Science Abstracts

American Journal of Human Genetics

J. J. Johnston et al., **Secondary Variants in Individuals Undergoing Exome Sequencing: Screening of 572 Individuals Identifies High-Penetrance Mutations in Cancer-Susceptibility Genes**, *Am J Hum Genet* 91.1 (July 13, 2012): 97–108 • Genome- and exome-sequencing costs are continuing to fall, and many individuals are undergoing these assessments as research participants and patients. The issue of secondary (so-called incidental) findings in exome analysis is controversial, and data are needed on methods of detection and their frequency. We piloted secondary variant detection by analyzing exomes for mutations in cancer-susceptibility syndromes in subjects ascertained for atherosclerosis phenotypes. We performed exome sequencing on 572 ClinSeq participants, and in 37 genes, we interpreted variants that cause high-penetrance cancer syndromes by using an algorithm that filtered results on the basis of mutation type, quality, and frequency and that filtered mutation-database entries on the basis of defined categories of causation. We identified 454 sequence variants that differed from the human reference. Exclusions were made on the basis of sequence quality (26 variants) and high frequency in the cohort (77 variants) or dbSNP (17 variants), leaving 334 variants of potential clinical importance. These were further filtered on the basis of curation of literature reports. Seven participants, four of whom were of Ashkenazi Jewish descent and three of whom did not meet family-history-based referral criteria, had deleterious *BRCA1* or *BRCA2* mutations. One participant had a deleterious *SDHC* mutation, which causes paragangliomas. Exome sequencing, coupled with multidisciplinary interpretation, detected clinically important mutations in cancer-susceptibility genes; four of such mutations were in individuals without a significant family history of disease. We conclude that secondary variants of high clinical importance will be detected at an appreciable frequency in exomes, and we suggest that priority be given to the development of more efficient modes of interpretation with trials in larger patient groups.

Y. Xue et al., **Deleterious- and Disease-Allele Prevalence in Healthy Individuals: Insights from Current Predictions, Mutation Databases, and Population-Scale Resequencing**, *Am J Hum Genet* 91.6 (December 7, 2012): 1022–1032 • We have assessed the numbers of potentially deleterious variants in the genomes of apparently healthy humans by using (1) low-coverage whole-genome sequence data from 179 individuals in the 1000 Genomes Pilot Project and (2) current predictions and databases of deleterious variants. Each individual carried 281-515 missense substitutions, 40-85 of which were homozygous, predicted to be highly damaging. They also carried 40-110 variants classified by the Human Gene Mutation Database (HGMD) as disease-causing mutations (DMs), 3-24 variants in the homozygous state, and many polymorphisms putatively associated with disease. Whereas many of these DMs are likely to represent disease-allele-annotation errors, between 0 and 8 DMs (0-1 homozygous) per individual are predicted to be highly damaging, and some of them provide information of medical relevance. These analyses emphasize the need for improved annotation of disease alleles both in mutation databases and in the primary literature; some HGMD mutation data have been recategorized on the basis of the present findings, an iterative process that is both necessary and ongoing. Our estimates of deleterious-allele numbers

are likely to be subject to both overcounting and undercounting. However, our current best mean estimates of ~400 damaging variants and ~2 bona fide disease mutations per individual are likely to increase rather than decrease as sequencing studies ascertain rare variants more effectively and as additional disease alleles are discovered.

Cell

Y. G. Kamberov et al., **Modeling Recent Human Evolution in Mice by Expression of a Selected EDAR Variant**, *Cell* 152.4 (February 14, 2013): 691–702 • An adaptive variant of the human Ectodysplasin receptor, EDARV370A, is one of the strongest candidates of recent positive selection from genome-wide scans. We have modeled EDAR370A in mice and characterized its phenotype and evolutionary origins in humans. Our computational analysis suggests the allele arose in central China approximately 30,000 years ago. Although EDAR370A has been associated with increased scalp hair thickness and changed tooth morphology in humans, its direct biological significance and potential adaptive role remain unclear. We generated a knockin mouse model and find that, as in humans, hair thickness is increased in EDAR370A mice. We identify new biological targets affected by the mutation, including mammary and eccrine glands. Building on these results, we find that EDAR370A is associated with an increased number of active eccrine glands in the Han Chinese. This interdisciplinary approach yields unique insight into the generation of adaptive variation among modern humans.

A. Soufi et al., **Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome**, *Cell* 151.5 (November 21, 2012): 994–1004 • The ectopic expression of transcription factors can reprogram cell fate, yet it is unknown how the initial binding of factors to the genome relates functionally to the binding seen in the minority of cells that become reprogrammed. We report a map of Oct4, Sox2, Klf4, and c-Myc (O, S, K, and M) on the human genome during the first 48 hr of reprogramming fibroblasts to pluripotency. Three striking aspects of the initial chromatin binding events include an unexpected role for c-Myc in facilitating OSK chromatin engagement, the primacy of O, S, and K as pioneer factors at enhancers of genes that promote reprogramming, and megabase-scale chromatin domains spanned by H3K9me3, including many genes required for pluripotency, that prevent initial OSKM binding and impede the efficiency of reprogramming. We find diverse aspects of initial factor binding that must be overcome in the minority of cells that become reprogrammed.

Cell Stem Cell

S.M. Buckley et al., **Regulation of Pluripotency and Cellular Reprogramming by the Ubiquitin-Proteasome System**, *Cell Stem Cell* 11.6 (December 7, 2012): 783–798 • Although transcriptional regulation of stem cell pluripotency and differentiation has been extensively studied, only a small number of studies have addressed the roles for posttranslational modifications in these processes. A key mechanism of posttranslational modification is ubiquitination by the ubiquitin-proteasome system (UPS). Here, using shotgun proteomics, we map the ubiquitinated protein landscape during embryonic stem cell (ESC) differentiation and induced pluripotency. Moreover, using UPS-targeted RNAi screens, we identify additional regulators of pluripotency and differentiation. We focus on two of these proteins, the deubiquitinating enzyme Psmd14 and the E3 ligase Fbxw7, and characterize their importance in ESC pluripotency and cellular reprogramming. This global characterization of the UPS as a key regulator of stem cell pluripotency opens the way for future studies that focus on specific UPS enzymes or ubiquitinated substrates.

A. Golipour et al., **A Late Transition in Somatic Cell Reprogramming Requires Regulators Distinct from the Pluripotency Network**, *Cell Stem Cell* 11.6 (December 7, 2012): 769–782 • Reprogramming of somatic cells to a pluripotent state via expression of Oct4, Klf4, Myc, and Sox2

is a multistep process involving phased changes in gene expression. Here, we focus on the later stages of reprogramming, termed maturation and stabilization. We show that the stabilization phase and the acquisition of pluripotency are dependent on the removal of transgene expression late in the maturation phase. Clonal analysis of cells undergoing reprogramming revealed subsets of stabilization-competent (SC) and stabilization-incompetent (SI) cells. SC clones acquire a competency gene-expression signature late in the maturation phase. Functional analysis of SC signature genes identified enhancers of the transition to the stabilization phase and a distinct subset of genes required for the maintenance of pluripotency. Thus, the acquisition and maintenance of pluripotency are regulated by distinct molecular networks, and a specific regulatory program not previously implicated in reprogramming is required for the transition to transgene independence.

E. Haimes et al., **Position Statement on the Provision and Procurement of Human Eggs for Stem Cell Research**, *Cell Stem Cell* 12.3 (March 7, 2013): 285–291 • The nature of compensation for women who donate eggs (oocytes) for research remains a contentious issue internationally. This position paper lays out the arguments for, and discusses the arrangements in which, a modest payment might be ethically justifiable.

L. B. Li et al., **Trisomy Correction in Down Syndrome Induced Pluripotent Stem Cells**, *Cell Stem Cell* 11.5 (November 2, 2012): 615–619 • Human trisomies can alter cellular phenotypes and produce congenital abnormalities such as Down syndrome (DS). Here we have generated induced pluripotent stem cells (iPSCs) from DS fibroblasts and introduced a TKNEO transgene into one copy of chromosome 21 by gene targeting. When selecting against TKNEO, spontaneous chromosome loss was the most common cause for survival, with a frequency of ~10(-4), while point mutations, epigenetic silencing, and TKNEO deletions occurred at lower frequencies in this unbiased comparison of inactivating mutations. Mitotic recombination events resulting in extended loss of heterozygosity were not observed in DS iPSCs. The derived, disomic cells proliferated faster and produced more endothelia in vivo than their otherwise isogenic trisomic counterparts, but in vitro hematopoietic differentiation was not consistently altered. Our study describes a targeted removal of a human trisomy, which could prove useful in both clinical and research applications.

S. Wakayama et al., **Successful Serial Recloning in the Mouse over Multiple Generations**, *Cell Stem Cell* 12.3 (March 7, 2013): 293–297 · Previous studies of serial cloning in animals showed a decrease in efficiency over repeated iterations and a failure in all species after a few generations. This limitation led to the suggestion that repeated recloning might be inherently impossible because of the accumulation of lethal genetic or epigenetic abnormalities. However, we have now succeeded in carrying out repeated recloning in the mouse through a somatic cell nuclear transfer method that includes a histone deacetylase inhibitor. The cloning efficiency did not decrease over 25 generations, and, to date, we have obtained more than 500 viable offspring from a single original donor mouse. The reprogramming efficiency also did not increase over repeated rounds of nuclear transfer, and we did not see the accumulation of reprogramming errors or clone-specific abnormalities. Therefore, our results show that repeated iterative recloning is possible and suggest that, with adequately efficient techniques, it may be possible to reclone animals indefinitely.

Develomental Cell

R. Freitas et al., **Hoxd13 Contribution to the Evolution of Vertebrate Appendages**, *Dev Cell* 23.6 (December 11, 2012): 1219–1229 • Fossil data suggest that limbs evolved from fish fins by sequential elaboration of their distal endoskeleton, giving rise to the autopod close to the tetrapod origin. This elaboration may have occurred by a simultaneous reduction of the distal ectodermal fold of fish fins. Modulation of 5'Hoxd gene transcription, through tetrapod-specific digit enhancers, has been suggested

as a possible evolutionary mechanism involved in these morphological transformations. Here, we overexpress hoxd13a in zebrafish to investigate the impact of increasing 5'Hoxd expression during fin development. This overexpression causes increased proliferation, distal expansion of chondrogenic tissue and finfold reduction. In addition, we also show that the tetrapod-specific 5'Hoxd enhancer CsC promotes similar expression in zebrafish fins and mouse limbs. Our results support the idea that modulation of 5'Hoxd gene expression, by acquisition of novel enhancer elements, offered the substrate for the evolution of fins and the origin of tetrapod limbs.

G. Lin et al., **Imparting Regenerative Capacity to Limbs by Progenitor Cell Transplantation**, *Dev Cell* 24.1 (January 14, 2013): 41–51 • The frog Xenopus can normally regenerate its limbs at early developmental stages but loses the ability during metamorphosis. This behavior provides a potential gain-of-function model for measures that can enhance limb regeneration. Here, we show that frog limbs can be caused to form multidigit regenerates after receiving transplants of larval limb progenitor cells. It is necessary to activate Wnt/β-catenin signaling in the cells and to add Sonic hedgehog, FGF10, and thymosin β4. These factors promote survival and growth of the grafted cells and also provide pattern information. The eventual regenerates are not composed solely of donor tissue; the host cells also make a substantial contribution despite their lack of regeneration competence. Cells from adult frog legs or from regenerating tadpole tails do not promote limb regeneration, demonstrating the necessity for limb progenitor cells. These findings have obvious implications for the development of a technology to promote limb regeneration in mammals.

Nature

1000 Genomes Project Consortium et al., **An Integrated Map of Genetic Variation from 1,092 Human Genomes**, *Nature* 491.7422 (November 1, 2102): 56–65 • By characterizing the geographic and functional spectrum of human genetic variation, the 1000 Genomes Project aims to build a resource to help to understand the genetic contribution to disease. Here we describe the genomes of 1,092 individuals from 14 populations, constructed using a combination of low-coverage whole-genome and exome sequencing. By developing methods to integrate information across several algorithms and diverse data sources, we provide a validated haplotype map of 38 million single nucleotide polymorphisms, 1.4 million short insertions and deletions, and more than 14,000 larger deletions. We show that individuals from different populations carry different profiles of rare and common variants, and that low-frequency variants show substantial geographic differentiation, which is further increased by the action of purifying selection. We show that evolutionary conservation and coding consequence are key determinants of the strength of purifying selection, that rare-variant load varies substantially across biological pathways, and that each individual contains hundreds of rare non-coding variants at conserved sites, such as motif-disrupting changes in transcription-factor-binding sites. This resource, which captures up to 98% of accessible single nucleotide polymorphisms at a frequency of 1% in related populations, enables analysis of common and low-frequency variants in individuals from diverse, including admixed, populations.

A. Scally et al., **Insights into Hominid Evolution from the Gorilla Genome Sequence**, *Nature* 483.7388 (March 7, 2012): 169–175 • Gorillas are humans' closest living relatives after chimpanzees, and are of comparable importance for the study of human origins and evolution. Here we present the assembly and analysis of a genome sequence for the western lowland gorilla, and compare the whole genomes of all extant great ape genera. We propose a synthesis of genetic and fossil evidence consistent with placing the human-chimpanzee and human-chimpanzee-gorilla speciation events at approximately 6 and 10 million years ago. In 30% of the genome, gorilla is closer to human or chimpanzee than the latter are to each other; this is rarer around

coding genes, indicating pervasive selection throughout great ape evolution, and has functional consequences in gene expression. A comparison of protein coding genes reveals approximately 500 genes showing accelerated evolution on each of the gorilla, human and chimpanzee lineages, and evidence for parallel acceleration, particularly of genes involved in hearing. We also compare the western and eastern gorilla species, estimating an average sequence divergence time 1.75 million years ago, but with evidence for more recent genetic exchange and a population bottleneck in the eastern species. The use of the genome sequence in these and future analyses will promote a deeper understanding of great ape biology and evolution.

PNAS

S. Achard et al., **Hubs of Brain Functional Networks Are Radically Reorganized in Comatose Patients**, *PNAS* 109.50 (December 11, 2012): 20608–20613 • Human brain networks have topological properties in common with many other complex systems, prompting the following question: what aspects of brain network organization are critical for distinctive functional properties of the brain, such as consciousness? To address this question, we used graph theoretical methods to explore brain network topology in resting state functional MRI data acquired from 17 patients with severely impaired consciousness and 20 healthy volunteers. We found that many global network properties were conserved in comatose patients. Specifically, there was no significant abnormality of global efficiency, clustering, small-worldness, modularity, or degree distribution in the patient group. However, in every patient, we found evidence for a radical reorganization of high degree or highly efficient "hub" nodes. Cortical regions that were hubs of healthy brain networks had typically become nonhubs of comatose brain networks and vice versa. These results indicate that global topological properties of complex brain networks may be homeostatically conserved under extremely different clinical conditions and that consciousness likely depends on the anatomical location of hub nodes in human brain networks.

N. Arinaminpathy et al., **Self-Boosting Vaccines and Their Implications for Herd Immunity**, *PNAS* 109.49 (December 4, 2012): 20154–20159 • Advances in vaccine technology over the past two centuries have facilitated far-reaching impact in the control of many infections, and today's emerging vaccines could likewise open new opportunities in the control of several diseases. Here we consider the potential, population-level effects of a particular class of emerging vaccines that use specific viral vectors to establish long-term, intermittent antigen presentation within a vaccinated host: in essence, "self-boosting" vaccines. In particular, we use mathematical models to explore the potential role of such vaccines in situations where current immunization raises only relatively short-lived protection. Vaccination programs in such cases are generally limited in their ability to raise lasting herd immunity. Moreover, in certain cases mass vaccination can have the counterproductive effect of allowing an increase in severe disease, through reducing opportunities for immunity to be boosted through natural exposure to infection. Such dynamics have been proposed, for example, in relation to pertussis and varicella-zoster virus. In this context we show how self-boosting vaccines could open qualitatively new opportunities, for example by broadening the effective duration of herd immunity that can be achieved with currently used immunogens. At intermediate rates of self-boosting, these vaccines also alleviate the potential counterproductive effects of mass vaccination, through compensating for losses in natural boosting. Importantly, however, we also show how sufficiently high boosting rates may introduce a new regime of unintended consequences, wherein the unvaccinated bear an increased disease burden. Finally, we discuss important caveats and data needs arising from this work.

E. Castle et al., **Neural and Behavioral Bases of Age Differences in Perceptions of Trust**, *PNAS* 109.51 (December 18, 2012): 20848–20852 • Older adults are

disproportionately vulnerable to fraud, and federal agencies have speculated that excessive trust explains their greater vulnerability. Two studies, one behavioral and one using neuroimaging methodology, identified age differences in trust and their neural underpinnings. Older and younger adults rated faces high in trust cues similarly, but older adults perceived faces with cues to untrustworthiness to be significantly more trustworthy and approachable than younger adults. This age-related pattern was mirrored in neural activation to cues of trustworthiness. Whereas younger adults showed greater anterior insula activation to untrustworthy versus trustworthy faces, older adults showed muted activation of the anterior insula to untrustworthy faces. The insula has been shown to support interoceptive awareness that forms the basis of "gut feelings," which represent expected risk and predict risk-avoidant behavior. Thus, a diminished "gut" response to cues of untrustworthiness may partially underlie older adults' vulnerability to fraud.

J. E. De Neve and A. J. Oswald, **Estimating the Influence of Life Satisfaction and Positive Affect on Later Income using Sibling Fixed Effects**, *PNAS* 109.49 (December 4, 2012): 19953–19958 • The question of whether there is a connection between income and psychological well-being is a long-studied issue across the social, psychological, and behavioral sciences. Much research has found that richer people tend to be happier. However, relatively little attention has been paid to whether happier individuals perform better financially in the first place. This possibility of reverse causality is arguably understudied. Using data from a large US representative panel, we show that adolescents and young adults who report higher life satisfaction or positive affect grow up to earn significantly higher levels of income later in life. We focus on earnings approximately one decade after the person's well-being is measured; we exploit the availability of sibling clusters to introduce family fixed effects; we account for the human capacity to imagine later socioeconomic outcomes and to anticipate the resulting feelings in current well-being. The study's results are robust to the inclusion of controls such as education, intelligence quotient, physical health, height, self-esteem, and later happiness. We consider how psychological well-being may influence income. Sobel-Goodman mediation tests reveal direct and indirect effects that carry the influence from happiness to income. Significant mediating pathways include a higher probability of obtaining a college degree, getting hired and promoted, having higher degrees of optimism and extraversion, and less neuroticism.

W. Li et al., **Identification of Oct4-Activating Compounds that Enhance Reprogramming Efficiency**, *PNAS* 109.51 (December 18, 2012): 20853–20858 • One of the hurdles for practical application of induced pluripotent stem cells (iPSC) is the low efficiency and slow process of reprogramming. Octamer-binding transcription factor 4 (Oct4) has been shown to be an essential regulator of embryonic stem cell (ESC) pluripotency and key to the reprogramming process. To identify small molecules that enhance reprogramming efficiency, we performed a cell-based high-throughput screening of chemical libraries. One of the compounds, termed Oct4-activating compound 1 (OAC1), was found to activate both Oct4 and Nanog promoter-driven luciferase reporter genes. Furthermore, when added to the reprogramming mixture along with the quartet reprogramming factors (Oct4, Sox2, c-Myc, and Klf4), OAC1 enhanced the iPSC reprogramming efficiency and accelerated the reprogramming process. Two structural analogs of OAC1 also activated Oct4 and Nanog promoters and enhanced iPSC formation. The iPSC colonies derived using the Oct4-activating compounds along with the quartet factors exhibited typical ESC morphology, gene-expression pattern, and developmental potential. OAC1 seems to enhance reprogramming efficiency in a unique manner, independent of either inhibition of the p53-p21 pathway or activation of the Wnt-β-catenin signaling. OAC1 increases transcription of the Oct4-Nanog-Sox2 triad and Tet1, a gene known to be involved in DNA demethylation.

A. Weiss et al., **Evidence for a Midlife Crisis in Great Apes Consistent with the U-Shape in Human Well-Being**, *PNAS* 109.49 (December 4, 2012): 19949–19952 · Recently, economists and behavioral scientists have studied the pattern of human well-being over the lifespan. In dozens of countries, and for a large range of well-being measures, including happiness and mental health, well-being is high in youth, falls to a nadir in midlife, and rises again in old age. The reasons for this U-shape are still unclear. Present theories emphasize sociological and economic forces. In this study we show that a similar U-shape exists in 508 great apes (two samples of chimpanzees and one sample of orangutans) whose well-being was assessed by raters familiar with the individual apes. This U-shaped pattern or "midlife crisis" emerges with or without use of parametric methods. Our results imply that human well-being's curved shape is not uniquely human and that, although it may be partly explained by aspects of human life and society, its origins may lie partly in the biology we share with great apes. These findings have implications across scientific and social-scientific disciplines, and may help to identify ways of enhancing human and ape well-being.

Science

M. Gymrek et al., **Identifying Personal Genomes by Surname Inference**, *Science* 339.6117 (January 18, 2013): 321–324 • Sharing sequencing data sets without identifiers has become a common practice in genomics. Here, we report that surnames can be recovered from personal genomes by profiling short tandem repeats on the Y chromosome (Y-STRs) and querying recreational genetic genealogy databases. We show that a combination of a surname with other types of metadata, such as age and state, can be used to triangulate the identity of the target. A key feature of this technique is that it entirely relies on free, publicly accessible Internet resources. We quantitatively analyze the probability of identification for U.S. males. We further demonstrate the feasibility of this technique by tracing back with high probability the identities of multiple participants in public sequencing projects.