

Géraldine Mathonnet · Damian Labuda  
Caroline Meloche · Tina Wambach  
Maja Krajinovic · Daniel Sinnett

## Variable continental distribution of polymorphisms in the coding regions of DNA-repair genes

Received: 14 August 2003 / Accepted: 1 October 2003 / Published online: 19 November 2003  
© The Japan Society of Human Genetics and Springer-Verlag 2003

**Abstract** DNA-repair pathways are critical for maintaining the integrity of the genetic material by protecting against mutations due to exposure-induced damages or replication errors. Polymorphisms in the corresponding genes may be relevant in genetic epidemiology by modifying individual cancer susceptibility or therapeutic response. We report data on the population distribution of potentially functional variants in XRCC1, APEX1, ERCC2, ERCC4, hMLH1, and hMSH3 genes among groups representing individuals of European, Middle Eastern, African, Southeast Asian and North American descent. The data indicate little interpopulation differentiation in some of these polymorphisms and typical  $F_{ST}$  values ranging from 10 to 17% at others. Low  $F_{ST}$  was observed in APEX1 and hMSH3 exon 23 in spite of their relatively high minor allele frequencies, which could suggest the effect of balancing selection. In XRCC1, hMSH3 exon 21 and hMLH1 Africa clusters either with Middle East and Europe or with Southeast Asia, which could be related to the demographic history of human populations, whereby human migrations and genetic drift rather than selection would account for the observed differences.

**Keywords** Human populations · Genetic polymorphisms · DNA repair proteins · Allele frequency · Genetic drift · Selection

### Introduction

DNA-repair systems are essential for the maintenance of integrity of the genetic material. They play a key role in protecting the genetic material against deleterious mutations leading to cancer as well as to neurodegeneration or aging. Interindividual differences in the DNA-repair efficiency appear to be genetically determined (Crompton and Ozsahin 1997; Mohrenweiser and Jones 1998; Pero et al. 1983; Pero et al. 1989). Presence of potentially functional polymorphisms within the coding regions of the APEX1, XRCC1, ERCC4, and ERCC2 (Shen et al. 1998), hMSH2 (Liu et al. 1995), as well as hMLH1, and hMSH3 genes (Benachenhou et al. 1998) is consistent with this view. The XRCC1 protein interacts with DNA ligase III, polymerase  $\beta$  and poly (ADP-ribose) polymerase (Caldecott et al. 1996), as well as apurinic endonuclease APEX1 (Demple et al. 1991). The proteins hMLH1 and hMSH3 contribute to the mismatch-repair machinery (e.g., Kolodner 1996). ERCC4 corresponds to a nuclease that acts at the 5' ends of DNA lesions (Bessho et al. 1997), while ERCC2 has a 5'-3' ATP-dependent helicase activity (Sung et al. 1993) associated with the TFIIH protein complex required both for transcription and repair (Lehmann 1995). Amino acids replacements within these proteins might affect the efficiency of DNA repair and thus modulate an individual's risks of developing cancer or influence the therapeutic response. The knowledge of the frequency profiles of such genetic variants across human populations is needed for genetic epidemiological or pharmacogenetics studies.

In particular, we wanted to know whether the distribution of different variants can be explained by recent out-of-Africa expansion, separating sub-Saharan Africa from Europe and the Middle East (ME) on one side and Southeast Asia (SEA) and America on the other (Cavalli-Sforza et al. 1994; Lahr and Foley 1994). It turns out that different polymorphisms in a plot of multidimensional scaling of the respective  $F_{ST}$ 's, demonstrate

G. Mathonnet · D. Labuda · C. Meloche · T. Wambach  
M. Krajinovic · D. Sinnett  
Centre de Recherche, Hôpital Sainte-Justine,  
Montréal, Québec, Canada

D. Labuda (✉) · M. Krajinovic · D. Sinnett  
Département de Pédiatrie, Université de Montréal,  
3175 Côte Sainte-Catherine, Montréal,  
Québec, H3T 1C5, Canada  
E-mail: damian.labuda@umontreal.ca  
Tel.: +1-514-3454931  
Fax: +1-514-3454731

different continental affinities, thus suggesting a role of selection or genetic drift confounded with demographic factors in their partitioning of genetic variants among human populations.

## Materials and methods

### DNA samples and genotyping

Genomic samples were either from earlier non-nominative DNA collections or were obtained, on a non-nominative basis, from consenting adults who provided the information about their ethnicity and country of origin following a protocol approved by the Institutional Review Board. Studied groups included individuals from the Middle East and North Africa ( $n=23$ ), referred to as Middle East; from Southeast Asia ( $n=24$ ); North America (Athabaskan speakers from Saskatchewan  $n=23$  and Algonquian speakers from Quebec and Ontario  $n=24$ ); individuals of sub-Saharan African descent ( $n=23$ ), referred to as Africans; as well as French Canadians from the Province of Quebec, Canada ( $n=323$ ), representing individuals of European descent. PCR was carried out in standard conditions in a volume of 50  $\mu$ l using 25 ng of genomic DNA. Amplification products were dot blotted onto a Hybond TM-N+ membrane (Amersham) and subsequently hybridized with ASO probes (Table 1) following the protocol described earlier (Labuda et al. 1999).

### Statistical analysis

$F_{ST}$  (see Weir and Cockerham 1984), a measure of allele frequency differences among population samples, was estimated per site or locus, for pairs of continental populations, for all of them as a group, or for combined groups of populations based on their pairwise analysis as indicated. It is defined as  $F_{ST} = \frac{H_T - H_S}{H_T}$  where  $H_S$ , the average heterozygosity among subpopulations, corresponds to  $H_S = \frac{1}{s} \sum_{j=1}^s H_j$ , and  $H_j = \left(1 - \sum_{i=1}^k p_{ij}^2\right) \frac{2n_j}{2n_j - 1}$  with  $s$  repre-

senting the number of subpopulations,  $k$  the number of alleles,  $p_i$  the frequency of allele  $i$  and  $n$  the number of individuals tested in subpopulation  $j$ , whereas  $H_T$ , total heterozygosity, is  $H_T = \frac{2}{s(s-1)} \sum_{j=1}^s \left(1 - \sum_{i=1}^k p_{ij} p_{il}\right)$ .  $F_{ST}$  was calculated with the help of the Arlequin package v.2.0 (Schneider et al. 2000). The same results were obtained with the GDA software v.1.0 (Lewis and Zaykin 2001). Because of a much larger size of our French Canadian sample representing Europe, which, when used as such for the  $F_{ST}$  estimation, strongly biased the resulting  $F_{ST}$  values, its size was artificially reduced to  $n=46$  chromosomes, a size comparable to all other studied populations.

Two-site haplotypes from the hMSH3 gene were obtained from the observed genotypes by inspection. Their estimated frequencies were subsequently confirmed using the EH linkage utility program, which was also used to assess the significance of linkage disequilibrium between the contributing polymorphisms. EH is available at ott@linkage.rockefeller.edu.  $D^* = D/D_{max}$  was calculated as described in Hartl and Clark (1989). The graphical display of the pairwise  $F_{ST}$ 's by multidimensional scaling was obtained using STATISTICA v.6.

## Results

### Population distribution of DNA-repair protein variants

Five subcontinental groups were surveyed for the presence of seven amino acid substitution polymorphisms in six proteins involved in DNA repair (Table 2). The population samples extended from sub-Saharan Africa through the ME to Europe, and further to SEA and North-America. In all but ERCC4, minor allele frequency exceeds 10% in at least one of the populations. We note that at ERCC4 there is no significant differentiation among populations, as indicated by zero  $F_{ST}$ . Two other sites, APEX1 and

**Table 1** Characteristics of DNA-repair polymorphisms and conditions of PCR-ASO hybridization genotyping assays. *Uppercase characters* indicate polymorphic sites

Locus	Amino acid change	Primers (5'-3')	Annealing (°C)	Product size (bp)	Variant site <sup>a</sup>	ASO <sup>b</sup> (5'-3')
hMLH1	Val219	F, CCTGTGTCCTCTGCTGTTG	52	146	G676	tccaagaCggagcg
	Ile219	R, GATTTTTTATATAGGTTATCGAC			A676	cgctccAtctttgga
	Gly322	R, GGTAAGTGCAGGTTACATAA			G967	ctgagagCcagtggt
hMSH3	Gln940	F, GTGAAGAGGAAAATCAAGG	55	248	A2835	aaaggacAgagtaca
	Arg940	R, ATACAGTTTACCGGATACTTA			G2835	tgtactcCgtccttt
	Ala1036	F, TCAGGCACAGTTTTGATC			G3124	gttctcgtgCgcctg
XRCC1	Thr1036	R, CCTCTCTTCTTGGACAA	60	485	A3124	caggcAcagcagaac
	Arg194	F, TACCCTCAGACCCACGAGT			C26304	cttcagcCggatcaa
	Trp194	R, GCCAGGGCCCTCCTTCAA			T26304	ttgatccAgctgaag
ERCC2	Arg399	F, CTTTGCCCTCAGATCACAC	60	206	G28152	tcccGgaggtaaaggc
	Gln399	R, CAGGATAAGGAGCAGGTTG			A28152	tcccAgaggtaaaggc
	Met241	R, CTCACCTGGTTGATGCACAG			T18067	ggccaTgctgctga
ERCC4	Lys751	F, TCAAACATCCTGTCCCTACT	52	344	A35931	gacgctgAagaggat
	Gln751	R, CTGCGATTAAAGCTGTGGA			C35931	gacgctgCagaggat
ERCC4	Ser379	F, CTAGGATCTCAGTGTTTCATT	55	401	T1117	cactttgAgttgctt
	Pro379	R, ATATGTACTGATGCTCGTGT			C1117	aagcaacCcaaagt
	Gln415	F, AAGGGGGCACAGGGAAACT			55	275
APEX1	Arg415	R, TTTGGGGGTCTTTAGGTC	50	395	G1244	acatgttCggtcatc
	Asp148	F, CTTACGGCATAGGTGAGAC			T2197	taggcgaTgaggagc
	Glu148	R, CCAGCATTCTTTTTGTTCC			G2197	taggcgaGgaggagc

<sup>a</sup>The nucleotide numbering was drawn from sequences reported in the following accession numbers: hMLH1, U07343; hMSH3, U61981; XRCC1, L34079; ERCC2, L47234; APEX1, M99703; ERCC4, L77890; (GenBank release 112.0)

<sup>b</sup>ASO allele-specific oligonucleotide

**Table 2** Population frequencies of DNA-repair genetic variants. *n.d.* not determined. Note that for the estimation of  $F_{ST}$  we set the European sample to be of similar size as other populations i.e.  $n=46$

Gene	Exon	Allele			Population minor allele frequency (%)						$F_{ST}$
		Mouse	Chimpanzee	Human <sup>a</sup>	European descent	Middle East	African descent	Southeast Asia	American Indians		
									Algonquian speakers	Athabaskan speakers	
XRCC1	6	Arg	Arg	Arg194Trp	5.0	13.0	6.5	28.6	47.9	42.9	0.173
APEX1	5	Glu	n.d	Asp148Glu	43.7	43.8	36.8	25.0	19.6	n.d.	0.032
ERCC2	23	Gln	Gln	Lys751Gln	39.1	32.1	13.6	0.0	21.1	7.1	0.120
ERCC4	8	Arg	Gln	Arg415Gln	8.2	5.9	2.2	2.1	4.2	n.d.	-0.004
hMLH1	8	Ile	Ile	Ile219Val	31.3	23.9	2.2	2.2	42.5	13	0.144
hMSH3	21	Arg	Arg	Arg940Gln	16.7	22.5	4.8	2.6	0	0	0.097
hMSH3	23	Asp	Ala	Thr1036Ala	29.6	42.5	35.7	23.7	7.1	20.0	0.057
Number of chromosomes analyzed <sup>b</sup>				522–640	32–46	38–46	36–46	38–48	30–46		

<sup>a</sup>The second listed is the minor allele for which frequency information is provided

<sup>b</sup>The number of typed chromosomes varied among loci within the indicated range

hMSH3 exon 23, also show little differentiation with nonzero but low  $F_{ST}$  values and are nonsignificant in population pairwise comparison. In contrast to ERCC4, in these two sites, the minor allele frequency is in the range of 0.1–0.45, thus suggesting a mechanism maintaining a relatively high and comparable heterozygosity across continental populations.

In other loci,  $F_{ST}$  estimates ranging from 10 to 17% (Table 2) are similar to those observed in a number of other genetic systems (Cavalli-Sforza et al. 1994; Fullerton et al. 2002; Relethford 2001). Minor allele frequencies (Table 2) as well as matrices of pairwise  $F_{ST}$ 's represented by multidimensional scaling in Fig. 1, indicate different geographic distribution of the analyzed variants. Strikingly, there is no clear division between sub-Saharan Africa and the remaining populations. For all polymorphic sites, Europe and the ME cluster together, at least looking at one dimension at a time. This clustering includes Africa in the case of XRCC1, APEX1, and hMSH3 exon 23. In contrast, in hMLH1 and ERCC4 Africa clearly joins SEA. In XRCC1, when the total population is considered to be composed of two subpopulations—one comprising Africa, the ME, and Europe and the other SEA and North America—the estimated  $F_{ST}$  of 0.241 is close to that estimated considering the total population composed of six subpopulations (Table 2). This indicates that greatest differentiation occurred along the line separating the above groups of populations. In hMLH1, if Africa and SEA are combined in a single subpopulation as opposed to others, the  $F_{ST}$  of 0.159 is almost identical to the  $F_{ST}$  evaluated considering six independent subpopulations.

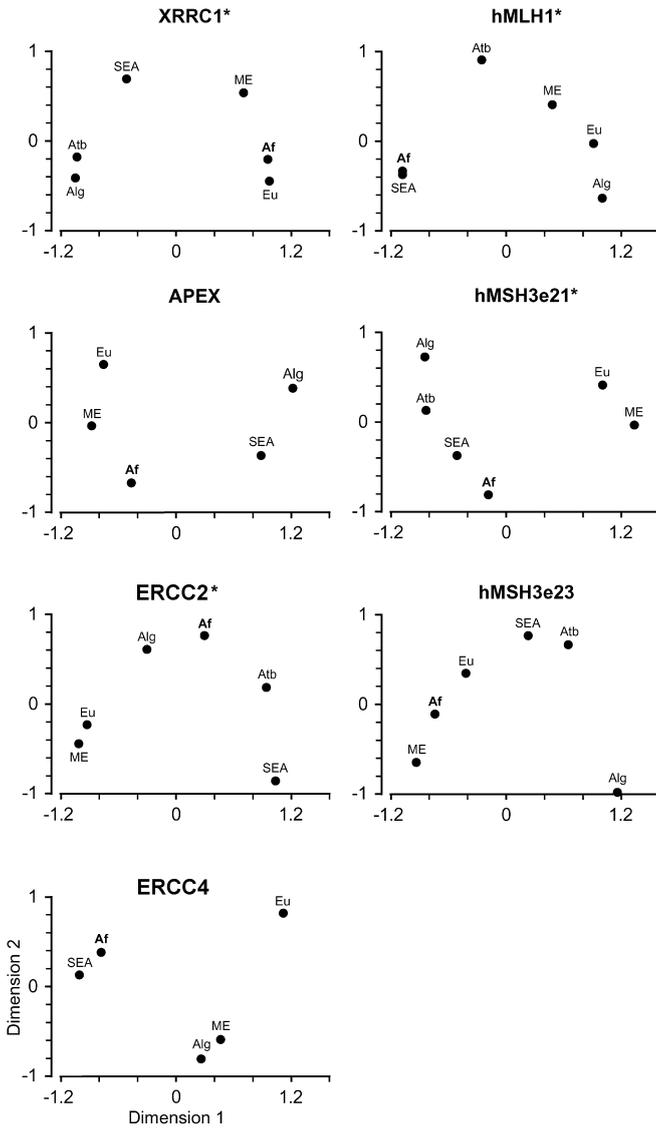
In hMSH3, differences between sites within the locus are observed. In contrast to the site in exon 21, that in exon 23 displays little population differentiation and high minor allele frequency. Interestingly, in spite of the increase in frequency of Gln940 allele in Europe and the ME, the frequency of exon 23 Ala1036 is maintained similar across Old World continents. Indeed, if Algon-

quian speakers are excluded from the analysis,  $F_{ST}$  of this polymorphism reduces to almost zero among the remaining five subpopulations. It is therefore interesting to evaluate the relationship between these two sites in hMSH3 (Table 3). Based on the linkage disequilibrium and the allelic state in different species, we can assign the origins of both polymorphisms to Africa through mutations Arg → Gln and Ala → Thr occurring on independent chromosomes and the appearance of Gln-Thr haplotype through a recombination. The value of the  $F_{ST}$  estimate for the hMSH3 locus two site haplotype (Table 3) drops to about 5%, less than each site considered separately (data not shown).

## Discussion

Under neutrality and in the absence of mutation, genetic variation across populations is expected to be determined by genetic drift only, which in turn is determined by the demographic history of populations. A priori, all loci in the genome have the same expected degree of differentiation, which may be used to detect the action of natural selection (Cavalli-Sforza 1966; Fullerton et al. 2002; Lewontin and Krakauer 1973). If the allele frequency data are available for a large set of putatively neutral loci, then an empirical distribution of  $F_{ST}$  values can be constructed to identify loci with unusual differentiation patterns.

Recently, Fullerton et al. (2002) compared  $F_{ST}$  estimates in *CAPN10*, a candidate susceptibility locus for type 2 diabetes, with those of 86 biallelic RFLPs from earlier studies. They found that several polymorphic sites within *CAPN10* had a relatively elevated  $F_{ST}$ , which could have been interpreted as the effect of selection. These authors observed the most pronounced diversification between sub-Saharan African and non-African populations that differentiated the population risk of type 2 diabetes attributable to the susceptibility haplotypes (Fullerton et al. 2002). This diversification of



**Fig. 1** Display of pairwise  $F_{ST}$ 's among studied populations by multidimensional scaling: *Af* Africa, *Eu* Europe, *SEA* Southeast Asia, *ME* Middle East, *Alg* and *Atb* Algonquian and Athabascan speakers from North America, respectively. *Asterisks* indicate polymorphisms where seven or more of population pairwise  $F_{ST}$  distances were statistically significant

allelic frequencies could have been caused by neutral demographic mechanisms, such as drift or migrations, population bottlenecks, and founder effects accompanying the out-of-Africa expansion (Cavalli-Sforza et al. 1994).

Interestingly, although our analysis of seven DNA-repair genetic polymorphisms did not show the greatest differentiation between Africans and non-Africans, certain observed continuities and discontinuities in allelic frequencies (Fig. 1) go along with two putative routes of the out-of-Africa expansion (Lahr and Foley 1994): the northern route connecting Africa to Europe through the ME and the earlier southern route linking Africa with Southeast Asia (Kivisild et al. 1999; Quintana-Murci et al. 1999).

In the case of the XRCC1 and hMSH3 exon 23 polymorphisms, the Africans show a tendency to cluster with Europe and the ME, while in hMLH1 they clearly cluster with Asia (seen also in hMSH3 exon 21 and ERCC4). In contrast, the distribution of two other polymorphisms, APEX1 and hMSH3, exon 23 does not seem to correlate with the history of human migrations. As noted before, in spite of tight linkage between hMSH3 exon 21 and exon 23 polymorphisms, their distribution patterns appear to follow different trajectories. APEX1 and hMSH3 exon 23 sites are characterized by an overall high heterozygosity and yet a low or not significantly different  $F_{ST}$ , suggesting a possibility of balancing selection as a force maintaining a high and relatively even frequency distribution among populations (e.g., Bürger 2000).

The interpretation evoking selection is more plausible, if the polymorphism affects the protein primary structure, when the substitution changes the nature of the amino acid residue and when this residue appears conserved in extant species. These conditions were not clearly met in the case of studied polymorphisms (Table 2). This is not unexpected given that at certain loci, a "favorable" allele, under certain circumstances, may become an "at risk" allele in a different context (Neel 1962), as illustrated by the effect of common polymorphisms in the MTHFR locus (Rosenberg et al. 2002). The  $F_{ST}$  of about 10% for 677T and 1268C polymorphisms based on Rosenberg et al. (2002) data

**Table 3** hMSH3 haplotypes involving sites G2835A (Arg940Gln) and A3124G (Thr1036Ala)

Haplotype (variant base, variant amino acid)	Population haplotype frequency (%)					
	European descent	Middle East	African descent	Southeast Asia	American Indians	
					Algonquian speakers	Athabascan speakers
G-A Arg-Th	68.6	57.5	61.9	76.3	92.9	80.0
A-A Gln-Thr	2.0	0	2.4	0	0	0
G-G Arg-Ala	14.7	20.0	33.3	21.1	7.1	20.0
A-G Gln-Ala	14.7	22.5	2.4	2.6	0	0
Linkage disequilibrium coefficient $D'$	0.804	1.0	0.220	1.0	0	0
$P$	< 0.001	0.01	0.98	0.7	n/a	n/a

are within the expected range (e.g., Fullerton et al. 2002). However, functional significance of MTHFR polymorphisms was independently documented in different studies. Therefore, here a high minor allele frequency across populations seems to be the best indicator of potential functional/selective significance, as in the APEX1 or hMSH3 exon 23 polymorphisms.

The results presented here were obtained with a particular population of European origin and a relatively small population sample of non-Europeans. Certain estimates may change if more samples or different populations representing continents or linguistic groups are surveyed. However, we would not expect a major change in the overall picture, since our estimates are similar to those obtained for a variety of European populations as well as for non-Europeans for loci whose data were available (Abdel-Rahman et al. 2000; Butkiewicz et al. 2001; Duell et al. 2001; Dybdahl et al. 1999; Fan et al. 1999; Fredman et al. 2002; Lunn et al. 1999).

In conclusion, this study illustrates how population genomics can be used to provide insight into the functional significance of certain DNA variants. The question remains why none of the seven studied polymorphisms demarcates sub-Saharan Africa from other continents, expected under a neutral scenario and assuming the recent out-of-Africa model of human population history. Finally, it is known that the incidence of different cancers differs between populations of different geographic and ethnic origin. It remains to be shown to what extent, beside environmental factors, the genetic differences in candidate genes, such as demonstrated here, are responsible for this variable incidence.

**Acknowledgements** We are grateful to our colleagues who shared samples from their collections and to all individuals who kindly consented to provide DNA for this study. We thank Chantal Richer, Hugues Sinnett, Patrick Beaulieu, and Vania Yotova for their assistance, and Dominika Kozubska for secretarial help. MK and DS are scholars of the Fonds de la Recherche en Santé du Québec, whereas GM and CM were recipients of studentships from the Fondation Hôpital Ste-Justine/Power Corporation Inc., and the Natural Sciences and Engineering Research Council, respectively. This work was supported by the Fondation Charles-Bruneau and, in part, by a research grant from the Canadian Institutes of Health Research to DL (MOP-12782).

## References

- Abdel-Rahman SZ, Soliman AS, Bondy ML, Omar S, El-Badawy SA, Khaled HM, Seifeldin IA, Levin B (2000) Inheritance of the 194Trp and the 399Gln variant alleles of the DNA repair gene XRCC1 are associated with increased risk of early-onset colorectal carcinoma in Egypt. *Cancer Lett* 159:79–86
- Benachenhou N, Guiral S, Gorska-Flipot I, Labuda D, Sinnett D (1998) High resolution deletion mapping reveals frequent allelic losses at the DNA mismatch repair loci hMLH1 and hMSH3 in non-small cell lung cancer. *Int J Cancer* 77:173–180
- Bessho T, Sancar A, Thompson LH, Thelen MP (1997) Reconstitution of human excision nuclease with recombinant XPF-ERCC1 complex. *J Biol Chem* 272:3833–3837
- Bürger R (2000) The mathematical theory of selection, recombination, and mutation, Wiley, New York
- Butkiewicz D, Rusin M, Enewold L, Shields PG, Chorazy M, Harris CC (2001) Genetic polymorphisms in DNA repair genes and risk of lung cancer. *Carcinogenesis* 22:593–597
- Caldecott KW, Aoufouchi S, Johnson P, Shall S (1996) XRCC1 polypeptide interacts with DNA polymerase beta and possibly poly (ADP-ribose) polymerase, and DNA ligase III is a novel molecular 'nick-sensor' in vitro. *Nucleic Acids Res* 24:4387–4394
- Cavalli-Sforza LL (1966) Population structure and human evolution. *Proc R Soc Lond B Biol Sci* 164:362–379
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes, Princeton University Press, Princeton
- Crompton NE, Ozsahin M (1997) A versatile and rapid assay of radiosensitivity of peripheral blood leukocytes based on DNA and surface-marker assessment of cytotoxicity. *Radiat Res* 147:55–60
- Demple B, Herman T, Chen DS (1991) Cloning and expression of APE, the cDNA encoding the major human apurinic endonuclease: definition of a family of DNA repair enzymes. *Proc Natl Acad Sci U S A* 88:11450–11454
- Duell EJ, Millikan RC, Pittman GS, Winkel S, Lunn RM, Tse CK, Eaton A, Mohrenweiser HW, Newman B, Bell DA (2001) Polymorphisms in the DNA repair gene XRCC1 and breast cancer. *Cancer Epidemiol Biomarkers Prev* 10:217–222
- Dybdahl M, Vogel U, Frenzt G, Wallin H, Nexø BA (1999) Polymorphisms in the DNA repair gene XPD: correlations with risk and age at onset of basal cell carcinoma. *Cancer Epidemiol Biomarkers Prev* 8:77–81
- Fan F, Liu C, Tavare S, Arnheim N (1999) Polymorphisms in the human DNA repair gene XPF. *Mutat Res* 406:115–120
- Fredman D, Siegfried M, Yuan YP, Bork P, Lehtvaslainen H, Brookes AJ (2002) HGVbase: a human sequence variation database emphasizing data quality and a broad spectrum of data sources. *Nucleic Acids Res* 30:387–391
- Fullerton SM, Bartoszewicz A, Ybazeta G, Horikawa Y, Bell GI, Kidd KK, Cox NJ, Hudson RR, Di Rienzo A (2002) Geographic and haplotype structure of candidate type 2 diabetes susceptibility variants at the calpain-10 locus. *Am J Hum Genet* 70:1096–1106
- Hartl DL, Clark AG (1989) Principles of population genetics, Sinauer, Sunderland
- Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, Ferak V, Villems R (1999) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* 9:1331–1334
- Kolodner R (1996) Biochemistry and genetics of eukaryotic mismatch repair. *Genes Dev* 10:1433–1442
- Labuda D, Krajcinovic M, Richer C, Skoll A, Sinnett H, Yotova V, Sinnett D (1999) Rapid detection of CYP1A1, CYP2D6, and NAT variants by multiplex polymerase chain reaction and allele-specific oligonucleotide assay. *Anal Biochem* 275:84–92
- Lahr MM, Foley RA (1994) Multiple dispersals and modern human origins. *Evol Anthropol* 3:48–60
- Lehmann AR (1995) Nucleotide excision repair and the link with transcription. *Trends Biochem Sci* 20:402–405
- Lewis PO, Zaykin D (2001) Genetic data analysis: Computer program for the analysis of allelic data. v. 1.0 (d16c)
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74:175–195
- Liu B, Nicolaides NC, Markowitz S, Willson JK, Parsons RE, Jen J, Papadopoulos N, Peltomaki P, de la Chapelle A, Hamilton SR, et al (1995) Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. *Nat Genet* 9:48–55
- Lunn RM, Langlois RG, Hsieh LL, Thompson CL, Bell DA (1999) XRCC1 polymorphisms: effects on aflatoxin B1-DNA adducts and glycophorin A variant frequency. *Cancer Res* 59:2557–2561

- Mohrenweiser HW, Jones IM (1998) Variation in DNA repair is a factor in cancer susceptibility: a paradigm for the promises and perils of individual and population risk estimation? *Mutat Res* 400:15–24
- Neel JV (1962) Diabetes mellitus: a “thrifty” genotype rendered detrimental by “progress”? *Am J Hum Genet* 14:353–362
- Pero RW, Bryngelsson C, Bryngelsson T, Norden A (1983) A genetic component of the variance of N-acetoxy-2-acetylaminofluorene-induced DNA damage in mononuclear leukocytes determined by a twin study. *Hum Genet* 65:181–184
- Pero RW, Johnson DB, Markowitz M, Doyle G, Lund-Pero M, Seidegard J, Halper M, Miller DG (1989) DNA repair synthesis in individuals with and without a family history of cancer. *Carcinogenesis* 10:693–697
- Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat Genet* 23:437–441
- Relethford JH (2001) *Genetics and the search for modern human origins*, Wiley, New York
- Rosenberg N, Murata M, Ikeda Y, Opare-Sem O, Zivelin A, Geffen E, Seligsohn U (2002) The frequent 5,10-methylenetetrahydrofolate reductase C677T polymorphism is associated with a common haplotype in whites, Japanese, and Africans. *Am J Hum Genet* 70:758–762
- Schneider S, Roessli D, Excoffier L (2000) Arlequin: A software for population genetics data analysis. v. 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland
- Shen MR, Jones IM, Mohrenweiser H (1998) Nonconservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans. *Cancer Res* 58:604–608
- Sung P, Bailly V, Weber C, Thompson LH, Prakash L, Prakash S (1993) Human xeroderma pigmentosum group D gene encodes a DNA helicase. *Nature* 365:852–855
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370