CHAPTER II

REVIEW OF RELATED LITERATURES

1.Characteristic and aetiological factor of NPC

The first reported of NPC was in 1837 which published in a paper entitled 'Primary Carcinoma of Nasopharynx : A table of cases'.[14] NPC was an endemic disease that the highest incidence group were found in Southern China, South-East Asia, Arctic population and tribes of North Africa.[1] NPC was a tumour of epidermal origin and has been divided from histopathology by WHO into three broad categories.

1. Keratinizing squamous carcinoma.
2. Non-keratinizing carcinoma.
3. Undifferentiated carcinoma of nasopharyngeal type.

Epidemiological studies of NPC reveal that major aetiological factors included the environmental factor, genetic factor and Epstein-Barr virus. (EBV) [15,16]

1.Environmental factor

The geographic differences in incidence of NPC indicated that the development of this cancer should be associated with environmental factor. Most studies have evaluated the two groups of environmental factors.

-Dietary exposures

Previous studies of traditional lifestyle of the Cantonese Chinese showed the evidence of association between NPC and salted food such as salted fish, salted leafy vegetables, salted root and salted egg.[16-19] In addition, fresh pork/beef meats and alcoholic beverage showed positive association.[20] It was thought that these food contained chemical carcinogen. Many studies have shown that salted fish contains high level of volatile nitrosamine inducing nasal tumours in animals.[21] However, the association between increased risk of NPC and dietary of nitrosamine and nitrite during childhood did not show the same result as in adult. [22]

-Non-dietary exposures

Much of the early research showed association between disease and some exposure material, which always used in high-risk area. Firstly, tobacco in the form of cigarettes, cigars and pipes have a moderately relationship.[23] Secondly, occupational exposure to products of combustion displayed a strong risk factor.[24] Thirdly, agricultural workers and carpenters were found to correlate with increase risk of NPC. Similarly, occupational exposure to formaldehyde only increases risk in squamous cell carcinoma.[25] However, cotton dust appeared as a protective factor, which the explanation remains unclear.[24]

2.Genetic susceptibility

Many of the migrant studies were reported to support the hypothesis of genetic risk factors. Having evidence to support this, higher trend in the risk was showed in emigrant Chinese to California, Australia and Singapore.[26-29] In addition, there were numerous reported of familial nasopharyngeal cancer.[30,31] Currently at least three genetic susceptibility loci have been documented. These included human leukocyte antigen (*HLA*), cytochrome P450 2E1 (*CYP2E1*) and polymeric immunoglobulin receptor (*PIGR*).

-Human leukocyte antigen (HLA)

HLA is the major histocompatibility complex in the human. The association between HLA and NPC has first been discovered in Singapore, which showed that the disease correlated with HLA A2 and B locus blank.[32] The latter was later identified as the B46 antigen. The linkage between the A2/B46 haplotype and NPC was substantiated in several studies carried out in Chinese patients in Malaysia, Hong Kong, China and California, U.S.A. As the physiological function of HLA is antigen presentation, one hypothesis is that these particular HLA class I antigens may not present EBV antigens efficiently, thereby allowing the EBV to persist in nasopharyngeal epithelium cells. One candidate antigen is the EBV LMP-2 gene product, which HLA A2.1 (A*0201) restricts in the generation of a cell-mediated response.[33] However, by applying high-resolution HLA

typing of A2, we showed that A* 0201 was presented in 39.5 percent of NPC cases compared with only 15.7 percent of normal control subjects, the reverse of what the model predicted. Therefore, A*0201 is unlikely to be the factor in the EBV clearance hypothesis, at least not in relation to LMP-2. In addition, HLA in non-Chinese patients with NPC is not associated with A2 and B46. These data suggest that HLA might not be the susceptibility gene per se, but there is the NPC susceptibility locus closely linked to HLA region. [34]

-Cytochrome P450 2E1 (*CYP2E1*)

CYP2E1 is an enzyme involved in the metabolic activation of various procarcinogen such as N-nitrosodimethylamine (NDMA) and tobacco specific nitrosamine N-nitro-sonornicotine (NNN).[35,36] Evidence from previous study suggested that NPC was correlated with nitrosamine exposure [37], salted fish [18] and tobacco smoke. [24,38] Furthermore, rats that were fed with nitrosamine-containing salted fish developed to nasal cavity tumours.[21] Interestingly, the association of *CYP2E1* and NPC was observed in Taiwan and Thailand. The Taiwanese correlation study showed that individuals homozygous for the variant allele (-/-) of *CYP2E1* detectable by *RsaI* digestion were at an increased risk for NPC development OR(95%CI)=2.6(1.2-5.7).[6,7] Thai presented OR (95%CI)=2.1(0.90-4.96).[8]

-Polymeric immunoglobulin receptor (*PIGR*)

The structure and function of *PIGR* will be discussed later. The first association between *PIGR* and NPC was reported in Kongruttanachok's thesis. [9] An intron polymorphism of *PIGR* showed association in Chinese patient (OR=2.81; 95%CI =1.52-5.34). Therefore, *PIGR* may be a susceptibility gene for NPC development and may help indicate the way that EBV entering nasopharyngeal epithelium. Nevertheless, further confirmation is needed especially the searching for a specific altering functional mutation. In addition, if the mutation can be found in other population, in addition to Chinese, the association must be discovered.

3. Epstein-Barr Virus (EBV)

EBV is a ubiquitous human herpesviruses which categorized in subfamily gammaherpesviridae. It has icosadeltahedral shape and 172 kilobases double strand DNA. Seroepidermiological studies indicated that EBV was widespread in all human population, in which the great majority of adults having antibodies against the virus.[39,40] NPC was shown to have the strongest association with this virus. [41] Serologically, high titer of IgA-EBV complex antibody was detected specifically from patients with NPC and people developing to NPC.[13] In addition, Southern Blot analysis of EBV DNA in this tumour showed monoclonal origin indicating that EBV infected the tumour cell before the clonal expansion.[42] Furthermore, EBV latent gene products, EBERs, EBNA-I, LMP-I, LMP-2 and BamHI, were detected in the cancer cells. [33] It was believed that EBV's protein played interference roles in cellular genes function.

EBV infection caused a variety of diseases that difference from primary to latent infection (Fig 1).

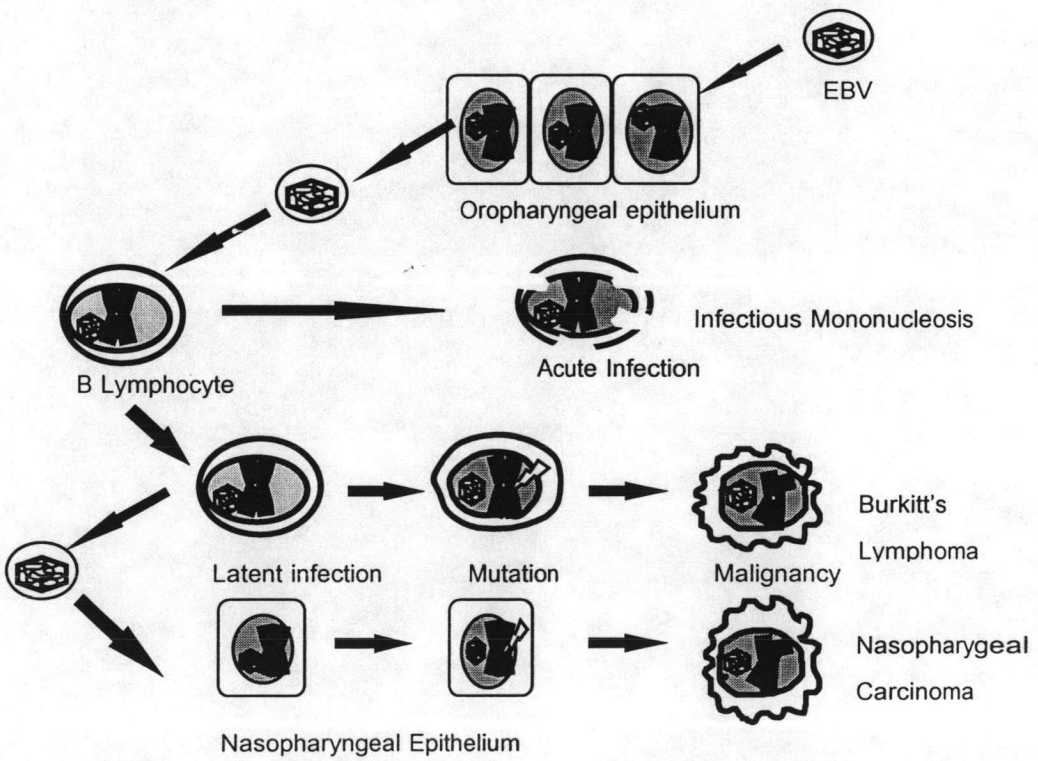

Figure 1 A model of the Epstein-Barr virus infection and disease development

Primary infection was occurred at oropharyngeal epithelium. Then, EBV replicated, lysed cells and infected B-lymphocytes by CR2 receptor, Cd21,C3d. After infection, the virus might reenter the lytic phase and caused the infectious mononucleosis disease. Within some infected cells, however, the virus became latent. With additional host mutations, the B cells would transform into several types of B cell tumors such as lymphomas of the immunosuppressed, endemic Burkitt's lymphoma and Hodgkin's disease. Furthermore, the latent EBV could be released from the B cell and infected other cell types, such as gland, smooth muscle and nasopharyngeal epithelium. Consequently, with the inclusion of specific host mutation cancer might be developed such as salivary gland tumor, leiomyosarcoma and NPC. Among these, NPC was the most specific to EBV and most frequent malignancy in the world.[43] Nasopharyngeal EBV receptor was an important protein to be identified since it played the important role in the initiation process of NPC but not EBV associated B cell lymphoma. Nevertheless, no protein has yet been proved to be responsible for this process.

2.Polymeric Immunoglobulin Receptor

The mucosal surfaces are generally covered by a physically monolayered epithelium which is constantly exposed to infectious agent such as bacteria, virus, fungi and parasites and also to soluble molecules from the diet or the environment. Facing all these exogenous challenges, numerous innate defense mechanisms have developed, and they cooperate intimately with a local adaptive immune system to inhibit epithelial colonization of pathogens and mucosal penetration.[44] The first line of specific immunological defense is mainly provided by secretory IgA (SIgA) and to some extent also secretory IgM (SIgM) antibodies. As a basis for this external antibody transport, local B-cell blasts and plasma cells produce polymeric IgA (pIgA, dimers and larger polymers) and to lesser extent pentameric IgM, which both can bind to the epithelial polymeric immunoglobulin receptor (pIgR), also known as the transmembrane secretory component (SC) expressing by secretory epithelium. [45,46] In addition, SIgA and SIgM are

the secretory IgA and IgM containing bound SC that occurs by cleavage of pIgR at the apical surface of the epithelium.

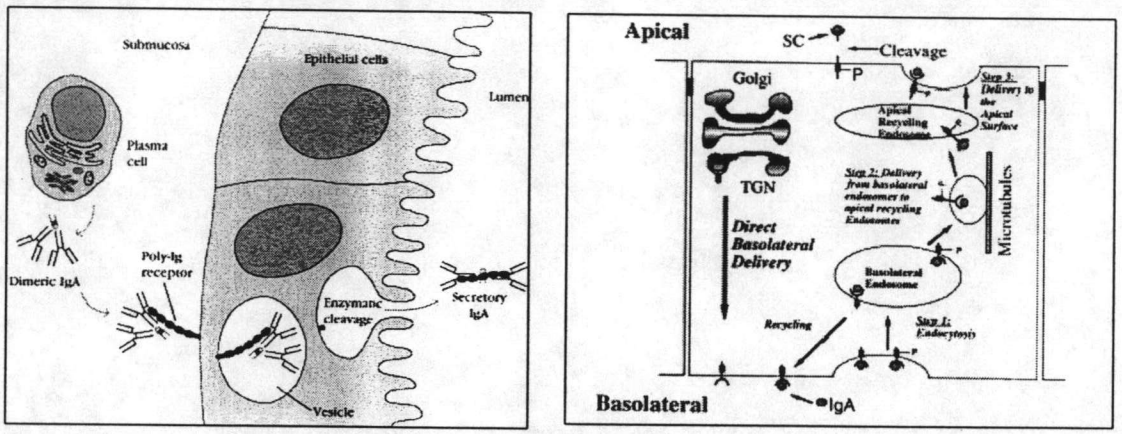

Figure 2 Schematic representation of transported polymeric IgA across epithelial cell by using pIgR. [47]

Function of pIgR

PIgR is a key protein in secretory immunity which mediates the active epithelial transfer of polymeric Ig to external secretion.[48,49] It is constitutively expressed on the basolateral surface of secretory epithelial cells. The receptor is synthesized in the rough endoplasmic reticulum as a 90 to 100 kDa transmembrane precursor protein that matures to 100 to 120 kDa after terminal glycosylation in the Golgi apparatus.[50] In the trans-most cisterna of this organelle, pIgR is sorted for delivery to the basolateral plasma membrane. At this cell surface, pIgR may interact with pIgA and pentameric IgM, although a substantial amount of pIgR is usually transported through the cells with out any bound ligand. [51] The epithelial transport of pIgR can be divided into three discrete steps: 1) binding and endocytosis at the basolateral surface; 2) transcytosis from basolateral to apical endosome; and 3) delivery to the apical cell surface followed by

proteolytic cleavage of the pIgR with release of its 80 kDa extracellular part into secretion. (Fig 2)

General structure and expression of *pIgR*

Transmembrane SC, or pIgR, belongs to the Ig superfamily and contains five extracellular disulfide-stabilized domains. In 1991 the complete genomic organization of the human *PIGR/SC* gene (chromosome 1, region 1q31-q42) [52] and shown to include 11 exons. [53] The first exon is not translated while the open reading frame is encoded by exons 2-11. Each of three extracellular Ig-like domain (D1, D4 and D5) is encoded by a separate exon, whereas two (D2 and D3) are encoded by the same exon. The pIgR protein has an N-terminal ligand-binding part, a single membrane-spanning segment, and a rather long (103 amino acids) cytoplasmic C-terminal segment. The five extracellular domains (D1-D5) are homologous in structure to the variable Ig domains. [50,53] This basic pIgR/SC structure exists in all characterized species, which include the rabbit, human, rat, bovine and mouse, [53-57] as well as the recently characterized pIgR from the marsupial Brushtail possum.[58] In addition, rabbit pIgR also occurs as a splice variant in which D2 and D3 are deleted. This low molecular variant is found in all exocrine rabbit tissues that express pIgR, [59] and it is functional in terms of pIgA binding [60,61] and epithelial trancytosis.[62] With this exception, the pIgR is remarkably well conserved among species—The degree of homology within the extracellular domain being in the order of D1>D3=D4>D5>D2, and the overall homology being 73%.[63] The particular high degree of D1 homology among species reflects the important function of this domain in ligand binding. The transmembrane segment and the cytoplasmic tail are also highly conserved, reflecting their importance for proper receptor function.[50] The pIgR protein is heavily glycosylated and may contain up to 25% carbohydrate. [63,64] The pattern of glycosylation, however, is not particularly well conserved,[63] and the ligand-binding function of pIgR does not depend on glycosylation.[65] Conversely, proper glycosylation is important for the ability of SC to interact with certain bacterial products such as toxin A from *Clostridium difficile*. [66]

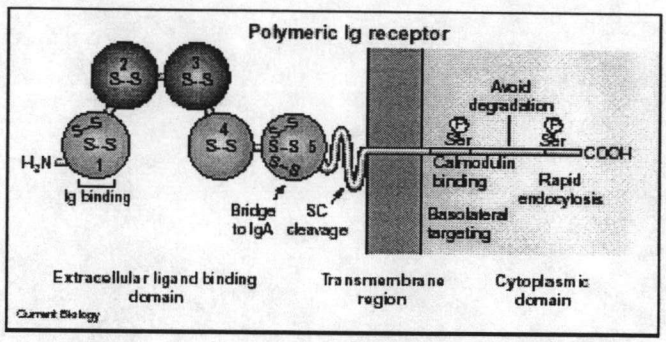

Figure 3 Structure of polymeric immunoglobulin receptor (pIgR). The pIgR is a transmembrane receptor with five extracellular Ig homology domains and a long cytoplasmic tail containing signals for intracellular sortingand endocytosis. Extracellular domains 1,4 and 5 are highly conserves among species, and contain signals for non-covalent and disulfide bonding to polymeric IgA. Domains 2 and 3 are less conserved, and are absent in some species variants of rabbit pIgR. Cleavage of full-length pIgR to release the soluble secretory component (SC) occurs within the linker connecting domain 5 to the transmembrane region.[67]

Regulation of *PIGR* expression

1. Cytokine

Several different human adenocarcinoma cell lines have been used to study the expression of human *PIGR* in epithelium. In addition to its constitutive expression, *PIGR* can be upregulated in the human colonic adenocarcinoma cell line HT-29 by interferon (IFN) - γ [68,69], interleukin (IL)-1 [70,71], IL-4 [69,70] and TNF- γ [70]. Moreover, IL-13 likewise mediates upregulation of *PIGR* in HT-29 cell. Consistent with the in vitro-induced *PIGR* upregulation obtained in cell cultures, increased expression of *PIGR* is seen in situ in various mucosal and glandular disease.[72] This is presumably caused by local production of IFN-γ and other cytokines secreated from activated local T cells and macrophages, and thus represent a biological link between local immune responses (including increased production of pIgA and pentameric IgM) and enhanced external transport of secretory antibodies.[45]

2.Regulatory DNA elements and transcription factors

Several research groups have sought to identify DNA elements that regulate both the constitutive and cytokine-enhanced transcription of the *PIGR*/SC gene by use of reporter gene assays, The transcription start site has been mapped, and the presence of a consensus TATA box 27 bases upstream of the initiation site was determined. [73,74] By transient transfection studies with the gastric adenocarcinoma cell line AGS, the study can identify a composite DNA element essential for the constitutive expression of *PIGR*,[73] this element consists of a 5'E-box and a partially overlapping 3'inverted repeat sequence (IRS). The E-box is the binding site for transcription factors that belong to the basic-helix-loop-helix/leucine-zipper (bHLH-LZ) family. Furthermore, mutation of E-box abolished protein binding in GEMSA and inhibited transcription from the reporter gene in transfected cells. Sequence analysis of the upstream promoter region of the human *PIGR* gene, as well as sequences downstream of transcription start, has identified a number of putative DNA elements that may contribute to hormone-regulated [75,76] and cytokine-enhanced [74] expression of this gene. Three interferon-stimulated response elements (ISREs), two upstream of transcription start (centered around position –133 and –100) and one in exon 1 (centered around position +13), are potential sites for gene induction mediated by IFN-γ and TNF-α.[74] The exon1 ISRE has also been implicated in TNF-α induced upregulation of *PIGR*. in a reporter gene construct extending from –280, its deletion abolished the effect of TNF-α but not that of IFN-γ . In nuclear run-on assays with HT-29 cells, we found that IL-4 provided better stimulation of *PIGR* transcription than IFN-γ and TNF-α, [77] and it has been reported that IL-4 and IFN-γ act synergisically.[69,78] Mapping of the IL-4 responsive element has identified a 570-bp region that is necessary for induction of the human *PIGR* gene by IL-4. This region contains an IL-4 dependent enhancer DNA element.

3.Other mediators

In addition to cytokines, other soluble factors may modulate the expression of *PIGR* both *in vitro* and *in vivo*. Butyrate, which is an abundant fermentation product in

the colon, can enhance the stimulatory effect of IL-1 and TNF-α in HT-29 cells, particulatly in the presence of IFN-γ , but rather decreases the stimulatory effect of IL-4, even in the presence of IFN-γ.[70] Furthermore, both constitutive and cytokine-enhanced *PIGR* expression appears to depend on adequate presence of vitamin A (retinoic acid) and the nutrition state of the subject.[79,80] In the rat kidney, *PIGR* mRNA levels were found to be upregulated by a vasopressin-coupled pathway in response to variations in water intake.[81]

3.Gene identification

There are four main approaches to map the disease gene.

1.Functional cloning

If the biochemical basis of an inherited disease is known, it may be possible to purify and characterize some of the gene products. Many methods, for example, gene specific oligonucleotide and specific antibodies can be used to identify the gene.

2.Positional cloning

Positional cloning identifies a disease gene based on no biochemical information except its approximate chromosomal location. There are two essential requirements for finding disease genes. The first is genetic markers, it can use to locate the disease gene. The second is families to establish linkage.

3.Candidate gene approach

Candidate gene approach is a new strategy for identifying disease gene. Relative gene with matching function of the disorder was selected to study.

4.Positional candidate gene approach

This strategy combines the positional and candidate gene together. It's powerful to identify disease gene in present day.

However, for complex diseases, functional cloning was almost impossible to be applied for the hereditary gene hunting technique. On the contrary, the two complementary positional cloning methods, linkage analysis and association study, are available either to screen the whole genome simultaneously or on candidate gene basis.[82]

1.Linkage analysis

Linkage analysis is the test for co-segregation of a marker and disease phenotype in the family. Genome wide screen approach, typing of 300 highly polymorphic markers for finding disease gene, is powerful in mapping mendelian disease and subsets of complex traits which show simple mendelian inheritance.[82] However, complex traits that involve many genes and environments are less powerful.

2.Association study

The study compares the marker frequencies between unrelated cases and controls within the same population. The approaches divide in two levels.

-Case/control in simplex families

One affected child in the family can be used in association study. From nuclear family data, the parental marker alleles not transmitted to an affected child or never transmitted to an affected sib pair, form a control population referred to as affected family-based controls.(AFBACs)[83] Other popular testing is the transmission disequilibrium test (TDT) which testing for 50% transmission ratio of associated allele from heterozygote parents to affected child. [84]

-Population level

Linkage disequilibrium (LD) or allelic association is a tool of mapping complex disease from observed relationship between marker alleles and disease phenotype. This approach can be quite powerful when disease chromosomes are descended from a single founder mutation and the markers are tightly linked to the disease locus. From the basis of LD, a large number of marker loci throughout the genome or a set of polymorphisms around a candidate gene were tested and hope that one of these marker loci would be close enough to a disease locus. Thus these closely marker alleles might be associated with the disease allele. In addition, different alleles of the marker may be associated with the disease in different populations, depending on population history.

Animal models of human disease can be informative for selecting candidate genes to map and for physiological studies of genetic and environmental factors. [85] An advantage of the animal model is the ability to analyze large numbers of offspring from single parents and to understand the expression of the gene and protein function.

A multistrategy approach to the mapping of complex traits is still appropriate: no single method is sufficient or optimal. The increasing availability of more markers across the genome, combined with multipoint analyses using closely linked markers will increase the power of linkage and association studies. [86,87]

4. Possible explanations for population association

From association study, allele D was found association with disease. Possible important causes include the following:

1.Direct causation

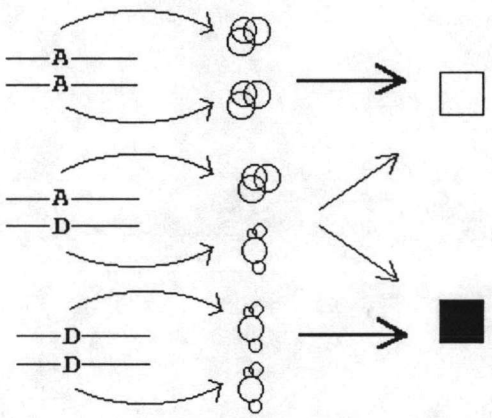

Figure 4 Model of direct causation. A and□ represent normal but D and■ represent disease.(http://www.mds.qmw.ac.uk/statgen/dcurtis/lectures/assoc1.html)

Positive association can occur if allele D is actually a cause of the disease. In this case allele D related with the disease in any population studied.

2.Population stratification

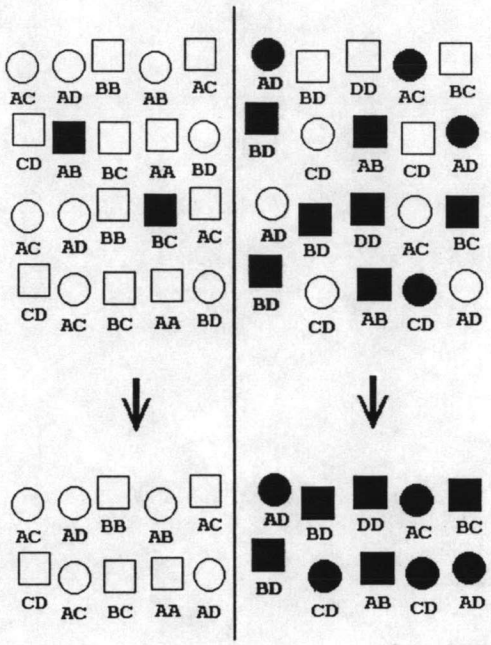

Figure 5 Model of population stratification.

(http://www.mds.qmw.ac.uk/statgen/dcurtis/lectures/assoc3.html)

Most disturbingly, positive association can also arise as an artifact of population admixture. The disease and marker allele D are both rare in the first group but common in the second. When cases and controls are sampled from the whole population the disease appears to be associated with the D allele, even though they are not related in either subgroup. On the contrary, the result of TDT is unaffected by population stratification because using family as a control group.

3.Linkage disequilibrium

Close linkage can produce association at the population level, disease bearing chromosomes were descended from one or a few ancestral chromosome. If linkage disequilibrium is the cause of the association, there should be a gene near the D locus that has mutations in people with this disease. Different alleles of the marker may be associated with the disease in different populations.

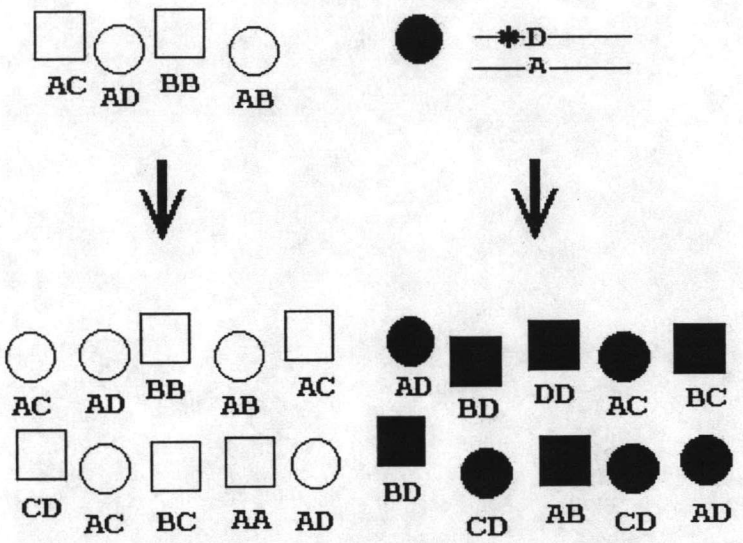

Figure 6 Model of linkage disequilibrium.

(http://www.mds.qmw.ac.uk/statgen/dcurtis/lectures/assoc2.html)

5. Polymorphic markers

Mapping human disease gene, recombination fraction between pair of disease would be the obvious way to construct but disease-disease mapping is not possible in human. Furthermore, people heterozygote for two different diseases are rare and there is low probability to have the children. For this reason mapping human disease gene depends on markers. There are several types of markers that available in recent day.

1.Restriction fragment length polymorphisms (RFLP)

RFLPs were initially typed by preparing Southern blots from restriction digested DNA. This technology required plenty of time and money. Nowadays this is less of a problem because RFLPs can usually be typed by PCR.[88] RFLPs have only two alleles, cut and uncut with enzyme, and maximum heterozygosity is 0.5.

2.Microsatellites or short tandem repeat (STR)

Microsatellites are the repetitive sequence, tri or tetra nucleotide repeats. These markers have many alleles by the number of repeat and comfortable to detect several markers in one reaction of multiplex PCR.

3.Single nucleotide polymorphism (SNP)

From Human Genome Project, the million of variants are single nucleotide polymorphism. Biallelic SNPs are the markers of choice because of their high frequency, low mutation rate and amenability to automation. However, there is less information in SNP than microsatellites.[89,90] Nevertheless, researcher are beginning to use genome wide association screens in complex disease. Genome wide SNP map in disease mapping has been initiated to co-ordinate an international effort to characterize at least 500,000 SNPs. [91]