The evolutionary significance of copy number variation in the human genome

G.H. Perry

Department of Human Genetics, University of Chicago, Chicago, IL (USA)

Accepted in revised form for publication by H. Kehrer-Sawatzki and D.N. Cooper, 22 July 2008.

Abstract. Copy number variation provides the raw material for gene family expansion and diversification, which is an important evolutionary force. Moreover, copy number variants (CNVs) can influence gene transcriptional and translational levels and have been associated with complex disease susceptibility. Therefore, natural selection may have affected at least some of the greater than one thousand CNVs thus far discovered among the genomes of phenotypically normal humans. While identifying and understanding particular instances of natural selection may shed

light on important aspects of human evolutionary history, our ability to analyze CNVs in traditional population genetic frameworks has been limited. However, progress has been made by adapting some of these frameworks for use with copy number data. Moving forward, these efforts will be aided by non-human organism studies of the population genetics of copy number variation, and by more direct comparisons of within-species copy number variation and between-species copy number fixation.

Copyright © 2009 S. Karger AG, Basel

Human genetic diversity is comprised of single nucleotide polymorphisms (SNPs), small insertion and deletion variants, short tandem repeat polymorphisms, retrotransposable element insertion variants (e.g., *Alus*), inversion variants, and copy number variants (CNVs). CNVs are larger-scale insertions and deletions that range from several kilobases (kb) to several megabases (Mb) in size (Feuk et al., 2006). While this component of genetic diversity has long been considered important by the scientific community (e.g., Ottolenghi et al., 1974; Awdeh and Alper, 1980; Trask et al., 1998; Buckland, 2003), recent advances in microarray and other genome-scale technologies have facilitated the discovery of more than 1,000 human CNVs (Redon et al., 2006; Sebat, 2007), leading to intensified interest and excitement. This review focuses on the potential evolutionary significance of copy number variation in the human genome, including results from previous studies and potential directions for future research.

The functional and evolutionary potential of copy number variation

Gene-containing CNVs may influence mRNA and protein expression levels (e.g., Aldred et al., 2005; Gonzalez et al., 2005; McCarroll et al., 2006; Stranger et al., 2007). Therefore, CNVs have the potential to affect downstream phenotypes and, ultimately, reproductive fitness (Kondrashov and Kondrashov, 2006; Hurles et al., 2008). However, one cannot simply assume a direct relationship between gene copy number and expression level. In part, this uncertainty may be attributable to the location of a gene-containing CNV with respect to that of the gene regulatory machinery (Cooper et al., 2007). For example, each duplicated segment of the starch-digesting amylase gene *AMY1* contains the regulatory sequences necessary for salivary-specific expression (Groot et al., 1990; Ting et al., 1992), and there is a significant positive correlation between *AMY1* copy number and amylase protein levels in saliva (Bank et al., 1992; Perry et al., 2007). In contrast, there is a single regulatory 'locus control region' upstream of the red (*OPN1LW*) and green (*OPN1MW*) opsin visual pigment genes on the X chromosome (Wang et al., 1992). Although mutations in the copy-number variable *OPN1MW* gene may result in color blindness (Nathans et al., 1986a, b; Wolf et al., 1999), only the copy nearest the locus control region is expressed to an ap-

G.H.P. is supported by NIH fellowship F32GM085998.

Request reprints from George H. Perry
Department of Human Genetics, University of Chicago
Chicago, IL 60637 (USA)
telephone: +1 773 834 1984; fax: +1 773 834 8470
e-mail: gerry@uchicago.edu

preciable extent, such that a male with a disrupted first gene but intact subsequent genes will have color blindness (Hayashi et al., 1999).

CNVs are reportedly associated with susceptibility to systemic autoimmunity diseases (Fanciulli et al., 2007; Yang et al., 2007), psoriasis (Hollox et al., 2008), and HIV infection and progression to AIDS (Gonzalez et al., 2005) among other complex diseases (Fellermann et al., 2006; Le Marechal et al., 2006). While these results provide further evidence of the potential functional relevance of CNVs in general, disease can be a powerful evolutionary force in its own right and could therefore affect patterns of CNV diversity via natural selection. For example, deletions of the hemoglobin genes *HBA1*, *HBA2*, *HBB*, or *HBD* result in thalassemia (Ottolenghi et al., 1974, 1976; Taylor et al., 1974; Orkin et al., 1979). Although homozygous deletion is typically fatal (Weatherall and Clegg, 1981), individuals heterozygous for these deletions receive protection against malaria infection, and thalassemia frequency is strongly correlated with malaria prevalence, even down to very local levels (Flint et al., 1986; Hill et al., 1988; Allen et al., 1997). This is a classic example of balancing selection in humans.

Higher copy numbers of the immunoregulatory and inflammatory cytokine *CCL3L1* gene are associated with lower risks of HIV infection and the progression to AIDS (Gonzalez et al., 2005). Interestingly, average *CCL3L1* copy number in Africans is nearly two times greater than in non-Africans: 5.95 versus 2.99 copies, respectively (Gonzalez et al., 2005). In a subsequent genome-wide study, the level of population differentiation at this locus was found to be extraordinary compared to that of other CNVs (Redon et al., 2006), suggesting that natural selection may have influenced *CCL3L1* copy number in humans. Because AIDS has only recently been a human disease, it is unlikely to have driven patterns of *CCL3L1* copy number in our species. However, other diseases for which susceptibility may also be associated with *CCL3L1* copy number may have had such an effect. Therefore, we stand to benefit from experiments that interrogate the detailed functional effects of different *CCL3L1* copy number genotypes (e.g., Dolan et al., 2007), which may lead to further medical and evolutionary insights. For example, Mamtani et al. (2008) recently reported an association between *CCL3L1* copy number and susceptibility to systemic lupus erythematosus, providing evidence that this CNV may affect diverse, multi-systemic, pathways that could have been subject to dynamic evolutionary pressures during human evolution.

Population genetic analyses of copy number variation

Due to current technological limitations in CNV ascertainment, and diversity in and uncertainty over CNV architectures, we face considerable challenges in obtaining reliable genotypes for CNVs and using traditional population genetic analyses to understand their evolutionary significance (Conrad and Hurler, 2007; Kidd et al., 2007; McCarroll and Altshuler, 2007; Perry et al., 2008a). Despite these

challenges, there have been some successful modifications of population genetic frameworks for use with CNV data. For example, rather than considering allele frequencies, Redon et al. (2006) analyzed directly the relative intensity log2 ratios for clones from their whole-genome array-based comparative genomic hybridization platform to highlight CNVs with relatively high levels of between-population differentiation. These CNVs, which include the *CCL3L1* CNV discussed above, are excellent candidates for further functional and evolutionary analyses.

In another study, we discovered that mean *AMY1* copy number is higher in populations with high-starch diets compared to populations with traditionally low-starch diets (Perry et al., 2007). In a subset of these populations, the level of differentiation at the *AMY1* locus is unusual compared to that for other genome-wide CNVs, suggesting that positive or directional selection may have favored higher *AMY1* copy numbers in at least some high-starch populations (Perry et al., 2007). Combined with findings from previous population genetic analyses of alleles responsible for lactase persistence (Bersaglieri et al., 2004; Tishkoff et al., 2007), this result demonstrates the importance of diet – and particularly the transition to agriculture – in human evolution.

Nozawa and colleagues (2007) compared gene- and pseudogene-containing CNVs to examine the evolutionary significance of olfactory receptor gene copy number variation. Previous studies have consistently shown that human CNVs are significantly enriched for genes with sensory perception (including olfactory receptors) and defense response functions (e.g. Cooper et al., 2007). While this enrichment could be interpreted as evidence of positive selection for variation at the copy number level of genes (Nguyen et al., 2006), it is also consistent with relatively stronger functional constraint on copy numbers of genes with other functions. Using CNV data from Redon et al. (2006), Nozawa et al. (2007) compared the proportion of functional olfactory receptor genes that are copy number variable to that for olfactory receptor pseudogenes, which are expected to reflect neutral patterns of diversity. A similar number of genes and pseudogenes were copy number variable, which is consistent with neutral evolution (e.g., genetic drift) on the copy numbers of functional olfactory receptors (Nozawa et al., 2007).

It will be interesting to revisit the Nozawa et al. (2007) olfactory receptor analysis once advances in CNV technologies facilitate improved breakpoint resolution and more accurate genotype estimates, for increased certainty of the specific copy-number-variable genes and to be able to consider the full frequency distributions, respectively. In addition, comparing the human results to those from other species in which olfactory receptors may have been subject to different evolutionary pressures including rodents, canines, and even other primates (Gilad et al., 2003, 2004) will be particularly informative. Finally, we may be enlightened by the results of such comparisons among human populations with different ecological histories.

In general, there are not large samples of copy-number-variable pseudogenes in the human genome for non-olfactory functional categories (Redon et al., 2006), which may

preclude widespread application of the Nozawa et al. (2007) test. As an alternative neutral proxy, one could analyze a set of intergenic regions carefully matched (e.g., for repetitive element densities and recombination rates) to the functional genes of interest. Of course, not all intergenic region CNVs are likely to be impervious to natural selection; for example, Stranger et al. (2007) identified six CNVs that were significantly correlated with the mRNA expression levels of genes >1 Mb distant. However, intergenic CNVs are still likely to better reflect neutrality than gene-containing CNVs, and thus would provide a suitable database for initial comparisons in a population genetics framework.

Patterns of copy number variation in non-human species

Widespread copy number variation has now been described in the genomes of chimpanzees, rhesus macaques, mice, rats, the fruitfly *Drosophila melanogaster*, and even the malaria parasite *Plasmodium falciparum* (Li et al., 2004; Perry et al., 2006; Cutler et al., 2007; Dopman and Hartl, 2007; Egan et al., 2007; Graubert et al., 2007; Anderson et al., 2008; Emerson et al., 2008; Guryev et al., 2008; Lee et al., 2008; Mok et al., 2008; She et al., 2008). Characterizing CNVs in non-human genomes not only helps us to understand better the evolutionary histories of these species (e.g., Nair et al., 2007), but also will enhance our knowledge of the functional and evolutionary significance of human CNVs.

Specifically, CNVs occur in orthologous regions of different primate genomes considerably more often than would be expected by chance, likely a result of shared genomic architectures that facilitate recurrent CNV genesis (Perry et al., 2006; Lee et al., 2008). Even in rats, 113 CNVs were discovered that occur in regions orthologous to human CNVs (Guryev et al., 2008). Especially in model organisms, these loci represent excellent opportunities to examine the potential functional significance of human CNVs. Moreover, between-species comparisons of the detailed phenotypic effects of orthologous-region CNVs will contribute to our understanding of the functional importance of CNV fine-scale architecture (e.g. specific breakpoints) and genetic background (e.g. nucleotide sequence variation).

Analyses of the patterns of copy number variation in non-human genomes are also expected to aid in the development of CNV-tailored population genetic analyses. In this respect, the relatively high level of neutral genetic diversity in *Drosophila* (Aquadro et al., 2001) makes this model organism an ideal candidate for CNV evolutionary analyses. In an initial study comparing the *Drosophila melanogaster* reference sequence strain to five wild-type strains, Dopman and Hartl (2007) identified an average of 436 CNVs per strain. These CNVs were then analyzed in the context of the detailed knowledge of functional elements in the *Drosophila* genome to test intriguing hypotheses concerning the biological and evolutionary significance of copy number variation. For example, genes with tissue-specific

rather than widespread expression are significantly more likely to be copy number variable in *Drosophila melanogaster*, and these tissue-specific genes are particularly enriched for midgut and male accessory gland expression, including genes involved in digestion, defense response, insecticide detoxification, and sperm competition (Dopman and Hartl, 2007). Detailed population genetic analyses focused on these CNVs may provide important insights into *Drosophila* evolutionary history and would contribute to our general understanding of the potential functional and evolutionary significance of copy number variation. In a more recent *Drosophila melanogaster* study, Emerson et al. (2008) identified four high-frequency duplications that contain one or more genes with toxin response/insecticide detoxification functions. These particular CNVs may have been affected by positive selection and are therefore excellent candidates for further interrogation.

The relationship between copy number variation and fixation

Inter-specific copy number differences (CNDs) are common among primate genomes (Locke et al., 2003; Fortna et al., 2004; Newman et al., 2005; Goidts et al., 2006; Wilson et al., 2006; Dumas et al., 2007), and may have been involved in the evolution of species- and lineage-specific phenotypes (Kehrer-Sawatzki and Cooper, 2007). In fact, with respect to base pair content, CNDs may account for a greater proportion of total human-chimpanzee genome divergence than single nucleotide substitutions (Cheng et al., 2005; Chimpanzee Sequencing and Analysis Consortium, 2005). Although the relationship between CNVs and CNDs is relatively complex since segmental duplications are prone to subsequent CNV genesis via non-allelic homologous recombination mechanisms (e.g., Cooper et al., 2007), we can still advance our understanding of the evolutionary significance of both CNVs and CNDs by analyzing them in concert.

The McDonald-Kreitman test (1991) compares ratios of fixation to polymorphism for functional and putatively neutral sites. An excess of functional fixation suggests that some differences may have been fixed by positive selection. Zhang (2007) adapted this test to compare CND:CNV ratios for intact olfactory receptor genes and pseudogenes (CNDs were based on comparison of the human and chimpanzee reference genome sequences; CNV data were from Redon et al. (2006)). Although there is a slight excess of intact gene fixation (16:116 for intact genes; 11:143 for pseudogenes), the ratios were not significantly different; therefore, the null hypothesis of neutrality could not be rejected (Zhang, 2007).

Recently, we extended this framework to consider CNDs and CNVs that encompass genes with different functional categories, based on Gene Ontology classifications (Perry et al., 2008b). In our study, CNVs were identified in 30 human and 30 chimpanzee individuals, using a whole-genome array-based comparative genomic hybridization platform. To

identify fixed CNDs, we used the same platform to compare one human and one chimpanzee individual, and then filtered out the gains and losses that overlapped human or chimpanzee CNVs. We compared the CND:human CNV ratio for each functional category to that for intergenic regions. Relative to intergenic regions (18:52) and compared to the ratios for other functional categories, the cell proliferation (6:8) and inflammatory response (5:7) categories have an excess of CNDs (Perry et al., 2008b). Although these results are also not statistically significant, the cell proliferation and inflammatory response CNDs are intriguing candidates for future studies aiming to characterize the genetic basis of adaptive phenotypic differences between humans and chimpanzees.

Comparisons such as those discussed above will help us understand better the obvious relationship between CNDs and CNVs and can identify among-taxa variation in evolutionary pressures on copy number. These analyses will become more powerful with technological advances – especially with improved breakpoint estimation and more precise knowledge of the functional elements contained within

each variant. However, it will be more difficult to alleviate an ascertainment bias inherent in human and non-human primate comparisons. Specifically, we are challenged to identify reliably deletions of unique sequence that were fixed in the human lineage. For example, our study (Perry et al., 2008b) was based on a human-specific platform; these sequences (fixed human-specific deletions) would not have been represented on the microarray. One could construct a multi-species microarray platform (e.g., Gilad et al., 2005) or identify CNDs based on genome sequence comparisons (e.g., Cheng et al., 2005; Chimpanzee Sequencing and Analysis Consortium, 2005), but to reliably identify human-specific deletions, both of these approaches would require high-quality finished genome sequences for all species of interest. Once these issues are circumvented, the annotation and characterization of functional elements contained within the regions that are deleted in humans may require unconventional approaches, but these results will be particularly interesting and could provide important insights into our evolutionary history.

References

- Aldred PM, Hollox EJ, Armour JA: Copy number polymorphism and expression level variation of the human alpha-defensin genes *DEFA1* and *DEFA3*. *Hum Mol Genet* 14:2045–2052 (2005).
- Allen SJ, O'Donnell A, Alexander ND, Alpers MP, Peto TE, et al: alpha+-Thalassemia protects children against disease caused by other infections as well as malaria. *Proc Natl Acad Sci USA* 94:14736–14741 (1997).
- Anderson JA, Song YS, Langley CH: Molecular population genetics of *Drosophila* subtelomeric DNA. *Genetics* 178:477–487 (2008).
- Aquadro CF, Bauer DuMont V, Reed FA: Genome-wide variation in the human and fruitfly: a comparison. *Curr Opin Genet Dev* 11:627–634 (2001).
- Awdeh ZL, Alper CA: Inherited structural polymorphism of the fourth component of human complement. *Proc Natl Acad Sci USA* 77:3576–3580 (1980).
- Bank RA, Hetteema EH, Muijs MA, Pals G, Arwert F, et al: Variation in gene copy number and polymorphism of the human salivary amylase isoenzyme system in Caucasians. *Hum Genet* 89:213–222 (1992).
- Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, et al: Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* 74:1111–1120 (2004).
- Buckland PR: Polymorphically duplicated genes: their relevance to phenotypic variation in humans. *Ann Medicine* 35:308–315 (2003).
- Cheng Z, Ventura M, She X, Khaitovich P, Graves T, et al: A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* 437:88–93 (2005).
- Chimpanzee Sequencing and Analysis Consortium: Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437:69–87 (2005).
- Conrad DF, Hurler ME: The population genetics of structural variation. *Nat Genet* 39:S30–S36 (2007).
- Cooper GM, Nickerson DA, Eichler EE: Mutational and selective effects on copy-number variants in the human genome. *Nat Genet* 39:S22–S29 (2007).
- Cutler G, Marshall LA, Chin N, Baribault H, Kassner PD: Significant gene content variation characterizes the genomes of inbred mouse strains. *Genome Res* 17:1743–1754 (2007).
- Dolan MJ, Kulkarni H, Camargo JF, He W, Smith A, et al: CCL3L1 and CCR5 influence cell-mediated immunity and affect HIV-AIDS pathogenesis via viral entry-independent mechanisms. *Nat Immunol* 8:1324–1336 (2007).
- Dopman EB, Hartl DL: A portrait of copy-number polymorphism in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 104:19920–19925 (2007).
- Dumas L, Kim YH, Karimpour-Fard A, Cox M, Hopkins J, et al: Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Res* 17:1266–1277 (2007).
- Egan CM, Sridhar S, Wigler M, Hall IM: Recurrent DNA copy number variation in the laboratory mouse. *Nat Genet* 39:1384–1389 (2007).
- Emerson JJ, Cardoso-Moreira M, Borevitz JO, Long M: Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science* 320:1629–1631 (2008).
- Fanciulli M, Norsworthy PJ, Petretto E, Dong R, Harper L, et al: *FCGR3B* copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat Genet* 39:721–723 (2007).
- Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, Wehkamp J, et al: A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am J Hum Genet* 79:439–448 (2006).
- Feuk L, Carson AR, Scherer SW: Structural variation in the human genome. *Nat Rev Genet* 7:85–97 (2006).
- Flint J, Hill AV, Bowden DK, Oppenheimer SJ, Sill PR, et al: High frequencies of alpha-thalassemia are the result of natural selection by malaria. *Nature* 321:744–750 (1986).
- Fortna A, Kim Y, McLaren E, Marshall K, Hahn G, et al: Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS Biol* 2:E207 (2004).
- Gilad Y, Man O, Paabo S, Lancet D: Human specific loss of olfactory receptor genes. *Proc Natl Acad Sci USA* 100:3324–3327 (2003).
- Gilad Y, Wiebe V, Przeworski M, Lancet D, Paabo S: Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol* 2:E5 (2004).
- Gilad Y, Rifkin SA, Bertone P, Gerstein M, White KP: Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. *Genome Res* 15:674–680 (2005).
- Goidts V, Armengol L, Schempp W, Conroy J, Nowak N, et al: Identification of large-scale human-specific copy number differences by interspecies array comparative genomic hybridization. *Hum Genet* 119:185–198 (2006).
- Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, et al: The influence of *CCL3L1* gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307:1434–1440 (2005).
- Graubert TA, Cahan P, Edwin D, Selzer RR, Richmond TA, et al: A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genet* 3:e3 (2007).
- Groot PC, Mager WH, Henriquez NV, Pronk JC, Arwert F, et al: Evolution of the human alpha-amylase multigene family through unequal, homologous, and inter- and intrachromosomal crossovers. *Genomics* 8:97–105 (1990).
- Guryev V, Saar K, Adamovic T, Verheul M, van Heesch SA, et al: Distribution and functional impact of DNA copy number variation in the rat. *Nat Genet* 40:538–545 (2008).

- Hayashi T, Motulsky AG, Deeb SS: Position of a 'green-red' hybrid gene in the visual pigment array determines colour-vision phenotype. *Nat Genet* 22:90–93 (1999).
- Hill AV, Bowden DK, O'Shaughnessy DF, Weatherall DJ, Clegg JB: Beta thalassemia in Melanesia: association with malaria and characterization of a common variant (IVS-1 nt 5 G→C). *Blood* 72:9–14 (1988).
- Hollox EJ, Huffmeier U, Zeeuwen PL, Palla R, Lascorz J, et al: Psoriasis is associated with increased beta-defensin genomic copy number. *Nat Genet* 40:23–25 (2008).
- Hurles ME, Dermitzakis ET, Tyler-Smith C: The functional impact of structural variation in humans. *Trends Genet* 24:238–245 (2008).
- Kehrer-Sawatzki H, Cooper DN: Structural divergence between the human and chimpanzee genomes. *Hum Genet* 120:759–778 (2007).
- Kidd JM, Newman TL, Tuzun E, Kaul R, Eichler EE: Population stratification of a common APO-BEC gene deletion polymorphism. *PLoS Genet* 3:e63 (2007).
- Kondrashov FA, Kondrashov AS: Role of selection in fixation of gene duplications. *J Theor Biol* 239:141–151 (2006).
- Le Marechal C, Masson E, Chen JM, Morel F, Ruzsniwski P, et al: Hereditary pancreatitis caused by triplication of the trypsinogen locus. *Nat Genet* 38:1372–1374 (2006).
- Lee AS, Gutierrez-Arcelus M, Perry GH, Vallender EJ, Johnson WE, et al: Analysis of copy number variation in the rhesus macaque genome identifies candidate loci for evolutionary and human disease studies. *Hum Mol Genet* 17:1127–1136 (2008).
- Li J, Jiang T, Mao JH, Balmain A, Peterson L, et al: Genomic segmental polymorphisms in inbred mouse strains. *Nat Genet* 36:952–954 (2004).
- Locke DP, Segraves R, Carbone L, Archidiacono N, Albertson DG, et al: Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res* 13:347–357 (2003).
- Mamtani M, Rovin B, Brey R, Camargo JF, Kulkarni H, et al: *CCL3L1* gene-containing segmental duplications and polymorphisms in *CCR5* affect risk of systemic lupus erythematosus. *Ann Rheum Dis* 67:1076–1083 (2008).
- McCarroll SA, Altshuler D: Copy number variation and association studies of human disease. *Nat Genet* 39:S37–S42 (2007).
- McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, et al: Common deletion polymorphisms in the human genome. *Nat Genet* 38:86–92 (2006).
- McDonald JH, Kreitman M: Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652–654 (1991).
- Mok BW, Ribacke U, Sherwood E, Wahlgren M: A highly conserved segmental duplication in the subtelomeres of *Plasmodium falciparum* chromosomes varies in copy number. *Malar J* 7:46 (2008).
- Nair S, Nash D, Sudimack D, Jaidee A, Barends M, et al: Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. *Mol Biol Evol* 24:562–573 (2007).
- Nathans J, Piantanida TP, Eddy RL, Shows TB, Hogness DS: Molecular genetics of inherited variation in human color vision. *Science* 232:203–210 (1986a).
- Nathans J, Thomas D, Hogness DS: Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science* 232:193–202 (1986b).
- Newman TL, Tuzun E, Morrison VA, Hayden KE, Ventura M, et al: A genome-wide survey of structural variation between human and chimpanzee. *Genome Res* 15:1344–1356 (2005).
- Nguyen DQ, Webber C, Ponting CP: Bias of selection on human copy number variants. *PLoS Genet* 2:e20 (2006).
- Nozawa M, Kawahara Y, Nei M: Genomic drift and copy number variation of sensory receptor genes in humans. *Proc Natl Acad Sci USA* 104:20421–20426 (2007).
- Orkin SH, Old JM, Weatherall DJ, Nathan DG: Partial deletion of beta-globin gene DNA in certain patients with beta 0-thalassemia. *Proc Natl Acad Sci USA* 76:2400–2404 (1979).
- Ottolenghi S, Lanyon WG, Paul J, Williamson R, Weatherall DJ, et al: The severe form of alpha thalassaemia is caused by a haemoglobin gene deletion. *Nature* 251:389–392 (1974).
- Ottolenghi S, Comi P, Giglioni B, Tolstoshev P, Lanyon WG, et al: Delta-beta-thalassemia is due to a gene deletion. *Cell* 9:71–80 (1976).
- Perry GH, Tchinda J, McGrath SD, Zhang J, Picker SR, et al: Hotspots for copy number variation in chimpanzees and humans. *Proc Natl Acad Sci USA* 103:8006–8011 (2006).
- Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, et al: Diet and the evolution of human amylase gene copy number variation. *Nat Genet* 39:1256–1260 (2007).
- Perry GH, Ben-Dor A, Tsalenko A, Sampas N, Rodriguez-Revena L, et al: The fine-scale and complex architecture of human copy-number variation. *Am J Hum Genet* 82:685–695 (2008a).
- Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, et al: Copy number variation and evolution in humans and chimpanzees. *Genome Res* 18:1698–1710 (2008b).
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al: Global variation in copy number in the human genome. *Nature* 444:444–454 (2006).
- Sebat J: Major changes in our DNA lead to major changes in our thinking. *Nat Genet* 39:S3–S5 (2007).
- She X, Cheng Z, Zollner S, Church DM, Eichler EE: Mouse segmental duplication and copy number variation. *Nat Genet* 40:909–914 (2008).
- Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, et al: Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315:848–853 (2007).
- Taylor JM, Dozy A, Kan YW, Varmus HE, Lie-Injo LE, et al: Genetic lesion in homozygous alpha thalassaemia (hydrops fetalis). *Nature* 251:392–393 (1974).
- Ting CN, Rosenberg MP, Snow CM, Samuelson LC, Meisler MH: Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev* 6:1457–1465 (1992).
- Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, et al: Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* 39:31–40 (2007).
- Trask BJ, Friedman C, Martin-Gallardo A, Rowen L, Akinbami C, et al: Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum Mol Genet* 7:13–26 (1998).
- Wang Y, Macke JP, Merbs SL, Zack DJ, Klaunberg B, et al: A locus control region adjacent to the human red and green visual pigment genes. *Neuron* 9:429–440 (1992).
- Weatherall DJ, Clegg JB: The Thalassemia Syndromes (Blackwell, Oxford 1981).
- Wilson GM, Flibotte S, Missirlis PI, Marra MA, Jones S, et al: Identification by full-coverage array CGH of human DNA copy number increases relative to chimpanzee and gorilla. *Genome Res* 16:173–181 (2006).
- Wolf S, Sharpe LT, Schmidt HJ, Knau H, Weitz S, et al: Direct visual resolution of gene copy number in the human photopigment gene array. *Invest Ophthalmol Vis Sci* 40:1585–1589 (1999).
- Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, et al: Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am J Hum Genet* 80:1037–1054 (2007).
- Zhang J: The drifting human genome. *Proc Natl Acad Sci USA* 104:20147–20148 (2007).