Human Lineage–Specific Amplification, Selection, and Neuronal Expression of DUF1220 Domains

Magdalena C. Popesco,1,2,3* Erik J. MacLaren,1,2,3* Janet Hopkins,1,2,3 Laura Dumas,2,3 Michael Cox,2,3 Lyne Meltesen,2,4 Loris McGavran,1,4 Loris McGavran,1,4 Loris McGavran,1,4 Gerald J. Wyckoff,5 James M. Sikela1,2,3‡

Extensive gene duplication in a species-specific manner, followed by divergence and functional specialization, can be an important factor in the evolution of phenotypic traits unique to that species. Copy number variations between human and chimpanzee have been discovered with the use of a draft sequence of the chimpanzee genome, although primate outgroup information is currently limited to draft sequence from only one other species, the macaque. Draft sequences are prone to misassembly of recently duplicated sequences, a limitation that is the most severe for the most evolutionarily recent (i.e., similar) duplications. A complementary approach is cDNA array-based comparative genomic hybridization (aCGH). We previously used cDNA aCGH to carry out genome-wide gene copy number comparisons between human and great ape species, and identified 134 genes showing human lineage-specific amplification and six genes showing HLS decreases.

To obtain an independent estimate of the copy number of each HLS gene, we determined the full insert sequences of the cDNAs (table S1) and used these as BLAT (http://genome.ucsc.edu) queries to search a recent human genome assembly (Build 35) (6) as well as available genome draft sequences from chimpanzee and macaque. The great majority (86.4%) of genes predicted by cDNA aCGH to have an HLS increase in copy number produced more BLAT hits (score >200) in the human genome than in either chimpanzee or macaque (table S2), and 44 of these (31%) had more than five BLAT hits in the human genome (Fig. 1A).

After removal of all BLAT hits predicted to be intronless, one gene, MGC8902 (cDNA IMAGE clone 843276), showed the most striking human-specific increase, with 49, 10, and 4 hits found in human, chimpanzee, and macaque, respectively (Fig. 1A). All human hits associated with MGC8902 (49/49) were predicted to be nonretroposed copies. It was also ranked as the fifth highest HLS aCGH signal out of the 134 genes predicted to have HLS increases in copy number (7), and contains six predicted DUF1220 domains (Fig. 1B).

The genomic sequences predicted to encode DUF1220 domains typically show a unique signature of an evenly spaced two-exon repeat unit (Fig. 1B). A recent report treats this exon pair as a new repeat that is part of a gene family termed NBPF (8). The repeat is inclusive of the DUF1220 domain but also contains additional protein-coding sequences that may not share all the biological and evolutionary characteristics of DUF1220.

---

1Human Medical Genetics and 2Neuroscience Programs, 3Department of Pharmacology, 4Department of Pathology, University of Colorado at Denver and Health Sciences Center, Aurora, CO 80045, USA. 5Division of Molecular Biology and Biochemistry, University of Missouri, Kansas City, MO 64110, USA.

*These authors contributed equally to this work.

†Present address: Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK.

‡To whom correspondence should be addressed. E-mail: jame.sikela@uchsc.edu

Fig. 1. Cross-species BLAT survey of HLS cDNAs and organization of the MGC8902 gene. (A) BLAT searches were performed using full cDNA insert sequences for 140 HLS genes (5) as queries. The IMAGE clones (http://image.llnl.gov) that yielded >5 BLAT hits in the human genome are shown. BLAT hits with span sizes exceeding the size of the cDNA query were scored as potentially containing introns. Potentially “intronless” BLAT hits are shown in white. The asterisk denotes BLAT hits associated with the ribosomal protein gene RLP23AP7, which had hit totals of 150, 144, and 133 for human, chimpanzee, and macaque, respectively. All of these were intronless. (B) The genomic exon/intron organization of MGC8902 and the predicted domain structure of the translated protein. A representative DUF1220 genomic repeat unit is also shown.

---

[Table S1](http://www.sciencemag.org) [Table S2](http://www.sciencemag.org)
It has been estimated that 34 different human genes encode DUF1220 domains (table S3) (www.ncbi.nlm.nih.gov/IEB/Research/Assembly). Pfam (Version 17.0) (9) predicts that 60 human DUF1220-containing proteins exist, containing a total of 271 DUF1220 domains (fig. S1) derived from 11 seed domains (10) (fig. S2A). Estimates based on cDNA sequences indicated that 22 genes exist, including six pseudogenes (8). None of these cDNAs showed perfect identity to human genomic sequences, raising the possibility that this count is an underestimate. Recent additional sequencing of chromosome 1 identified at least 15 gene sequences that encode DUF1220 domains, although several sequence gaps still remain in DUF1220-encoding regions (11).

The amino acid sequences of each of the 11 DUF1220 seed domains were next used as BLAT queries against genome sequences from several species (table S4). The 11 seed domains showed no matches outside of mammals, and 10 of the 11 were primate-specific, with the highest number of copies always found in human (fig. S2B). The remaining seed domain (O75042) was found in primate and nonprimate mammals, usually as a single domain encoded by a single-copy gene [Myomegalin/ PDE4DIP (12)] that also encodes a spindle-associated domain (fig. S2C). The most human BLAT hits were observed with three domains found in one predicted protein, Q8IX62. One of these (Q8IX62_Human/17-83) had 90 hits in human but only 16 and 11 in chimp and macaque, respectively. Of the human hits, 37 (41%) were 100% matches (fig. S2B), by far the highest frequency of identical human matches found for any of the 11 seed domains. For macaque, a similar number of BLAT hits was obtained whether the January 2005 or January 2006 sequence assembly was used (table S5).

To provide an independent estimate of the frequency of this domain, we carried out quantitative polymerase chain reaction (QPCR) analysis on multiple individuals from each of several species, using a primer and probe set with sequences that are identical to sequences in the human and chimp genomes (fig. 2 and table S6). Consistent with BLAT results and previously reported aCGH data, QPCR analysis also found that the human genome had significantly more copies than that of any other species (P < 0.01) and, generally, the evolutionarily closer the species was to human, the more DUF1220 domains were encoded in its genome. Interhominoid cDNA aCGH data reported previously (5) for cDNA IMAGE 843276 [average log2 ratio: bonobo (3) = –1.79; chimpanzee (4) = –1.98; gorilla (3) = –1.19; orangutan (3) = –2.74] are in very close agreement (r = 0.9886) with the cross-species QPCR data presented in Fig. 2. Taken together, aCGH, BLAT, and QPCR data indicate that the number of DUF1220 copies is highly expanded in humans, reduced in African great apes, further reduced in orangutan and Old World monkeys, single-copy in nonprimate mammals, and absent in nonmammalian species. Some intraspecies copy number variability was apparent, although a survey of a limited number of individuals (22 individuals from diverse human populations) revealed no population-specific trends (table S6).

Human genomic locations predicted from the BLAT analysis using the 11 seed domains (fig. S3, A and B) were in general agreement with fluorescence in situ hybridization (FISH) analysis (fig. S4) and two recent reports (8,11), positioning the majority of DUF1220 (NBPF) sequences at 1q21.1, a complex genomic region immediately adjacent to the pericentromeric C-band 1q12. Additional sequences are found at 1p13.3 and 1p36. After eliminating redundant (overlapping) positions, we identified 212, 37, and 30 unique DUF1220-positive BLAT hits in human, chimp, and rhesus, respectively, along with only one each for mouse and rat.

Fig. 2. QPCR-based estimation of the number of DUF1220 domains found within different species. QPCR was carried out to survey the frequency of DUF1220 domain (Q8IX62 17-33) sequences in various primate species. Corresponding numerical values can be found in table S6.

Fig. 3. Ka/Ks values of DUF1220-encoding sequences. (A) Sequences with Ka/Ks values 0 through 2.0. Rodent-primate and primate-primate comparison means are shown (arrows); human-specific comparisons produce an even higher mean. Of these comparisons, 3106 have a Ka/Ks value of 0, and 2035 have a Ka/Ks value of 2.0 or greater; the scale here is limited so that the major patterns can be seen. (B) Sequences with Ka/Ks values 2.25 through 8.0; all of these comparisons are primate homologous comparisons, and most are human-human pairwise comparisons.
Evolutionary analysis (13) was performed with a nonredundant set of the human, chimp, rhesus, mouse, and rat DUF1220 nucleotide sequences derived from the BLAT searches described above (table S7). These sequences were filtered for frame-shift insertions and aligned, and the resulting 256 sequences were used for construction of a phylogenetic tree (fig. S5). In addition, Ka/Ks ratios (ratios of the rate of amino acid substitution to silent substitution) were determined for each pairwise combination of sequences (fig. S5). The domain that is found as a single copy in nonprimate mammals, O75042, is the likely ancestral domain, consistent with a phylogenetic analysis of the 11 DUF1220 seed domains (www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF06758), with the primate-specific domains appearing more recently. The ladder-like nature of the phylogenetic tree suggests that serial domain amplification and subsequent divergence are the rule in this large set of repeats. Also, 33% (10,583/32,131) of pairwise comparisons showed a Ka/Ks ratio of 1 or greater—a traditional signature of positive selection (14).

On average, primate-primate homologous comparisons had a higher ratio of nonsynonymous to synonymous changes (Ka/Ks mean = 0.91) than did rodent-primate homologous comparisons (Ka/Ks mean = 0.61), indicative of either a higher level of positive selection or a relaxation of functional constraint (fig. 3A). The average Ka/Ks value for primate homologous comparisons was unusually high (0.91) relative to general estimates of primate evolutionary rate (15), with two human-versus-rhesus comparisons producing the highest values (>7.1) (Fig. 3B). In contrast, Ka/Ks analysis of non-DUF1220 sequences from a DUF1220-containing gene did not appear to show evidence of positive selection (8).

Western blot analysis was carried out on a panel of normal adult human tissues, using an affinity-purified antibody directed against a 20-amino acid peptide derived from a primate-specific DUF1220 domain. A heavy band was visible at ~36 kD in heart, brain, spleen, skeletal muscle, and small intestine (fig. 4A), which was blocked in all tissues, except in the skeletal muscle, by the adsorption control (fig. S6A). This same band was faintly present in kidney, lung, stomach, colon, and rectum. In addition, other heavy bands were visible between 25 and 40 kD throughout the tissue panel. The same ~36 kD band was highly expressed in frontal lobe, temporal lobe, parietal lobe, occipital lobe, and cerebellum, whereas it was absent in placenta (fig. 4B).

Using double-label immunofluorescence, we analyzed normal adult brain regions from several individuals with the same affinity-purified DUF1220 antibody. DUF1220 sequences were consistently found in neurons but not in glia. In the cerebellum, preferential expression was observed in Purkinje cells, where signals were restricted to cell bodies (cytoplasms) and dendrites (fig. 4, C to E, and fig. S6, B to D). In addition to labeling in the cerebellum, neuron-specific DUF1220 signals were present in the cortical layers of the hippocampus (fig. 4, F and G). DUF1220 domains were also abundantly expressed in neurons within the neocortex (frontal, parietal, occipital, and temporal lobes), thought to be critical to higher cognitive functions (fig. 4, H to K).

Although the precise function of genes encoding DUF1220 domains and the domains themselves is at present unknown, the pattern of amplification and location of expression have led us to speculate that the domains and the genes that encode them may be important to cognitive function. In light of the strong DUF1220 expression we observed in neurons of the neocortex, it is intriguing that multiple independent evolutionary processes [brain enlargement, neocortex expansion (16), gene duplication, and domain amplification] can be seen as having individually and cumulatively contributed to increasing the
Reducing the Racial Achievement Gap: A Social-Psychological Intervention

Geoffrey L. Cohen, Julio Garcia, Nancy Apfel, Allison Master

Two randomized field experiments tested a social-psychological intervention designed to improve minority student performance and increase our understanding of how psychological threat mediates performance in chronically evaluative real-world environments. We expected that the risk of confirming a negative stereotype aimed at one’s group could undermine academic performance in minority students by elevating their level of psychological threat. We tested whether such psychological threat could be lessened by having students reaffirm their sense of personal adequacy or “self-integrity.” The intervention, a brief in-class writing assignment, significantly improved the grades of African American students and reduced the racial achievement gap by 40%. These results suggest that the racial achievement gap, a major social concern in the United States, could be ameliorated by the use of timely and targeted social-psychological interventions.

The drive for self-integrity—seeing oneself as good, virtuous, and efficacious—is a fundamental human motivation (1–3). Membership in valued social groups is often a major source of individuals’ sense of self-integrity (4, 5). Consequently, negative characterizations of one’s group can prove threatening, especially in chronically evaluative environments.

Because people subjected to widely known negative stereotypes impugning the intelligence of their group are aware of these negative characterizations, they may worry that performing poorly could confirm the stereotype of their group (6–8). This situation can create chronic stress at school and work, by burdening people with an extra psychological threat not experienced by those outside their group. If too severe, stress can undermine performance (6–10). Indeed, simply observing a group member who might confirm a negative stereotype about one’s group can induce threat, undermining performance (5).

One potentially effective way to buffer people against threat and its consequences, we suggest, is to allow them to reaffirm their self-integrity (2, 3). Self-affirmations, by buttressing self-worth, can alleviate the stress arising in threatening performance situations (11). They can take the form of reflections on personally important, overarching values, such as the importance of family or a self-defining skill (2, 3).

The research reported here tested whether a self-affirmation intervention designed to lessen threat would enhance the academic achievement of negatively stereotyped minority students. The intervention rested on three assumptions: First, people are motivated to maintain self-integrity; second, because group memberships are an important source of self-integrity, negative group characterizations can pose a chronic threat to self-integrity; third, such threat, if too severe, can undermine performance.

School settings can be stressful to almost all students regardless of race. However, for African American students, the academic environment involves an extra degree of threat not experienced by nonminority students, due to the negative stereotype about the intelligence of their race. This threat, on average, raises stress to levels that are debilitating to performance (6–9). Accordingly, we expect that a self-affirmation intervention would be particularly effective at improving their academic performance. We would, in fact, expect this intervention to improve the performance of all groups of individuals subjected to a threat sufficiently pervasive and intense to impede that entire group’s average performance.

This prediction was tested in two randomized double-blind field experiments (12). The second, a replication study, occurred a year after the first and involved a different cohort of students. Participants were seventh-graders from middle- to lower-middle-class families at a suburban northeastern middle school whose student body was divided almost evenly between African Americans and European Americans. The experiments involved 119 African American students and 124 European American students distributed roughly evenly across the two studies. All the teachers who participated taught the same academic subject (one not typically related to gender stereotypes). This subject was the intervention–targeted course in both studies; it was the one in which the intervention was administered.

In the full term of each year, students were randomly assigned, at the level of the individual student, to the affirmation condition or the control condition. For each teacher and classroom period, there were approximately equal numbers of participants in each condition. Teachers were blind to students’ condition assignment and unaware of the