Introduction

Nasopharyngeal carcinoma (NPC) is a malignancy of epithelial cell origin that occurs in the retro-nasal cavity (1). The 5-year survival rate is 88% for patients at Stage I, but dramatically drops down to 28% for those at late Stage IVB (2), suggesting that early diagnosis is a favorable prognostic factor. However, the diagnosis is largely delayed due to the nonspecific clinical presentations. Therefore, dissecting the risk factors for NPC could provide fundamental elements to develop approaches for NPC prediction so as to improve early diagnosis.

It has been proposed that the etiology of NPC is a multi-stage process involving genetic components, infection of Epstein-Barr virus (EBV) and exposures to environmental carcinogens (3,4). Lines of evidence have implicated the link between genetic lesions and NPC risk. Firstly, the incidence of NPC is prevalent in southern China, northern Africa, and Alaska (5), showing remarkable geographic distribution. Moreover, the second and third generation Chinese emigrants in American have higher NPC incidence than the local Caucasian (6). Secondly, 10% of NPC patients have family history, either in Chinese or Caucasian populations (7-9). In addition, Wee et al. proposed an interesting hypothesis of NPC origin from the ancient Bai-Yue Chinese tribe, based on the worldwide concordance of NPC incidence rates and history of Chinese migrations (10). Many genetic studies have been conducted to address susceptibility genes of NPC, by using linkage approach or association analysis, where some risk genes have been identified (11,12). Apart from genetic susceptibility, NPC has been widely recognized as an EBV-associated cancer, mainly attributed to the observations of elevated EBV DNA load and the EBV-related antibodies in peripheral blood, as well as clonal EBV strain in tumor cells, in NPC patients (1,13).

Herein, we will summarize the current findings of
NPC risk loci, mainly focusing on those consistently being replicated and also propose future challenges in uncovering the mystery of NPC proneness in the endemic area.

**Research updates on genetic susceptibility loci of NPC**

Linage analysis and association study are two major strategies to identify genes leading to NPC risk. Four linkage studies have been reported, with the latest one in 2008, where susceptibility loci including 6p21 (14), 4p15.1−q12 (15), 3p21.31−21.2 (16), and 5p13 (17) have been implicated in NPC families of Chinese origins. The results are not concordant, since none of the studies provided supporting linkage evidence for the others. This might be partially explained by the heterogeneity of population or NPC subtype, or simply lack of statistical power (12). Case-control association study is a more common approach to dissect susceptibility genes of NPC, which is to test the association of genetic variants and NPC risk. Most case-control association studies of NPC were conducted as a candidate gene-based design, which requires prior knowledge or hypothesis on the functional relevance of the candidate genes to NPC development. Genes involved in the immune regulation, metabolism of carcinogens, DNA damage and repair, as well as tumorigenesis have been examined for their associations with NPC. More detailed information can be found in another review paper (11). By contrast, genome-wide association study or GWAS is a hypothesis-free approach, allowing association tests at genome-wide level (18). For NPC, four GWASs have been conducted in Chinese populations in Malaysia, Taiwan, and southern China (19-22), which revealed susceptibility loci of ITGA49 at chromosome 3p22.2, HLA-A and GABBR1 at 6p22.1, HLA-B/C, and MICA at 6p21.33, HLA-DQ/DR at 6p21.32, MECOM at 3q26.2, CDKN2A/2B at 9q21.3, and TNFRSF19 at 13q12.12, respectively. More detailed information has been summarized in another review paper (12).

**HLA loci**

Since NPC has been associated with EBV, association of immune-related genes especially HLA and NPC have been intensively studied. The association of HLA loci with NPC risk is the most consistent finding on the whole. This includes the first linkage study involving 30 sibships of NPC families from southern China, Singapore and Malaysia, which revealed a recessive susceptibility gene conferring an increased risk of 20.9 (95% CI=5.1 to infinite) for NPC (14). A GWAS in Taiwan Chinese also pointed to a polymorphism downstream of the HLA-A gene (rs2517713, \(P=5.54\times 10^{-12}\) (22). Another GWAS with much larger sample size in southern Chinese (5,090 cases and 4,957 controls plus 279 trios) further confirmed the strong genetic effect of HLA on NPC susceptibility, by revealing three independent associations at rs2860580 (\(P_{\text{combined}}=4.88\times 10^{-67}\), OR=0.58), rs2894207 (\(P_{\text{combined}}=3.42\times 10^{-13}\), OR=0.61), and rs28421666 (\(P_{\text{combined}}=2.49\times 10^{-18}\), OR=0.67) (20). The two top SNPs rs2517713 and rs2894207 from the two studies are in complete linkage (\(r^2=0.99\)), suggesting that they might be tagging the same causal variant in this region. Moreover, imputation analysis suggested that the top SNP rs2860580 tagged HLA-A*1101 as a protective allele (OR=0.56; \(P=3\times 10^{-18}\)) (20), which is consistent with the previous association studies on HLA alleles and NPC in Chinese (23,24). Both GWASs reported independent HLA associations, suggesting the complexity of causal lesions at the HLA region. HLA is the most gene dense region in human genome, encoding more than 250 genes including several key immune response genes (25). It's also a region with strong linkage disequilibrium and under strong selections (25). Better understanding the mechanism of the independent associations at this region requires further efforts on fine-mapping the HLA alleles and haplotypes (26).

**TERT/CLPTM1L locus**

TERT/CLPTM1L has been reported as susceptibility locus for multiple cancers (27,28). Variations at the locus have been associated with NPC risk, including rs401681 in Hong Kong Chinese (OR=0.77; \(P=1\times 10^{-4}\)) with moderate sample size and candidate gene approach (29), rs402710 in Thailand population (OR=0.56; \(P=3\times 10^{-18}\)) (20), which is consistent with the previous association studies on HLA alleles and NPC in Chinese (23,24). Both GWASs reported independent HLA associations, suggesting the complexity of causal lesions at the HLA region. HLA is the most gene dense region in human genome, encoding more than 250 genes including several key immune response genes (25). It's also a region with strong linkage disequilibrium and under strong selections (25). Better understanding the mechanism of the independent associations at this region requires further efforts on fine-mapping the HLA alleles and haplotypes (26).
different genetic lesions conferring NPC risk. Consistently, an imputation-based fine-mapping study in this region demonstrated six independent associations with risk of multiple cancers, suggesting the pleiotropic mechanisms at this locus (34). Therefore, at least TERT and CLPTM1L at the locus should be paid attentions. TERT or telomerase reverse transcriptase is the catalytic subunit of the telomerase complex, which is a ribonucleoprotein polymerase in maintaining telomeric ends (35,36). It’s well documented that telomerase expression and activity may be involved in tumorigenesis (37). As a key component in NPC, shorter telomere length was reported in tumors as compared with that in the para-tumor tissues and in chronic nasopharyngitis (38). Moreover, latent membrane protein 1 (LMP1), a known oncopgenic protein encoded by EBV, was able to induce the elongation of telomere in NPC cell line (39). On the other hand, CLPTM1L is required for Ras-induced oncogenic transformation and anchorage-independent growth, and its depletion in lung cancer cells resulted in smaller tumors in the xenograft model, which strongly suggested the tumorigenic role of CLPTM1L (40). Again, further investigations are needed to pinpoint the exact genetic lesions in the locus leading to NPC.

**Other loci**

Besides those summarized previously (11,12), association studies have been continuously conducted between candidate genes and NPC risk, such as those related to DNA damage repair and oxidative stress pathways. Alleles in nitric oxide synthase (NOS; NOS3−786C, NOS3+894T, and NOS2−277G) and glutathione-S transferases (GSTs; GSTT1 (del/del genotype) were prevalent in Tunisians NPC patients (41). However, the association of GSTT1 gene and NPC risk was not significant in a large-scale meta-analysis consisted of 1,295 cases and 1,967 controls (42). Large-scale meta-analysis may provide better statistical power to obtain a pooled estimate of effect to detect or reject associations. Two separated meta-analyses identified the association of GSTM1 del/del genotype with increased risk of NPC (42,43). Recently, a meta-analysis involving 9,705 NPC cases and 11,041 controls from 34 case-control studies supported the susceptibility of polymorphism of XRCC1 (Arg399Gln), MMP-1 (1G/2G), CYP2E1 (Rsal), MMP2 (−1306C>T) and TP53 (Arg72Pro) with NPC and reject the association of XRCC1 (Arg194Trp and Arg280His) and MDM2 (309T>G) (44). Alternatively, pathway-based studies could comprehensively evaluate susceptibilities of genes according to their joint effects as functional units (45).

The polymorphisms of XRCC1 and APE1 in base exclusion repair pathway were shown jointly contributing to the increased NPC risk in Chinese population (46). In addition, polymorphism of cytokine-related genes have been examined for their associations with NPC, such as IL10 (47,48), IL1A (49), IL16 (50) and IL18 (51,52). However, more validation efforts are required to solidify the associations, as that for IL-1A is controversial with TT genotype (889C>T) being either protective or risk for NPC in two studies (49,53).

The genetic study of NPC has been extended to test associations between polymorphisms and NPC progression. The Argonaute 2 gene (AGO2) polymorphism rs3928672 was demonstrated to confer risk for lymph mode metastasis in southern Chinese (54). The CELF2 (rs3740194) (55), TP53 (Arg72Pro) (56,57), ERCC1 [Cys8092Ala (58,59) and Gln504Lys (60)] and XRCC1 (Arg399Gln) (59) were shown to be potential prognostic biomarkers, though the effect of ERCC1 (Cys8092Ala) was contradictory for its lack of association with relapse-free survival or overall survival in another study (61).

**Future challenges**

Accumulating studies have identified many susceptibility loci or genes associated with NPC, and emphasized the possible participation of EBV with the risk genes such as HLA-A, TERT and TNFRSF19, supporting the hypothesis of multifactorial involvement for NPC development. On one hand, many efforts are awaited to precisely locate the causal variants in the risk loci and work out their mechanisms leading to NPC, especially the HLA region with the greatest genetic effect. On the other hand, the estimated genetic effect sizes for NPC risk are less than two according to GWASs, meaning that these are the “low-hanging fruit” and more substantial genetic susceptibility genes remain to be identified (62). With sufficient knowledge of risk genes of NPC, we might be able to develop better prediction model for NPC risk.

**More loci to be identified to explain the missing heritability of NPC**

Like other complex diseases, NPC is mostly a sporadic disease, with a small proportion (<10%) of familial cases, where the genetic susceptibility has been confirmed. The spectrum of genetic lesions ranges from low to high...
frequencies for risk alleles, and from high to low genetic effects or penetrance, correspondingly (62). Susceptibility variants identified by linkage studies are in general of high penetrance but less frequent in population, while those revealed by GWASs are largely common variants with low penetrance. However, for complex diseases including NPC, there is still a large missing gap to fully explain the disease heritability (62). The contribution of the genetic loci on NPC was estimated under the threshold model, which is assuming a normal distribution of liability (risk) toward the threshold trait and individual with liability above a certain threshold actually having the trait (63). In an empirical estimation of NPC prevalence rate of 1/1,000 in Cantonese, the seven SNPs reported in the previous GWAS (rs1572072, rs9510787, rs1412829, rs28421666, rs2894207, rs2860580 and rs6774494) (20) could jointly explain only 2.1% genetic variance, meaning that more additional loci are yet to be discovered. Toward further discovery, well characterized phenotype (i.e., the precise subtype), and minimized impact of population stratification are important to ensure sufficient statistical power. More importantly, we need to extend wider coverage of the variations spectrum and risk factors.

**Contribution of X chromosome loci to NPC risk**

A common feature of NPC incidence is the male preponderance, other than geographic and ethnic proneness. In most populations, the incidence rate is 2- to 3-fold higher in males than females (64,65). The sexual difference may be partially attributed to the unequal exposures of environmental risk factors between males and females, like smoking, diet habit, and so on. At genetic aspect, the involvement of X-chromosome variations in NPC development has been hypothesized (10,26). However, these remain unexploited at chromosomal level for NPC, and previous GWASs have excluded X-chromosome variations from analyses. For other complex diseases such as Grave’s disease (66), Schizophrenia (67) and fasting insulin and height (68), novel X-linked susceptibility genes have been identified, accounting for their “missing heritability”. The major challenge for genetic association study of X-chromosome could be the random X-inactivation in female, which helps balance the total allele dosages between genders by silencing one copy of the female allele (69,70). The unique feature of X chromosome imposes difficulty in association study and makes X-linked association less straightforward to interpret as compared to that for the autosomal chromosomes. To address these, several statistical methods have been proposed, such as estimation of the variance explained by X chromosome (71), and association test on X-linked loci (72-74). For the methodology including the data pre-processing, quality control, association test and result interpretation has been summarized elsewhere (75,76).

**Contribution of rare variations to NPC risk**

DNA variations are the basis of genetic susceptibility, where two major disputable hypotheses have been proposed. The ‘common disease common variant’ hypothesis suggests that a causal variant underlying a risk locus with common frequency might contribute minor to moderate effects on a nearby gene or gene with long-range LD (77). The recent GWASs have identified many common variants for different diseases (loci could be retrieved at GWAS catalog; https://www.ebi.ac.uk/gwas/). Four GWAS of NPC revealed some risk loci, but whether the common variants are functional or tagging the nearby causal variants remain unclear. The ‘rare variant hypothesis’ proposed that a significant proportion of the inherited susceptibility to common diseases may be due to the joint effects of a certain low frequency variants from different genes, each conferring a moderate risk effect (78). Such rare variant might render founder effects and better explain the familial clustering of complex disease including NPC. Moreover, Rivas et al., used the next-generation sequencing technology to fine-map the previous GWAS loci and identified several rare variants in the coding regions with greater genetic effects for inflammatory bowel diseases (79), supporting the ‘synthetic association’ concept that the common loci by GWAS might tag many rare variants with greater functional variations for different affected individuals (80). In any scenario, further investigations of rare variants in NPC are pending to figure out their contribution to the susceptibility, which would come soon at the era of high throughput and next-generation sequencing technologies.

**Role of EBV infection in the NPC prevalence**

NPC incidence is high in some certain areas, where the environmental factors such as diets and the viral exposures are diverse, in addition to the different genetic background. These variable factors might independently or jointly lead to the specific prevalence of the disease. NPC is rare but well known as EBV-related malignancy, however, EBV infection is very common in general population. These lead to a hypothesis that there might be an NPC specific
EBV subtype contributing to the high incidence of NPC in the endemic regions. Many studies have been carried out to identify the NPC-related EBV subtype or strain [please refer to references (81,82)]. Most recently, we identified a polymorphism at EBV encoded gene, RPMS1 (locus 155391 G>A), to be significantly associated with NPC in southern and northern Chinese (OR=5.27, 95% CI=4.31–6.44, P<0.001); moreover, the frequencies of the EBV variant are significantly correlated with the incidence rates worldwide; in addition, the variant is likely associated with NPC but not other EBV-related diseases (83). These provide strong evidence of the existence of NPC-specific EBV subtype. Moreover, the threshold model showed that the EBV polymorphism (155391G>A) can explained 5.5% of the variance, two times more than that by the genetic loci as mentioned earlier. These suggest that the contribution of EBV to NPC prevalence or proneness should receive more attentions.

Taken together, NPC is a complex disease with multifactorial contributions from genetic susceptibility, environmental exposures and EBV infection; although some progresses have been made in identifying genetic and viral factors to NPC risk, there is still a huge gap to fully explain the heritability or prevalence of NPC. Efforts are awaited to bridge the gap, until when NPC risk prediction could be helpful for effective population screening of individuals with high NPC-risk. As moving forward, a genetic risk score model integrating the seven GWAS loci, environmental risk factors (consumption of salted fish and preserved vegetables and cigarette smoking) and family history of NPC showed discriminatory ability of 0.74 according to the area under the receiver-operating characteristic curves (AUC) (84). We included EBV subtype [RPMS1, locus 155391 G>A, reference (83)] in the model and observed a significant improvement in the discriminatory ability (AUC=0.88) in a small cohort, though further validations are pending.

Acknowledgements

Funding: This work was partly supported by the National Natural Science Foundation of China (81101544 and 81222035), 973 program of China (2011CB504302), and 863 program of China (2012AA02A206).

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

68. Tukiainen T, Pirinen M, Sarin AP, et al. Chromosome X-wide association study identifies Loci for fasting insulin and height and evidence for incomplete dosage


Cite this article as: Bei JX, Zuo XY, Liu WS, Guo YM, Zeng YX. Genetic susceptibility to the endemic form of NPC. Chin Clin Oncol 2016;5(2):15. doi: 10.21037/ccco.2016.03.11