Polymorphisms of LIG4 and XRCC4 involved in the NHEJ pathway interact to modify risk of glioma

ARTICLE *in* HUMAN MUTATION · MARCH 2008

Impact Factor: 5.14 · DOI: 10.1002/humu.20645 · Source: PubMed

HUMAN MUTATION 29(3), 381–389, 2008

RESEARCH ARTICLE

Polymorphisms of *LIG4* and *XRCC4* Involved in the NHEJ Pathway Interact to Modify Risk of Glioma

Yanhong Liu,[1] Keke Zhou,[2] Haishi Zhang,[2] Yin Yao Shugart,[3,4] Lina Chen,[3] Zhonghui Xu,[1] Yu Zhong,[1] Hongliang Liu,[1] Li Jin,[1] Qingyi Wei,[5] Fengping Huang,[2] Daru Lu,[1]* and Liangfu Zhou[2]*

[1]State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China; [2]Neurosurgery Department of Huashan Hospital, Fudan University, Shanghai, China; [3]Department of Social Medicine, University of Bristol, Bristol, United Kingdom; [4]Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland; [5]Department of Epidemiology, M.D. Anderson Cancer Center, University of Texas, Houston, Texas

Communicated by Michael Dean

Although the role of environmental risk factors in the etiology of gliomas remains to be elucidated, accumulative epidemiological evidence suggests that genetic factors, such as variants in genes involved in DNA repair, may also play an important role. LIG4 and XRCC4 are known to form a complex and are functionally linked in the repair of double-stranded DNA breaks. To determine whether *LIG4* and *XRCC4* polymorphisms are associated with susceptibility to glioma and whether there are interactions between *LIG4* and *XRCC4*, we conducted a case–control study of 771 glioma patients and 752 cancer-free controls, assessed the associations between glioma risk and 20 tagging SNPs, and evaluated their potential gene–gene interactions using the multifactor dimensionality reduction (MDR), interaction dendrogram, and entropy analysis. In the single-locus analysis, only one variant, the *LIG4* SNP2 rs3093739:T>C (P-permutation = 0.009) was significantly associated with risk of developing glioma. Haplotype analysis revealed an association of glioma risk with genetic variants in *LIG4* block 1 (global P = 0.011), and *XRCC4* blocks 2 and 4 (both global P<0.0001). Moreover, the MDR analysis suggested a significant three-locus interaction model involving *LIG4* SNP4 rs1805388:C>T, *XRCC4* SNP12 rs7734849:A>T, and SNP15 rs1056503:G>T. Further dendrogram and graph analysis indicated a more-than-additive effect among these three loci. These results suggested that these variants may contribute to glioma susceptibility. Hum Mutat 29(3), 381–389, 2008. **© 2007 Wiley-Liss, Inc.**

KEY WORDS: tagging SNP; brain tumor; nonhomologous end-joining repair; haplotype; gene–gene interaction

INTRODUCTION

Glioma is a central nervous system (CNS) neoplasm derived from glial cells that surround and support neurons [Kleihues and Cavenee, 2000], and it is one of the most common brain tumors worldwide, accounting for the majority (~78%) of primary malignant brain tumors [Parkin et al., 2002]. Although there have been significant advances in the treatment of other types of cancers, only modest progress has been made in brain tumor treatment to date. Although exposure to ionizing radiation (IR) is believed to be associated with risk of glioma [Little et al., 1998; Bondy et al., 2001; Neglia et al., 2006], the role of environmental factors in the etiology of gliomas remains to be elucidated. Accumulative epidemiological evidence suggests that genetic factors, such as variants in genes involved in DNA repair, may also play an important role in the etiology of glioma [Chen et al., 2000b; Inoue et al., 2003; Wang et al., 2004; Parhar et al., 2005; Yang et al., 2005; Rao, 2007; Bohr et al., 2007].

IR induces various types of DNA damage, particularly double-strand breaks (DSBs) that are the major threats to the genomic integrity of cells, because DSBs may lead to either chromosomal aberrations or rearrangement events that may further lead to altered apoptosis [Khanna and Jackson, 2001]. Consequently, efficient mechanisms have evolved to repair such DNA damage. In eukaryotes, there are two major pathways for DSB repair: homologous recombination (HR) and nonhomologous end-joining (NHEJ), which differ in their requirement for a homologous template DNA and in the fidelity of DSB repair. Although HR is a

The Supplementary Material referred to in this article can be accessed at http://www.interscience.wiley.com/jpages/1059-7794/suppmat.

Received 28 April 2007; accepted revised manuscript 15 August 2007.

*Correspondence to: Daru Lu, State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, 220 Handan Rd., Shanghai 200433, China. E-mail: drlu@fudan.edu.cn; or Liangfu Zhou, Neurosurgery Department of Huashan Hospital, Fudan University, 12 Wulumuqi Zhong Rd., Shanghai 200040, China. E-mail: lfzhouc@online.sh.cn

Grant sponsor: China National Key Basic Research Program; Grant number: 2002CB512902; Grant sponsor: National "211" Environmental Genomics Grant.

Yanhong Liu and Keke Zhou contributed equally to this work.

DOI 10.1002/humu.20645
Published online 28 December 2007 in Wiley InterScience (www.interscience.wiley.com).

main pathway in yeast cells, NHEJ predominantly acts in mammalian cells [van Gent et al., 2001].

Five critical proteins, namely Ku70, Ku80, DNA-PKcs, DNA ligase IV (LIG4; MIM# 601837), and X-ray repair cross-complementing group 4 (XRCC4; MIM# 194363), function in the NHEJ pathway. In the NHEJ process [Critchlow et al., 1997; Chen et al., 2000a; Lieber et al., 2003], a DSB is first recognized by the DNA-PK complex (consisting of Ku70/Ku80 and DNA-PKcs), and then the LIG4/XRCC4 complex is recruited to perform the end-joining reaction. As a major protein involved in NHEJ, LIG4 forms a heterodimer with XRCC4 to execute the final rejoining step of NHEJ. XRCC4 is a nuclear phosphoprotein that interacts directly with LIG4, stabilizing and stimulating the LIG4 activity [Critchlow et al., 1997]. In addition, XRCC4 acts as a molecular bridge that links LIG4 to other components of the NHEJ apparatus [Leber et al., 1998]. Inactivation of either LIG4 or XRCC4 in mice is embryonically lethal, with an excessive neuronal apoptosis [Gao et al., 1998]. Mouse embryonic cells with disruption of either gene showed reduced proliferation, radiation hypersensitivity, chromosomal instability, and severely impaired V(D)J recombination [Barnes et al., 1998; Frank et al., 1998]. In addition, defects in human LIG4 cause the LIG4 syndrome (MIM# 606593) that is characterized by immunodeficiency and developmental and growth delay [O'Driscoll et al., 2001; Ben-Omran et al., 2005].

While several studies have investigated the role of the *LIG4* gene in the susceptibility of cancer [Goode et al., 2002; Kuschel et al., 2002; Roddam et al., 2002; Girard et al., 2004; Han et al., 2004], there are few studies on its partner, the *XRCC4* gene [Fu et al., 2003; Garcia-Closas et al., 2006; Allen-Brady et al., 2006]. Furthermore, there are no reported studies on the association between genetic variants in *LIG4* or *XRCC4* and glioma risk. Hence, we tested the hypothesis that variants in the *LIG4* and *XRCC4* genes may modulate risk of glioma. Because *LIG4* and *XRCC4* functionally form a heterodimer complex, we also tested the hypothesis that there are potential interactions between *LIG4* and *XRCC4* in the etiology of glioma.

MATERIALS AND METHODS

Patients

The study population and subject characteristics were previously described elsewhere [Liu et al., 2007]. In brief, this was a hospital-based case–control study including 771 glioma cases and 752 cancer-free controls. All the subjects were of Han Chinese origin and recruited from Shanghai and the surrounding provinces, including Zhejiang, Jiangsu, and Anhui in east China. Patients newly diagnosed with histopathologically confirmed glioma were consecutively recruited between October 2004 and May 2006 in the Department of Neurosurgery at Huashan Hospital of Fudan University without the restrictions of age, sex, and histology. The exclusion criteria included self-reported cancer history and previous radiotherapy and chemotherapy for unknown disease conditions. In addition, between July and August 2006, pathological and medical records of all cases were reviewed by trained abstractors to validate all diagnoses and to ascertain that there was no subsequent evidence of an occult primary tumor that was undiagnosed at the time of recruitment. The control subjects consisted of trauma outpatients (20%) recruited from the Emergency Medical Centre and hospital visitors (80%) who attended their annual checkup at the same hospital. These trauma outpatients did not differ from the annual checkup subjects by demographic data (data not shown). The exclusion criteria for the control subjects included known central nervous system–related diseases, self-reported history of cancer of any sites, and the history of radiotherapy and chemotherapy for unknown disease conditions. All the control subjects were frequency-matched to the cases on age (± 5 years), sex, and residence area (urban or rural).

Subjects were informed of the objectives, procedures, and voluntary nature of the study in the cover letter attached to the questionnaire. Signed informed consent to participate in the study was obtained before completion and return of the questionnaire. Consent for acquiring and reviewing medical records was obtained using a separate signed consent form. Among 900 eligible glioma patients and 950 eligible control subjects we identified, 771 case patients and 752 control subjects agreed to participate in this study; thus participation rates for cases and controls were 85.7% and 79.2%, respectively.

A structured brain tumor study questionnaire developed by the Department of Epidemiology at M.D. Anderson Cancer Center (provided by Dr. Melissa Bondy) served as a model to develop our own questionnaire in China, which is shorter than the original English version. Each participant was scheduled for a face-to-face interview with the structured questionnaire that elicited information on demographic factors, occupational IR exposure histories, and health characteristics. Occupational IR exposure history included work as a pilot, flight attendant, astronaut, uranium miner, industrial and nuclear power plant worker, or X-ray medical worker, radiology technologist, specialist physician, or other worker exposed to IR in the workplace.

After interview, a one-time sample of approximately 3 to 5 ml of venous blood was collected from each of the study participants. Genomic DNA was extracted from white blood cell fractions using the Qiagen Blood Kit (Qiagen, Chatsworth, CA). The study was approved by the Fudan University Ethics Committee for Human Subject Research.

Selection of the Tagging SNPs

To date, the International HapMap project (www.hapmap.org) has produced genotypes of over 3.8 million SNPs in 270 individuals of three different populations, making it possible to select tagging SNPs (tSNPs) with reasonable coverage for genome-wide association studies. In this study, we selected tSNPs by using genotype data obtained from the unrelated HapMap Chinese Han in Beijing (CHB) individuals (HapMap Data Rel 19/Phase, Oct 05, 2005, on NCBI B34 assembly, dbSNP b124).

For *LIG4* (13q22–q34; NT_00952.14), which contains two exons and spans approximately 8 kb, HapMap resequencing data available for *LIG4* covered the gene at a density of one SNP every 0.8 kb (10 SNPs). For *XRCC4* (5q13–q14; NT_006713.14) with 8 exons and approximately 276 kb in length, HapMap data had a density of one SNP every 2.5 kb (109 SNPs). Using the Haploview program (www.broad.mit.edu/mpg), we selected the tSNPs that had a minor allele frequency (MAF) greater than 0.05 in CHB. Four HapMap-based tSNPs of *LIG4* were selected by using an r^2 threshold of 0.8; we also included one additional functional SNP (A3 V rs1805389:C > T), which was in strong linkage disequilibrium (LD) with one of the selected tSNPs, T9I rs1805388:C > T [Roddam et al., 2002; Girard et al., 2004]. The five SNPs (four tSNPs and one additional functional SNP) had a SNP resolution of one SNP per 1.6 kb. In the case of *XRCC4* tSNP selection, because of the extensive haplotype diversity and weak LD across the entire gene (including 109 available SNPs and seven LD blocks in the HapMap CHB data), we used an r^2 threshold of 0.6, following the suggestions by Carlson et al. [2004] due to our

financial constraints. We did not find any common (MAF >0.05) and potential functional SNPs (e.g., nonsynonymous SNPs, promoter SNPs, and SNPs located within the enhancer region) in *XRCC4*. Thus, 15 tSNPs of *XRCC4* were selected at a resolution of one SNP per 18.3 kb. Overall, a total of 20 SNPs were chosen for this study, of which two were nonsynonymous SNPs, one synonymous SNP, two in the UTR region and 15 intronic variants (see Supplementary Table S1, available online at http://www.interscience.wiley.com/jpages/1059-7794/suppmat).

Genotyping Assays

All SNPs were genotyped using the fluorogenic 5′ nuclease TaqMan assay (ABI, Foster City, CA). The TaqMan primers and FAM- or VIC-labeled probes were designed using the Primer Express Oligo Design software v2.0 (ABI PRISM) and are available upon request. PCR reactions were performed in a reaction mixture of 5 μL containing 5 ng DNA, 2.5 μL 2 × TaqMan Universal PCR Master Mix, and 0.083 μL 40 × Assay Mix. Amplification conditions on ABI 9700 were as follows: 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 seconds, and 60°C for 1 minute. We scanned the completed PCR plates with ABI 7900HT Sequence Detector in the end point mode using the Allelic Discrimination Sequence Detector Software. The PCR assay for both case patients and control subjects were arrayed together in 384-well plates, with eight no-template controls (NTCs) and eight duplicated samples in each 384-well format as quality controls. On average, 97% of genotypes were successfully determined for all SNPs (see Supplementary Table S1).

Statistical Methods

Goodness-of-fit to the Hardy-Weinberg equilibrium (HWE) expectation in control subjects was assessed by the χ^2 test for each SNP. The Akaike information criterion (AIC) was used to determine the genetic model for each SNP [Akaike, 1974]. Genotype frequencies of case patients and control subjects were compared using the χ^2 test. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated by unconditional logistic regression analysis with adjustment for age and sex. We used 18 years old as the age cut point and divided the subjects into two groups (children, ≤ 18 years old, and adults >18 years old). We stratified the case patients into three subgroups of glioblastoma, astrocytomas except for glioblastoma (including diffuse astrocytomas, anaplastic astrocytomas, or other astrocytomas), and other gliomas (including oligodendrogliomas, ependymomas, or mixed glioma). To reduce the potential of spurious findings due to multiple testing, we applied 10,000 permutations to the empirically-derived P-values. All statistical tests were two-sided.

The LD among the polymorphisms was examined using Lewontin's standardized coefficient D' and LD coefficient r^2 [Lewontin, 1988]. We used the Haploview program to estimate the pairwise LD between markers and partition haplotype blocks. LD blocks defined by the method of LD confidence interval [Gabriel et al., 2002]. A pair of SNPs is defined to have "strong LD" if the one-sided upper 95% CI boundary on D' is >0.98 and the lower boundary is >0.7. The HAPLO.STATS package developed by Schaid et al. [2002] (www.mayo.edu/hsr/Sfunc.html) in the software language R was used for the haplotype analysis. This method, based on the generalized linear model framework, allows adjustment for possible confounding variables and provides both global and haplotype-specific tests. Haplotypes with a frequency of less than 0.03 were pooled into a combined group. Empirical P-values, based on 10, 000 simulations, were computed for the global score test and each of the haplotype-specific score tests.

Finally, a multistrategy of analyzing gene–gene interactions was employed. We first used the multifactor dimensionality reduction (MDR) method to detect and characterize locus–locus and gene–gene interaction models [Ritchie et al., 2001]. Then we further applied interaction dendrograms and graphs based on entropy (measurement of randomness) estimates as described by Jakulin and Bratko [2003] to confirm, visualize, and interpret the interactions models identified by MDR.

The MDR approach is described in detail by Ritchie et al. [2001], and reviewed by Moore and Williams [2002] and Moore et al. [2006]. At the heart of the MDR approach is a constructive induction algorithm that creates a new attribute by pooling genotypes from multiple SNPs. This method includes a combined cross-validation/permutation-testing procedure that minimizes false-positive results by multiple examinations of the data. Models that are true-positives are likely to be generalized to independent datasets and will have estimated testing accuracies of greater than 0.5. In addition to the testing accuracy, we also report the cross-validation consistency (CVC), a measure of how many times out of 10 divisions of the data that MDR found the same best model. Among this set of best multifactor models, the combination of genetic factors that maximizes the testing accuracy and/or the highest CVC is selected and further evaluated using permutation testing. The MDR analysis was performed by using the version 0.5.1 of the open-source MDR software package that is freely available online (www.epistasis.org/software.html).

Then, interaction dendrogram was employed. Jakulin and Bratko [2003] define the following dissimilarity measure, D(A,B), that is used by a hierarchical clustering algorithm to build a dendrogram. The value of 1,000 is used as an upper bound to scale the dendrograms.

$$D(A,B) = I(A;B;C)^{-1} \text{ if } I(A;B;C)^{-1} < 1,000 \text{ otherwise})$$

The attributes (i.e., SNPs) that are strongly interacting will hence appear close together at the leaves of the tree, and those that do not cointeract will appear far from one another. The color branches convey the type of the interaction, green is a weak interaction, red is a strong interaction, and blue is a negative interaction.

Finally, the third approach, Interaction graphs based on information theory, was used. Consider two attributes, A and B, and a class label C. Let H(X) be the Shannon entropy of X. The information gain (IG) of A, B, and C can be written as Equation (1) and defined in terms of Shannon entropy (Eqs. (2) and (3)).

$$IG(ABC) = I(A;BC) - I(A;B) \quad (1)$$

$$I(A;BC) = H(AC) + H(BC) - H(A,BC) \quad (2)$$

$$I(A;B) = H(A) + H(B)H(A,B) \quad (3)$$

The first term in Equation (1), I(A;BC), measures the interaction of A and B. The second term, I(A;B), measures the dependency or correlation between A and B. If this difference is positive, then there is evidence for an attribute interaction that cannot be linearly decomposed. If the difference is negative, then the information between A and B is redundant. If the difference is zero, then there is evidence of conditional independence or a mixture of synergy and redundancy. Briefly, synergy means positive interactions, while redundancy means negative interactions.

Interaction graphs were compromised of a node for each attribute (i.e. SNPs) with pairwise connections between them. The percentage of entropy removed (i.e., IG) by each attribute was visualized for each node and the percentage of entropy removed for each pairwise product of attributes was visualized for each connection. Both the interaction dendrograms and interaction graphs were created using the Orange machine learning software package, which is written in Python and provided for free as open-source [Curk et al., 2005].

RESULTS

Sample Characteristics

The distributions of selected characteristics between glioma patients and controls were previously described elsewhere [Liu et al., 2007]. Overall, case patients and control subjects were adequately matched on age and sex (P = 0.935, and 0.780, respectively). For occupational IR exposure histories, 21 glioma cases and nine cancer-free controls reported a history of occupational exposure to IR (P = 0.047). Also, case patients were significantly more likely to report a family history of cancer than the controls in their first-degree relatives (P = 0.010). Of the 771 cases, 295 had glioblastoma, 242 had astrocytomas except for glioblastoma, and 234 had other gliomas.

Individual SNP Association Analysis

Genotyping results for the quality control analysis were 100.0% (32/32) concordant, positive (duplicates), or negative (NTC). The SNP IDs, locations, and allele frequencies are given in Supplementary Table S1. All genotype distributions in the controls were consistent with those expected from the HWE (all P > 0.05) (see Supplementary Table S1). The genotype distributions of 20 selected tSNPs in *LIG4* and *XRCC4* in the cases and controls are summarized in Supplementary Table S2. In the single-locus analyses, the allele frequencies of three SNPs were significantly different between the cases and the controls: *LIG4* SNP2 rs3093739:T>C (P = 0.002), *XRCC4* SNP7 rs13161662:A>G (P = 0.004), and *XRCC4* SNP8 rs2731858:G>A (P = 0.023), but only the association between *LIG4* SNP2 and glioma risk remained significant after 10,000 permutations (P = 0.009). In addition, a marginal evidence was observed for *XRCC4* SNP7 after 10,000 permutations (P = 0.054).

Further logistic regression analyses revealed that in the dominant-effect model as assessed by the AIC, compared with wild-type carriers, significant protective effects were associated with *XRCC4* SNP7 genotypes (adjusted OR = 0.75, 95% CI = 0.61–0.92 for *AG/GG*), whereas significantly increased risks were associated with *LIG4* SNP2 genotypes (adjusted OR = 1.41, 95% CI = 1.13–1.76 for *TC/CC*), and *XRCC4* SNP8 (adjusted OR = 1.28, 95% CI = 1.04–1.59 for *GA/AA*). Meanwhile, in the recessive-effect model (assuming that only the rare homozygotes affect the risk of gliomas), binary logistic regression analyses revealed that a significant protective effect were associated with variant homozygotes of SNP9 rs7715771:A>T in *XRCC4*. Subjects carrying the *TT* genotype had a significantly decreased risk of gliomas (adjusted OR = 0.44, 95% CI = 0.20–0.89) compared with those carrying the combined *AA/AT* genotype (see Supplementary Table S2).

We further evaluated the associations of the *LIG4* SNP2 and *XRCC4* SNP7 variant genotypes with glioma risk stratified by selected variables and histological types. As shown in Table 1, compared with the common homozygous genotype, the effect of combined variant genotypes were both more evident in female subjects (adjusted OR = 1.82, 95% CI = 1. 25–2.65 for *LIG4* SNP2 *TC/CC*, and adjusted OR = 0.62, 95% CI = 0.44–0.87 for *XRCC4* SNP7 *AG/GG*), and cases with glioblastoma (adjusted OR = 1.52, 95% CI = 1.13–2.04 for *LIG4* SNP2 *TC/CC* and adjusted OR = 0.70, 95% CI = 0.51–0.93 for *XRCC4* SNP7 *AG/GG*). However, the increased risk associated with the *LIG4* SNP2 *TC/CC* variant genotype was more pronounced in the adults (adjusted OR = 1.47, 95% CI = 1.16–1.86), whereas the decreased risk associated with the *XRCC4* SNP7 *AG/GG* was more evident in the children (adjusted OR = 0.45, 95% CI = 0.24–0.84).

Haplotype Block Structure and LD Analysis

Figure 1 shows plots of the pairwise LD (*D′*) values for the tSNPs and LD structures of each gene. The LD plot indicates that, for *LIG4*, two blocks with high LD were identified, block 1 (SNPs 1–2; size = 82 bp) encompassing the upstream region intron 1, and block 2 (SNPs 3–5; size = 3 kb) spanning the majority of the gene from exon 2 through the 3′ UTR (Fig. 1A); for *XRCC4*, we identified the following four regions of strong LD: block 1 (SNPs 1–3; size = 18 kb) covering the upstream region 5′ UTR; intron 1, block 2 (SNPs 4–7; size = 70 kb) spanning the majority of the gene from intron 2 through intron 6; block 3 (SNPs 8–11; size = 89 kb) including intron 6 and intron 7; and block 4 (SNPs 12–15; size = 9 kb) ranging from intron 7 to the 3′ UTR predominantly (Fig. 1B).

TABLE 1. **Stratified Analyses Between the *LIG4* SNP2 and *XRCC4* SNP7 Genotypes and Glioma Risk**

	***LIG4* SNP2 (rs3093739 T>C)**			***XRCC4* SNP7 (rs13161662 A>G)**		
		Adjusted OR (95% CI)[a]			**Adjusted OR (95% CI)[a]**	
Variables	**Number[b]**	***CC* vs. *TT***	***TC/CC* vs. *TT***	**Number[b]**	***GG* vs. *AA***	***AG/GG* vs. *AA***
Age (years)						
Children (≤18)	**83/84**	**0.19 (0.03–1.74)**	**0.92 (0.46–1.85)**	**85/83**	**0.31 (0.07–1.27)**	**0.45 (0.24–0.84)**
Adults (>18)	**682/663**	**2.21 (1.20–4.08)**	**1.47 (1.16–1.86)**	**667/652**	**0.76 (0.49–1.18)**	**0.81 (0.65–1.00)**
Sex						
Male	**459/459**	**1.54 (0.74–3.20)**	**1.17 (0.86–1.59)**	**450/457**	**0.45 (0.25–0.83)**	**0.88 (0.66–1.16)**
Female	**294/288**	**4.93 (1.61–15.06)**	**1.82 (1.25–2.65)**	**290/278**	**1.22 (0.62–2.42)**	**0.62 (0.44–0.87)**
Histological types						
Glioblastoma	**292/747**	**2.36 (1.19–4.66)**	**1.52 (1.13–2.04)**	**286/735**	**0.92 (0.52–1.39)**	**0.70 (0.51–0.93)**
Astrocytomas except for glioblastoma[c]	**241/747**	**1.28 (0.54–3.04)**	**1.20 (0.86–1.66)**	**238/735**	**0.69 (0.36–1.30)**	**0.82 (0.60–1.11)**
Other gliomas[d]	**232/747**	**1.72 (0.76–3.87)**	**1.49 (1.08–2.06)**	**228/735**	**0.38 (0.16–0.85)**	**0.71 (0.52–0.96)**

[a]Adjusted for age and sex, accordingly.
[b]Numbers of cases/controls.
[c]Astrocytomas except for glioblastoma, including diffuse astrocytomas, anaplastic astrocytomas, and other astrocytomas.
[d]Other gliomas including oligodendrogliomas, ependymomas, or mixed gliomas.

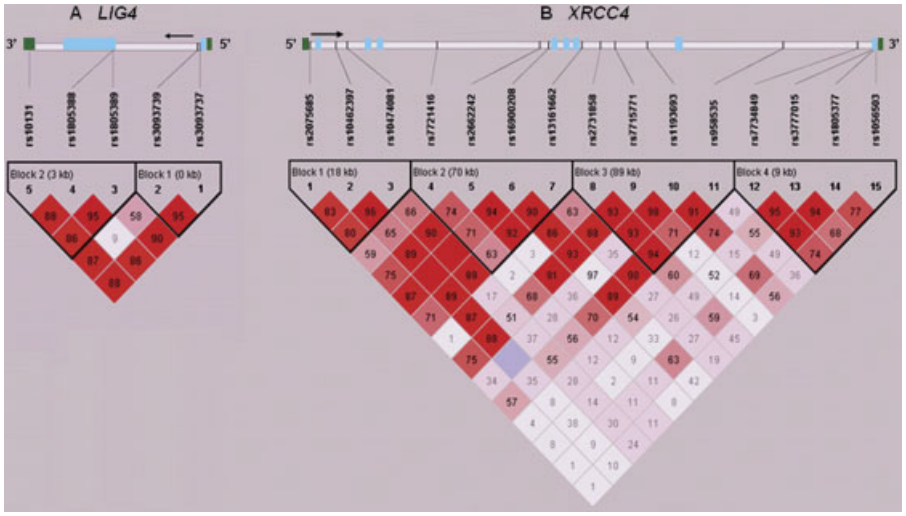

FIGURE 1. **Graphical representation of the SNP locations and LD structure of *LIG4* and *XRCC4*. The SNP distribution and haplotype-block structure across the *LIG4* (A) and *XRCC4* (B) genes are shown, respectively. The arrows illustrate the transcription direction of the genes. Dark-shaded boxes represent the UTRs, and light-shaded boxes represent coding exons. The exact SNP positions are listed in Supplementary Table S1. Haplotype blocks were defined according to the criteria laid out by Gabriel et al. [2002]. The measure of LD (*D′*) among all possible pairs of SNPs is shown graphically according to the shade of color, where white represents very low *D′* and dark represents very high *D′*. The numbers in squares are *D′* values (D′ × 100). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]**

Haplotype Analysis

Table 2 summarizes the associations between frequencies of the haplotypes and risk of glioma. For *LIG4*, two risk haplotypes, 1-B (adjusted OR = 1.30, 95% CI = 1.04–1.61) and 1-C (adjusted OR = 1.52, 95% CI = 1.08–2.13), were found for block 1. Consistent with the individual SNP analyses, of the four observed haplotypes in block 1, both haplotypes 1-B and 1-C carried the SNP2 variant that was individually associated with an increased risk. No association was observed for the *LIG4* block 2. For *XRCC4*, one risk haplotype 2-C (adjusted OR = 1.64, 95% CI = 1.23–2.17) and one protective haplotype 2-E (adjusted OR = 0.21, 95% CI = 0.12–0.37) were found for block 2; and in block 4, two risk haplotypes, 4-B (adjusted OR = 1.28, 95% CI = 1.03–1.59) and 4-C (adjusted OR = 1.61, 95% CI = 1.19–2.22), and haplotype 4-D, which was associated with a 10-fold reduction in risk of glioma (adjusted OR = 0.10, 95% CI = 0.05–0.19), were identified. No haplotypes in *XRCC4* block 1 or block 3 were found to be significantly associated with risk of glioma. On the other hand, the global test score showed statistically significant differences in the haplotype profile between cases and controls for *LIG4* block 1 (global P = 0.011) and *XRCC4* blocks 2 and 4 (both global P < 0.0001).

Locus–Locus and Gene–Gene Interactions

A total of 20 tSNPs in *LIG4* and *XRCC4* were included in the MDR analysis. Figure 2A summarizes the best interaction models obtained from the MDR analysis. In the one-locus model, *XRCC4* SNP7 was the best attribute for predicting glioma risk (testing accuracy = 50.65%; CVC = 6; P = 0.172). But the best interaction model was a three-locus model (i.e., *LIG4* SNP4 rs1805388:C > T, and *XRCC4* SNP12 rs7734849:A > T and SNP15 rs1056503:G > T), with an improved testing accuracy to 63.14% (CVC = 9; P = 0.001). The two-locus model (i.e., *XRCC4* SNP12 and SNP15) and the four-locus model (i.e., *XRCC4* SNP4 rs7721416:G > A, SNP5 rs2662242:T > C, SNP12, and SNP15) also had significant probability values (both P = 0.001). However, these two-locus and four-locus models had lower CVC (7.0 and 5.0, respectively) or lower testing accuracy (61.68% and 66.55%, respectively) than that in the three-locus model (Fig. 2A).

The MDR analysis indicated that the three-locus model was the best model but did not specify whether there was a synergistic relationship. Therefore, we applied the interaction dendrogram and graph to determine this relationship. Figure 2B illustrates the interaction dendrogram for these models. The hierarchical cluster analysis placed *LIG4* SNP4 and *XRCC4* SNP12 and SNP15 on the same branch, and their distinguishingly closer position in the diagram clearly shows that the three-locus model may have a synergistic interaction effect on modulating risk of glioma. *XRCC4* SNP4 and SNP5 were on another branch, revealing an interaction between them. *XRCC4* SNP7 was located on a different remote branch, suggesting that this SNP may have no relationship with other SNPs.

Consistent with the results in the MDR and interaction dendrogram, the interaction entropy graph (Fig. 2C) supported the three-locus model that had an interaction effect in the absence of a main effect. We found that small percentages of the entropy in the case–control status were explained by *LIG4* SNP4 (0.08%), *XRCC4* SNP12 (0.06%) or SNP15 (0.02%) when considered independently, but a large percentage of the entropy was explained by the pairwise interactions between the SNPs of the three-locus model (0.13%, 0.34%, and 8.73%, respectively), which indicated a synergistic interaction. In addition, the *XRCC4* SNP7 had a main effect (0.34%) that was independent of the other SNPs.

DISCUSSION

This haplotype-based study represents a comprehensive evaluation of associations between variants of the *LIG4* and *XRCC4* genes and glioma risk. Of the 20 selected tSNPs, two SNPs (*LIG4* SNP2 and *XRCC4* SNP7) showed a significant association with

TABLE 2. Associations Between LIG4 and XRCC4 Haplotypes in Each Block and Glioma Risk

Gene and block	Haplotype	Haplotype frequency			Logistic regression[a]		Global test[b]
		Total	Cases	Controls	OR (95% CI)	*P*	
LIG4							
Block 1 (SNPs 1–2)							P = 0.011
	1-A (*CT*)	0.832	0.812	0.853	1.00		
	1-B (*GC*)	0.113	0.124	0.101	1.30 (1.04–1.61)	0.023	
	1-C (*CC*)	0.051	0.060	0.042	1.52 (1.08–2.13)	0.014	
	1-D (*GT*)	0.004	0.004	0.004	0.72 (0.23–2.27)	0.584	
Block 2 (SNPs 3–5)							P = 0.253
	2-A (*CCG*)	0.658	0.645	0.676	1.00		
	2-B (*CCA*)	0.121	0.126	0.112	1.14 (0.91–1.43)	0.251	
	2-C (*TTG*)	0.112	0.108	0.113	1.01 (0.80–1.28)	0.907	
	2-D (*CTG*)	0.105	0.115	0.091	1.32 (0.98–1.75)	0.068	
	Others[c]	0.007	0.006	0.008	1.33 (0.49–3.57)	0.563	
XRCC4							
Block 1 (SNPs1–3)							P = 0.512
	1-A (*GCC*)	0.793	0.791	0.794	1.00		
	1-B (*TCC*)	0.099	0.094	0.104	0.91 (0.71–1.16)	0.470	
	1-C (*TTG*)	0.056	0.058	0.054	1.07 (0.78–1.49)	0.642	
	1-D (*TCG*)	0.034	0.039	0.029	1.39 (0.92–2.08)	0.124	
	Others[c]	0.019	0.018	0.019	0.97 (0.55–1.72)	0.923	
Block 2 (SNPs 4–7)							P < 0.0001
	2-A (*GTAA*)	0.532	0.565	0.500	1.00		
	2-B (*GTAG*)	0.238	0.231	0.242	0.85 (0.70–1.02)	0.077	
	2-C (*ACAA*)	0.088	0.115	0.058	1.64 (1.23–2.17)	0.0007	
	2-D (*ACGA*)	0.054	0.066	0.042	1.39 (0.97–1.96)	0.069	
	2-E (*GCAA*)	0.032	0.010	0.057	0.21 (0.12–0.37)	< 0.0001	
	Others[c]	0.058	0.013	0.101	0.13 (0.08–0.22)	< 0.0001	
Block 3 (SNPs 8–11)							P = 0.391
	3-A (*GAAG*)	0.473	0.465	0.482	1.00		
	3-B (*GAAA*)	0.177	0.177	0.177	1.03 (0.83–1.27)	0.802	
	3-C (*GTAA*)	0.126	0.119	0.131	0.93 (0.74–1.18)	0.543	
	3-D (*AAAA*)	0.120	0.128	0.112	1.19 (0.94–1.49)	0.139	
	3-E (*AAGA*)	0.074	0.080	0.067	1.25 (0.93–1.67)	0.131	
	Others[c]	0.030	0.030	0.031	0.99 (0.60–1.64)	0.979	
Block 4 (SNPs 12–15)							P < 0.0001
	4-A (*ATAG*)	0.667	0.701	0.633	1.00		
	4-B (*TTGT*)	0.147	0.172	0.121	1.28 (1.03–1.59)	0.028	
	4-C (*TCGT*)	0.074	0.094	0.054	1.61 (1.19–2.22)	0.002	
	4-D (*ATAT*)	0.046	0.007	0.085	0.10 (0.05–0.19)	< 0.0001	
	Others[c]	0.065	0.027	0.107	0.33 (0.23–0.46)	< 0.0001	

[a]Adjusted for age and sex.
[b]Generated by 10,000 permutation test.
[c]Haplotypes with a frequency less than 0.03 were pooled into a combined group.

glioma risk. The haplotype analysis revealed that significant differences in the haplotype profile for *LIG4* block 1 and *XRCC4* blocks 2 and 4. Moreover, the MDR analysis identified a significant three-locus interaction model involving *LIG4* SNP4 and *XRCC4* SNP12 and SNP15. The interaction dendrogram and graph analyses helped us further interpret the nature of the interaction models and revealed interactions among these three loci. This is, to the best of our knowledge, the first and largest study that describes tSNPs in *LIG4* and *XRCC4* and their potential relationship with glioma risk. None of the polymorphisms in our study had been examined previously in glioma association studies. Furthermore, our study is among few that have examined the gene–gene interaction in a specific repair pathway in the etiology of glioma.

Human XRCC4 is a 336–amino acid protein that has a long helical stem domain (block 2, residues 119–203) responsible for multimerization and interaction with LIG4 [Junop et al., 2000; Sibanda et al., 2001]. For LIG4, interaction with XRCC4 occurs via the region that lies between the two C-terminal BRCT domains (block 2, residues 751–800) [Girard et al., 2004]. *LIG4* SNP4 T9I was located in exon 2 (block 2), *XRCC4* SNP12 was in intron 7 and SNP15 S307S was in exon 8 (block 4). While none of these SNPs is located in the interaction regions where the two proteins are known to interact, it is unknown which SNPs are biologically responsible for the observed statistical interaction, because our statistical interaction did not provide any information for the link of the typed SNPs to the untyped SNPs that may be biologically plausible and disease-causing.

Several previous studies have assessed SNPs in *LIG4* for their associations with risk of cancers but their results were mixed [Roddam et al., 2002; Fu et al., 2003; Garcia-Closas et al., 2006; Wu et al., 2006]. It is reported that the two linked polymorphisms in the N-terminal of LIG4—A3 V and T9I, a region shown to be essential for the *LIG4* ligation activity—were significantly associated with a reduction in the risk of multiple myeloma [Roddam et al., 2002], although neither SNP is thought to alter the structural conformation of the LIG4 protein. However, other studies did not confirm such an association [Fu et al., 2003; Garcia-Closas et al., 2006; Wu et al., 2006]. Here we also failed to find a significant association of these two variants with risk of glioma. However, our MDR analysis revealed that *LIG4* SNP4 (T9I) interacted with *XRCC4* SNP12 and SNP15 to modulate the risk of glioma, a finding with biological plausibility, although the contributory role of A3 V and T9I in cancer risk needs further validation. In addition, it is noted that the MAF of the *LIG4* SNP3 (A3 V) was 0.053, as reported in the NCBI dbSNP, and 0.07 in a European study [Roddam et al., 2002] but was much higher in our Chinese study population (MAF = 0.115, 0.118, for cases and

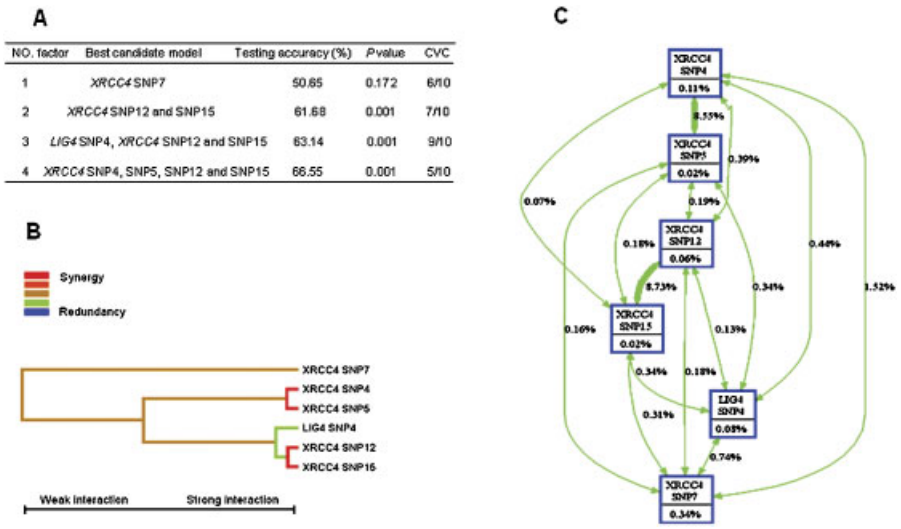

FIGURE 2. **The MDR models, interaction dendrogram, and interaction graph for gene–gene interactions on glioma risk. A: Summary of the MDR interaction models. B: Interaction dendrogram. The color indicates the strength of the dependence, green is weak, and red is strong. The hierarchical cluster analysis placed *LIG4* SNP4, *XRCC4* SNP12 and SNP15 on the same branch but *XRCC4* SNP4 and SNP5 on another branch; *XRCC4* SNP7 was located at a different remote branch. C: Orange Canvas Interaction entropy graph. Each SNP is shown in a box with the percent of entropy below the label (main effect). Two-way interactions between SNPs are depicted as an arrow accompanied by a percent of entropy explained by that interaction (interaction effect). In the interaction graph, the most important attribute, *XRCC4* SNP7 polymorphism, alone eliminates 0.34% of class entropy and has the largest univariate effect. Only small percentages of the entropy in the case–control status explained by *LIG4* SNP4 (0.08%), *XRCC4* SNP12 (0.06%), or *XRCC4* SNP15 (0.02%), when considered independently, but a large percentage of entropy was explained by their pairwise interactions (0.13%, 8.73%, and 0.34%, respectively), indicating a synergistic interaction.**

controls, respectively), which is compatible with a previous report in Taiwanese (MAF = 0.133) [Fu et al., 2003].

Although a number of studies have investigated the association between cancer risk and the polymorphisms of *LIG4*, little is known about *XRCC4* in terms of the potential impact of its genetic variation on repair capability and consequent risk of carcinogenesis. Two SNPs of *XRCC4*, one splicing-site polymorphism (SNP14 rs1805377:A>G) and one intronic polymorphism (SNP1 rs2075685:G>T), have been studied, and the results are conflicting [Fu et al., 2003; Garcia-Closas et al., 2006; Allen-Brady et al., 2006]. We found no evidence of an association of these two *XRCC4* variants with risk of glioma.

The failure to replicate the published results of some single-locus analyses might be the inclusion of different diseases and ethnic groups or geographical areas. Another explanation for the lack of consistency in published association studies on complex diseases is the complex underlying genetic architecture, in which locus–locus and gene–gene interactions are the expected phenomenon rather than the exception [Moore and Williams, 2002]. Thus, genetic association studies that ignore epistasis or gene–gene interactions are likely to reveal only a partial component of the genetic etiology. It is conceivable that genetic factors for glioma may also be multifactorial. Therefore, testing for joint effect is necessary in the search for glioma susceptibility factors. However, such complex multifactor interactions are often difficult to detect and characterize by using traditional parametric statistical methods, such as logistic regression analysis, because of the insufficient power and sparseness of the data in a high order of dimensions. To address this problem, we adopted a multistrategy for the detection, characterization, and interpretation of gene–gene interactions between *LIG4* and *XRCC4*.

MDR is a nonparametric and model-free approach to detect and characterize nonlinear interactions among discrete genetic and environmental attributes [Ritchie et al., 2001]. With MDR, multilocus genotypes are pooled into high-risk and low-risk groups, effectively reducing the dimensionality of the genotype predictors (i.e., attributes) from N dimensions to one dimension. Although MDR is a promising new approach for overcoming some of the limitations of logistic regression and has been successfully applied in a variety of common human diseases [Ritchie et al., 2001; Moore and Williams, 2002; Cho et al., 2004; Tsai et al., 2004; Ma et al., 2005], the MDR method is not the panacea for dealing with multifactorial, dimensional data. Perhaps the greatest challenge is the interpretation of epistasis models. Therefore, we have further explored interaction graphics and interaction dendrograms analysis for the statistical interpretation of epistasis effects. As the results showed, the three-locus model (i.e., *LIG4* SNP4 and *XRCC4* SNP12 and SNP15) in *LIG4* and *XRCC4* genes, which are known to form a dimeric complex, is the best model for the prediction of the risk of glioma. It should be noted that, although the three SNPs were not associated with glioma in the single-locus analyses, they might act in concert to modulate the risk of glioma.

The most significant finding in this study was the allelic and haplotypic association of *XRCC4* SNP7 rs13161662:A>G (located in block 2) with glioma risk. In the single-locus association analysis, *XRCC4* SNP7 exhibited a statistically significant main effect (P = 0.004). Also, haplotype analysis revealed that *XRCC4* block 2 was strongly associated with glioma risk (global P<0.0001). Consistent with allelic association results, MDR further confirmed the main effect of *XRCC4* SNP7. It was the strongest one-locus model for predicting glioma risk in the MDR interaction model. Further interaction entropy analysis showed that *XRCC4* SNP7 had a main effect that was independent of the other SNPs. Thus, this significant association appeared to be consistent across all analyses, which strongly suggests that *XRCC4* SNP7 or a genetic variant in LD with it, is involved in the etiology of glioma.

Perhaps the most striking finding was the association between *XRCC4* block 4 and glioma risk. In the haplotype analysis, *XRCC4* block 4 was associated with glioma risk (global $P < 0.0001$), and two risk haplotypes, 4-B and 4-C, and a highly protective haplotype 4-D were found to be associated with 10-fold reduction in the effect on glioma risk. More interestingly, haplotype block 4 did not contain any SNPs that were associated with glioma risk in the single-locus analysis. Subsequent gene–gene interaction analysis helped explain this paradigm. The best interaction model revealed that *XRCC4* SNP12 and SNP15 (both located in block 4) interact with *LIG4* SNP4, which was more evident in the interaction entropy analysis. Only small percentages of the entropy were removed by *LIG4* SNP4, *XRCC4* SNP12 or SNP15 when considered independently (consistent with the single-locus analysis results), but a large percentage of the entropy was explained by the pairwise interactions between these three loci, indicating a synergistic interaction. Taken together, these consistent and robust results provided evidence of epistasis or gene–gene interaction, suggesting that the effect of one gene may be too weak to be detected, particularly when the effect of another functionally-linked gene was not accounted for.

A strength of this study was the relatively large sample size (771 glioma patients and 752 control subjects) among glioma association studies published to date. Several studies have investigated putative associations between various SNPs and glioma [Chen et al., 2000b; Inoue et al., 2003; Wang et al., 2004; Parhar et al., 2005; Yang et al., 2005], but the number of glioma patients in these studies has rarely exceeded 500.

Despite the strengths and biologic plausibility of the associations observed in our study, there are limitations inherent in most retrospective association studies, including our current study. First, our results may overestimate the true size of effect or identify spurious associations due to possible population stratification. In this study, the frequency of glioblastomas (38.26%) is substantially lower than those in most reports from the west (about 50%), which may have led to selection bias. However, this bias is unlikely to be of significance in this report, because our frequency of glioblastomas was similar to a Hong Kong study (33%), suggesting that the racial factors may be affecting the incidence and relative frequency of glioblastoma [Tham and Saw, 1989]. Moreover, the observed allelic frequencies of the genotyped SNPs showed consistent results with the International HapMap Project data and NCBI dbSNP database. The study population was of homogenous ethnic background: Han Chinese in east China; given the relatively high response rates in both cases and controls (85.7% vs. 79.2%), we believe that, bias due to selection, if any, should not be substantial.

Second, because of the relatively low number of subjects exposed to possible IR in the study population, we were not able to evaluate the gene–environment interactions underlying risk of glioma, though exposure to IR is a well-established risk factor for glioma. Generally speaking, epidemiologic studies of occupational exposure to IR face a number of obstacles with respect to assessment of the dose–response relationship in the low-dose region [Boice et al., 2000]. Some believe that epidemiologic studies of low-dose IR are not feasible, because precise estimation of what appear to be small risk require "impracticably" large samples. Given these uncertainties, no single population can securely provide enough information to distinguish the absence of effects from the small effects at low doses.

Finally, because the SNPs examined in this study were chosen to maximize SNP tagging for genetic variation rather than for functionality of a gene, the observed associations between glioma risk and tSNPs should be interpreted in the context of the presence of LD with untyped SNPs that are possibly functional. Further studies are required to identify potential causal variants by more detailed genetic mapping, including sequencing and functional characterization of variants. Meanwhile, because this study is in a Chinese population, the finding merits further investigation in larger and/or other ethnic groups.

In summary, our study demonstrated that some representative tSNPs of *LIG4* and *XRCC4* may modulate the risk of glioma. In particular, our results support the notion that some association or genetic effect that may not be detectable in a single-locus analysis can be unraveled when locus–locus and gene–gene interactions are considered. Therefore, our findings highlight the importance of taking into account such interactions in the etiology of glioma.

ACKNOWLEDGMENTS

We thank Xilan Mei, Jian Yu, and Mei Chong for recruiting the subjects and data management, and Yin Wang and Wenting Wu for laboratory assistance. We also thank Dr. Melissa Bondy for providing the M.D. Anderson brain tumor questionnaire. We are most indebted to the participants and staff at the Huashan Hospital of Fudan University whose help made this study possible.

REFERENCES

Akaike H. 1974. A new look at the statistical model identification. IEEE Trans Automat Contr 19:716–723.

Allen-Brady K, Cannon-Albright LA, Neuhausen SL, Camp NJ. 2006. A role for XRCC4 in age at diagnosis and breast cancer risk. Cancer Epidemiol Biomarkers Prev 15:1306–1310.

Barnes DE, Stamp G, Rosewell I, Denzel A, Lindahl T. 1998. Targeted disruption of the gene encoding DNA ligase IV leads to lethality in embryonic mice. Curr Biol 8:1395–1398.

Ben-Omran TI, Cerosaletti K, Concannon P, Weitzman S, Nezarati MM. 2005. A patient with mutations in DNA ligase IV: clinical features and overlap with Nijmegen breakage syndrome. Am J Med Genet A 137:283–287.

Bohr VA, Ottersen OP, Tonjum T. 2007. Genome instability and DNA repair in brain, ageing and neurological disease. Neuroscience 145:1183–1186.

Boice JD Jr, Blettner M, Auvinen A. 2000. Epidemiologic studies of pilots and aircrew. Health Phys 79:576–584.

Bondy ML, Wang LE, El-Zein R, de Andrade M, Selvan MS, Bruner JM, Levin VA, Alfred Yung WK, Adatto P, Wei Q. 2001. Gamma-radiation sensitivity and risk of glioma. J Natl Cancer Inst 93:1553–1557.

Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. 2004. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 74:106–120.

Chen L, Trujillo K, Sung P, Tomkinson AE. 2000a. Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase. J Biol Chem 275:26196–26205.

Chen P, Wiencke J, Aldape K, Kesler-Diaz A, Miike R, Kelsey K, Lee M, Liu J, Wrensch M. 2000b. Association of an ERCC1 polymorphism with adult-onset glioma. Cancer Epidemiol Biomarkers Prev 9:843–847.

Cho YM, Ritchie MD, Moore JH, Park JY, Lee K-U, Shin HD, Lee HK, Park KS. 2004. Multifactor-dimensionality reduction shows a two-locus interaction associated with Type 2 diabetes mellitus. Diabetologia 47:549–554.

Critchlow SE, Bowater RP, Jackson SP. 1997. Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. Curr Biol 7:588–598.

Curk T, Demsar J, Xu QK, Leban G, Petrovic U, Bratko I, Shaulsky G, Zupan B. 2005. Microarray data mining with visual programming. Bioinformatics 21:396–398.

Frank KM, Sekiguchi JM, Seidl KJ, Swat W, Rathbun GA, Cheng HL, Davidson L, Kangaloo L, Alt FW. 1998. Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. Nature 396:173–177.

Fu Y-P, Yu J-C, Cheng T-C, Lou MA, Hsu G-C, Wu C-Y, Chen S-T, Wu H-S, Wu P-E, Shen C-Y. 2003. Breast cancer risk associated with genotypic polymorphism of the nonhomologous end-joining genes: a multigenic study on cancer susceptibility. Cancer Res 63:2440–2446.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. 2002. The structure of haplotype blocks in the human genome. Science 296:2225–2229.

Gao Y, Sun Y, Frank KM, Dikkes P, Fujiwara Y, Seidl KJ, Sekiguchi JM, Rathbun GA, Swat W, Wang J, Bronson RT, Malynn BA, Bryans M, Zhu C, Chaudhuri J, Davidson L, Ferrini R, Stamato T, Orkin SH, Greenberg ME, Alt FW. 1998. A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. Cell 95:891–902.

Garcia-Closas M, Egan KM, Newcomb PA, Brinton LA, Titus-Ernstoff L, Chanock S, Welch R, Lissowska J, Peplonska B, Szeszenia-Dabrowska N, Zatonski W, Bardin-Mikolajczak A, Struewing JP. 2006. Polymorphisms in DNA double-strand break repair genes and risk of breast cancer: two population-based studies in USA and Poland, and meta-analyses. Hum Genet 119:376–388.

Girard PM, Kysela B, Harer CJ, Doherty AJ, Jeggo PA. 2004. Analysis of DNA ligase IV mutations found in LIG4 syndrome patients: the impact of two linked polymorphisms. Hum Mol Genet 13:2369–2376.

Goode EL, Dunning AM, Kuschel B, Healey CS, Day NE, Ponder BA, Easton DF, Pharoah PP. 2002. Effect of germ-line genetic variation on breast cancer survival in a population-based study. Cancer Res 62:3052–3057.

Han J, Hankinson SE, Ranu H, De Vivo I, Hunter DJ. 2004. Polymorphisms in DNA double-strand break repair genes and breast cancer risk in the Nurses' Health Study. Carcinogenesis 25:189–195.

Inoue R, Isono M, Abe M, Abe T, Kobayashi H. 2003. A genotype of the polymorphic DNA repair gene MGMT is associated with de novo glioblastoma. Neurol Res 25:875–879.

Jakulin A, Bratko I. 2003. Analyzing attribute dependencies. In: Lavrac N, Gamberger D, Blockeel H, Todorovski L, editors. Proceedings of the 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), Cavtat-Dubrovnik, Croatia, September 22–26, 2003. Lecture Notes in Computer Science, vol 2838, Lecture Notes in Artificial Intelligence. New York: Springer. p 229–240.

Junop MS, Modesti M, Guarne A, Ghirlando R, Gellert M, Yang W. 2000. Crystal structure of the Xrcc4 DNA repair protein and implications for end joining. EMBO J 19:5962–5970.

Khanna KK, Jackson SP. 2001. DNA double-strand breaks: signaling, repair and the cancer connection. Nat Genet 27:247–254.

Kleihues P, Cavenee WK. 2000. Pathology and genetics of tumors of the nervous system. Lyon: IARC Press.

Kuschel B, Auranen A, McBride S, Novik KL, Antoniou A, Lipscombe JM, Day NE, Easton DF, Ponder B, Pharoah P, Dunning A. 2002. Variants in DNA double-stranded break repair genes and breast cancer susceptibility. Hum Mol Genet 11:1399–1407.

Leber R, Wise TW, Mizuta R, Meek K. 1998. The XRCC4 gene product is a target for and interacts with the DNA-dependent protein kinase. J Biol Chem 273:1794–1801.

Lewontin RC. 1988. On measures of gametic disequilibrium. Genetics 120:849–852.

Lieber MR, Ma Y, Pannicke U, Schwarz K. 2003. Mechanism and regulation of human non-homologous DNA end-joining. Nat Rev Mol Cell Bio 4:712–720.

Little MP, de Vathaire F, Shamsaldin A, Oberlin O, Campbell S, Grimaud E, Chavaudra J, Haylock RG, Muirhead CR. 1998. Risks of brain tumour following treatment for cancer in childhood: modification by genetic factors, radiotherapy and chemotherapy. Int J Cancer 78:269–275.

Liu Y, Zhang H, Zhou K, Chen L, Xu Z, Zhong Y, Liu H, Li R, Shugart YY, Wei Q, Jin L, Huang F, Lu D, Zhou L. 2007. Tagging SNPs in nonhomologous end-joining pathway genes and risk of glioma. Carcinogenesis 28:1906–1913.

Ma DQ, Whitehead PL, Menold MM, Martin ER, Ashley-Koch AE, Mei H, Ritchie MD, DeLong GR, Abramson RK, Wright HH, Cuccaro ML, Hussman JP, Gilbert JR, Pericak-Vance MA. 2005. Identification of significant association and gene-gene interaction of GABA receptor subunit genes in autism. Am J Hum Genet 77:377–388.

Moore JH, Williams SM. 2002. New strategies for identifying gene-gene interactions in hypertension. Ann Med 34:88–95.

Moore JH, Gilbert JC, Tsai CT, Chiang FT, Holden T, Barney N, White BC. 2006. A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. J Theor Biol 241:252–261.

Neglia JP, Robison LL, Stovall M, Liu Y, Packer RJ, Hammond S, Yasui Y, Kasper CE, Mertens AC, Donaldson SS, Meadows AT, Inskip PD. 2006. New primary neoplasms of the central nervous system in survivors of childhood cancer: a report from the Childhood Cancer Survivor Study. J Natl Cancer Inst 98:1528–1537.

O'Driscoll M, Cerosaletti KM, Girard P-M, Dai Y, Stumm M, Kysela B, Hirsch B, Gennery A, Palmer SE, Seidel J, Gatti RA, Varon R, Oettinger MA, Neitzel H, Jeggo PA, Concannon P. 2001. DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency. Mol Cell 8:1175–1185.

Parhar P, Ezer R, Shao Y, Allen JC, Miller DC, Newcomb EW. 2005. Possible association of p53 codon 72 polymorphism with susceptibility to adult and pediatric high-grade astrocytomas. Mol Brain Res 137:98–103.

Parkin DM, Whelan SL, Feraly J, Teppo L, Thomas DB. 2002. Cancer incidence in five continents. Lyon: IARC Press.

Rao KS. 2007. Mechanisms of disease: DNA repair defects and neurological disease. Nat Clin Pract Neurol 3:162–172.

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. 2001. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 69:138–147.

Roddam PL, Rollinson S, O'Driscoll M, Jeggo PA, Jack A, Morgan GJ. 2002. Genetic variants of NHEJ DNA ligase IV can affect the risk of developing multiple myeloma, a tumour characterised by aberrant class switch recombination. J Med Genet 39:900–905.

Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. 2002. Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet 70:425–434.

Sibanda BL, Critchlow SE, Begun J, Pei XY, Jackson SP, Blundell TL, Pellegrini L. 2001. Crystal structure of an Xrcc4-DNA ligase IV complex. Nat Struct Biol 8:1015–1019.

Tham KT, Saw D. 1989. Central nervous system tumours in Hong Kong: consideration of racial factors affecting incidence. J Hong Kong Med Assoc 41:164–168.

Tsai CT, Lai LP, Lin JL, Chiang FT, Hwang JJ, Ritchie MD, Moore JH, Hsu KL, Tseng CD, Liau CS, Tseng YZ. 2004. Renin-angiotensin system gene polymorphisms and atrial fibrillation. Circulation 109:1640–1646.

van Gent DC, Hoeijmakers JHJ, Kanaar R. 2001. Chromosomal stability and the DNA double-stranded break connection. Nat Genet 2:196–206.

Wang L-E, Bondy ML, Shen H, El-Zein R, Aldape K, Cao Y, Pudavalli V, Levin VA, Yung WKA, Wei Q. 2004. Polymorphisms of DNA repair genes and risk of glioma. Cancer Res 64:5560–5563.

Wu X, Gu J, Grossman HB, Amos CI, Etzel C, Huang M, Zhang Q, Millikan RE, Lerner S, Dinney CP, Spitz MR. 2006. Bladder cancer predisposition: a multigenic approach to DNA-repair and cell-cycle-control genes. Am J Hum Genet 78:464–479.

Yang P, Kollmeyer TM, Buckner K, Bamlet W, Ballman KV, Jenkins RB. 2005. Polymorphisms in GLTSCR1 and ERCC2 are associated with the development of oligodendrogliomas. Cancer 103:2363–2372.