***HLA* and *KIR* Associations of Cervical Neoplasia**

Xiao Bao[1,2,3], Aimee L. Hanson[4], Margaret M. Madeleine[5], Sophia S. Wang[6], Stephen M. Schwartz[5], Felicity Newell[2], Ulrika Pettersson-Kymmer[7,8], Kari Hemminki[9,10], Sven Tiews[11], Winfried Steinberg[11], Janet S. Rader[12], Felipe Castro[13], Mahboobeh Safaeian[14], Eduardo L. Franco[15], François Coutlée[16], Claes Ohlsson[17,18], Adrian Cortes[4], Mhairi Marshall[2], Pamela Mukhopadhyay[4], Katie Cremin[4], Lisa G. Johnson[5], Suzanne M. Garland[19,20], Sepehr N. Tabrizi[19,20], Nicolas Wentzensen[14], Freddy Sitas[21,22,23], Cornelia Trimble[24], Julian Little[25], Maggie Cruickshank[26], Ian H. Frazer[27], Allan Hildesheim[14], Matthew A. Brown[2*], Emma L. Duncan[2*], YingPu Sun[1,3*], Paul J. Leo[2*]

*These authors contributed equally to this work.

1. Center for Reproductive Medicine, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, 450052, China.

2. Institute of Health and Biomedical Innovation, School of Biomedical Sciences, Queensland University of Technology (QUT), Translational Research Institute, Princess Alexandra Hospital, Woolloongabba, QLD, 4102, Australia.

3. Henan Key Laboratory of Reproduction and Genetics, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, 450052, China.

4. University of Queensland Diamantina Institute, Translational Research Institute, Princess Alexandra Hospital, Woolloongabba, QLD, 4102, Australia.

5. Program in Epidemiology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

6. Department of Population Sciences, Beckman Research Institute, City of Hope, Duarte, CA, USA.

7. Department of Pharmacology and Clinical Neuroscience, Umeå University, Sweden.

8. Department of Public Health and Clinical Medicine, Umeå University, Sweden.

9. Division of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.

10. Center for Primary Health Care Research, Lund University, Sweden.

11. MHC Laboratory for Cytopathology, Dr.Steinberg GmbH, Soest, Germany.

12. Department of Obstetrics and Gynecology, Medical College of Wisconsin, Milwaukee, Wisconsin 53226, USA.

13. Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany.

14. Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA.

15. Division of Cancer Epidemiology, McGill University, Montreal, QC, Canada.

16. Département de Microbiologie, Infectiologie et Immunologie, Centre Hospitalier de l'Université de Montréal, Montréal, QC, Canada.

17. Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.

18. Centre for Bone and Arthritis Research, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.

19. Western Pacific Regional Human Papillomavirus Laboratory Network, Department of Microbiology and Infectious Diseases, The Royal Women's Hospital, Locked Bag 300, Grattan St & Flemington Rd, Parkville, Victoria, 3052, Australia.

20. Department of Obstetrics and Gynaecology, University of Melbourne, Murdoch Childrens Research Institute, The Royal Children's Hospital, Flemington Rd, Parkville, Victoria, 3052, Australia.

21. Cancer Council NSW, Sydney, NSW, Australia.

22. Sydney School of Public Health, University of Sydney, Camperdown, NSW, Australia.

23. School of Public Health and Community Medicine, University of New South Wales, Kensington, NSW, Australia.

24. Center for Cervical Dysplasia, Johns Hopkins University, Baltimore, Maryland, United States.

25. School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, Canada.

26. Division of Medical Education, University of Aberdeen, Aberdeen, Scotland.

27. Faculty of Medicine and Biomedical Sciences, The University of Queensland, Translational Research Institute, Princess Alexandra Hospital, Woolloongabba, QLD, 4102, Australia.

Summary

Human leukocyte antigen and killer immunoglobulin-like receptor alleles were assessed for association with cervical neoplasia. Our findings suggest *HLA-C1* group alleles protect against HPV16-related cervical neoplasia, mainly through a *KIR*-mediated mechanism.

FOOTNOTES

Conflict of Interests

The authors have no conflict of interest to declare.

Funding

This work was supported by: National Health and Medical Research Council (Australia) Senior Principal Research Fellowship to MAB; the Australian Cancer Research Foundation; National Cancer Institute at the National Institutes of Health [grants P01CA042792 and R01CA112512]; National Health and Medical Research Council (Australia) [grant 387701]; Cancer Council New South Wales (Australia); Canadian Institutes of Health Research [grant MOP-42532]; Réseau sida et maladies infectieuses du Fonds de recherche du Québec – Santé; Swedish Research Council; the Swedish Foundation for Strategic Research; the LUA-ALF research grants in Gothenburg and Umeå; Lundberg Foundation; Torsten and Ragnar Söderberg's Foundation; Novo Nordisk Foundation; European Commission [grant HEALTH-F2-2008-201865-GEFOS, BBMRI.se]; Swedish Society of Medicine; Kempe-Foundation (grant JCK-1021); the Medical Faculty of Umeå University; County Council of Västerbotten (grant Spjutspetsanslag VLL:159:33-2007); JL holds a Tier 1 Canada Research Chair in Human Genome Epidemiology; We acknowledge Royal Women's Hospital Clinical Research Foundation for funding.

Corresponding Authors:

Professor YingPu Sun,

Center for Reproductive Medicine,

Henan Key Laboratory of Reproduction and Genetics,

The First Affiliated Hospital of Zhengzhou University,

Zhengzhou, Henan, China.

Telephone: +86 (0371) 66913635

Fax: +86 (0371) 66913635

syp2008@vip.sina.com

Associate Professor Paul Leo,

Institute of Health and Biomedical Innovation,

School of Biomedical Sciences,

Queensland University of Technology (QUT),

Translational Research Institute,

Princess Alexandra Hospital,

Woolloongabba, Brisbane, Australia.

Telephone: +61 7 34437366

Fax: +61 7 34437099

p2.leo@qut.edu.au

ABSTRACT

Background: Cervical cancer is the fourth most common cancer in women, and we recently reported human leukocyte antigen (*HLA*) alleles showing strong associations with cervical neoplasia risk and protection. HLA ligands are recognised by killer immunoglobulin-like receptors (KIRs) expressed on a range of immune cell subsets, governing their proinflammatory activity. We hypothesized that the inheritance of particular *HLA-KIR* combinations would increase cervical neoplasia risk.

Methods: Here, we used *HLA* and *KIR* dosages imputed from SNP genotype data from 2,143 cervical neoplasia cases and 13,858 healthy controls of European decent.

Results: Four novel HLA alleles were identified in association with cervical neoplasia: *HLA-DRB3*9901* (OR=1.24, P=2.49×10^{-9}), *HLA-DRB5*0101* (OR=1.29, P=2.26×10^{-8}), *HLA-DRB5*9901* (OR=0.77, P=1.90×10^{-9}) and *HLA-DRB3*0301* (OR=0.63, P=4.06×10^{-5}), due to their linkage disequilibrium with known cervical neoplasia-associated *HLA-DRB1* alleles. We also found homozygosity of *HLA-C1* group alleles is a protective factor for HPV16-related cervical neoplasia (*C1/C1,* OR=0.79, P=0.005). This protective association was restricted to carriers of either *KIR2DL2* (OR=0.67, P=0.00045) or *KIR2DS2* (OR=0.69, P=0.0006).

Conclusions: Our findings suggest that *HLA-C1* group alleles play a role in protecting against HPV16-related cervical neoplasia, mainly through a *KIR*-mediated mechanism.

Keywords: Cervical neoplasia; human leukocyte antigens (HLA); killer immunoglobulin-like receptors (KIRs); HPV16-related cervical neoplasia.

BACKGROUND

Cervical cancer is the fourth most common cancer in women, with over 500,000 new cases presenting world-wide in 2012, and accounting for 7.5% of female cancer deaths [1]. Its impact is particularly high in young women as the second commonest cancer affecting women aged 20-39 years [2]. Cervical cancer results from chronic infection with human papillomavirus (HPV), with the HPV genome detected in nearly all cervical cancers. Of the different HPV types, HPV16 and –18 are the most frequently involved, and together account for approximately 70% of cervical cancers world-wide [3]. Whilst infection with HPV is essentially universal, most cervical HPV infections are cleared by the immune system [4, 5], and only ~1% of women with cervical HPV infection develop cervical cancer [6].

Genetic factors strongly influence persistence of HPV infection and risk of cervical cancer. HPV persistence is associated with the monogenic disorders epidermodysplasia verruciformis and WHIM syndrome, from mutations in *EVER1/2* and *CXCR4*, respectively [7, 8]. The only robust common variant genetic associations with cervical cancer are with genes of the major histocompatibility complex (MHC), in particular human leukocyte antigens (HLA). We demonstrated that haplotypes *HLA-DRB1*1501/HLA-DQB1*0602/HLA-DQA1*0102* and *HLA-DQA1*0301/HLA-DRB1*0401* increase the risk of HPV-associated cervical neoplasia, and that the allele *HLA-B*15* and haplotype *HLA-DRB1*1301/HLA-DQB1*0603* are protective. Of note, *HLA-DRB1*1301/HLA-DQA1*0103/HLA-DQB1*0603* is associated with protection from oral and pharyngeal cancer, particularly HPV-positive cases [9]. We showed that the HLA risks of cervical neoplasia were determined by amino acids at positions 13 and 71 in pocket 4 of HLA-DRB1 and position 156 in HLA-B [10]. Common genetic variant contribution to cervical neoplasia susceptibility is substantial (36%) [10], although a large component of the heritability has yet to be elucidated.

HLA proteins are critical for antigen presentation to effector cells of the adaptive immune system [11]. HLA Class I complexes are expressed on all nucleated cells and present endogenous, intracellular-derived antigens, as well as pathogen-derived peptides (as with viral infection), with residue length of 8-10 amino acids. These are recognized by $CD8^+$ T-cells that can engage foreign peptides through their T-cell receptor. Conversely, natural killer (NK) cells, a component of the innate immune system, are able to respond to downregulated surface HLA, a consequence of the immune evasion strategy of some viruses to avoid $CD8^+$ T-cell recognition. Once activated, these lymphocytes can kill the antigen-presenting cell via release of cytotoxic granules. HLA Class II complexes typically present extracellular-derived antigens, such as bacterial pathogens. After endocytosis, these proteins are processed and presented on the cell surface bound to MHC Class II, to initiate an immune response from $CD4^+$ cells.

Interaction between HLA, viral epitopes, and killer-immunoglobulin-like receptors (KIR) expressed on NK cells lead either to activation or inhibition of NK cell cytotoxic activity. KIRs are expressed on all NK cells, and a minority of T-cells (including some $CD4^+$, $CD8^+$ and $\gamma\delta$ cells). Seventeen *KIR* genes have been identified, encoded within the leukocyte receptor complex (LRC) on chromosome 19q13.4, all of which share significant homology (85-99% DNA sequence similarity) [12, 13]. They are encoded in variable gene content haplotypes with activating and inhibitory counterparts. Different inhibitory and activating KIRs demonstrate specificity for different HLA subgroups, providing fine-tuning of NK and KIR-bearing T-cell responses [14].

Variability at the *KIR* locus includes allelic, gene combination (haplotypic), and expression level differences, the latter under significant epigenetic control [15]. *KIR* genes are inherited in haplotypes of vastly diverse content, ranging from 4-14 receptor-encoding loci, with over 50 distinct haplotypes

based on gene content alone [13, 16]; with ~700 allelic variants have been reported [16, 17]. Group 1 HLA-C (*HLA-C1*) allotypes have an asparagine at residue 80, and are ligands for the inhibitory receptors, *KIR2DL2* and *KIR2DL3*, which segregate as alleles of a single locus, and *KIR2DS2* [18]. The remaining HLA-C allotypes (group 2, *HLA-C2*) have a lysine at position 80, and are ligands for *KIR2DL1* (an inhibitory receptor) and *KIR2DS1* (the homologous activating receptor). HLA-B *Bw4* allotypes serve as ligands for *KIR3DL1* and *KIR3DS1* (Fig. 1).

The immunological mechanisms involved in clearance of HPV are poorly understood. NK cells are crucial for clearing viral infection and for anti-tumour immunity, and are thought to be important in HPV control [19]. Regulation of NK cell responses depend on *KIR* genotype, *HLA* genotype, heterozygosity versus homozygosity for each of these, interaction of *HLA* and *KIR*, and changes in KIR and HLA expression. The development of imputation-based methods for *HLA* typing, and more recently for *KIR* typing, has enabled the use of SNP-typed datasets to investigate genetic associations with these loci in large cohorts. Here, *HLA* and *KIR* gene imputation and association tests were performed using genotype data from 2,143 cervical neoplasia cases and 13,858 healthy controls of European descent, to test whether *HLA-KIR* combinations are associated with cervical neoplasia risk.

METHODS AND MATERIALS

Study Population

Phenotypic information, including cervical histology and HPV genotype, appears in Supplementary Table S1. As described in Leo *et al.* [10], all cases needed to have CIN2/3 (if CIN2, age over 30 to limit analysis to women who failed to clear HPV) or cervical cancer (HPV type and histology not always known). HPV DNA types from tumor tissues were categorized into six groups: a) HPV16-positive (but not HPV18); b) HPV18-positive (but not HPV16); c) neither HPV16 nor HPV18; d) both HPV16 and HPV18; e) negative for any HPV (noting that usually samples are only tested for HPV16 and HPV18); and f) those not tested for HPV genotype [10].

Genotyping

Case samples were SNP microarray genotyped in-house. 791 cases were genotyped using Illumina OmniExpress BeadChips ('Omni') and 1,352 cases using Illumina Human660-Quad BeadChips ('660Q'). Controls were genotyped using Illumina Immunochip BeadChips ('Ichip'; Fig. 2). Bead intensity data were processed and normalized for each sample, and genotypes called within participating studies using GenomeStudio, and verified manually and corrected where necessary. Standard quality control measures were performed [10], with particular care taken in comparison of the performance of different chip types.

HLA imputation

HLA alleles were inferred using HLA*IMP:03 (http://www.biorxiv.org/content/early/2016/12/09/091009). Individuals with posterior probability of HLA alleles < 0.6 were excluded from downstream association testing (Supplementary Fig. S1). HLA groups (C1, C2, Bw4, Bw6) were inferred from known allele classifications. HLA amino acids were inferred by SNP2HLA [20], using a reference panel from the Type 1 Diabetes Genetics Consortium (N=5,225). Amino acids imputed by SNP2HLA with r^2 <0.5 were excluded, and samples where the allele dosage at any HLA type exceeded 2.5 were removed.

KIR imputation

KIR genes and haplotypes were imputed with KIR*IMP [21] using SNP genotypes across the *KIR* locus. This requires certain informative 'key' SNPs for accurate imputation, which vary according to the SNP chip used (see Supplementary Table S2). To avoid imputation bias caused by different numbers of key SNPs genotyped by the Omni and 660-Quad BeadChips, the two groups of cases were imputed separately, with a separate comparator group for each obtained by randomly dividing the control group into two: Ichip control group 1 (Ichip1, n=6,703) and group 2 (Ichip2, n=6,725; Fig 2). Imputed data for *KIR* and *HLA* loci were compared between the two control groups. There were no significant differences between the two control groups for 271 KIR SNPs (noting that most of these were only available on Ichip, not on chips used for case genotyping, hence the smaller number available for case-control analyses). Of 256 *HLA* alleles, 10 alleles (3.9%) were significantly different (P<0.05) between the two control groups. Most (8/10) were rare (≤0.05); the remaining two had frequency 0.08. None of these SNPs were significant in subsequent analyses. The 77 key SNPs for Omni cases and Ichip1, and the 66 key SNPs for 660Quad cases and Ichip2, were re-clustered manually (converting array intensity data for each allele of the SNP concerned to genotype calls) before imputation to improve genotyping accuracy (Supplementary Table S3). Individuals with posterior probability of accurate *KIR* imputation for each KIR allele <0.6 were excluded in the downstream association testing.

Statistical Methods

Population stratification was assessed via principal component analysis of genome-wide genotypes using Shellfish (http://www.stats.ox.ac.uk/~davison/software/shellfish/shellfish.php). Association analyses were performed using custom R scripts. Three models were applied to test *HLA-C1*, *HLA-C2*, *HLA-Bw4, HLA-Bw6* and *HLA* alleles, and their combination with *KIR* genes: a) dosage model, treating genotypes as 0, 1, or 2 copies; b) dominant model, treating 1 or 2 copy genotypes as present, 0 as

absent; and c) recessive model to study the difference of homozygotes and heterozygotes of HLA-B and HLA-C, treating 2 copy genotypes as present, 0 or 1 copy as absent. Meta-analysis of combination between Omni-Ichip1 and 660Q-Ichip2 was performed using METAL software (http://csg.sph.umich.edu/abecasis/metal/). For tests over all HLA alleles, the multiple testing correction methodology was employed [22, 23], using a correlation matrix derived from SNP2HLA imputation with all 1027 *HLA* alleles and amino acids. This analysis estimated 206 independent loci, implying a Bonferroni correction to $P=2.4\times10^{-4}$ for a type I error rate of 5%. For the *HLA-KIR* interaction using 12 *KIR* alleles we applied the most conservative Bonferroni correction ($P=(0.05/(206\text{x}12) = 2.0\times10^{-5}$ for statistical significance).

RESULTS

Quality Control

2,143 cases and 13,428 controls passed quality control. Amongst cases, 736 were squamous cell carcinoma, 542 adenocarcinoma, and 865 with histology unspecified. Four principal components were used as covariates to control for population stratification, calculated using 11,980 common SNPs shared by the Illumina OmniExpress, 660-Quad and Immunochip SNP microarrays. To assess population stratification, we used a subset of 333 "null" SNPs outside the MHC region included on each chip type, and avoiding SNPs included on Immunochip, because of their potential immunogenetic significance, associated with reading and learning disability, schizophrenia and psychosis. From this, the genomic inflation factor overall was calculated as 1.018 (Supplementary Fig. S2). No divergences were observed between cases and controls, or different genotype platforms (Supplementary Fig. S3). All *HLA* loci were imputed above 95% accuracy (Supplementary Table S4), and all *KIR* loci imputed above 80% accuracy (Supplementary Table S3).

MHC Findings

HLA Association Analysis

Consistent with our recent study [10], we found increased and decreased risk of cervical neoplasia associated with *HLA* haplotypes (Supplementary Table S5), and determined these associations are carried by amino acids at positions 13 and 71 in pocket 4 of HLA-DRB1, and position 156 in HLA-B. Using these in-depth imputation methods across *HLA* and *KIR* loci, we detected further novel associations from *HLA-DRB3* and *–DRB5*.

Novel risk associations with cervical neoplasia were identified with *HLA-DRB3*9901* (OR=1.24, $P=2.49\times10^{-9}$; Table 1) and *HLA-DRB5*0101* (OR=1.29, $P=2.26\times10^{-8}$). Although *HLA-DRB3*9901* is not in linkage disequilibrium with any individual cervical neoplasia *HLA* risk allele ($P<0.05$), adjusting for the association with *HLA-DRB1* amino acid positions 37, 71 and 96 attenuates the association with *HLA-DRB3*9901* ($P>0.01$). *HLA-DRB5*0101* is in positive linkage disequilibrium with the HLA Class II risk allele, *HLA-DRB1*1501* ($r^2=0.99$), and protective allele, *HLA-DRB5*9901* ($r^2=0.93$; Table 1 and Fig. 3). No residual association *HLA-DRB5*0101* was observed after controlling for the association of *HLA-DRB1*1501*, *HLA-DRB5*9901*, or *HLA-DRB1* amino acid positions 13 and 71 ($P<0.01$).

Novel inverse associations with cervical neoplasia were observed with *HLA-DRB5*9901* (OR=0.77, $P=1.90\times10^{-9}$) and *HLA-DRB3*0301* (OR=0.63, $P=4.06\times10^{-5}$). *HLA-DRB5*9901* is in positive linkage disequilibrium with the HLA Class II risk alleles, *HLA-DRB1*1501* ($r^2=0.93$) and *HLA-DRB5*0101* ($r^2=0.93$; Table 1 and Fig. 3). No residual association at these two *HLA* loci was observed after controlling for the association of *HLA-DRB1* amino acid 71 ($P<0.01$; Table 1).

HLA alleles were associated with disease when assessing HPV genotype, without regard to histologic classification. Comparing HPV16-related cases (N=667) with all controls (N=13,428), risk associations were observed with *HLA-DRB3*9901* (OR=1.43, $P=1.27\times10^{-8}$; Supplementary Table S6). Controlling for the association with *HLA-DRB1* amino acid positions 13 and 37 controls for the association of *HLA-DRB3*9901* ($P>0.01$). For HPV18-related cervical neoplasia (cases=166), the strongest associated *HLA* allele was *HLA-DPA1*0103* (OR=1.89, $P=0.00039$; Supplementary Table S7). Considering histopathology, reduced risk was seen in *HLA-DRB5*9901* with squamous cell carcinoma (OR=0.72, $P=6.57\times10^{-5}$). No residual association was observed after controlling for the association of *HLA-DRB1* amino acids 13, 17 and 96 ($P<0.01$). Increased risk was seen in *HLA-DRB3*9901* with adenocarcinoma (OR=1.31, $P=7.38\times10^{-5}$; Supplementary Table S8). No residual association was observed after controlling for the association of *HLA-DRB1* amino acids 13, 71 and 96 ($P<0.01$).

HLA-Bw4/Bw6 and HLA-Cw1/Cw2 Association Analysis

Dominant and dosage models showed a nominal risk association of *HLA-Bw4* in HPV16-related cervical neoplasia (OR=1.24, $P=0.014$ and OR=1.16, $P=0.011$, respectively; Table1 and Supplementary Table S9). The dosage model suggests that *HLA-Bw4* has a weak inverse association with HPV18-related cervical neoplasia (OR=0.78, $P=0.04$). No association was found between *HLA-Bw4* or *Bw6* with cervical cancer overall, squamous cell carcinoma or adenocarcinoma.

All three models suggested *HLA-C1/2* alleles are associated with HPV16-related cervical neoplasia (Supplementary Table S10). *HLA-C1* and *HLA-C2* are mutually exclusive, meaning one can have two copies of either or one copy of each. Both *HLA-C1* and *HLA-C2* can interact with specific KIR and lead to inhibitory or activating signaling. To investigate the role of *HLA-C1* or *HLA-C2*, and their combination with KIR, we employed a recessive model to distinguish *HLA-C* homozygotes and heterozygotes. The frequency of individuals with two copies of *HLA-C1* alleles was lower in HPV16-related cervical neoplasia cases (35.7%) than controls (41.3%) ($P=0.005$, OR=0.79; Table 2). The

frequency of two copies of *HLA-C2* alleles did not differ between the groups. No association was found between *HLA-C1* or *HLA-C2* with cervical cancer overall, HPV18-related cervical cancer, squamous cell carcinoma or adenocarcinoma.

KIR Findings

KIR imputation concordance was explored by comparison of *KIR* haplotype frequencies between four datasets (Supplementary Fig. S4), comparison with published population prevalences (Supplementary Fig. S5), and control-control association test (group1 vs. group2, using the randomly divided Ichip dataset; Supplementary Table S11). For the control-control association test of 16 imputed *KIR* loci, four (*KIR2DS3, KIR2DL1, KIR2DP1* and *KIR2DL5*) were inconsistent due to the different key SNP numbers in each group, and were not analysed further. Prevalences of the other 12 *KIR* genes were concordant between groups. No association was found between *KIR* genes with cervical cancer overall, or with HPV18-related cervical cancer. Weak protective association was seen with *KIR2DL2* with HPV16-related cervical cancer (P_{meta}=0.04, OR_{omni}=0.87, OR_{660Q}=0.83).

KIR-HLA-Bw4/Bw6 and KIR-HLA-C1/2 combination

No *KIR-HLA-Bw4/Bw6 or KIR-HLA-C1/2* combination was associated with cervical neoplasia, HPV18-related cervical neoplasia, squamous cell carcinoma or adenocarcinoma. *HLA-Bw4* alleles were associated with increased risk of HPV16-associated cervical neoplasia, this association being restricted to *KIR3DL1* carriers (P_{meta}=0.0085; OR=1.22). No association was seen in individuals presenting with Bw6 *HLA-B* alleles.

KIR2DL3 and KIR2DL2 bind HLA-C1 allotypes, with KIR2DL2 binding with greater affinity [18]. The protective association for HPV16-related cervical neoplasia of *HLA-C1/C1* was restricted to individuals carrying *KIR2DL2* (P_{meta}=0.00045; OR=0.67), or *KIR2DS2* (P_{meta}=0.0006; OR=0.69), these

KIR alleles often being found together on KIR haplotypes. In our dataset, *KIR2DL2* and *KIR2DS2* are in near complete linkage disequilibrium (r^2=0.99). *KIR2DL2* and *KIR2DL3* were not associated with cervical neoplasia in individuals who were lacking *HLA-C1/C1*. HLA-C2 can interact with either KIR2DS1 or KIR2DL1. No association with HPV16-related cervical neoplasia was seen in individuals with *KIR2DS1* and *HLA-C2*. *KIR2DL1* could not be investigated in this study due to the low *KIR2DL1* imputation accuracy.

KIR and HLA allele combinations

The strongest associations of any combination of *KIR* and *HLA* alleles with cervical neoplasia was *HLA-B*5501* and *KIR2DS2* (P_{meta}=5.97×10^{-5}; OR_{omni}=0.61, OR_{660Q}=0.19) and *HLA-B*5501* and *KIR2DL2* (P_{meta}=0.00013, OR_{omni}=0.6, OR_{660Q}=0.2, Table 3 and Supplementary Fig. S6). *HLA-B*5501* was not associated with cervical neoplasia independent of this interactive association with *KIR* genes.

DISCUSSION

Here, we report novel associations of combinations of *HLA* and *KIR* alleles with cervical neoplasia. We also extend our previous findings of protective and risk haplotypes associated with cervical neoplasia, demonstrating association of *HLA-DRB3* and *HLA-DRB5* alleles with the disease, and confirm that these are due to linkage disequilibrium with HLA-DRB1 amino acids [10]. In our previous GWAS, no significant association was noted at the LRC on chromosome 19q13, but many SNPs at this locus failed quality control because the complex genetic structure of the locus leads to reduced accuracy of genotype calling by automated algorithms. In the current study, careful manual checking of genotype calling was performed, *KIR* genotypes were imputed, and analysed in combination with *HLA* alleles.

We used imputation methods to infer *HLA* and *KIR* genotypes from SNP microarray data to perform one of the largest *HLA-KIR* association studies reported to date. By comparing our imputed *KIR* haplotype and *KIR* gene prevalences with published data generated using direct genotyping approaches, we confirmed that concordance between genotyped and imputed data is high. We further compared our imputation findings with direct *HLA* and *KIR* genotype data in 86 1000 Genome study samples. The concordance for *HLA* genes in 2-digit and 4-digit resolution was 99.77% and 99.42%, respectively (data not shown). Whilst we were able to impute 16 *KIR* genes, the imputation of four genes was inconsistent between our two cohorts, and therefore not considered. Of the remaining twelve, two (*KIR3DP1, KIR2DL4*) are framework *KIR* genes present in all individuals, and thus association with disease cannot be calculated for these genes. For the remaining ten *KIR* genes, we did not find differences between cases and controls in isolation, but demonstrate suggestive associations in combination with specific *HLA* alleles.

We identified three new HLA allelic associations with cervical neoplasia, *HLA-DRB5*0101* and *–DRB3*9901* as risk factors, and *HLA-DRB3*301* as a protective factor. *HLA-DRB3*, *–DRB4* and *–DRB5* are paralogues of *HLA-DRB1* with which they are in linkage disequilibrium, and their expression level is one fifth that of *HLA-DRB1* (RefSeq, Jul 2008). Consistent with the strong linkage disequilibrium across this locus, the associations of the *HLA-DRB3* and *HLA-DRB5* alleles with cervical neoplasia in this study were due to linkage disequilibrium with *HLA-DRB1* alleles and amino acid variants.

Homozygosity of the HLA-C1 genotype group (*C1/C1*) overall was associated with protection from HPV16-related cervical neoplasia. This protective association was restricted to carriers of either *KIR2DL2* or *KIR2DS2*, suggesting it operates through a KIR-mediated mechanism. A role for *HLA-C1* genotypes in cervical neoplasia is supported by previous studies, although the direction of

association has not been consistent, and studies have generally been small. A study of cervical cancer-affected parent-case trios showed that HLA-C1 was over-transmitted in women with invasive cervical cancer (P=0.04), particularly in the subgroup of women infected with HPV16 or –18 types (P=0.008) [24]. A study using unrelated cases and controls of European ancestry showed *HLA-C1/KIR2DL2* and *HLA-C1/KIR2DL3* pairs were decreased in patients with HPV-positive cervical lesions, and increased in HPV high-risk infected patients, compared with HPV low-risk group [25]. Whilst these studies are not definitive, they suggest involvement of *HLA-C1* group alleles in modulating the risk of cervical neoplasia, perhaps through their function as KIR ligands.

Several small candidate gene studies of the *KIR* locus and cervical cancer have been performed. Consistent with our findings, a Western Australian cohort study of 147 cases demonstrated weak association of *KIR2DL2* and *KIR2DS2* with high-grade CIN (CIN2 and 3) overall (P=0.046 and P=0.049, respectively) [26]. A study of Eastern U.S. and Costa Rican individuals (196 cases, 330 controls) indicated *KIR3DS1* increased, and *HLA-C2* alleles reduced, the risk of cervical cancer [27]. Our data do not support these findings. Other studies from Sweden (65 cases) [28], Brazil (79 cases) [29] and Korea (132 cases) [30] have not reported positive associations, although their samples sizes were too small to identify anything other than essentially monogenic risk-associations.

The strongest *HLA-KIR* combination association with cervical neoplasia observed in this study was between *HLA-B*5501* and *KIR2DL2/DS2* (P_{meta}=5.97×10^{-5}), although this did not meet the conservative Bonferroni-corrected significance threshold of P=2.0×10^{-5} for *HLA-KIR* interactions. Whilst *HLA-B*5501* frequency did not differ between cases (3.5%, 71/2,015) and controls (3.4%, 426/12,458), suggestive associations of the combination of *HLA-B*5501* and either *KIR2DS2* or *KIRSDL2* were observed. The signal was mainly contributed to by the 660Q-Ichip2 data. No imputation bias was found in *KIR2DS2*, *KIR2DL2* and *HLA-B*5501* between case-case and control-

control tests, suggesting that this is not due to imputation artefact, but in the absence of clear replication, further studies are required to determine its significance. *HLA-B55* has been reported to be associated with cervical neoplasia, but this has not been replicated in larger studies [31]. KIR2DS2 and KIR2DL2 bind a range of HLA-C1 group allotypes, which include HLA-B46 and HLA-B73, but not HLA-B55. HLA-B55 is an HLA-Bw6 allele, a group that does not bind KIR. It is possible that the association we observed here is due to linkage disequilibrium with other HLA types, or it may be a non-informative suggestive association, particularly given that it was only found in comparison with one of the two imputation sets.

CONCLUSIONS

Cervical neoplasia may be associated with *HLA-C1* group alleles that interact with *KIR* in controlling NK cell activation. Further HLA associations were demonstrated and shown to be driven by linkage disequilibrium with known amino acid components of HLA-DRB1 allelic associations of disease, further supporting that these variants are key to HLA-associations of cervical neoplasia. No definitive *KIR* associations were noted with cervical neoplasia, although we only examined *KIR* gene carriage and not allelic variation within *KIR* genes, for which larger and more densely sequenced reference datasets for imputation will be required. Further studies of *KIR* allelic variation are warranted, particularly given the role of KIR-bearing cells in immunity to viral infection, and the suggestive association of HLA-C1 variants observed with disease.

Acknowledgements

Jessica Darlington-Brown coordinated specimen collection for the NSW component. We thank Dr David Pennisi for proof-reading and copy editing the manuscript.

REFERENCES

1. Ferlay J SI, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Available at: http://globocan.iarc.fr. Accessed 8/Sep 2017.

2. Fidler MM, Gupta S, Soerjomataram I, Ferlay J, Steliarova-Foucher E, Bray F. Cancer incidence and mortality among young adults aged 20-39 years worldwide in 2012: a population-based study. Lancet Oncol **2017**; 18:1579-89.

3. Li N, Franceschi S, Howell-Jones R, Snijders PJ, Clifford GM. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication. Int J Cancer **2011**; 128:927-35.

4. Dunne EF, Park IU. HPV and HPV-associated diseases. Infect Dis Clin North Am **2013**; 27:765-78.

5. Stewart BW, Wild C, International Agency for Research on Cancer, World Health Organization. World cancer report 2014. Lyon, France

Geneva, Switzerland: International Agency for Research on Cancer

WHO Press, **2014**.

6. Schiffman M, Glass AG, Wentzensen N, et al. A long-term prospective study of type-specific human papillomavirus infection and risk of cervical neoplasia among 20,000 women in the Portland Kaiser Cohort Study. Cancer Epidemiol Biomarkers Prev **2011**; 20:1398-409.

7. Ramoz N, Rueda LA, Bouadjar B, Montoya LS, Orth G, Favre M. Mutations in two adjacent novel genes are associated with epidermodysplasia verruciformis. Nature genetics **2002**; 32:579-81.

8. Hernandez PA, Gorlin RJ, Lukens JN, et al. Mutations in the chemokine receptor gene CXCR4 are associated with WHIM syndrome, a combined immunodeficiency disease. Nature genetics **2003**; 34:70-4.

9. Lesseur C, Diergaarde B, Olshan AF, et al. Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. Nature genetics **2016**; 48:1544-50.

10. Leo PJ, Madeleine MM, Wang S, et al. Defining the genetic susceptibility to cervical neoplasia-A genome-wide association study. PLoS Genet **2017**; 13:e1006866.

11. Science NYG. The Major Histocompatibility Complex and Its Functions. Immunobiology: The Immune System in Health and Disease, **2001**.

12. Martin AM, Freitas EM, Witt CS, Christiansen FT. The genomic organization and evolution of the natural killer immunoglobulin-like receptor (KIR) gene cluster. Immunogenetics **2000**; 51:268-80.

13. Pyo CW, Wang R, Vu Q, et al. Recombinant structures expand and contract inter and intragenic diversification at the KIR locus. BMC Genomics **2013**; 14:89.

14. Fauriat C, Ivarsson MA, Ljunggren HG, Malmberg KJ, Michaelsson J. Education of human natural killer cells by activating killer cell immunoglobulin-like receptors. Blood **2010**; 115:1166-74.

15. Schenk A, Bloch W, Zimmer P. Natural Killer Cells--An Epigenetic Perspective of Development and Regulation. Int J Mol Sci **2016**; 17:326.

16. Norman PJ, Hollenbach JA, Nemat-Gorgani N, et al. Defining KIR and HLA Class I Genotypes at Highest Resolution via High-Throughput Sequencing. Am J Hum Genet **2016**; 99:375-91.

17. Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res **2015**; 43:D423-31.

18. Winter CC, Gumperz JE, Parham P, Long EO, Wagtmann N. Direct binding and functional transfer of NK cell inhibitory receptors reveal novel patterns of HLA-C allotype recognition. J Immunol **1998**; 161:571-7.

19. Villa LL. Human papillomaviruses and cervical cancer. Adv Cancer Res **1997**; 71:321-41.

20. Jia X, Han B, Onengut-Gumuscu S, et al. Imputing amino acid polymorphisms in human leukocyte antigens. PLoS One **2013**; 8:e64683.

21. Vukcevic D, Traherne JA, Naess S, et al. Imputation of KIR Types from SNP Variation Data. Am J Hum Genet **2015**; 97:593-607.

22. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb) **2005**; 95:221-7.

23. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet **2004**; 74:765-9.

24. Martin MP, Borecki IB, Zhang Z, et al. HLA-Cw group 1 ligands for KIR increase susceptibility to invasive cervical cancer. Immunogenetics **2010**; 62:761-5.

25. Rizzo R, Gentili V, Rotola A, Bortolotti D, Cassai E, Di Luca D. Implication of HLA-C and KIR alleles in human papillomavirus infection and associated cervical lesions. Viral Immunol **2014**; 27:468-70.

26. Brestovac B, Wong ME, Tjendera R, Costantino PJ, Mamotte C, Witt CS. Human papillomavirus, high-grade intraepithelial neoplasia and killer immunoglogulin-like receptors: a Western Australian cohort study. Infect Agent Cancer **2013**; 8:33.

27. Carrington M, Wang S, Martin MP, et al. Hierarchy of resistance to cervical neoplasia mediated by combinations of killer immunoglobulin-like receptor and human leukocyte antigen loci. J Exp Med **2005**; 201:1069-75.

28. Arnheim L, Dillner J, Sanjeevi CB. A population-based cohort study of KIR genes and genotypes in relation to cervical intraepithelial neoplasia. Tissue Antigens **2005**; 65:252-9.

29. Marangon AV, Guelsin GA, Visentainer JE, et al. The association of the immune response genes to human papillomavirus-related cervical disease in a Brazilian population. Biomed Res Int **2013**; 2013:146079.

30. Song MJ, Lee CW, Kim JH, et al. Association of KIR genes and HLA-C alleles with HPV-related uterine cervical disease in Korean women. Tissue Antigens **2013**; 81:164-70.

31. Krul EJ, Schipper RF, Schreuder GM, Fleuren GJ, Kenter GG, Melief CJ. HLA and susceptibility to cervical neoplasia. Hum Immunol **1999**; 60:337-42.

Table 1. Analysis of novel cervical neoplasia associated *HLA* alleles, conditioned on *HLA* alleles and HLA amino acids.

	HLA Alleles	*DRB3*9901*	*DRB5*0101*	*DRB5*9901*	*DRB3*0301*
	Unconditioned *P*-values	2.49×10^{-9}	2.26×10^{-8}	1.90×10^{-8}	4.06×10^{-6}
	Odds ratio	1.24	1.29	0.77	0.63
Conditioned *P*-values	*DRB3*9901*	NA	4.19×10^{-5}	3.66×10^{-5}	0.001033
	*DRB5*0101*	3.07×10^{-5}	NA	0.45	4.43×10^{-5}
	*DRB5*9901*	3.21×10^{-5}	0.66	NA	4.50×10^{-5}
	*DRB3*0301*	1.24×10^{-6}	8.7×10^{-8}	7.41×10^{-8}	NA
	*DRB1*1501*	1.99×10^{-5}	0.21	0.37	5.28×10^{-5}
	*B*0702*	9.64×10^{-6}	0.0040	0.0040	6.49×10^{-5}
	*DQB1*0602*	6.94×10^{-6}	0.0059	0.0043	1.81×10^{-5}
	*DRB1*1302*	3.18×10^{-6}	4.33×10^{-8}	4.45×10^{-8}	0.14
	DRB1_11	0.0018	0.055	0.056	0.0063
	DRB1_13	4.73×10^{-6}	0.01	0.0096	0.013
	DRB1_37	0.11	0.01	0.0092	0.0063
	DRB1_71	0.013	0.04	0.037	0.84
	DRB1_96	0.49	0.00058	0.0006	0.001
	DRB1_13 & 71	2.26×10^{-5}	0.37	0.94	0.54
	DRB1_71 & 96	0.84	0.046	0.086	0.7

Table 2. *KIR-HLA-Bw4/ Bw6* and *KIR-HLA-C1/C2* type combinations associated with HPV16-related cervical neoplasia.

Genetic factor	Frequency HPV16+ cases	Frequency controls	OR	95%CI	*P*-values
HLA-C1/C1	232/649 (35.7%)	5434/13148 (41.3%)	0.79	0.67-0.93	0.005
HLA-C1/C2	326/649 (50.2%)	6054/13148 (46%)	1.18	1.01-1.38	0.039
HLA-C2/C2	91/649 (14.0%)	1660/13148 (12.6%)	1.13	0.90-1.42	0.29
2DL2	295/633 (46.6%)	6658/13073 (50.9%)	0.84	0.72-0.99	0.04
2DL3	596/647 (92.1%)	11942/13188 (90.6%)	1.22	0.91-1.63	0.25
2DS2	310/648 (47.8%)	6751/13167 (51.3%)	0.87	0.74-1.02	0.08
2DL2-C1/C1	96/624 (15.4%)	2754/12932 (21.3%)	0.67	0.54-0.84	0.00045
C1/C1 in 2DL2+	96/291 (33.0%)	2754/6587 (41.8%)	0.69	0.51-0.88	0.0029
C1/C1 in 2DL2-	131/333 (39.3%)	2599/6345 (41.0%)	0.93	0.75-1.17	0.56
2DL3-C1/C1	216/638 (33.9%)	4890/13044 (37.5%)	0.85	0.72-1.01	0.09
2DL2-C1/C2	158/624 (25.3%)	3027/12932 (23.4%)	1.11	0.92-1.33	0.29
2DL3-C1/C2	291/638 (45.6%)	5432/13044 (41.6%)	1.18	1.00-1.38	0.053
2DS2-C1/C1	101/639 (15.8%)	2791/13023 (21.4%)	0.69	0.55-0.85	0.0006
C1/C1 in 2DS2+	101/306 (33.0%)	2791/6677 (41.8%)	0.68	0.54-0.87	0.0024
C1/C1 in 2DS2-	131/333 (39.3%)	2599/6346 (41.0%)	0.93	0.75-1.17	0.56

2DS2-C1/C2	165/639 (25.8%)	3068/13023 (23.6%)	1.13	0.94-1.35	0.23
2DS1-C2/C2	22/652 (3.4%)	573/13199 (4.3%)	0.77	0.50-1.19	0.15
HLA-Bw4	406/623 (65.2%)	7459/12375 (60.3%)	1.24	1.04-1.46	0.014
3DL1-HLA-Bw4	388/623 (62.3%)	7111/12375 (57.5%)	1.22	1.03-1.44	0.0085
HLA-Bw4 in 3DL1+	388/601 (64.5%)	7111/11805 (60.2%)	1.20	1.02-1.43	0.034
HLA-Bw4 in 3DL1-	18/22 (81.8%)	348/570 (61.1%)	2.8	0.96-8.6	0.059
3DS1+HLA-Bw4	156/623 (25.0%)	2773/12371 (22.4%)	1.16	0.96-1.39	0.09

Frequencies of *HLA-B, HLA-C* and *KIR-HLA-B, C* combinations among individuals with HPV16 infected cervical cancer cases and healthy controls are shown. *HLA-C1C1* indicates two group 1 *HLA-C* alleles, *HLA-C2C2* indicates two group 2 *HLA-C* alleles, and *HLA-C1C2* indicates one of each. *HLA-Bw4* indicates one or two group *Bw4* alleles. *P*-values were calculated by using R code glm model with principle components 1-4, combination *P*-values were calculated from meta-analysis between Omni-Ichip1 and 660Q-Ichip2 datasets; a positive odds ratio indicates a protective association with HPV16 infection.

CI: 95% Confidence Interval; HLA-C1: HLA-C01, 03, 07, 08, 12, 13, 14, 16 and HLA-B4601, 7301; HLA-C2: HLA-C02, 04, 05, 06, 15, 17, 18; HLA-Bw4: HLA-B05, 5102, 5103, 13, 17, 27, 37, 38, 44, 47, 49, 51, 52, 53, 57, 58, 59, 63, 77; OR: Odds Ratio.

Table 3. *KIR-HLA* combinations are associated with cervical neoplasia.

Combination			OR		Sample size		FRQ of HLA		FRQ of KIR		KIR+HLA+		KIR-HLA-		KIR+HLA-		KIR-HLA+	
HLA alleles	KIR genes	*P*-values	Omni-Ichip1	660Q-Ichip2	CASE	CON	CASE	CON	CASE	CON	CASE	CON	CASE	CON	CASE	CON	CASE	CON
*HLA-B*5501*	*KIR2DS2*	$5.97x10^{-5}$	0.61	0.19	2019	12650	0.034	0.034	0.52	0.51	21	240	923	5994	1028	6230	47	186
*HLA-B*5501*	*KIR2DL2*	0.00013	0.6	0.21	1983	12563	0.034	0.033	0.511	0.508	21	233	922	5993	993	6151	47	186
*HLA-DQB1*0601*	*KIR2DL3*	0.0032	0.14	0.17	2062	12975	0.012	0.0097	0.905	0.905	17	118	188	1215	1850	11634	7	8
*HLA-DRB1*1502*	*KIR2DL3*	0.0042	0.2	0.18	1920	12349	0.012	0.0104	0.908	0.908	16	119	170	1133	1727	11088	7	9
*HLA-B*3801*	*KIR2DS4TOTAL*	0.0047	0.16	0.19	2056	12800	0.025	0.023	0.956	0.955	46	287	85	575	1919	11931	6	7
*HLA-B*3801*	*KIR3DL1ex4*	0.0047	0.16	0.16	2055	12807	0.025	0.023	0.956	0.954	46	287	85	576	1918	11937	6	7
*HLA-B*3801*	*KIR3DL1ex9*	0.0048	0.16	0.19	2053	12805	0.026	0.023	0.956	0.954	47	290	85	575	1915	11933	6	7
*HLA-DQA1*0103*	*KIR2DL3*	0.0078	0.57	0.5	2078	13106	0.084	0.120	0.905	0.906	147	1422	170	1086	1733	10453	28	145

Number of individuals with cervical neoplasia cases and healthy control present *KIR-HLA* combinations are shown. Combination *P*-values were calculated from meta-analysis between Omni-Ichip1 and 660Q-Ichip2 datasets. *P*-values of each dataset were calculated by using R code glm model with principal components 1-4. CON: controls; OR: Odds Ratio.

FIGURE LEGENDS

Figure 1. KIR proteins are classified by the number of extracellular immunoglobulin domains (2D or 3D) and by whether they have a long (L) or short (S) cytoplasmic domain. Inhibitory KIRs and KIR2DL4 have immunoreceptor tyrosine-based inhibitory motifs (ITIM) in their cytoplasmic domains. Activating KIRs possess a basic amino acid in the transmembrane domain, which allows interactions with the accessory molecule, DAP-12, delivering activating signals through its immunoreceptor tyrosine-based activating (ITAM) motif. The ligands for several KIR proteins are subsets of HLA class I proteins. KIR2DL4 has a charged amino acid and ITIM motifs, and it interacts with the accessory protein, FcεRI-γ, which sends an activating signal via its ITAM similar to DAP-12. Note that HLA-Bw6 alleles are not known to be KIR ligands.

Figure 2. *HLA* and *KIR* gene imputation and association tests were performed using genotype data from 2,143 cervical neoplasia cases and 13,858 healthy controls of European descent. 791 individuals were genotyped using Illumina OmniExpress Beadchip ("Omni") and 1,352 individuals using Illumina Human660-Quad BeadChip ("660Q"). All 13,428 controls were genotyped using Illumina Immunochip BeadChip. To avoid imputation bias caused by different number of key SNPs between Omni and 660Q, these two groups of cases were imputed separately, and compared with two control groups – Immunochip control group 1 (Ichip1, n=6,703) and group2 (Ichip2, n=6,725) – obtained by randomly dividing the controls. 77 key SNPs for Omni cases and Ichip1 and 66 key SNPs for 660Q cases and Ichip2 were re-clustered and used for imputation. Case-control analyses were conducted separately, followed by meta-analysis.

Figure 3. Pairwise linkage disequilibrium (r^2) plot of *HLA* alleles associated with cervical neoplasia. HLA alleles are clustered according to their pairwise linkage disequilibrium on both the x– and y-axes. On the y-axis, alleles are labelled as to whether they are risk or protective alleles in the overall cervical neoplasia dataset, and on the x-axis according to whether they are HLA Class I or II alleles.

Figure 1.

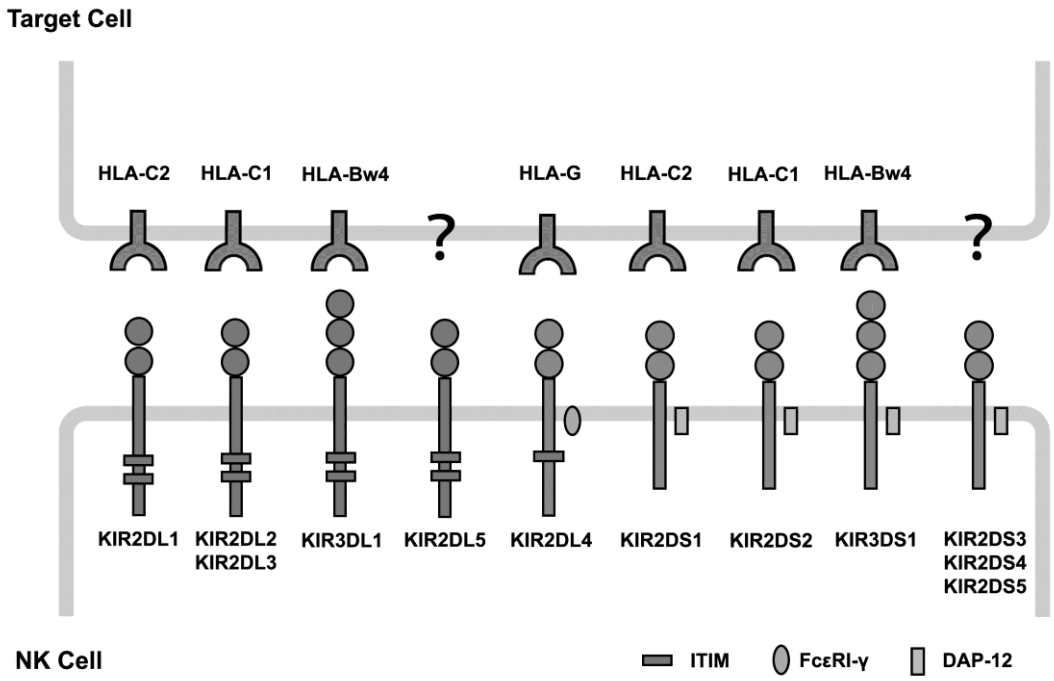

Figure 2.

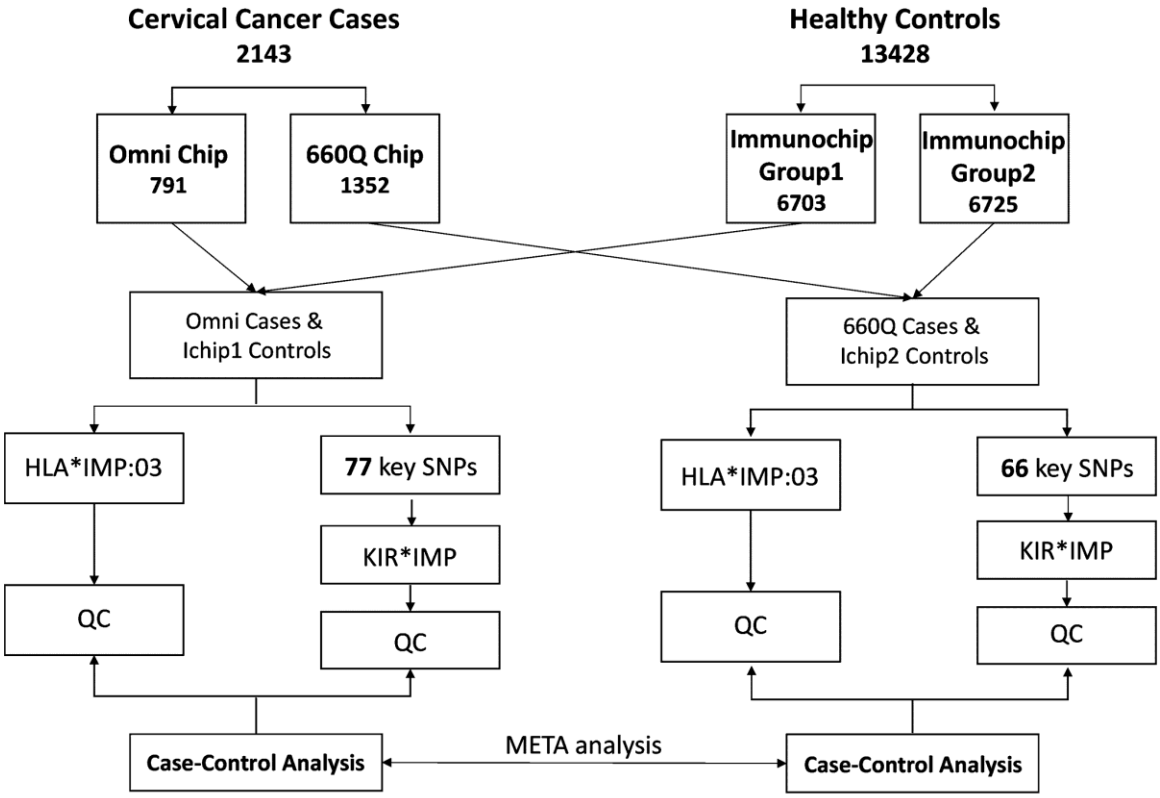

Figure 3.