Genetics Select

Cell

The complete human genome sequence presents researchers with unique opportunities not only for identifying the mechanisms underlying a battery of genetic diseases but also for revealing what makes us human. This issue's Genetics Select highlights recent developments in the field of human genetics that promise to direct future studies in the fields of comparative genomics and human disease research.

Humans, Your Genes Are Numbered

Genomics-scale analyses are now used to study evolution, comparative genetics, and human disease. As such analyses become more common, proper annotation of protein-coding genes will be essential, and a tough challenge is making this determination on a genome-wide basis. It is not feasible to tackle this problem experimentally gene by gene, and assigning labels based on sequence conservation is difficult. In a new study, Clamp et al. (2007) address this challenge by establishing criteria to define protein-coding genes in the human genome, and they develop a computational protocol that enables them to rapidly classify human genes according to their genomic characteristics. Current catalogs consider the presence of open reading frames (ORFs) to be indicative of a protein-coding gene, even if the sequence is not conserved in other mammals. This mode of thinking predicts that human evolution necessitated the advent of new genes. Through a new method of genome analysis, Clamp et al. reject this practice, proposing that nonconserved, or orphan, ORFs might arise randomly during genomic evolution. Indeed, their analysis suggests that there might be only 168 human-specific protein-coding genes in the human genome. This new methodology trims the calculated number of human protein-coding genes from ~24,500 to ~20,500 and establishes criteria for future evaluation of putative human protein-coding genes. The authors suggest that areas of the genome that have not been adequately annotated, especially those regions encoding short polypeptides or containing noncoding genes, where marked changes might be more readily observed, are ripe areas for future study. Overall, the new work indicates surprisingly that relatively little genomic innovation might have occurred during human evolution. Importantly, the genomics algorithm used in this study should be readily applicable to other sequenced genomes.

M. Clamp et al. (2007). Proc. Natl. Acad. Sci USA ***104****, 19428–19433.*

Splicing Creates a Species Distinction

Photograph of Amos, a male chimpanzee living at the Yerkes National Primate Research Center Field Station. Photograph courtesy of Dr. Lisa Parr.

The differences between humans and chimpanzees, although visually striking, are relatively small on a genomic scale. Although chimpanzees are our closest relatives, we display many species-specific morphological, immunological, and behavioral differences. How are these differences established? Certainly some genes are expressed differentially, but in most cases, the same protein products are produced, leading to only slight alterations in cellular signaling events. Conversely, alternative splicing provides a framework for marked diversification in protein expression profiles. To determine the extent of alternative splicing differences between humans and chimps, Calarco et al. (2007) used a strategy that coupled comparative genomics with microarray profiling. They found that ~6%–8% of orthologous exons display substantial differences in alternative splicing; these changes showed little overlap with differentially expressed genes, suggesting that different layers of regulation exist to modulate protein expression between the two species. Splicing changes appeared to arise from nucleotide substitutions in *cis*-acting splicing elements as well as from expression changes in specific splicing factors; indeed, the splicing factor SRp40 is itself differentially spliced in the two species. Interestingly, these changes in alternative splicing affect a diverse set of proteins whose biological functions range from overall gene regulation to the immune response. Splicing changes were also detected in proteins whose dysfunction has been implicated in human disease, for example, splicing changes in the glutathione transferase Gsto2 might contribute to Alzheimer's disease etiology. These preliminary results suggest that in examining alternative splicing differences between human and nonhuman primates (and also between humans and mice, the most frequently used genetic model organism), we might gain much-needed insight into human-specific molecular events, both normal and disease-related.

J.A. Calarco et al. (2007). Genes Dev. ***21****, 2963–2975.*

Evolution in Fits and Spurts

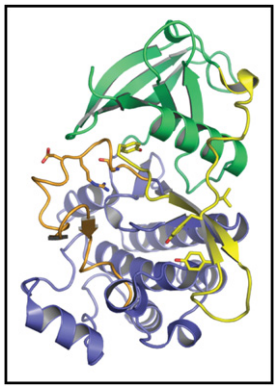
Although most human genes are highly conserved, human genomic architecture is more complex than in most other species, including most nonhuman primates. The human genome harbors many areas indicative of recent duplication events. These duplications, often large highly similar stretches of DNA, are interspersed throughout the genome. They can underlie numerous genomic alterations leading to polymorphisms and copy number variation as well as genetic innovation. Why do these differences exist between human and nonhuman primates, and how do they influence human genetics? Jiang et al. (2007) addressed these questions by analyzing genome-wide human segmental duplications. They note that whereas some chromosomes contain relatively few structural differences compared to other mammalian species, others display extensive intrachromosomal (and less frequently interchromosomal) duplications. Their results suggest that such duplications might have served as "seeding"

events for further genomic rearrangements that triggered important evolutionary events. Indeed, the majority of such duplications occurred following the split of humans and chimpanzees from the macaque lineage. A subset of these changes, focal points for genomic restructuring, are termed core duplicons; these core duplicons contain a greater than expected number of genes and expressed sequence tags (ESTs). Strikingly, the cores contain rapidly evolving and highly divergent genes and gene families; several show evidence of significant selective pressure. The importance and the functions of these genetic loci remain incompletely understood. Future studies, however, should reveal further clues about the unique nature of the human genome and the influence that these loci had on specific evolutionary paths. The new work indicates that extensive and nonrandom structural changes underlie human-specific genomic innovation, including the expansion of new gene families. These same regions are creating genomic instability that predisposes to autism, mental retardation, and developmental delay in humans. Although these changes account for a relatively small portion of the human genome, they might have driven tremendous evolutionary change.
Z. Jiang et al. (2007). Nat. Genet. ***39****, 1361–1368.*

Measuring Meiotic Crises

Although alterations in protein expression are contributing factors for many human diseases, several diseases also result from meiotic recombination defects that generate structural changes, duplications, deletions, or inversions in the genome. A subset of meiotic errors arises from nonallelic homologous recombination (NAHR) events that drive copy number variation of dosage-sensitive genes. Several NAHR loci are associated with two diseases: one resulting from a duplication and the other from the reciprocal deletion event, e.g., Charcot-Marie-Tooth disease type 1A (CMT1A, duplication) and hereditary neuropathy with liability to pressure palsies (HNPP, deletion) at 17p11.2-12. The frequency of such events is inherently difficult to measure, and in some cases only pathogenic deletion events have been identified. The reason for this discrepancy is currently unknown: perhaps some duplication events might cause very mild defects and therefore never be clinically ascertained; conversely, some duplication events might cause such severe defects that they are never observed; or duplications might occur less frequently than their seemingly reciprocal deletion events. However, healthy individuals harbor an approximately equivalent proportion of nonpathogenic duplication and deletion events. To address this question, Turner et al. (2007) developed a break point-specific PCR-based assay that can accurately detect and discern NAHR duplication and deletion events. Using pooled human sperm samples, they analyzed the meiotic error rate at four NAHR disease hotspots. At each breakpoint, duplications and deletions were identified; surprisingly, in each case, deletions occurred at a higher rate than duplications. Moreover, the different hotspots display significant differences in recombination frequency, suggesting that the underlying genetic architecture directs disease etiology. Why should pathogenic deletions arise more frequently? This question remains open. Nevertheless, this study suggests that although deletions and duplications can be reciprocal events, the rates of such changes need not be equivalent. Overall, these findings indicate that we have much to learn about understanding, detecting, and diagnosing diseases with known or suspected NAHR involvement. Encouragingly, this assay seems well suited for the clinic and so could be valuable for genetic counseling.
D.J. Turner et al. (2007). Nat. Genet. ***40****, 90–95.*

Taking Charge on the Genetics Superhighway



Ribbon structure of FLT3 showing the localization of various driver and passenger mutations in the context of the activation loop, P loop, N lobe, and JM domain. Image courtesy of Dr. Gary Gilliand.

At the time of clinical presentation, the tumors of human cancer patients contain numerous mutations, often numbering in the tens of thousands. Some mutations cause aberrations that promote cancer initiation or progression, so-called driver mutations. However, recent data indicate that some clonal mutations in tumors lack apparent function. These changes have been termed passenger mutations, and if improperly categorized, they can create numerous complications for clinicians and cancer researchers. Although whole-genome sequencing can identify the landscape of genetic changes that occur within a tumor, the challenge lies in identifying driver mutations that are the targets for therapeutic intervention. Acute myelogenous leukemia (AML) patients often harbor mutations in the tyrosine kinase growth factor receptor *FLT3*. Although many mutations are known to promote promiscuous kinase activity, in other cases the functional consequences of the mutations remain unknown. Fröhling et al. (2007) sequenced the *FLT3* gene in leukemic cells derived from 222 AML patients and characterized the identified mutations using a variety of cellular and biochemical assays. Through their screen, they identified four new candidate driver mutations whose activity could promote leukemogenesis, along with five passenger mutations that had no effect on FLT3 function. These results caution that although a specific gene is implicated in cancer, it can harbor mutations that do not influence cancer progression. More importantly, high-throughput DNA sequencing promises to be a powerful tool to identify relevant disease genes, and it seems likely that its use will drive further research into a variety of diseases for which causative mutations remain elusive. As an increasing number of studies identify risk alleles, this technique will be valuable in locating drivers while dismissing passengers and in directing the next steps in disease research.
S. Fröhling et al. (2007). Cancer Cell ***12****, 501–513.*

Sara B. Cullinan