Blood Cells, Molecules, and Diseases 42 (2009) 129–135

Contents lists available at ScienceDirect

Blood Cells, Molecules, and Diseases

journal homepage: www.elsevier.com/locate/ybcmd

Review

The human major histocompatibility complex and childhood leukemia: An etiological hypothesis based on molecular mimicry

Malcolm Taylor [a,*], Adiba Hussain [a], Kevin Urayama [b], Anand Chokkalingam [b], Pamela Thompson [a], Elizabeth Trachtenberg [c], Patricia Buffler [b]

[a] Cancer Immunogenetics, School of Cancer Sciences, University of Manchester, Manchester, UK
[b] School of Public Health, University of California, Berkeley, USA
[c] HLA Immunogenetics and Molecular Diagnostics Laboratories, Children's Hospital and Research Center, Oakland, California, USA

ARTICLE INFO

Article history:
Submitted 16 October 2008
Available online 2 December 2008

(Communicated by M. Lichtman, M.D.,
17 October 2008)

Keywords:
Childhood leukaemia
HLA
MHC mapping
Etiology
Reverse immunogenetics
Molecular mimicry

ABSTRACT

The extended human major histocompatibility complex (MHC) is a gene-rich region of about 7.6 Mb on chromosome 6, and includes a high proportion of genes involved in the immune response. Among these are the two Human Leukocyte Antigen (*HLA*) gene clusters, *class I* and *class II*, which encode highly polymorphic classical *HLA-A*, *B*, *C* and *HLA-DR*, *DQ* and *DP* genes, respectively. The protein products of the classical *HLA* genes are heterodimeric cell surface molecules that bind short peptides derived from non-self and self proteins, including infections and auto-antigens. The presentation of these HLA-anchored peptides to T lymphocytes triggers a cascade of responses in immune-associated genes that leads to adaptive immunity. Associations between *HLA class II* alleles and childhood leukemia have been reported in a number of studies. This could be due to the role of *HLA* allele-restricted peptide binding and T cell activation, or linkage disequilibrium to an MHC-linked "leukemia gene" in the pathogenesis of childhood leukemia. Efforts are currently in progress to resolve these questions, using large leukemia case-control sample series such as the UK Childhood Cancer Study (UKCCS) and the Northern California Childhood Leukemia Study (NCCLS). Here we review the background to these studies, and present a novel hypothesis based on the paradigm of *HLA*-associated auto-immune disease that might explain an infection-based etiology of childhood leukemia.

Contents

Introduction

Leukemia is the most common malignant disease among children under the age of 15 years in socio-economically developed parts of the world, including Europe, North America, and Australasia [1,2]. Despite considerable phenotypic and genetic diversity, the numerically dominant leukemia subtype is B cell precursor (BCP) ALL which has a characteristic but unexplained age-incidence peak between 2 and 5 years of age [3]. Cytogenetic and molecular analyses have revealed that leukemia cells carry non-random chromosome rearrangements juxtaposing or amplifying genes associated with leukemia initiation [4]. In BCP ALL, two types of rearrangement account for >50% of cases: ~25% of patients carry a reciprocal translocation between chromosomes 12 and 21 [t(12;21)(12p13;21q22)], leading to the in-frame fusion of the *TEL* (*ETV6*) gene on chromosome 12p13 and the *AML1* (*RUNX1*) gene on 21q22, respectively. A further 30% of cases have a high hyperdiploid (HeH) karyotype, where the modal chromosome number, characterized by non-random gain of chromosomes X, 4, 6, 8,

* Corresponding author. Cancer Immunogenetics, St Mary's Hospital, Manchester, M13 0JH, UK.
E-mail address: gmtaylor@manchester.ac.uk (M. Taylor).

1079-9796 © 2008 Elsevier Inc. Open access under CC BY-NC-ND license.
doi:10.1016/j.bcmd.2008.10.009

10, 14, 17, 18, and 21, exceeds 50 [4]. Molecular data indicate that initiating rearrangements are necessary, but not sufficient, to give rise to clinical leukemia [4]. In BCP ALL, progression from pre-leukemia to overt clinical disease requires additional genetic lesions which are thought to be acquired under the influence of an abnormal, usually post-natal, immune response to infection [5].

There has been much debate about the type and mechanism of infectious causation in childhood leukemia prompted by the failure to date to detect a causative infectious agent [5–7]. Molecular evidence has shown that the frequency of children with leukemia-initiating chromosomal rearrangements at birth exceeds the number developing clinical leukemia by two orders of magnitude [8], and are not increased following exposure of the pre-term newborn to diagnostic radiation [9]. This suggests that progression from pre-leukemia to clinical disease involves transition through a bottleneck influenced by the timing of exposure to post-natal infection. Paradoxically, early exposure to infection seems to confer protection, whilst delayed exposure is thought to lead to an aberrant immune response facilitating progression to leukemia [5]. Whilst this "second-hit" seems to require the lapse of sufficient time to accumulate additional somatic mutations, it is not yet clear whether the infection conferring protection following early exposure also leads to increased risk if exposure is delayed. Here we review some of the recent studies on the role of the human major histocompatibility complex (MHC) in the etiology of childhood leukemia, and propose a novel hypothesis for infectious causation.

The human major histocompatibility complex

In the absence of direct evidence for the role of an infectious agent in childhood leukemia, our interests have focused on proxy measures of immunity to infection that can be attributed to genetically-determined differences in immune function. Central to these studies are the genes of the human MHC, and specifically the Human Leukocyte Antigen (*HLA*) genes.

The extended human major histocompatibility complex (xMHC) on the short arm of chromosome 6 (6p21.3) spans about 7.6 megabase DNA [10]. It includes 421 loci, of which 252 are expressed, and is partitioned into the class I, II and III regions, the class III region being located between I and II, and the most gene-dense region of the human genome (Fig. 1). The MHC contains a number of clustered and duplicated genes, the foremost of which are the *HLA class I* and *class II* gene clusters. The *class I* gene cluster includes the highly polymorphic expressed classical *HLA-A*, *B* and *C* genes; the *class II* gene cluster includes the polymorphic classical *HLA-DR*, *DQ* and *DP* genes. Both gene clusters encode cell surface heterodimeric proteins whose primary function is to bind short peptide fragments derived from non-self and self proteins in a closed (class I) or open (class II)-ended peptide binding groove whose walls are formed by two alpha helices, the floor consisting of a beta-pleated sheet. Peptides anchored to class I molecules are targeted by the T cell receptors (TCR) of CD8+ cytotoxic T cells; those bound by class II molecules are presented to and activate CD4+ T cells [11]. In both cases, the secure anchoring of peptides (T cell epitopes) by an HLA allotypic molecule is an absolute requirement for the formation of a tri-molecular TCR–peptide–HLA complex leading to T cell activation. This binding is influenced by the charge and size characteristics of a series of polymorphic peptide binding pockets lining the peptide or antigen-binding groove of the HLA molecule. The affinity of a given antigenic peptide for a polymorphic HLA allotype enables T cells to constantly sample the individual's antigenic environment for the presence of pathogens requiring a speedy protective immune response to eliminate the pathogen. Rapid evolution of pathogen sequences is thought to have exerted the significant selective forces that have given rise to the extremely high level of *HLA* polymorphisms in the population [12]. However, there is considerable flexibility and overlap in peptide binding between allotypes, due to the sharing of peptide pocket polymorphisms between alleles, leading to within-locus *HLA* allele clusters known as supertypes [13]. Despite this overlap in the specificity of peptide binding, a relatively small number (4–6) of peptide binding pockets impose constraints on peptide sequences that can be accommodated by HLA molecules [14]. It would

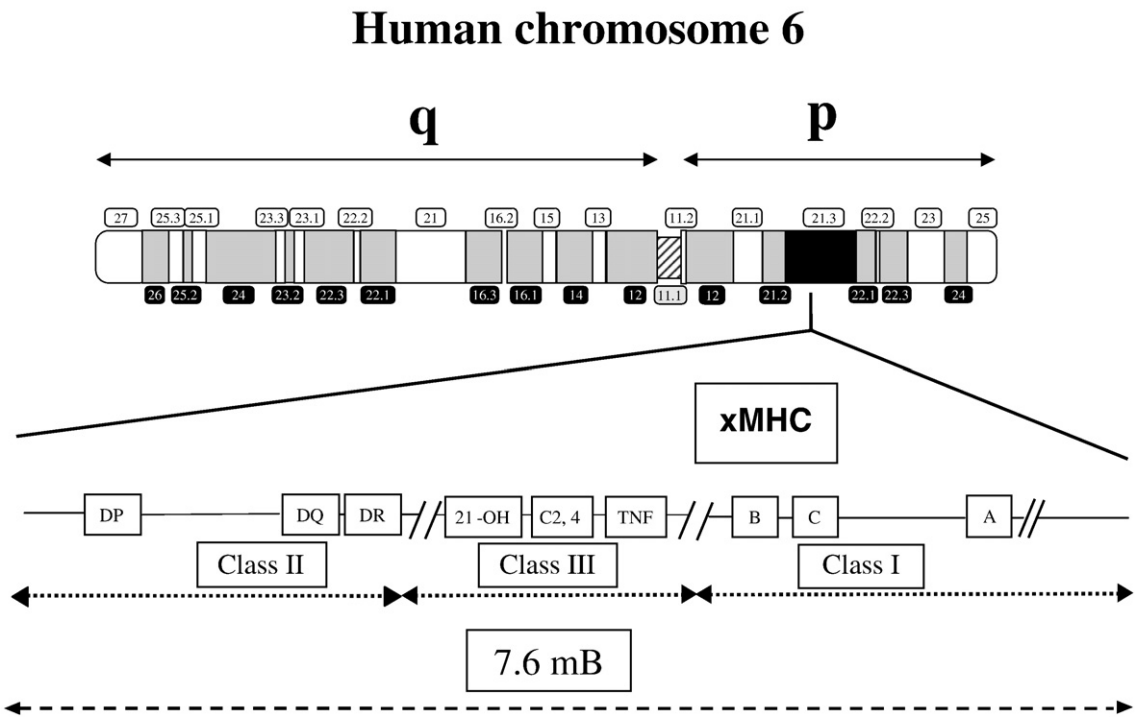

Fig. 1. Schematic map (not to scale) of the extended human MHC (xMHC) on the short (p) arm of human chromosome 6. The HLA class I gene cluster includes the 3 expressed classical class I genes, HLA-A, B and C, and is located telomeric of the class III region. The HLA class II gene cluster is located centromeric of the HLA class III region.

be over-simplistic to suggest that this constitutes the central control of immune response gene cascade; however, the formation of a TCR–peptide–HLA tri-molecular complex is at the epicenter of T cell interactions with non-self and self antigens, and probably holds the answers to many *HLA*-associated diseases [11]. Thus *HLA* allele associations with childhood leukemia are likely to provide important insights into the role of infection in progression from pre-leukemia to clinical disease.

In addition to the *HLA* genes, the xMHC includes some 67 other genes, comprising about 28% of expressed *MHC* genes, that are involved in either adaptive or innate immunity [10]. These include the genes of the immunoglobulin superfamily, inflammation-associated genes, such as tumor necrosis factor and lymphotoxin, leukocyte maturation, complement genes, non-classical *HLA* genes, and immune regulation and stress response genes. Furthermore, the region also contains a large number of genes not related to the immune system that play a role in cell cycle regulation, apoptosis, transcription and translational machinery, signaling, and other cellular processes. Although a detailed discussion of gene function is beyond the scope of this review, none of these *MHC* genes can yet be excluded as candidates in childhood leukemia susceptibility.

The MHC and childhood leukemia

Historically, the clinical significance of the *HLA* genes is their role as a barrier to the free exchange of organs and tissues between individuals. However, it is now clear that the primary biological function of *HLA* genes is the orchestration of the immune response to pathogens through the formation of TCR–peptide–HLA complexes. The fact that the *HLA* genes are associated with more diseases than any other region of the genome may reflect a common underlying pathogenetic pathway in which sub-version of the immune response can arise as a consequence of interactions between infection or other non-self antigens and HLA molecules. It is thus noteworthy that the majority of *HLA*-associated diseases have an autoimmune pathogenesis [12] possibly caused by the triggering of cross-reactive T cell responses to self proteins by pathogens. The breakdown of immune tolerance to self as a result of exposure to infection, leading to the destruction of self tissues recognized as auto-immune disease, has been termed molecular or pathogen mimicry and may be an underlying characteristic of many *HLA*-associated diseases [15]. In this paper, we consider how molecular mimicry might influence the risk of childhood leukemia in the context of an infectious etiology.

Prompted by the work of Lilly and colleagues in the 1960s showing that susceptibility to Gross retroviral leukemia in the mouse was linked to the *H-2* complex [16,17], childhood leukemia was one of the first human diseases to be investigated for associations with *HLA* antigens. These early studies focused on *HLA* class I antigens, but produced inconsistent results [18], probably due to small patient numbers influenced by selection, limited diagnostic criteria, and the use of serological *HLA* typing methods. Later studies of the HLA-D or class II region using serological typing suffered from the same problems, but suggested an association between class II alleles and survival [19].

By the 1990s, radical improvements in the diagnostic classification and treatment of childhood leukemias, better survival rates, and progress in the design of molecular epidemiological studies made it possible to re-examine the relationship between HLA and childhood leukemia. Importantly, the development of high resolution molecular typing methods driven by the needs of donor-recipient hematopoietic stem cell transplantation [20] made it possible to analyze inter-individual differences in the key antigen-binding sites of *HLA* alleles as a proxy measure of interactions with foreign (and self) peptides. More recent developments in high-throughput single nucleotide polymorphism (SNP) genotyping have extended this capability to the identification of conserved haplotype blocks across the entire xMHC [21] which should allow the construction of a detailed MHC map to localize all *MHC*-associated childhood leukemia genes, as recently demonstrated for type 1 diabetes [22]. Ultimately, the dual strategy of identifying the functional contribution of *HLA* alleles in the causation of childhood leukemia, and mapping of additionally important *MHC* genes will help to pinpoint the role of infection, and enable progress to be made toward the elusive goal of childhood leukemia prevention.

Confirmation of many *HLA* associations with a diversity of diseases has nevertheless proven to be extremely challenging, even in relatively common auto-immune diseases [23]. These challenges can largely be attributed to methodologic issues related to between-study heterogeneity in case definition, selection bias, statistical power to detect associations, and confounding due to population stratification. Additionally, the high levels of linkage disequilibrium (the non-random association between alleles of different genetic loci) within the MHC, giving rise to extended haplotypes, can make it difficult to identify the *MHC* genes primarily responsible for disease [24,25]. This is further complicated by the observation that linkage disequilibrium differs to varying degrees between race/ethnicity groups [21], emphasizing the critical nature of identifying racially homogenous study populations and implementing adjustment methods to account for differences in genetic structure.

In this paper, we describe recent and ongoing work that has been initiated to elucidate the role of the xMHC in childhood leukemia etiology using data and biospecimens from two large population-based case-control studies, the United Kingdom Childhood Cancer Study (UKCCS) and the Northern California Childhood Leukemia Study (NCCLS). The UKCCS was a carefully designed population-based case-control study set up in 1992, largely in response to societal concerns about the role of certain environmental carcinogens in the causation of childhood cancer and leukemia in the UK [26]. A key issue for consideration by the UKCCS was the perceived intrinsic vulnerability of parental germ cells, the fetus and newborn to changing patterns of radiation, and chemical carcinogen exposure and infection. A backdrop to the study, organized under the aegis of the UK Coordinating Committee on Cancer Research (UKCCCR), was the provocative suggestion, by Gardner et al. [27], that the Seascale leukemia cluster associated with the nuclear reprocessing plant at Sellafield, may have arisen by radiation-induced germ-line mutation. The need to address this and a number of other compelling hypotheses led to the initiation of this 5-year nationwide study [26].

Initiated in 1995, the NCCLS is a major ongoing US-based population-based case-control study which was also designed to investigate environmental and genetic causes of childhood leukemia largely in response to the large number of childhood cancer clusters that came to public attention, requiring investigation by health authorities [28]. Incident childhood leukemia cases are ascertained from major pediatric centers serving the 35-county Northern and Central California study area. Controls are randomly selected from the birth certificate files and matched to cases on date of birth, gender, maternal race, and Hispanic status. Among the vast amount of exposure data collected are detailed accounts of childhood infections, child's vaccination history, family structure, number of siblings and day-care and other school attendance. These data were obtained with special emphasis on exposure during time windows of development. Biospecimens collected in the NCCLS include constitutive DNA specimens from buccal cell scrapings and saliva, as well as pre-treatment blood and bone marrow from case subjects.

Reverse immunogenetics and molecular mimicry (UKCCS)

One of the major difficulties in establishing an etiological role for infection in childhood leukemia has been in the identification of a causal agent. A child is exposed to a blizzard of infection and allergen-derived antigenic peptides during the early post-natal period, many of which will have been cleared by the immune system before the development of childhood leukemia. Whilst it is technically possible to

analyze T cell memory – the TCR limb of the tri-molecular complex – of previous encounters with infection in children, this is logistically difficult in the prodromal phase of childhood leukemia. The alternative is to use reverse immunogenetics – the *HLA* component of the tri-molecular complex – to answer this question. Reverse immunogenetics aims to identify the peptide "footprint" of the *HLA* allele associated with disease (childhood leukemia) in order that this can be used to make predictions about the sequence of complementary peptide (i.e. infectious) ligands which bind to that HLA allotype. In fact, software programs able to predict the presence of HLA allotype binding peptides, often known as T cell epitope prediction algorithms, are now widely available, and have been used extensively in vaccine design [29,30]. Whilst it is now relatively easy to identify T cell epitopes in infections using, for example, T cell epitope mapping tools such as TEPITOPE [31], this requires prior knowledge of the infection and the availability of sequence information.

The obvious limitation of this approach in childhood leukemia is that the causal infection may be any of numerous microbes to which children are commonly exposed. Using a "reverse" approach however, we previously reported in the UKCCS that childhood BCP ALL is associated with alleles at the *HLA-DPB1* locus, including *DPB1*0201*, which have a glutamic acid (E) residue at position 69 of the DPβ subunit lining the P4 peptide pocket of the DPα/β heterodimer [32]. This endows *DPB1* alleles having this polymorphic residue ("*DPβ69E*" alleles) with a negatively charged P4 pocket that accommodates antigenic peptides with positively charged amino acids, such as lysine (K), at the P4 anchor position. Since the *HLA-DP* locus is only weakly linked to the *DR/DQ* loci, we believe that the association of BCP ALL with *DPB1*0201* is a primary association. Since peptide binding by different *HLA* alleles can overlap in specificity, it is possible to cluster the very large number of *DP* alleles (>100) into a small number of supertypes based on predictions about their shared peptide binding properties. In our leukemia case-control series it was possible to cluster about 90% of the *DP* alleles into 6 supertypes, 3 supertypes being *DPβ69E* alleles (*DP2*, *DP6* and *DP8*), and 3 being *DPβ69K* alleles (*DP1*, *DP3* and *DP4*). We reported that the *DP2* supertype is associated with susceptibility to BCP ALL [33], whereas none of the 3 *β69K* supertypes were associated with susceptibility. In fact *DP1* strongly protects against BCP ALL, possibly as a result of its ability to bind and present BCP ALL-associated onco-peptides such as TEL-AML1 and AML1 to T cells [34].

To explain the infection-mediated promotion of childhood ALL, Greaves [5] has hypothesized that a transient inflammatory response mediated by cytokines selectively suppresses normal hematopoiesis, leaving a pre-leukemia clone with a selective growth advantage. The observed association of BCP ALL with *DP2* [33] suggests that such an inflammatory reaction may involve a T cell response to a limited repertoire of infection-derived T cell epitopes. Furthermore, the consistent association of *HLA-class II* alleles with autoimmune diseases could be interpreted in the context of childhood BCP ALL as suggesting that T cells activated by certain types of infection in children with *DP2* have a greater propensity to develop auto-immunity. If the target of this transient or chronic auto-immune response was a bone marrow (i.e. self) antigen that was mimicked by infection, this could accommodate the hypothesis advanced by Greaves [5].

To illustrate a possible mechanism, we cite the well-documented example of molecular mimicry in treatment-resistant Lyme disease (TRLD) arthritis. This is a late manifestation of Lyme disease caused by infection with the spirochaete *Borrelia burgdorferei*, in which the association with *HLA-DRB1*0401* suggests an auto-immune etiology [35]. Although patients with TRLD may have no evidence of *B. burgdorferei*, they may have antibodies to its outer surface protein (OspA), indicating previous exposure to the spirochaete. Identification of a *DRB1*0401*-binding nonamer peptide in *B. burgdorferei*, $OspA_{165-173}$ prompted a human peptide sequence homology search that revealed a human lymphocyte functional antigen $alpha_{L332-340}$ (LFA-1, CD11a) peptide with 6 of the 9 amino acids identical to $OspA_{165-173}$ (YVIEGT*SKQ*) [36]. This peptide is located in the I-region of CD11α, a subunit that mediates the binding of LFA-1 to its ligand, ICAM-1. Evidence that the OspA and LFA-1α peptides both bind to DRB1*0401 and elicit cross-reactive T cell responses strongly suggests that $OspA_{165-173}$ mimics $LFA\text{-}1a_{332-340}$ in causing TRLD [37,38].

Based on the TRLD paradigm [35], we employed a bio-informatic approach to predict infection-derived peptides that might be responsible for activating a *DP2*-restricted CD4+ T cell response leading to BCP ALL. Diaz et al [39] have identified the principal amino acid anchors at relative peptide positions P1, 4, 6 and 9 in natural DP2 ligands leading, in theory, to 480 potential ligands (Fig. 2). For these

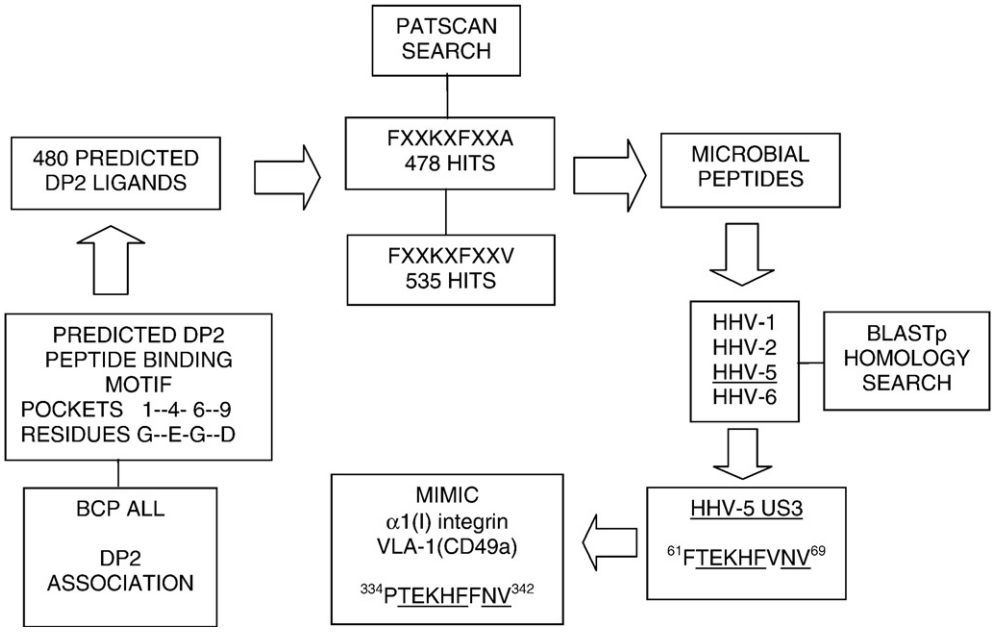

Fig. 2. Use of reverse immunogenetics to predict putative T cell epitopes (HLA-DP ligands) from the DP2 supertype associated with childhood BCP ALL. The two ligands used in the PATSCAN search are based on amino acid residues found most frequently in pockets 1, 4, 6 and 9 of *DPB1*0201* (see text for details). HHV-5 US3 was selected empirically by carrying out BLASTp searches of human viral and bacterial sequences identified by PATSCAN.

we selected residues present in each pocket at highest frequency, including phenylalanine (F) anchored by the hydrophilic pockets 1 and 6, positively charged lysine (K) anchored by pocket 4, and valine (V) or alanine (A) anchored by pocket 9 of DP2. These sequences give rise to two generic peptides, FXXKXFXXV and FXXKXFXXA (where X is an unknown amino acid residue), predicted to have a strong affinity for DP2. To identify microbial peptides with these sequences we used the pattern-scanning algorithm, PATSCAN [40] to search the SWISS-prot protein database [41]. This identified a total of 1013 DP2 ligands, including peptides of human and microbial origin. Selected microbial peptides identified in this way included *Bordetella pertussis*, *Candida albicans*, *Clostridium tetani*, *Escherichia coli*, *Haemophilus influenzae*, *Helicobacter pylori*, *Mycobacterium tuberculosis*, *Mycoplasma pneumoniae*, *Staphylcoccus aureas*, Variola virus and *Yersina pestis*. Among 4 herpes virus sequences with predicted DP2 ligands (HHV-1, -2, -5 and -6), we selected a peptide from the HHV-5 (human cytomegalovirus) unique short early protein, $US3_{61-69}$ (FTEKHFVNV) [42]. CMV is a potential candidate agent for childhood leukemia since it is known to elicit strong CD8+ CTL responses in asymptomatic infants, leading to high levels of long term immunological memory [43]. Using the $US3_{61-69}$ peptide sequence, we searched among non-redundant proteins using the Basic Local Alignment Search Tool (BLASTp) [44], for homologous human sequences. The best match was a peptide from the inserted (I) domain of alpha 1 integrin (CD49a), $\alpha I_{167-175}$, *P*TEKHFFNV which shares 7 of 9 amino acids with $US3_{61-69}$ [45].

The integrins are a widely expressed and diverse family of heterodimeric adhesion receptors that play a major role in cell–matrix and cell–cell interactions [46]. Alpha 1 integrin non-covalently associates with beta 1 integrin to form the $\alpha 1/\beta 1$ complex, VLA-1 (CD49a/CD29) which is expressed by bone marrow stromal myofibroblasts [47]. These cells support the growth of human B cell precursors [48] which are induced to proliferate by activated CD4+ T cells [49]. In light of our empirical analysis, we hypothesize that early HCMV [HHV-5] infection might lead to the development of *DP2*-restricted T cells to HCMV $US3_{61-69}$ that can, in certain situations, cross-react with VLA-1 $\alpha I_{167-175}$. Chronic stimulation of T cells by the VLA-1 $\alpha I_{167-175}$ peptide after clearance of the HCMV infection could lead to the secretion of bone-marrow suppressing cytokines, or direct attack on stromal myofibroblasts by cytotoxic T cells [50]. Either way, this could compromise the adhesion of normal B cell progenitors to stroma [51–53] and could account for the selective survival of pre-leukemia cells, as envisaged by Greaves [5] (Fig. 3). Our hypothesis, that BCP ALL arises as an indirect result of a transient auto-immune induced inflammatory reaction due to molecular mimicry, might explain why this subtype appears to be associated with delayed infection. Children with the *DP2*-supertype may be at higher risk of developing auto-immunity to a *DP2*-restricted T cell epitope such as VLA-1 $\alpha I_{167-175}$, which, with some delay following clearance of the agent, results in immune attack on bone marrow stroma.

Whilst this scenario is currently speculative, it provides a realistic framework for further hypothesis testing involving the confirmation of *HLA* allele associations in childhood leukemia, *in silico* prediction of HLA-restricted infectious and homologous self peptides, and the identification in patients of cross-reactive antibodies and memory T cells to these non-self and self peptides. Based on the results of these studies, it should be possible to recapitulate the impact of infection-induced auto-immune T cells on the growth of normal and malignant B cells *in vitro* in bone marrow stromal cultures to determine if the model is correct.

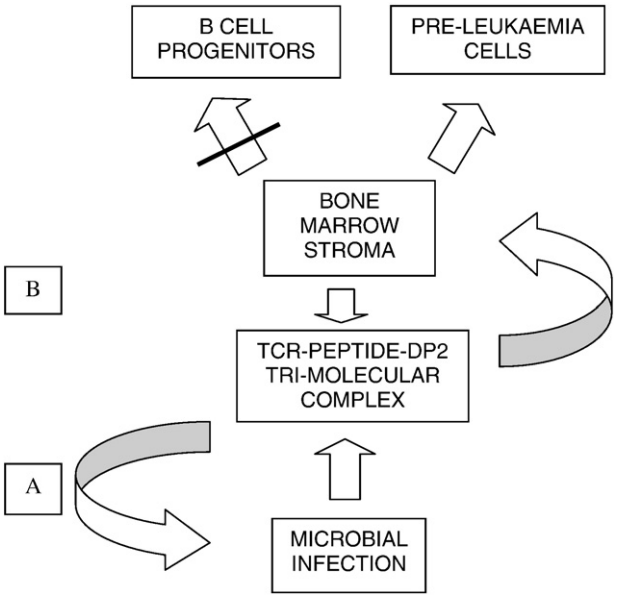

Fig. 3. Schematic representation of proposed events leading to BCP ALL. Microbial infection (HCMV?) elicits T cell response to ^{61}FTEKHFVNV^{69} which cross-reacts with CD49a (^{334}PTEKHFFNV^{342}) on bone marrow stroma. Although the infection is cleared by T cell-mediated immunity [A], bone marrow stroma continues to provide chronic cross-reactive DP2-restricted T cell stimulation [B], thus propagating auto-immune damage to stroma. Since pre-leukemia cells may be less dependent than normal B cell progenitors on interactions with stroma, they have a selective growth advantage.

MHC association mapping (NCCLS)

Recent advancements in high-throughput genotyping technologies, together with the rich genomic information provided by the Human Genome Project and International HapMap Project initiatives, have allowed investigators to effectively interrogate whole genomes in association studies. These approaches have been aided by the understanding that segments of the genome are arranged into distinct haplotype blocks defined by the level of linkage disequilibrium exhibited between neighboring genetic markers [54]. Taking advantage of this knowledge, a highly specific panel of SNPs can be assembled to represent the variation within each haplotype block, which can then be used in combination with select known functional SNPs and other SNPs based on linear spacing in association studies to map regions of the genome involved in disease susceptibility. Genome-wide mapping strategies can also be applied on a smaller scale, for example, to cover only the xMHC region to address more targeted hypotheses. This approach, using both case-control and family-based study designs, has proven effective for other immune-related complex diseases such as several classes of autoimmune diseases including lupus [55], multiple sclerosis [23], and type 1 diabetes [22]. Despite the accumulating support for a role of immunologic factors in the etiology of childhood leukemia [5,6], as well as the aforementioned reports of epidemiologic associations of childhood leukemia with specific HLA loci [32,34], no studies have yet comprehensively reported on the mapping of associations across the complete MHC region. Such an approach can confirm previously-reported childhood leukemia associations with specific HLA loci, help narrow down potential causal variants within these confirmed HLA loci, and also identify new associations with other loci within the MHC.

To begin exploring the role of genetic susceptibility behind several of the documented immune-related associations observed between childhood leukemia and exposure to infections and allergens, we have initiated an association mapping project of the MHC region in the NCCLS. Genotyping of non-Hispanic White and Hispanic subjects (the two largest race/ethnicity populations enrolled in the NCCLS) is being pursued using the Illumina MHC Mapping Panel (Illumina Inc., San Diego, CA). This panel [56] comprises a highly validated series of 1293 evenly spaced SNPs spanning an approximately 4 megabase region of the MHC bordered by the *TRIM27* and *MLN* genes at the telomeric and centromeric ends, respectively. There is an average 3.8 kilobase

spacing between each SNP, covering all major regions of the MHC, including the classical *class I*, *II* and *III* regions, the extended *class II* region and part of the extended *class I* region. The panel set was designed with a strong emphasis on haplotype tagging SNPs, SNPs that are highly informative about nearby SNPs in linkage disequilibrium. This is a notably important feature since the chance of detecting an association is significantly influenced by the ability of the SNP or combination of SNPs to adequately represent the haplotypic diversity of the region. Higher resolution mapping is also possible, via a second Illumina MHC specific panel, the MHC Exon-Centric Panel, which contains 1228 SNPs and emphasizes regions near and within coding exons. Combining the two panels yields a total of 2360 SNP genotypes with an average 2 kilobase spacing between the SNP loci. These Illumina MHC panels offer a cost-effective approach to scan the MHC region. We expect that the mapping effort begun here will help direct future efforts in selecting other specific classical *class I* and *II* HLA loci for allelic typing and the implementation of reverse immunogenetics methodology as described above.

It should be noted that the recent ability to test large numbers of SNP loci in large and diverse study populations has resulted in two major areas of concern, specifically, issues of increased probabilities of false positive findings and population stratification, a type of confounding which may be present due to unaccounted population substructure and/or recent admixture [57] in a multi-ethnic study population. [58–60]. Whilst a statistically significant difference in frequency of a particular SNP or haplotype between cases and controls provides initial evidence for a role in disease susceptibility, replication in independent populations is required to provide credibility for the finding and justification for more focused study to determine the gene and causal variants involved. In terms of false positives, in the approach outlined here, it is important to note that the goal of association mapping is to localize regions of interest. Individual markers with significant *p*-values are less pertinent for association mapping than a group of neighboring markers for whom *p*-values are significant. Thus, we are not concerned with identifying variants of interest, so much as regions that should be explored further in this and other study populations.

Whilst there has been a great deal of discussion regarding criteria for establishing causality in genetic epidemiology studies [61], it is widely agreed that replication is a critical step [62]. A result that has been replicated in one or more study populations is unlikely to be due solely to chance or possible population stratification. We intend to carefully interrogate regions identified through the MHC association mapping effort in other comparable study populations. Such efforts will be facilitated through the Childhood Leukemia International Consortium (CLIC, www.clic.berkeley.edu), a recently assembled consortium that currently includes 14 different epidemiologic studies of childhood leukemia representing populations from 10 different countries including the United Kingdom, Canada, Australia, New Zealand, France, Italy, Brazil, Germany, South Korea, and the United States. Both the UKCCS and the NCCLS are members this group, and plans are in progress to replicate *MHC* findings within CLIC studies.

Conclusions

Technologies are now available for a detailed examination of the role of the complete human xMHC in childhood leukemia. This will assist the dual objectives of defining the contribution of specific HLA molecules in the natural history of childhood leukemia, and identification of additional *xMHC* genes that contribute to the disease. The collection of comprehensive datasets on the timing and types of environmental exposures, including several surrogate measures of exposure to infections (e.g. daycare attendance, reported childhood infections, birth order, etc.) [6], will facilitate an analysis of gene-environment interactions with specific *HLA* alleles. Based on the results of *HLA class II* molecular typing, and the central role of this region in the pathogenesis of auto-immune disease, it is possible to make preliminary predictions about mechanisms involving molecular mimicry that might lead to childhood BCP ALL. Whilst still in the realm of speculation, this can be assisted by peptide pattern-scanning, and sequence homology searches. We have selected empirically a viral (HCMV US3) and self (alpha 1 integrin) peptide with sufficient homology to provide a hypothesis for the infection-induced suppression of normal bone marrow by molecular mimicry. Although further modeling will identify additional candidates, and the mapping efforts described may identify additional regions of interest in the MHC, our hypothesis integrates the *HLA* molecular data with the biological modeling of the early events in the development of childhood leukemia, that should be testable using epidemiological and biological methods.

Acknowledgments

We are grateful to the Kay Kendall Leukaemia Fund (UK), Children with Leukaemia Foundation (UK), and the National Cancer Institute (1R03CA125823-01A1; USA) for funding this work, and to the children and parents participating in the UK Childhood Cancer Study and Northern California Childhood Leukemia Study. This paper is based on a presentation at a Focused Workshop on “Immunity and Infection in the Etiology of Leukemia: a Workshop” sponsored by The Leukemia & Lymphoma Society and held in Sausalito, CA, Nov. 14–16, 2007.

References

[1] M. Smith, L. Ries, J. Gurney, J. Ross, Leukemia, in: L. Ries, M. Smith, J. Gurney, M. Linet, T. Tamra, J. Young, G. Bunin (Eds.), Cancer Incidence and Survival among Children and Adolescents: United States SEER Program 1975–1995, National Cancer Institute, SEER Program, Bethesda, MD, 1999, pp. 17–34.
[2] E. Steliarova-Foucher, C. Stiller, P. Kaatsch, F. Berrino, et al., Geographical patterns and time trends of cancer incidence and survival among children and adolescents in Europe since the 1970s (the ACCISproject): an epidemiological study, Lancet 364 (2004) 2097–2105.
[3] M.F. Greaves, S.M. Colman, M.E. Beard, K. Bradstock, et al., Geographical distribution of acute lymphoblastic leukaemia subtypes: second report of the collaborative group study, Leukemia 7 (1993) 27–34.
[4] M.F. Greaves, J. Wiemels, Origins of chromosome translocations in childhood leukaemia, Nat. Rev., Cancer 3 (2003) 639–649.
[5] M. Greaves, Infection, immune responses and the aetiology of childhood leukaemia, Nat. Rev. Cancer 6 (2006) 193–203.
[6] R.J. McNally, T.O. Eden, An infectious aetiology for childhood acute leukaemia: a review of the evidence, Br. J. Haematol. 127 (2004) 243–263.
[7] S.M. O'Connor, R.S. Boneva, Infectious etiologies of childhood leukemia: plausibility and challenges to proof, Environ. Health Perspect. 115 (2007) 146–150.
[8] H. Mori, S.M. Colman, Z. Xiao, A.M. Ford, et al., Chromosome translocations and covert leukemic clones are generated during normal fetal development, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 8242–8247.
[9] P.F. Ravetto, R. Agarwal, M.L. Chiswick, S.W. D'Souza, et al., Absence of leukaemic fusion gene transcripts in preterm infants exposed to diagnostic X rays, Arch. Dis. Child. Fetal Neonatal Ed. 88 (2003) F237–244.
[10] R. Horton, L. Wilming, V. Rand, R.C. Lovering, et al., Gene map of the extended human MHC, Nat. Rev. Genet. 5 (2004) 889–899.
[11] D.R. Madden, The three-dimensional structure of peptide–MHC complexes, Annu. Rev. Immunol. 13 (1995) 587–622.
[12] J. Trowsdale, HLA genomics in the third millennium, Curr. Opin. Immunol. 17 (2005) 498–504.
[13] A. Sette, J. Sidney, HLA supertypes and supermotifs: a functional perspective on HLA polymorphism, Curr. Opin. Immunol. 10 (1998) 478–482.
[14] J. Hammer, T. Sturniolo, F. Sinigaglia, HLA class II peptide binding specificity and autoimmunity, Adv. Immunol. 66 (1997) 67–100.
[15] U. Christen, M.G. von Herrath, Initiation of autoimmunity, Curr. Opin. Immunol. 16 (2004) 759–767.
[16] J. Klein, Genetic control of virus susceptibility, Chapter 16, Biology of the Mouse Histocompatibility-2 Complex, Springer-Verlag, Berlin, 1975, pp. 389–410.
[17] F. Lilly, E.A. Boyse, L.J. Old, Genetic basis of susceptibility to viral leukaemogenesis, Lancet 2 (1964) 1207–1209.
[18] R. Harris, S.D. Lawler, R.T. Oliver, The HLA system in acute leukaemia and Hodgkin's disease, Br. Med. Bull. 34 (1978) 301–304.
[19] J.T. Casper, M. Marrari, V. Piaskowski, S.J. Lauer, et al., Association between HLA-D region antigens and disease-free survival in childhood non-T, non-B acute lymphocytic leukemia, Blood 60 (1982) 698–702.
[20] E.W. Petersdorf, C. Anasetti, P.J. Martin, J.A. Hansen, Tissue typing in support of unrelated hematopoietic cell transplantation, Tissue Antigens 61 (2003) 1–11.

[21] P.I. de Bakker, G. McVean, P.C. Sabeti, M.M. Miretti, et al., A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC, Nat. Genet. 38 (2006) 1166–1172.
[22] S. Nejentsev, J.M. Howson, N.M. Walker, J. Szeszko, et al., Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A, Nature 450 (2007) 887–892.
[23] M.R. Lincoln, A. Montpetit, M.Z. Cader, J. Saarela, et al., A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis, Nat. Genet. 37 (2005) 1108–1112.
[24] M.M. Miretti, E.C. Walsh, X. Ke, M. Delgado, et al., A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms, Am. J. Hum. Genet. 76 (2005) 634–646.
[25] E.J. Yunis, C.E. Larsen, M. Fernandez-Vina, Z.L. Awdeh, et al., Inheritable variable sizes of DNA stretches in the human MHC: conserved extended haplotypes and their fragments or blocks, Tissue Antigens 62 (2003) 1–20.
[26] UK Childhood Cancer Study Investigators, The United Kingdom Childhood Cancer Study: objectives, materials and methods, Br. J. Cancer 82 (2000) 1073–1102.
[27] M.J. Gardner, M.P. Snee, A.J. Hall, C.A. Powell, et al., Results of case-control study of leukaemia and lymphoma among young people near Sellafield nuclear plant in West Cumbria, BMJ 300 (1990) 423–429.
[28] J.S. Chang, S. Selvin, C. Metayer, V. Crouse, et al., Parental smoking and the risk of childhood leukemia, Am. J. Epidemiol. 163 (2006) 1091–1100.
[29] M.N. Davies, D.R. Flower, Harnessing bioinformatics to discover new vaccines, Drug Discov. Today 12 (2007) 389–395.
[30] A.S. De Groot, J.A. Berzofsky, From genome to vaccine—new immunoinformatics tools for vaccine design, Methods 34 (2004) 425–428.
[31] H. Bian, J. Hammer, Discovery of promiscuous HLA-II-restricted T cell epitopes with TEPITOPE, Methods 34 (2004) 468–475.
[32] G.M. Taylor, S. Dearden, P. Ravetto, M. Ayres, et al., Genetic susceptibility to childhood common acute lymphoblastic leukaemia is associated with polymorphic peptide-binding pocket profiles in HLA-DPB1*0201, Hum. Mol. Genet. 11 (2002) 1585–1597.
[33] G.M. Taylor, A. Hussain, T.J. Lightfoot, J.M. Birch, et al., HLA-associated susceptibility to childhood B-cell precursor ALL: definition and role of HLA-DPB1 supertypes, Br. J. Cancer 98 (2008) 1125–1131.
[34] M. Taylor, C. Harrison, T. Eden, J. Birch, et al., HLA-DPB1 supertype-associated protection from childhood leukaemia: relationship to leukaemia karyotype and implications for prevention, Cancer Immunol. Immunother. 57 (2008) 53–61.
[35] R.A. Kalish, J.M. Leong, A.C. Steere, Association of treatment-resistant chronic Lyme arthritis with HLA-DR4 and antibody reactivity to OspA and OspB of *Borrelia burgdorferi*, Infect. Immun. 61 (1993) 2774–2779.
[36] D.M. Gross, T. Forsthuber, M. Tary-Lehmann, C. Etling, et al., Identification of LFA-1 as a candidate autoantigen in treatment-resistant Lyme arthritis, Science 281 (1998) 703–706.
[37] A.C. Steere, B. Falk, E.E. Drouin, L.A. Baxter-Lowe, et al., Binding of outer surface protein A and human lymphocyte function-associated antigen 1 peptides to HLA-DR molecules associated with antibiotic treatment-resistant Lyme arthritis, Arthritis Rheum. 48 (2003) 534–540.
[38] C. Trollmo, A.L. Meyer, A.C. Steere, D.A. Hafler, et al., Molecular mimicry in Lyme arthritis demonstrated at the single cell level: LFA-1 alpha L is a partial agonist for outer surface protein A-reactive T cells, J. Immunol. 166 (2001) 5286–5291.
[39] G. Diaz, B. Canas, J. Vazquez, C. Nombela, et al., Characterization of natural peptide ligands from HLA-DP2: new insights into HLA-DP peptide-binding motifs, Immunogenetics 56 (2005) 754–759.
[40] M. Dsouza, N. Larsen, R. Overbeek, Searching for patterns in genomic data, Trends Genet. 13 (1997) 497–498.
[41] A. Bairoch, B. Boeckmann, S. Ferro, E. Gasteiger, Swiss-Prot: juggling between evolution and stability, Brief Bioinform. 5 (2004) 39–55.
[42] T.R. Jones, E.J. Wiertz, L. Sun, K.N. Fish, et al., Human cytomegalovirus US3 impairs transport and maturation of major histocompatibility complex class I heavy chains, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 11327–11333.
[43] H. Komatsu, A. Inui, T. Sogo, T. Fujisawa, et al., Large scale analysis of pediatric antiviral CD8+ T cell populations reveals sustained, functional and mature responses, Immun. Ageing 3 (2006) 11.
[44] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[45] R. Briesewitz, M.R. Epstein, E.E. Marcantonio, Expression of native and truncated forms of the human integrin alpha 1 subunit, J. Biol. Chem. 268 (1993) 2989–2996.
[46] R.O. Hynes, Integrins: versatility, modulation, and signaling in cell adhesion, Cell 69 (1992) 11–25.
[47] F. Deschaseaux, P. Charbord, Human marrow stromal precursors are alpha 1 integrin subunit-positive, J. Cell. Physiol. 184 (2000) 319–325.
[48] I. Moreau, V. Duvert, C. Caux, M.C. Galmiche, et al., Myofibroblastic stromal cells isolated from human bone marrow induce the proliferation of both early myeloid and B-lymphoid cells, Blood 82 (1993) 2396–2405.
[49] N. Renard, V. Duvert, D. Blanchard, J. Banchereau, et al., Activated CD4+ T cells induce CD40-dependent proliferation of human B cell precursors, J. Immunol. 152 (1994) 1693–1701.
[50] I.S. Misko, S.M. Cross, R. Khanna, S.L. Elliott, et al., Crossreactive recognition of viral, self, and bacterial peptide ligands by human class I-restricted cytotoxic T lymphocyte clonotypes: implications for molecular mimicry in autoimmune disease, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 2279–2284.
[51] T.B. Geijtenbeek, Y. van Kooyk, S.J. van Vliet, M.H. Renes, et al., High frequency of adhesion defects in B-lineage acute lymphoblastic leukemia, Blood 94 (1999) 754–764.
[52] J.L. Liesveld, J.M. Winslow, K.E. Frediani, D.H. Ryan, et al., Expression of integrins and examination of their adhesive function in normal and leukemic hematopoietic cells, Blood 81 (1993) 112–121.
[53] A. Manabe, K.G. Murti, E. Coustan-Smith, M. Kumagai, et al., Adhesion-dependent survival of normal and leukemic human B lymphoblasts on bone marrow stromal cells, Blood 83 (1994) 758–766.
[54] J.D. Wall, J.K. Pritchard, Haplotype blocks and linkage disequilibrium in the human genome, Nat. Rev. Genet. 4 (2003) 587–597.
[55] M.M. Fernando, C.R. Stevens, P.C. Sabeti, E.C. Walsh, et al., Identification of two independent risk factors for lupus within the MHC in United Kingdom families, PLoS Genet. 3 (2007) e192.
[56] A. Oliphant, D.L. Barker, J.R. Stuelpnagel, M.S. Chee, BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping, Biotechniques Suppl (2002) 56–58, 60–51.
[57] M.L. Freedman, D. Reich, K.L. Penney, G.J. McDonald, et al., Assessing the impact of population stratification on genetic association studies, Nat. Genet. 36 (2004) 388–393.
[58] Framework for a fully powered risk engine, Nat. Genet. 37 (2005) 1153.
[59] D.C. Thomas, J.S. Witte, Point: population stratification: a problem for case-control studies of candidate-gene associations? Cancer Epidemiol. Biomarkers Prev. 11 (2002) 505–512.
[60] S. Wacholder, N. Rothman, N. Caporaso, Counterpoint: bias from population stratification is not a major threat to the validity of conclusions from epidemiological studies of common polymorphisms and cancer, Cancer Epidemiol. Biomark. Prev. 11 (2002) 513–520.
[61] J.P. Ioannidis, P. Boffetta, J. Little, T.R. O'Brien, et al., Assessment of cumulative evidence on genetic associations: interim guidelines, Int. J. Epidemiol. 37 (2008) 120–132.
[62] S.J. Chanock, T. Manolio, M. Boehnke, E. Boerwinkle, et al., Replicating genotype–phenotype associations, Nature 447 (2007) 655–660.