A Genetic Basis for Infectious Mononucleosis: Evidence From a Family Study of Hospitalized Cases in Denmark

Klaus Rostgaard, Jan Wohlfahrt, and Henrik Hjalgrim

Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark

(See the Editorial Commentary by Balfour on pages 1690–1.)

Background. Circumstantial evidence from genome-wide association and family studies of various Epstein-Barr virus–associated diseases suggests a substantial genetic component in infectious mononucleosis (IM) etiology. However, familial aggregation of IM has scarcely been studied.

Methods. We used data from the Danish Civil Registration System and the Danish National Hospital Discharge Register to study rate ratios of IM in a cohort of 2 823 583 Danish children born between 1971 and 2011. Specifically, we investigated the risk of IM in twins and in first-, second-, and third-degree relatives of patients with IM. In the analyses, IM was defined as a diagnosis of IM in a hospital contact. Effects of contagion between family members were dealt with by excluding follow-up time the first year after the occurrence of IM in a relative.

Results. A total of 16 870 cases of IM were observed during 40.4 million person-years of follow-up from 1977 to 2011. The rate ratios and the associated 95% confidence intervals were 9.3 (3.0–29) in same-sex twins, 3.0 (2.6–3.5) in siblings, 1.9 (1.6–2.2) in children, 1.4 (1.3–1.6) in second-degree relatives, and 1.0 (0.9–1.2) in third-degree relatives of IM patients. The rate ratios were very similar for IM in children (aged 0–6 years) and older children/adolescents (aged 7–19 years).

Conclusions. We found evidence of familial aggregation of IM that warrants genome-wide association studies on IM disease etiology, especially to examine commonalities with causal pathways in other Epstein-Barr virus–related diseases.

Keywords. family study; children; cohort; contagion; genetics.

Epstein-Barr virus (EBV) is ubiquitous, and >90% of the world's adult population is chronically infected with the virus [1]. When primary infection with EBV occurs in childhood, it is usually asymptomatic or only accompanied by mild symptoms. Primary infection later in life, in contrast, is often (25%–77%) accompanied by so-called infectious mononucleosis (IM) [2, 3]. Clinically, IM is typically characterized by fever, tonsillitis, lymphadenopathy, and fatigue. The median

duration is 16 days, although the fatigue may persist considerably longer [4]. IM is common in Western countries, where it is a leading cause of disruption of studies [5], and the leading infectious cause of time lost for army recruits in the United States [6].

Immunologically, IM responders differ from those with asymptomatic infection by experiencing a massive transient expansion of (mainly) EBV-specific cytotoxic T lymphocytes. The IM symptoms are attributed to excessive cytokine secretion by the cytotoxic T lymphocytes in conjunction with lysis of EBV-infected cells [7]. Interestingly, IM is associated with risk of other and more severe EBV-related diseases in white individuals, most prominently Hodgkin lymphoma (HL) and multiple sclerosis (MS), which are also characterized by other immunological peculiarities regarding EBV [8–11].

Familial accumulation is well established for HL and MS individually [12, 13], but HL and MS have also been

shown to cluster with each other within families [14, 15], consistent with a mutual set of risk factors for the 2 conditions. We hypothesize that these common risk factors include constitutionally determined "inadvertent" immunologic responses to EBV, which, in addition to HL and MS, may also lead to IM. This interpretation of the association with HL and MS suggests that there should be a substantial genetic component and familial clustering in IM [16].

One recommendation from a National Institutes of Health meeting on EBV vaccine research was the "prevention of infectious mononucleosis and EBV-associated cancers, facilitated by identification of disease-predictive surrogate markers." [6]. Therefore, understanding the etiology of IM, including evidence of genetic predisposition, could be important to efforts to prevent a host of EBV-related diseases—for example, through design of vaccines and biomarker-based screening for early-stage EBV-associated malignancies. Whereas there is an abundance of studies on genetic influences on other EBV-related diseases and traits [17–22], the number and scope of studies of familial or genetic exposures on the risk of IM itself is small [23–26]. To examine whether IM clusters in families beyond contagion, we therefore assessed the relative risks of IM after IM diagnosis in a twin or a first-, second-, or third-degree relative in the cohort of all Danes born in the period 1971–2011.

MATERIAL AND METHODS

The nationwide Danish Civil Registration System (CRS) was implemented on 1 April 1968. All Danish citizens have since been assigned a unique identification number (the CRS number), by which the CRS continuously monitors individual vital status, emigration status, identity of parents, and residence [27]. Use of the CRS number also enables easy linkage to health register records. Using the CRS, we established a cohort of all Danes born between 1 January 1971 and 31 December 2011 with a mother in the CRS, $n = 2823583$. For each cohort member, we identified relatives (both inside and outside the cohort) in the Danish Family Relations Database [28]. This database is based on the parent–child links registered in the CRS. For persons born in Denmark since 1950, we can identify almost all of their parents and nearly all of those of their siblings and half-siblings who were born since 1950, provided these relatives reside in Denmark. Coverage is lower for more distantly related relatives: 94% of children born after 1990 have a known grandparent, compared with 64% of children born between 1977 and 1989 [28, 29]. Twins were identified as individuals having the same mother and born on dates at most 1 day apart. To not count the same information twice, we used the convention that the older relative in a pair of relatives exposes the younger relative. Thus, first-degree relatives of a cohort member were defined as older opposite-sex twins, parents, and older siblings;

second-degree relatives as older half-siblings, grandparents, and aunts/uncles; and third-degree relatives as older first cousins. For twin pairs and other pairs with the same birthday, one was randomly chosen as the (older) relative.

The Danish National Hospital Discharge Register was established in 1977 and has since recorded 99.9% of all discharges from Danish nonpsychiatric hospitals [30, 31]. For each hospital contact, the register contains information on dates of admission and discharge, and discharge diagnoses classified according to the *International Classification of Diseases, Eighth Revision* (*ICD-8*) from 1977 to 1993, and the tenth revision (*ICD-10*) thereafter. Inpatient hospitalizations have been almost completely recorded since 1977. The registration of other types of hospital contacts (outpatient and emergency ward) was introduced in 1994 and similarly complete from 1995 onward [30]. In the register, we identified all hospital contacts with IM as main, secondary, or underlying diagnosis, classified as 075* in *ICD-8* and B27* in *ICD-10*. In all analyses, we considered a person incident with IM at the earliest occurrence of such a hospital contact.

The study was approved by the Danish Data Protection Board (J.nr 2008-54-0472).

According to Danish law, informed consent is not required for purely register-based studies.

Statistical Analysis

The cohort was followed for IM from 1 January 1977 or date of birth, whichever occurred last, until the date of IM, death, emigration, age 20, or 31 December 2011, whichever occurred first. Patients remain highly contagious for several months after IM (see Odumade et al and references therein [4]). To avoid inflation of the estimated rate ratios (RRs) associated with having an affected relative, we therefore excluded follow-up during the first year after IM in a relative in all analyses.

Log-linear Poisson regression was used to model incidence rates and RRs and thereby assess the associations between risk of IM and having an affected relative (ie, a relative having had IM). The variables investigated included having an affected relative of a given type and years since the most recent occurrence of IM in such type of relative examined as a log-linear trend. All these variables were time-dependent, that is, potentially changing when new relatives were born or became affected. All RRs were adjusted for birth cohort (calendar year of birth), sex-specific age (1-year groups), sex-specific period (calendar year), and having a relative of the type considered in the exposure (eg, the RR when having an affected older sibling was adjusted for "having an older sibling").

Associations were assessed for all ages (0–19 years) and by age group (0–6 and 7–19 years). All analyses were performed using SAS statistical software version 9.3 (SAS Institute, Cary, North Carolina). Two-sided *P* values and 95% confidence intervals (CIs) were based on Wald tests. Tests for effect modification

were performed by inclusion of interaction terms and were all likelihood-ratio based. To maximize power, the test for homogeneity by age (0–6 vs 7–19) of the RRs associated with having affected relatives were based solely on estimates regarding the 4 major groups of affected relatives (parents, siblings, half-siblings, and uncles/aunts) in a joint model. Similarly, all tests for effect modification were performed without stratification by age group.

RESULTS

Overall, 16 870 IM cases occurred during 40.4 million person-years of follow-up in our cohort. Overall, 88% of cases had IM as the main diagnosis in the relevant hospital contact, and 85% were inpatient admissions. In all the following, we exclude follow-up time during the first year since IM in any affected relative, thereby excluding 7.4% of the outcome events among exposed, corresponding to 0.38% of all outcome events in the cohort.

The table shows the RR of IM according to type of affected relative. The RR varied from a large and statistically highly significant, but imprecise, 9.3 for same-sex twins to 1.00 for third-degree relatives. Mostly, the estimated RRs associated with having an affected relative were a little larger in children aged 0–6 years than in children aged 7–19 years ($P = .03$). RRs associated with having different types of affected relatives with the same degree of relatedness could not be assumed to be homogeneous ($P < .0001$). Equivalent maternal and paternal RRs could not be

assumed to be homogeneous ($P = .006$, data not shown). This seemed due to different RRs associated with having affected maternal and affected paternal half-siblings. Other equivalent maternal and paternal familial associations (ie, through maternal/paternal parents, uncles and aunts, grandparents, and cousins) could be assumed similar ($P = .06$, data not shown). There was no variation by sex of the exposed in the RR associated with having a given type of affected relative ($P = .59$, data not shown).

Next we assessed variation in associations by time since IM in the affected relative. Based on the above results, we only considered first-degree affected relatives as an input to the discussion. The changes in RR per year since IM in a relative were RR = 1.01 (95% CI, .99–1.04; $P = .34$) for parents as affected relative, and RR = 0.94 (95% CI, .91–.97; $P = .0002$) for siblings as affected relative. Following on from this, we also assessed the RRs when follow-up started at least 5 years after the relative was diagnosed. As expected from the above trends, this yielded slightly higher estimates for the affected parent association than those shown in Table 1 (data not shown), and mostly somewhat lower estimates for the affected sibling association: 0–19 years, RR = 2.57 (95% CI, 2.08–3.17); 7–19 years, RR = 2.46 (95% CI, 1.98–3.06); 0–6 years, RR = 4.39 (95% CI, 1.82–10.6).

DISCUSSION

This is the first study to address familial clustering of IM, apart from a twin study [26] by Hwang et al. In the cohort of all

Table 1. Incidence Rate Ratios of Hospitalization With Infectious Mononucleosis, by Type of Relative Previously Hospitalized With Infectious Mononucleosis, Denmark, 1977–2011

Type of Affected Relative	Age 0–19 y			Age 0–6 y			Age 7–19 y		
	Events	Exposed	RR (95% CI)	Events	Exposed	RR (95% CI)	Events	Exposed	RR (95% CI)
Same-sex twin	3	104	9.31 (2.96–29.3)	0	23		3	99	10.5 (3.31–33.2)
First-degree	359	41 125	2.30 (2.07–2.55)	78	32 355	2.30 (1.83–2.88)	281	30 500	2.30 (2.04–2.59)
Opposite-sex twin	0	50		0	7		0	50	
Sibling	169	11 325	3.02 (2.59–3.51)	11	3742	3.74 (2.07–6.77)	158	10 815	2.93 (2.50–3.43)
Parent	192	29 882	1.89 (1.64–2.18)	68	28 676	2.20 (1.73–2.80)	124	19 746	1.75 (1.47–2.09)
Mother	101	14 260	2.17 (1.78–2.64)	35	13 724	2.35 (1.68–3.28)	66	9293	2.08 (1.63–2.65)
Father	94	15 788	1.65 (1.35–2.03)	33	15 109	1.98 (1.40–2.79)	61	10 568	1.52 (1.18–1.96)
Second-degree	323	58 389	1.43 (1.28–1.59)	71	51 931	1.27 (1.00–1.61)	252	41 361	1.48 (1.30–1.67)
Half-sibling	59	6348	1.81 (1.40–2.34)	12	3743	3.16 (1.78–5.59)	47	5674	1.64 (1.23–2.19)
Maternal half-sibling	36	2766	2.57 (1.85–3.57)	8	1541	5.36 (2.66–10.8)	28	2516	2.25 (1.55–3.28)
Paternal half-sibling	23	3588	1.21 (.80–1.83)	4	2208	1.71 (.64–4.58)	19	3162	1.15 (.73–1.81)
Grandparent	5	3303	0.61 (.25–1.47)	2	2999	0.82 (.20–3.27)	3	1656	0.52 (.17–1.62)
Uncle/aunt	262	48 888	1.40 (1.24–1.58)	57	45 366	1.14 (.87–1.48)	205	34 074	1.50 (1.30–1.72)
Third degree	138	37 482	1.00 (.85–1.19)	19	20 747	1.02 (.65–1.61)	119	31 952	1.00 (.83–1.20)
First cousin	138	37 482	1.00 (.85–1.19)	19	20 747	1.02 (.65–1.61)	119	31 952	1.00 (.83–1.20)

Events are number of infectious mononucleosis patients with an affected relative of said type. Exposed is number of children with an affected relative of said type at some time during follow-up in the given age interval.

Abbreviations: CI, confidence interval; RR, rate ratio.

Danes born in 1971–2011, having relatives who had been hospitalized with IM was associated with an increased risk of being hospitalized with IM before the age of 20 years. The relative risk of IM varied by degree of relatedness, from an imprecise RR of 9.3 in same-sex twins to 1.0 for third-degree relatives. The observed familial clustering is unlikely to be a chance phenomenon, but rather points to shared environmental or genetic risk factors for IM or a combination thereof.

Sociodemographic affluence has long been considered to be associated with an increased risk of IM through simple postponement of primary EBV infection until adolescence [32]. Therefore, the observed associations in our study could be partly due to sociodemographic correlation between relatives. However, the inconspicuous RR associated with affected third-degree relatives (cousins) and also second-degree relatives excluding maternal half-siblings suggests that this is a negligible contribution. Also, we observed similar RRs for children (aged 0–6 years) and older children/adolescents (aged 7–19 years) within all categories of exposed relatives, contradicting the proposed "postponement of infection" mechanism. Hence, IM risk appears to vary little by sociodemographic characteristics in an affluent and fairly equal society such as Denmark in 1977 onward. Accordingly, sociodemographic correlation between relatives seems an implausible explanation for our results. Beyond affluence, shared environmental characteristics or behavior also seem insufficient to completely explain the familial clustering of IM. The median time since parental IM in IM cases was 20 years (data not shown), and the associated RR did not vary by time since parental IM.

Because the RRs associated with affected parents, uncles, or cousins presumably are not to any substantial degree due to similarities in behavior or environment, the observed estimates may be interpreted as estimates of the genetic effects of having affected first-, second-, and third-degree relatives. Having seen the familial clustering of IM in first- and second-degree relatives, one would expect the RR in third-degree relatives of affected persons to be at least slightly elevated, and some leeway is provided in its upper confidence limit of 1.20. Hwang et al found the RR for monozygotic vs dizygotic twins to be 1.9 (95% CI, 1.2–5.3) [26], which fits the overall picture from the present study of modest genetic associations with IM risk.

At the same time, as our results suggest genetic risk factors for IM, they also indicate that these act in the context of environmental exposures. This is evidenced by the variation in RR within degrees of relatedness and within maternal/paternal half-siblings, and the declining RR by time since IM in an affected sibling. The most likely explanation for the latter trend is that the closer in age siblings are, the more similar their environments are likely to be. Accordingly, a diagnosis of IM in an older sibling is suggestive of an environment conducive of IM, and the closer in age to the affected sibling, the higher

the likelihood that an exposed sibling shares this particular environment. Alternative explanations seem unconvincing. For example, the higher the birth order of a child, the smaller the likelihood of hospitalization upon IM, because more experienced parents are more relaxed about disease in their children. The larger the difference in birth order, the larger the likely age differential between the siblings. Hence, a decreasing tendency for hospital-seeking by birth order could translate into a decreasing relative risk of hospitalization for IM by time since hospitalization for IM in an older sibling. However, when we adjusted our analyses for birth order, the resulting changes in estimates were inconspicuous, leaving explanations based on sibship structure unlikely (data not shown). The relationship between small age interval and similarity of environment may also explain why having an affected sibling is associated with a larger RR than having an affected parent. It would also contribute especially to elevate the RRs for twins. This point can be illustrated by calculating RR_{sst}, the expected RR in same-sex twins entirely due to genetics, under assumptions regarding the monozygotic (MZ) and dizygotic (DZ) twins in our study. If we assume $RR_{MZ}/RR_{DZ} = 1.9$ (from [26]) and $RR_{DZ} = 1.9$ (from the observed parent–offspring association), and $RR_{sst} = (RR_{MZ} + RR_{DZ})/2$, then $RR_{sst} = 2.8$, a small fraction of the observed RR of 9.3. The difference in RR between having affected maternal and paternal half-siblings may be also explained the same way: that maternal half-siblings presumably have much more in common than paternal half-siblings regarding environment and behavior.

The present study has several strengths and weaknesses to consider. We performed a purely register-based study, thus by design avoiding recall bias. The free access to hospitalization in Denmark [30] limits trivial sociodemographic biases in the ascertainment of outcomes and exposures in the study. Secular trends were adjusted for using very detailed sex-specific information on calendar period and birth cohort. There was no suggestion that our results were confounded by birth order characteristics. Inflation of familial RRs due to contagion by the affected relative was dealt with by discarding follow-up during the year following IM in a relative.

Many family studies have considered outcomes, and by implication, exposures that were very rare. In that case, the fact that the reference groups of no presumed exposure are polluted by a minimal fraction of unrecognized exposed has no practical statistical consequence. Not so in this study. The fact that we only ascertained hospitalized cases of IM and only did so during a limited time period, rather than from the birth of each relative, means that most IM cases in relatives of our study cohort go unnoticed by us. Figures from the Centers for Disease Control and Prevention suggest the lifetime risk of IM to be at least 3% [33]. Other estimates of IM lifetime risk can be obtained from the prevalence of having had IM in population controls in HL

case-control studies; for example, 4.7% in the United Kingdom [34] and 4.1% in Denmark [35] (authors' personal observation). Even with this small lifetime risk of IM, the number of relatives of various types would lead to noticeably deflated RRs in the present analysis.

We have not been able to locate a study on the validity of the diagnosis of IM in the Danish National Patient Register. Nor was it possible to validate the diagnoses based on registers. The proportion of false-negative IM diagnoses is likely to be nonnegligible, based on the experience of the diagnostic precision within many specialties in the register [36]. Importantly, however, misclassification of IM, which is both outcome and exposure here, would normally deflate association estimates and is therefore unlikely to explain our findings of familial aggregation. It is not obvious how our results will generalize to nonhospitalized IM. However, the distribution of hospitalized IM cases by sex and age seems no different from the distribution of self-reported IM in the Danish Blood Donor Study [37] (personal observation), whereas one would expect hospitalized IM patients to be older than nonhospitalized IM patients based on the dogma that later infection means more severe disease [38].

Our study supports the notion that there is a noticeable genetic component to the familial aggregation of IM—that is, that it cannot be entirely explained by contagion and sociodemographic characteristics. Even modest RRs from family studies may translate into substantial RRs for those with the least favorable genetic profile compared with those with the most favorable genetic profile [39]. Besides human leukocyte antigen (HLA) loci, which determine immune responses on chromosome 6, loci on chromosomes 1, 2, 7, 12, 19, and 22 have been associated with IM risk [22–25]. Therefore, thorough genetic association studies of IM cannot be restricted to the HLA region. At least in white individuals, the genetic associations should be stronger than suggested by our RRs of having an affected parent, uncle/aunt, or first cousin. In the present context of large associations with environmental factors or gene–environment interactions, GWA studies would probably be statistically most efficient using a within-family paired design, whereas access to an existing infrastructure of genotyped controls may suggest other designs.

In conclusion, we found familial aggregation of IM, which may be explained by a mixture of genetics, shared environmental characteristics, and shared behaviors between relatives. The presumed genetic components are large enough to warrant GWA studies on disease etiology, especially to examine commonalities with causal pathways in other EBV-related diseases so as to provide clues to their prevention and treatment.

Notes

References

1. Balfour HH Jr, Sifakis F, Sliman JA, Knight JA, Schmeling DO, Thomas W. Age-specific prevalence of Epstein-Barr virus infection among individuals aged 6–19 years in the United States and factors affecting its acquisition. J Infect Dis **2013**; 208:1286–93.
2. Crawford DH, Macsween KF, Higgins CD, et al. A cohort study among university students: identification of risk factors for Epstein-Barr virus seroconversion and infectious mononucleosis. Clin Infect Dis **2006**; 43:276–82.
3. Balfour HH Jr, Odumade OA, Schmeling DO, et al. Behavioral, virologic, and immunologic factors associated with acquisition and severity of primary Epstein-Barr virus infection in university students. J Infect Dis **2013**; 207:80–8.
4. Odumade OA, Hogquist KA, Balfour HH Jr. Progress and problems in understanding and managing primary Epstein-Barr virus infections. Clin Microbiol Rev **2011**; 24:193–209.
5. Macsween KF, Higgins CD, McAulay KA, et al. Infectious mononucleosis in university students in the United Kingdom: evaluation of the clinical features and consequences of the disease. Clin Infect Dis **2010**; 50:699–706.
6. Cohen JI, Fauci AS, Varmus H, Nabel GJ. Epstein-Barr virus: an important vaccine target for cancer prevention. Sci Transl Med **2011**; 3:107fs7.
7. Silins SL, Sherritt MA, Silleri JM, et al. Asymptomatic primary Epstein-Barr virus infection occurs in the absence of blood T-cell repertoire perturbations despite high levels of systemic viral load. Blood **2001**; 98:3739–44.
8. Hjalgrim H, Friborg J, Melbye M. The epidemiology of EBV and its association with malignant disease. In: Arvin A, Campadelli-Fiume G, Mocarski E, et al., eds. Human herpesviruses—biology, therapy, and immunoprophylaxis. 1st ed. Cambridge: Cambridge University Press, **2007**:929–59.
9. Ascherio A, Munger KL. Epstein-Barr virus infection and multiple sclerosis: a review. J Neuroimmune Pharmacol **2010**; 5:271–7.
10. Mueller NE, Lennette ET, Dupnik K, Birmann BM. Antibody titers against EBNA1 and EBNA2 in relation to Hodgkin lymphoma and history of infectious mononucleosis. Int J Cancer **2012**; 130:2886–91.
11. Levin LI, Chang ET, Ambinder RF, et al. Atypical prediagnosis Epstein-Barr virus serology restricted to EBV-positive Hodgkin lymphoma. Blood **2012**; 120:3750–5.
12. Goldin LR, Pfeiffer RM, Gridley G, et al. Familial aggregation of Hodgkin lymphoma and related tumors. Cancer **2004**; 100:1902–8.
13. Nielsen NM, Westergaard T, Rostgaard K, et al. Familial risk of multiple sclerosis: a nationwide cohort study. Am J Epidemiol **2005**; 162:774–8.
14. Hjalgrim H, Rasmussen S, Rostgaard K, et al. Familial clustering of Hodgkin lymphoma and multiple sclerosis. J Natl Cancer Inst **2004**; 96:780–4.
15. Landgren O, Kerstann KF, Gridley G, et al. Re: familial clustering of Hodgkin lymphoma and multiple sclerosis. J Natl Cancer Inst **2005**; 97:543–4.
16. Ramagopalan SV, Meier UC, Conacher M, et al. Role of the HLA system in the association between multiple sclerosis and infectious mononucleosis. Arch Neurol **2011**; 68:469–72.
17. Urayama KY, Jarrett RF, Hjalgrim H, et al. Genome-wide association study of classical Hodgkin lymphoma and Epstein-Barr virus status-defined subgroups. J Natl Cancer Inst **2012**; 104:240–53.
18. Cozen W, Li D, Best T, et al. A genome-wide meta-analysis of nodular sclerosing Hodgkin lymphoma identifies risk loci at 6p21.32. Blood **2012**; 119:469–75.
19. Enciso-Mora V, Broderick P, Ma Y, et al. A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at

2p16.1 (REL), 8q24.21 and 10p14 (GATA3). Nat Genet **2010**; 42: 1126–30.

20. Sawcer S, Hellenthal G, Pirinen M, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature **2011**; 476:214–9.

21. Rubicz R, Yolken R, Drigalenko E, et al. A genome-wide integrative genomic study localizes genetic factors influencing antibodies against Epstein-Barr virus nuclear antigen 1 (EBNA-1). PLoS Genet **2013**; 9: e1003147.

22. Vollmer-Conna U, Piraino BF, Cameron B, et al. Cytokine polymorphisms have a synergistic effect on severity of the acute sickness response to infection. Clin Infect Dis **2008**; 47:1418–25.

23. Li YH, Wen FQ, Xiao ZH, Chen YX, Zhang ZX, Chen LL. Genetic polymorphism of GST gene in children with infectious mononucleosis and acute lymphocytic leukemia [in Chinese]. Zhongguo Dang Dai Er Ke Za Zhi **2012**; 14:260–3.

24. Hatta K, Morimoto A, Ishii E, et al. Association of transforming growth factor-beta1 gene polymorphism in the development of Epstein-Barr virus–related hematologic diseases. Haematologica **2007**; 92:1470–4.

25. Helminen M, Lahdenpohja N, Hurme M. Polymorphism of the interleukin-10 gene is associated with susceptibility to Epstein-Barr virus infection. J Infect Dis **1999**; 180:496–9.

26. Hwang AE, Hamilton AS, Cockburn MG, et al. Evidence of genetic susceptibility to infectious mononucleosis: a twin study. Epidemiol Infect **2012**; 140:2089–95.

27. Pedersen CB. The Danish Civil Registration System. Scand J Public Health **2011**; 39(7 suppl):22–5.

28. Schnack TH, Poulsen G, Myrup C, Wohlfahrt J, Melbye M. Familial coaggregation of cryptorchidism, hypospadias, and testicular germ cell cancer: a nationwide cohort study. J Natl Cancer Inst **2010**; 102:187–92.

29. Munch TN, Rostgaard K, Rasmussen ML, Wohlfahrt J, Juhler M, Melbye M. Familial aggregation of congenital hydrocephalus in a nationwide cohort. Brain **2012**; 135(pt 8):2409–15.

30. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. Scand J Public Health **2011**; 39(7 suppl):30–3.

31. Andersen TF, Madsen M, Jorgensen J, Mellemkjoer L, Olsen JH. The Danish National Hospital Register. A valuable source of data for modern health sciences. Dan Med Bull **1999**; 46:263–8.

32. Niederman JC, Evans AS. Epstein-Barr virus. In: Evans AS, Kaslow RA, eds. Viral infections of humans: epidemiology and control. 4th ed. New York: Plenum Publishing Corporation, **1997**:253–83.

33. Hardy AM. Incidence and impact of selected infectious diseases in childhood. Vital Health Stat 10 **1991**; Oct:1–22.

34. Alexander FE, Lawrence DJ, Freeland J, et al. An epidemiologic study of index and family infectious mononucleosis and adult Hodgkin's disease (HD): evidence for a specific association with EBV+ve HD in young adults. Int J Cancer **2003**; 107:298–302.

35. Hjalgrim H, Smedby KE, Rostgaard K, et al. Infectious mononucleosis, childhood social environment, and risk of Hodgkin lymphoma. Cancer Res **2007**; 67:2382–8.

36. Mosbech J, Jorgensen J, Madsen M, Rostgaard K, Thornberg K, Poulsen TD. The national patient registry. Evaluation of data quality [in Danish]. Ugeskr Laeger **1995**; 157:3741–5.

37. Pedersen OB, Erikstrup C, Kotze SR, et al. The Danish Blood Donor Study: a large, prospective cohort and biobank for medical research. Vox Sang **2012**; 102:271.

38. Tattevin P, Le TY, Minjolle S, et al. Increasing incidence of severe Epstein-Barr virus-related infectious mononucleosis: surveillance study. J Clin Microbiol **2006**; 44:1873–4.

39. Hopper JL. Disease-specific prospective family study cohorts enriched for familial risk. Epidemiol Perspect Innov **2011**; 8:2.