and potato plants – not from other organisms. The creator of Innate potatoes, J.R. Simplot Company, has introduced a second-generation potato that uses similar RNAi-based technology to resist late blight and allow for storage at colder temperatures. The company expects FDA and EPA approvals for this new variety within a year; if that occurs, commercial planting would begin by 2017.

REFERENCES: FDA news release: FDA concludes Arctic Apples and Innate Potatoes are safe for consumption http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm439121.htm

J.R. Simplot Company's USDA petition for the second generation Innate potato https://www.aphis.usda.gov/brs/aphisdocs/14_09301p_dpra.pdf

Deciphering the mysteries of cotton

assembling the genome of this economically important crop

Upland cotton (*Gossypiym hirsutum*) provides more than 90% of the world's spinnable cotton fibers and has an estimated global economic impact of $500 billion. The species has a complex evolutionary history: it was created from a chance fertilization event 1-2 million years ago between two ancestral diploid species of cotton, the A species from Africa and the D species from the Americas (AA+DD→ AD). The fertilization event was quickly followed by a complete duplication of the genome (AD --> AADD), so the resulting tetraploid hybrid contained twice the genetic information of either ancestral diploid parent. This process, known as polyploidy, has occurred throughout history in over 80% of plant species. Many flowering plants have undergone multiple rounds of genome duplication. Human domestication and breeding further shaped the cotton genome, selecting for genetic variants that produced higher fiber yield and quality.

In 2015, two groups, each composed of scientists from the United States and China (the world's largest cotton-producing country), published complementary drafts of the 2.5 billion base-pair Upland cotton genome. The genome was compared to the

Citrus greening disease

bacterial infection destroys citrus orchards

Citrus greening is a bacterial disease that attacks the vascular system of citrus trees, blocking the transport of sugars produced by the leaves during photosynthesis. Caused by the *Candidatus Liberibacter* bacteria, the tree essentially starves to death. There is no cure and infected trees produce small and bitter-tasting green oranges.

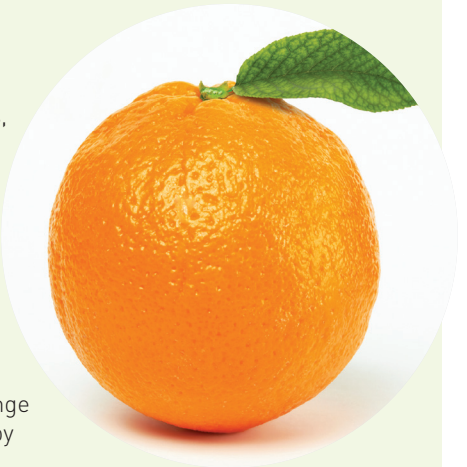
An aphid-like insect known as the Asian citrus psyllid carries the bacteria from tree to tree. Because there is often a long period between infection and onset of symptoms, the bacteria can spread widely before being detected. In Florida, the insect was first identified in 1998 but citrus greening was not observed until 2005. Today, nearly every orchard in Florida has been infected. Citrus greening was recently also detected in California. Florida and California produce the majority of citrus consumed in America. California oranges are primarily eaten fresh, while Florida oranges are generally squeezed into juice. Over the last ten years, citrus greening has cost Florida growers an estimated $3 billion in lost revenue and orange juice production has dropped by nearly one-third to its lowest level in 50 years.

Over $200 million has been spent to find a cure for citrus greening, including new combinations of bactericides, more disease tolerant rootstocks, and using insecticides and even parasitic wasps to attack and kill the psyllids. Scientists have discovered that using solar radiation to heat trees to 100 degrees Fahrenheit inside plastic tents can kill some of the bacteria within already-infected trees, prolonging the tree's lifespan by a few years. Borrowing a page from

recently published reference genomes for the presumptive A and D parental species. The tetraploid cotton has lost twice as many genes from the sequence contributed by the A parent than from the portion of the genome provided by the D parent. It is not known exactly why this has occurred, or if this selective loss is associated with differences in fiber quality or the domestication process.

To extend this work, the National Science Foundation recently awarded a $2.4 million grant to a partnership of five institutions, including HudsonAlpha, with the goal of producing a more finished, finer resolution genome. Once that version is complete, the information will assist farmers in breeding cotton varieties with longer, stronger fibers.

REFERENCES: Li F. et al. Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insight into genome evolution. Nature Biotechnology (2015) 33: 524-30 doi:10.1038/nbt.3208.
Zhang T. et al. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nature Biotechnology (2015) 33: 531-7 doi:10.1038/nbt.3207.



cancer-attaching viruses, scientists have even engineered a modified citrus virus to produce molecules that destroy the disease-causing bacteria.

Southern Gardens, a major Florida citrus grower, has developed a genetically modified orange that resists the disease by incorporating a spinach defensin gene. Defensins are naturally occurring proteins that kill bacteria by binding to and puncturing their cell walls. They are not harmful to humans and do not change the appearance or taste of the oranges. The Environmental Protection Agency has recently granted an experimental use permit to allow Southern Gardens to plant 150 acres in Florida and 50 acres in Texas with the modified trees. The company believes that the results of the field test should be known within three years.

REFERENCES: Grafton-Cardwell E.E., Stelinksi LL, Stansly P.A. Biology and Management of Asian Citrus Psyllid, Vector of the Huanglongbing Pathogens. Annual Review of Entomology (2013) 58:413-32 doi:10.1146/annurev-ento-120811-153542
National Geographic overview of citrus greening and the GM orange: http://news.nationalgeographic.com/news/2014/09/140914-florida-orange-citrus-greening-gmo-environment-science/

AGRISCIENCE - NEW FINDINGS

In brief

Applications of genome sequencing

Comparing wild and domesticated soybeans

Successful crop domestication is often achieved at the expense of genetic diversity. While domesticated versions of plants are usually more productive than their wild counterparts, they contain a tiny fraction of the genetic variation present in the ancestral species. Scientists in China recently sequenced *Glycine soja*, the wild relative of cultivated soybean (*Glycine max)*. A comparison between the wild and domesticated genomes revealed hundreds of genes in regions exhibiting copy number variation between the two strains. Many of these genes are associated with agriculturally-important traits like pest resistance, flowering time, and seed oil concentration. Breeders can take advantage of the information revealed by this genome comparison to develop new strains of soybean that will thrive under diverse environmental conditions.

HudsonAlpha researchers Jeremy Schmutz and Jane Grimwood, Ph.D. contributed to sequencing the cultivated soybean.

REFERENCE: Li Y. et al. De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology* (2014) 32:1045-52 doi:10.1038/nbt.2979

Analyzing breeding patterns from parchment samples

Parchment is the skin of calves, goats or sheep that has been dehaired, stretched and scraped. Before the widespread use of paper, parchment was the primary material for written documents and records. Recently, next generation sequencing techniques were used to characterize the origin of two British parchments from the 17th and 18th centuries. DNA was extracted from 2x2 cm pieces cut from the documents. Based on nuclear and mitochondrial DNA analysis, both parchments were identified as sheepskin. The ratio of X chromosome to autosomal DNA sequence fragments determined the contributing sheep were both females.

By comparing the genetic markers in the parchments with those from modern breeds of sheep, scientists were able to focus on the most likely regions where the parchments were produced – the 17th century parchment from a black-faced sheep in northern Britain and the 18th century belonging to a sheep derived from the Midlands and southern Britain. This proof of principle study suggests that sequence analysis of historic documents may help understand the complex process of livestock breeding that has taken place across Europe for the last 1,000 years.

REFERENCE: Teasdale M.D. et al. Paging through history: parchment as a reservoir of ancient DNA for next generation sequencing. Philosophical Translations Royal Society (2014) B370:20130379 doi:10.1098/rstb.2013.0379.

Developing high yield plants optimized for changing environmental conditions is a cornerstone of meeting our food, te

Agricultural

CONIFERS members of the order Pinales

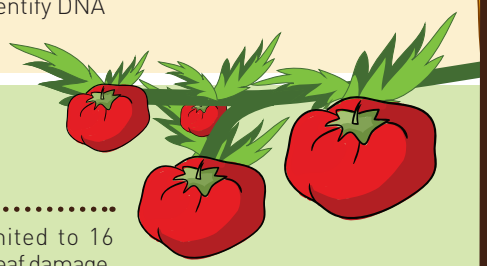
Genome size (millions of bases) >20,000 Mb
Number of Genes ~ 30,000

Conifers are a family of cone-bearing, primarily evergreen plants that include the cedar, fir, pine, and spruce. Researchers are identifying networks of genes involved in wood formation as well as adaptation to temperature change and pest resistance. A key part of forest management is replanting harvested lands with seedlings that have with superior growth characteristics, high wood quality and environmental adaptability. Genomics offers a rapid way to identify the best candidates for high-performance trees.

BREAD WHEAT *Triticum aestivum*

Genome size (millions of bases) 17,000 Mb
Number of Genes ~ 124,000

The wheat genome is almost 6 times larger than the human genome and largely consists of repetitive chunks of non-essential DNA. Determining agriculturally useful genetic variants is challenging. Researchers from Kansas State University have developed a wheat diversity "catalog," of genetic variation present in 62 varieties from around the world. This helps identify DNA changes associated with key traits.



TOMATO *Solanum lycopersicum*

Genome size (millions of bases) 900 Mb
Number of Genes ~ 34,000

Commercially-grown tomato plants are limited to 16 hours of light per day or they suffer a fatal form of leaf damage. Recently, researchers linked the CAB-13 gene to light tolerance. Commercial tomatoes have a partially inactive version of CAB-13. When the wild version was crossed into domesticated varieties, plants thrived under continuous light and tomato yield increased by 20 percent.

POTATO *Solanum tuberosum*

Genome size (millions of bases) 844 Mb
Number of Genes ~ 39,000

Potatoes are an important source of starch, protein, antioxidants and vitamins, but are susceptible to a wide range of pests and disease such as *Phytophthora infestans*, which caused the Irish Potato Famine in 1840. When the potato genome was sequenced, researchers identified over 800 genes associated with disease and pest resistance. The functions of these genes are under study, in the hopes of creating hardier, less susceptible varieties of potato.

l, textile and bioenergy needs. Using genomic information to shape agricultural processes is key to accomplishing this challenge.

Genomics

APPLE *Malus domestica*

Genome size (millions of bases) 745 Mb
Number of Genes ~ 57,000

In February 2015, the USDA approved the commercial planting of genetically engineered apples that resist browning when sliced or bruised. Scientists silenced the activity of a set of apple genes that produce the enzymes involved in browning. The so-called "Arctic apple" will be available in small, test-market quantities in late 2016. It will be many years before the nonbrowning fruit is widely distributed.

CORN *Zea Mays*

Genome size (millions of bases) 2,300 Mb
Number of Genes ~ 32,000

Plant growth is controlled by a complicated network of genes and their interaction with the environment. Determining which genes are activated under different conditions helps scientists decipher the relationship between genes and the environment. Many of these networks have been determined for corn, including those that influence leaf and ear size, water control and drought responsiveness. These findings open the door to selecting optimal genetic combinations for future varieties.

COTTON *Gossypium hirsutum*

Genome size (millions of bases) 2,300 Mb
Number of Genes ~ 70,000

The fluffy fibers harvested from the cotton plant are in fact hair-like structures that originate from the seeds enclosed in the cotton boll. There are over 20,000 fibers per seed. Cotton researchers want to understand the genetic and environmental regulators of fiber production. Scientists and farmers hope to identify genetic variants that optimize fiber length, quantity and quality and use them to guide the development of new cotton strains.

If you want to know more:

HudsonAlpha's 2015 Biotech 201 series focused on agricultural genomics. Videos of each session are available for viewing on the HudsonAlpha YouTube Channel. Links can be found on the HudsonAlpha website at: **hudsonalpha.org/education/lifelong-learning/biotech-series**

The subject of genetically modified foods is multilayered, emotionally charged, and complex. "Panic-Free GMOs" is a multi-part special series from Grist, an online environmental news magazine. It provides a deep dive into the science and a balanced review of the issues associated with GM foods. **http://grist.org/series/panic-free-gmos/**

HudsonAlpha's Genome Sequencing Center

The Genome Sequencing Center (GSC) at HudsonAlpha is led by Faculty Investigators Jeremy Schmutz and Jane Grimwood, Ph.D. Over the last 10 years, this team of researchers worked on the genomes of nearly 100 organisms, ranging from switchgrass to humans to leaf blight. The GSC is a partner in the Joint Genome Institute, which is funded by the US Department of Energy and focused on initiatives related to clean energy generation. Additionaly, research in the GSC lays the foundation for genomics-enabled breeding across several important crop species. Below are some of their most recent accomplishments.

Soybean – The GSC led the effort to sequence the soybean reference genome in 2010 and have recently assembled a diversity catalog showcasing genetic variation for more than 70 ancestral and current soybean cultivars. This effort seeks to broaden the genetic base of commercially available soybeans. Modern soybean strains represent only a fraction of the genetic variation present across the entire species. During domestication, breeders selected plants using a very narrow set of criteria, like higher yields. Valuable traits for pest resistance and climate hardiness were inadvertantly tossed aside. Identifying the full spectrum of genetic diversity allows important ancestral traits to be reclaimed.

Citrus – The GSC was part of an international team that analyzed and compared the genome sequences of 10 diverse citrus varieties, including sweet and sour orange, mandarin and pumelo. Modern citrus has very little genetic diversity and is vulnerable to the effects of disease and environmental stress. By inferring past events that gave rise to these common citrus varieties, researchers hope to identify strategies that restore genetic diversity, producing healthy and high-yielding trees.

Sorghum – Sorghum is gaining recognition as an important feedstock source for biofuel production. It is an excellent candidate because it doesn't require as much water or fertilizer as other grasses. The GSC, in collaboration with several research centers, was recently awarded a grant from the US Department of Energy to catalog the natural variation in sorghum DNA. Once researchers identify genes that regulate water and nutrient use, they can develop next-generation feedstock plants that maximize growth, producing large amounts of biomass for energy conversion.

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

KEY TECHNOLOGIES

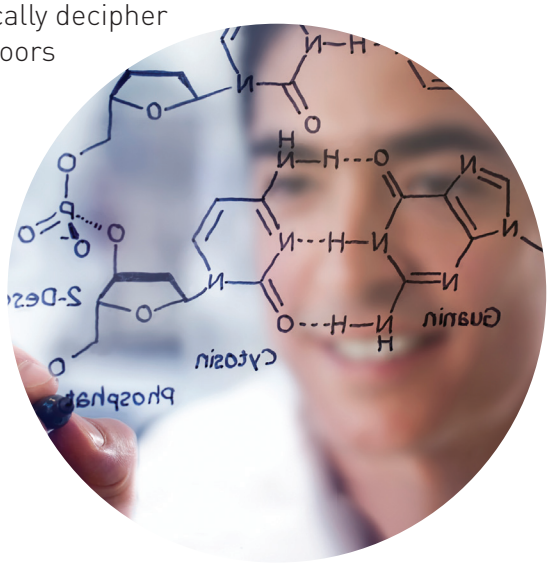
DNA Sequencing

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

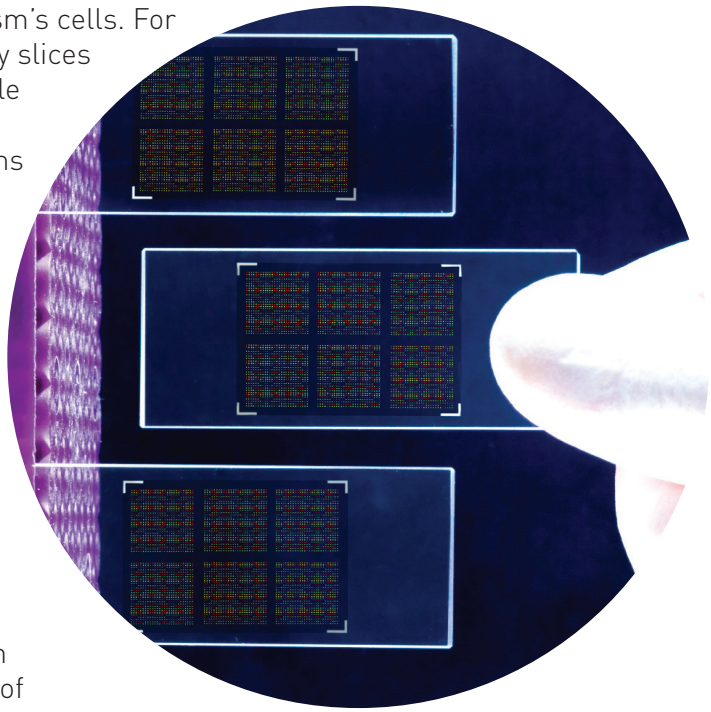
As sequencing techniques identify the genetic recipes of an organism, understanding the function of those genes becomes increasingly important. Many of the fundamental measurements used by molecular biologists seek to determine the presence, absence or relative amounts of RNA produced by a gene. Initially,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

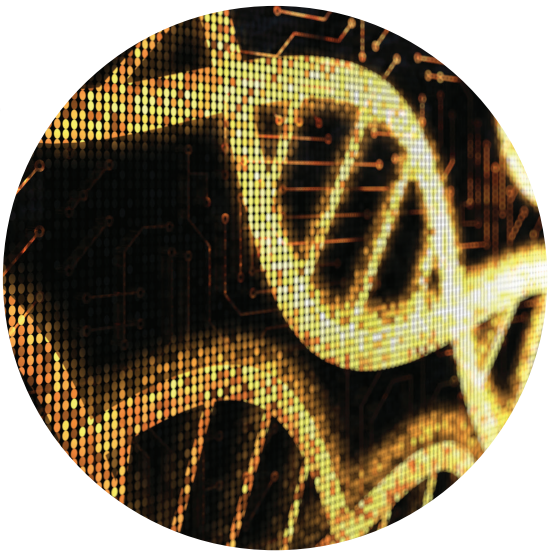


Bioinformatics

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.

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



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

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

APPLICATION

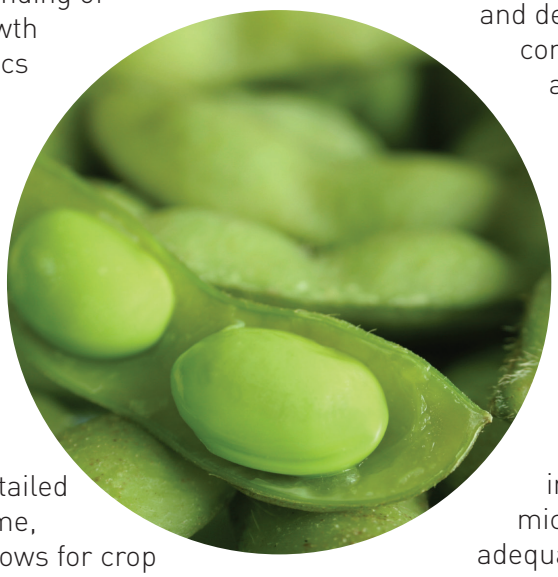
Agriculture

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.

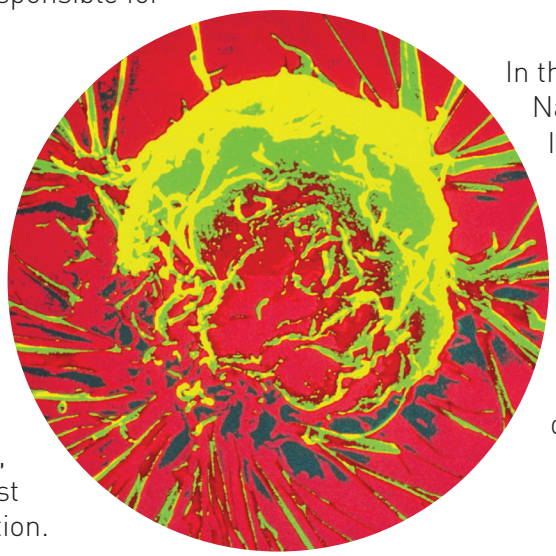
The idea that all cancers are genetic in nature and occur as a stepwise addition of mutations, many of which are initiated by environmental factors, is a useful addition to a discussion on common causes of disability and premature loss of life in a Health class (COS objective 10). These concepts should 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.

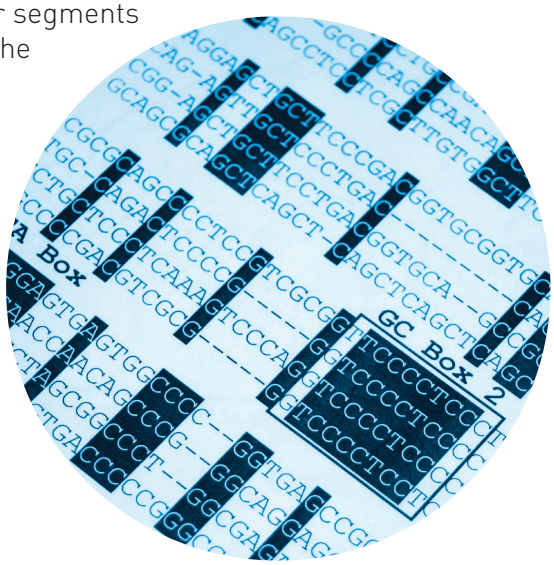
Microarray experiments are currently too cumbersome to perform in a clinic, so it is not likely they will be used routinely to diagnosis patients.

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.

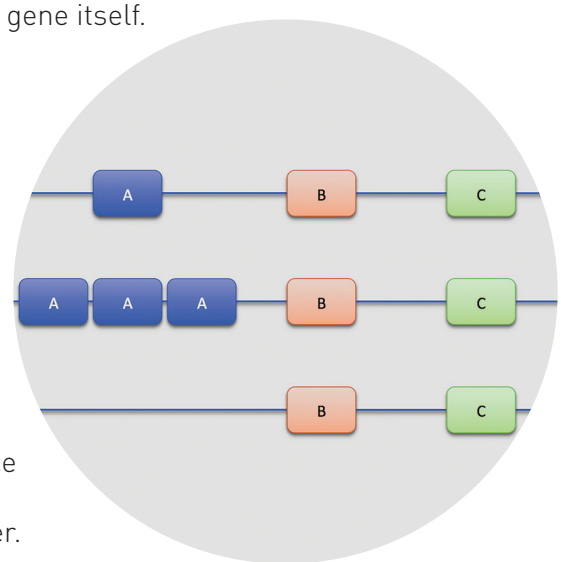
Relating genetic variation to human disease and inheritance is identified in the Biology COS under objective 8 and is described in detail in the Genetics COS objectives 2 and 5, particularly as it connects with genetic patterns of inheritance and multiple alleles. Genetic variation as it relates to human 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.

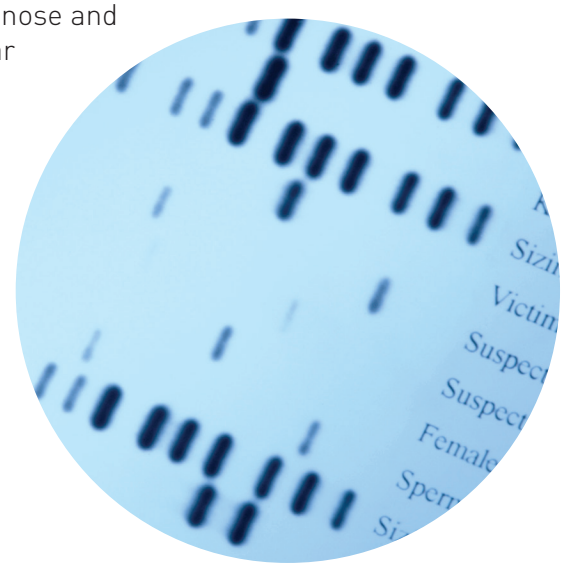


Criminal Justice and Forensics

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.

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.

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.

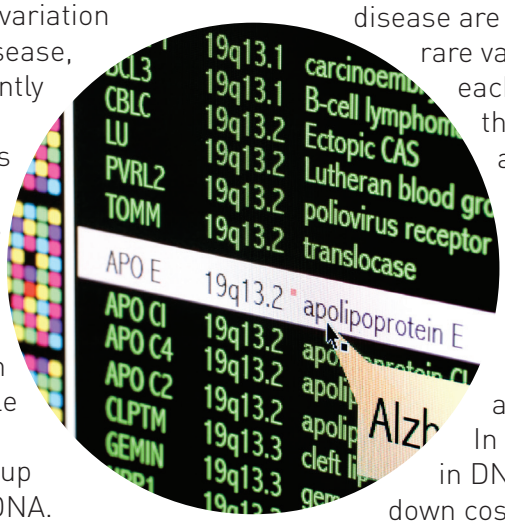
Identifying Genetic Influence on Disease

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.

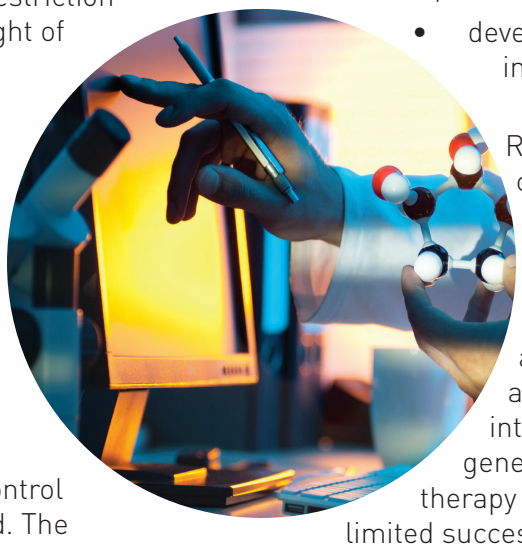
There are over 200 pharmaceutical products that either recommend genetic testing or point to the influence of genetic variability on the drug's response.

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.

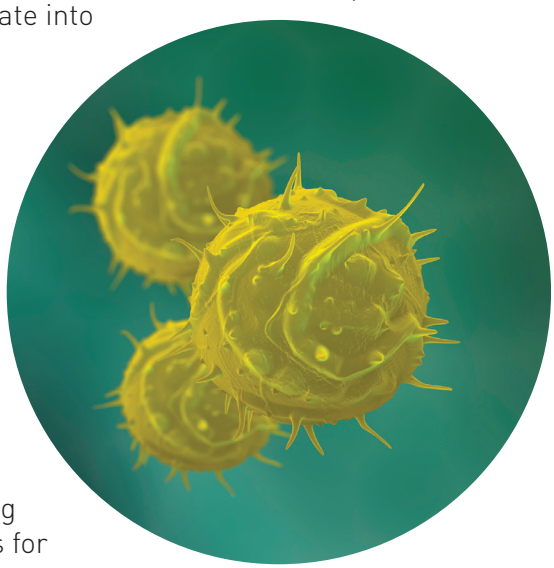
Stem cells

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.

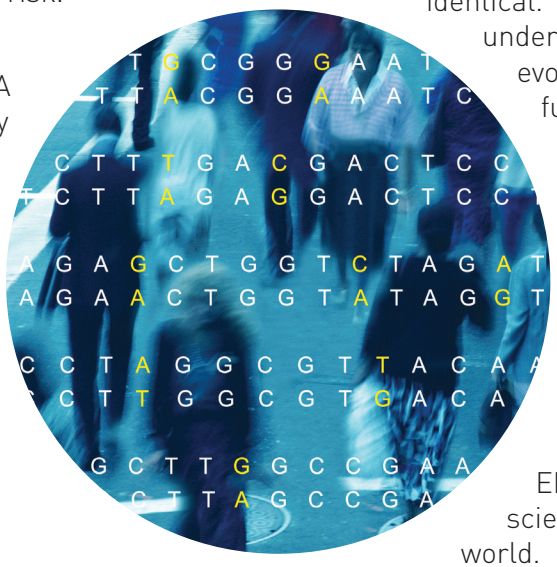
Studying the Genome to Understand the Sequence

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.

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.

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.

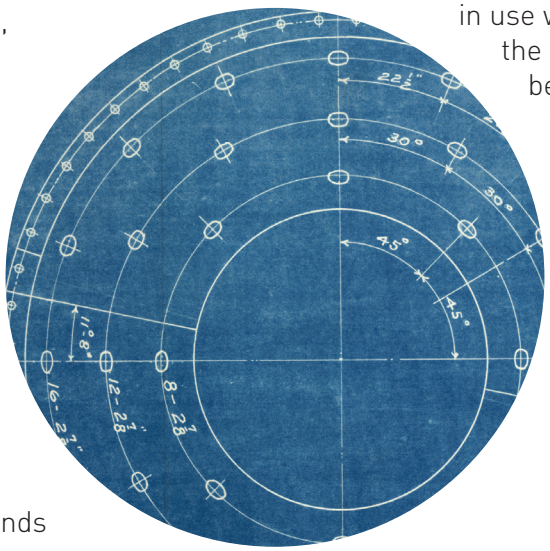
Just like the HGP, information generated from HapMap and ENCODE is freely accessible by scientists and the public around the world.

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.

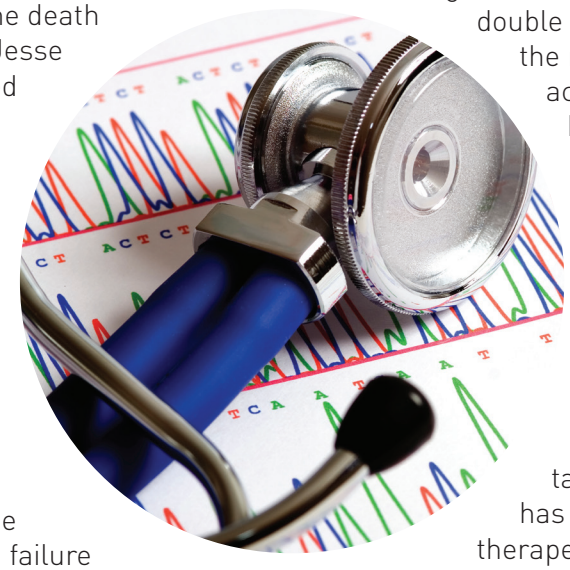
THERAPEUTIC APPROACHES

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.

Science snapshots references and image credits

References for Science Snapshot (pg 9)

1. Yuen R.K.C. et al. Whole-genome sequencing of quartet families with autism spectrum disorder. Nature Medicine (2015) 21:185-191 doi:10.1038nm.3792.

2. Cirulli E.T. et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science (2015) 347:1436-41 doi:10.1126/science.aaa3650.

3. Lin T. et al. Genomic analyses provide insights into the history of tomato breeding. Nature Genetics (2014) 46:1220-1226 doi:10.1038ng.3117.

4. The Cancer Genome Atlas Research Network. Genomic Classification of Cutaneous Melanoma. Cell (2016) 161:1681-96 doi:10.1016/j.cell.2015.05.044. The Cancer Genome Atlas Research Network. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. New England Journal of Medicine (2015) 372:2481-98 doi:10.1056/NEJMoa1402121. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature (2015) 517:576-82 doi:10.1038/nature13129. The Cancer Genome Atlas Research Network. Integrated Genomic Characterization of Papillary Thyroid Carcinoma. Cell (2014) 159:676-90 doi:10.1016/j.cell.2014.09.050.

5. May T. et al. Can targeted genetic testing offer useful health information to adoptees? Genetics in Medicine (2015) 17:533-35 doi:10.1038/gim.2015.58.

6. Balestra F.R., von Tobel L., and Gonczy P. Paternally contributed centrioles exhibit exceptional persistence in C. elegans embryos. Cell Research (2015) 25:642-4 doi:10.1038/cr.2015.49.

7. King T.E. et al. Identification of the remains of King Richard III. Nature Communications (2014) 5:5631 doi:10.1038/ncomms6631.

8. Carvalho D.O. et al. Suppression of a field population of Aedes aegypti in Brazil by sustained release of transgenic male mosquitoes. PLOS neglected tropical diseases (2015) doi:10.1371/journal.pntd.0003864.

9. Fields A.T. et al. Facultative parthenogenesis in a critically endangered wild vertebrate. Current Biology (2015) 25:R439-47 doi:10.1016/j.cub.2015.04.018.

10. Lamichhaney S. et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature (2015) 518:371-5 doi:10.1038/nature14181.

Image Credits

page 9 – Centrioles - NHGRI Digital Media Database, Darryl Leja http://www.genome.gov

page 10 – Pseudomonas fluorescens – acrylic by Catehulhu http://www.deviantart.com/art/Pseudomonas-Fluorescens-174966485

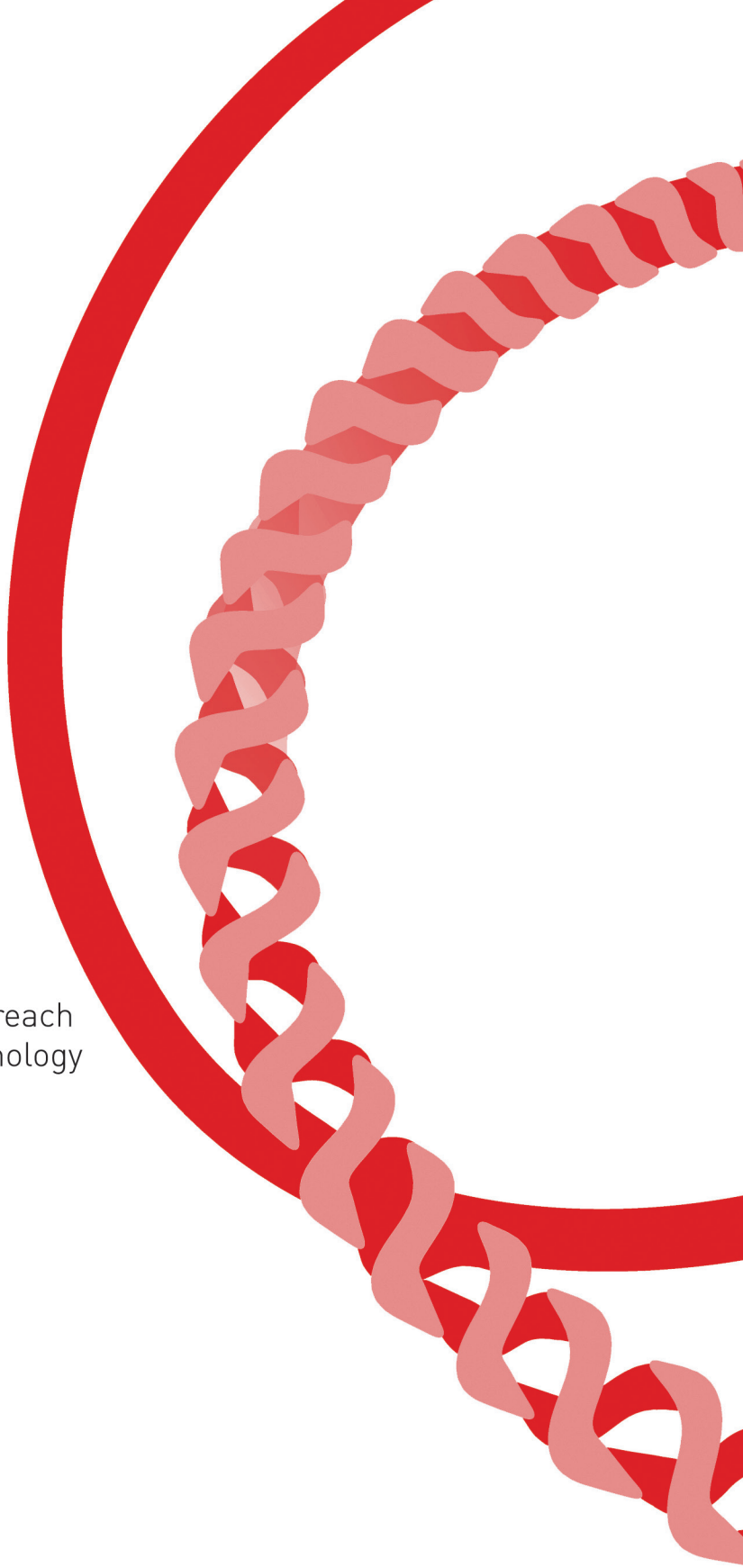
page 11 – Plankton mix from the Pacific Ocean – Copyright Christian Sardet/CNRS/Tara Expeditions. Used with permission.

page 15 Two Spotted Octopus - By Jeremyse at en.wikipedia [Public domain], via Wikimedia Commons

page 16 Double Helix with Data – NHGRI Digital Media Database, Jonathan Bailey http://www.genome.gov

page 16 Figure: The relationship between the number of stem cell divisions in the lifetime of a given tissue and the lifetime risk of cancer in that tissue. Used with permission from Science magazine.

page 19 Scribe buying parchment. - Hamburg Bible, 1255 (Copenhagen, Royal Library. MS 4, 2o) f. 183v.



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