

RESEARCH ARTICLE

Open Access



Variants in *CXCR4* associate with juvenile idiopathic arthritis susceptibility

Terri H. Finkel^{1,2,18*}, Jin Li^{3†}, Zhi Wei^{4†}, Wei Wang⁴, Haitao Zhang³, Edward M. Behrens^{1,2}, Emma L. Reuschel^{1,2}, Sophie Limou⁵, Carol Wise⁶, Marilyn Punaro⁷, Mara L. Becker⁸, Jane E. Munro^{9,10}, Berit Flatø¹¹, Øystein Førre¹¹, Susan D. Thompson¹², Carl D. Langefeld¹³, David N. Glass^{12^}, Joseph T. Glessner³, Cecilia E. Kim³, Edward Frackelton³, Debra K. Shivers¹, Kelly A. Thomas³, Rosetta M. Chiavacci³, Cuiping Hou³, Kexiang Xu³, James Snyder³, Haijun Qiu³, Frank Mentch³, Kai Wang^{3,14}, Cheryl A. Winkler⁵, Benedicte A. Lie¹⁵, Justine A. Ellis^{9,16} and Hakon Hakonarson^{2,3,17*}

Abstract

Background: Juvenile idiopathic arthritis (JIA) is the most common chronic rheumatic disease among children, the etiology of which involves a strong genetic component, but much of the underlying genetic determinants still remain unknown. Our aim was to identify novel genetic variants that predispose to JIA.

Methods: We performed a genome-wide association study (GWAS) and replication in a total of 1166 JIA cases and 9500 unrelated controls of European ancestry. Correlation of SNP genotype and gene expression was investigated. Then we conducted targeted resequencing of a candidate locus, among a subset of 480 cases and 480 controls. SUM test was performed to evaluate the association of the identified rare functional variants.

Results: The *CXCR4* locus on 2q22.1 was found to be significantly associated with JIA, peaking at SNP rs953387. However, this result is subjected to subpopulation stratification within the subjects of European ancestry. After adjusting for principal components, nominal significant association remained ($p < 10^{-4}$). Because of its interesting known function in immune regulation, we carried out further analyses to assess its relationship with JIA. Expression of *CXCR4* was correlated with *CXCR4* rs953387 genotypes in lymphoblastoid cell lines ($p = 0.014$) and T-cells ($p = 0.0054$). In addition, rare non-synonymous and stop-gain sequence variants in *CXCR4*, putatively damaging for *CXCR4* function, were significantly enriched in JIA cases ($p = 0.015$).

Conclusion: Our results suggest the association of *CXCR4* variants with JIA, implicating that this gene may be involved in the pathogenesis of autoimmune disease. However, because this locus is subjected to population stratification within the subjects of European ancestry, additional replication is still necessary for this locus to be considered a true risk locus for JIA. This cell-surface chemokine receptor has already been targeted in other diseases and may serve as a tractable therapeutic target for a specific subset of pediatric arthritis patients with additional replication and functional validation of the locus.

Keywords: Juvenile idiopathic arthritis, Genome-wide association study, *CXCR4*, Targeted resequencing

* Correspondence: tfinkel@nemours.org; hakonarson@email.chop.edu

†Equal contributors

^Deceased

¹Division of Rheumatology, The Children's Hospital of Philadelphia, 19104 Philadelphia, PA, USA

²Department of Pediatrics, University of Pennsylvania School of Medicine, 19104 Philadelphia, PA, USA

Full list of author information is available at the end of the article



Background

Pediatric arthritis is the leading cause of acquired disability in children, afflicting about one in 1000 children worldwide [1, 2], all ethnicities and both genders, with onset as early as the first year of life. Classification schemes for pediatric arthritis are under evolution, akin to the recent classification changes for adult rheumatoid arthritis [3]; juvenile rheumatoid arthritis (JRA) is the term used historically in North America, while juvenile idiopathic arthritis (JIA) is the preferred name elsewhere, and is now used increasingly worldwide. JIA is defined as a group of chronic arthritides of unknown etiology, occurring in children from 0 to 16 years of age [4]. Morbidity associated with JIA can be life-long – with as many as 50 % of children with JIA entering adulthood with active disease [1, 2] – and represents a significant medical, financial, and emotional burden for patients, their families, and society. In the United States alone, arthritis and rheumatic diseases impact more than 46 million adults and 300,000 children, at a cost of \$128 billion annually in direct and indirect medical costs [5–7]. Adults with JIA have lower rates of employment than matched controls, and health-related quality of life is diminished in adults with JIA, particularly in those with active disease [8]. Prompt recognition of the disease is important in preventing permanent disability; however, lack of specific confirmatory testing often delays diagnosis. The optimal management of JIA remains complicated and poorly defined, despite recent advances in therapy [1, 2].

The etiology of JIA is largely unknown. To our knowledge, there are no data supporting a major role for environmental exposures [9]. This does not preclude a role of the environment in the pathogenesis of JIA, but research to identify environmental risk factors is lacking. On the other hand, strong contribution of genetic components has been implicated from twin to family studies [10]: monozygotic twins have a concordance rate between 25 and 40 %; the calculated sibling recurrence risk ratio ($\lambda_s = 15\text{--}30$) is similar to that calculated for type 1 diabetes. Sibling pairs tend to show concordance for age of onset, subtype and course; and a subset of patients with JIA exhibits a heritable predisposition to develop this disease with an autosomal dominant pattern of inheritance.

Yet, in comparison to other autoimmune diseases of similar prevalence, the genetic etiology of JIA remains largely elusive. The major histocompatibility complex (MHC), in particular, the *HLA-DRB1* locus, has been established as having the strongest influence on susceptibility to JIA [11], contributing ~20 % of the proportion of sibling recurrent risk [12]. Non-MHC loci are important as well, with 16 loci now associated with JIA at genome-wide significance. Fourteen of these were identified for the first time by a recent Immunochip analysis

[13], a hypothesis-driven approach that focused upon genes with known associations with immune disorders [14]. To comprehensively search for genes related to JIA and given that the pathophysiological mechanisms underlying JIA are unknown, we took an unbiased approach of genome-wide association study (GWAS) and performed replication studies in independent cohorts, including a total of 1166 cases and 9500 controls after quality control (QC) filtering. We subsequently performed targeted resequencing at identified candidate locus of *CXCR4* gene among a subset of 480 cases and 490 controls. Here we report that variants in gene *CXCR4* associate with JIA.

Methods

Participants

The JIA cases in our study were recruited from five sites in USA, Australia, and Norway: Texas Scottish Rite Hospital for Children (TSRHC; Dallas, Texas), Children's Mercy Hospitals and Clinics (CMHC; Kansas City, Missouri), the Children's Hospital of Philadelphia (CHOP; Philadelphia, Pennsylvania), the Murdoch Childrens Research Institute (MCRI; Royal Children's Hospital, Melbourne, Australia), and Oslo University Hospital (OUH; Oslo, Norway). (Table 1, Additional file 1: Table S1). A subset of subjects from these sites has been described previously [15–19]. JIA diagnosis was made according to the International League of Associations for Rheumatology (ILAR) revised criteria [4] and confirmed using the JIA CalculatorTM software (URLs) [20], an algorithm-based tool adapted from the ILAR criteria. All JIA cases were of age of onset <16 years old.

The clinical data of JIA case in the CHOP cohort were collected from the JIA Registry maintained within the CHOP Division of Rheumatology; clinical data of case samples from TSRHC, CMHC, MCRI, and OUH were

Table 1 Demographic and clinical characteristics of our JIA dataset

JIA Subtypes ^A	% female	Age (years) ^B
Oligoarthritis, persistent	73 %	5.9 (2.9,9.7)
Oligoarthritis, extended	82 %	3.95 (2.4,7.8)
Polyarthritis, RF negative	79 %	7.48 (3.2,11.3)
Polyarthritis, RF positive	95 %	13.5 (10.3,15.1)
Systemic arthritis	67 %	6.6 (3.2,11.1)
Enthesitis-related arthritis	40 %	11.9 (9.0,14.1)
Psoriatic arthritis	71 %	9.9 (7.0,13.2)
Undifferentiated arthritis	55 %	9.8 (4.5,13.8)
Total	68 %	8.3 (3.8,12.2)

^ARevised ILAR criteria.

^BAge (years): median (25 % percentile, 75 % percentile)

drawn from medical records provided by the respective sites and stored in a de-identified database at the Center for Applied Genomics of the CHOP Research Institute.

The control subjects used are unrelated and disease-free children recruited within the CHOP Healthcare Network. Control subjects had no history of JIA or other chronic illnesses and were screened as negative for a diagnosis of autoimmune diseases, based on data from CHOP's electronic health record and by intake questionnaires obtained by the recruiting staff from the Center for Applied Genomics. A total of 6500 pediatric controls passed stringent quality control (QC) filtering, as detailed below; post-QC, cases and controls were matched based on the multidimensional scaling (MDS) analysis [21, 22]. For OUH cohort, the 3000 well-characterized subjects from the Wellcome Trust Case-control Consortium (WTCCC) [21] were used as controls.

We combined TSRHC and CMHC samples to form the discovery cohort, and kept CHOP, MCRI and OUH cohorts as three independent replication cohorts.

Ethics statement

The study was approved by the institutional review boards of TSRHC, CMHC, CHOP, MCRI, OUH, and CCHMC, and was compliant with HIPAA regulations. Parental written informed consent was obtained from all participants prior to inclusion in this study for the purpose of DNA collection and genotyping.

Genotyping

All samples except those in the OUH replication cohort were genotyped using Illumina HumanHap550 BeadChip or the Human610-Quad arrays. The 530,000 SNPs that are overlapped by these two platforms were included in the study. Samples in the OUH replication cohort were genotyped using the Affymetrix GeneChip 500 K Mapping Array Set.

Quality control

We employed the MDS algorithm implemented in PLINK to infer population structure, with the 924 individuals from the HapMap project as reference. Only samples of genetically inferred European ancestry were kept for further analysis. We then applied QC filters to exclude samples of poor genotyping quality prior to association analysis. A sample was excluded if the genotype call rate was <95 % or if the sample showed excess or deficient heterozygosity (inbreeding coefficient $|F| > 0.1$). Cryptic relatedness or erroneous duplicates were evaluated using pair-wise identity-by-descent estimation, and the sample with lower genotype call rate was removed from each identified relative pair. In our study, we also eliminated SNPs with genotype call rate <98 %, with minor allele frequency (MAF) <1 % in either cases

or controls, or if there was significant departure from Hardy-Weinberg equilibrium ($p < 0.0001$). In the discovery cohort, a total of 518,907 genotyped SNPs passed QC and were included in analysis.

Principal component analysis (PCA)

After the above sample and SNP QC, we conducted principal component analysis, trying to resolve relationship within-European samples. Using PLINK, we performed LD pruning and SNP exclusion so that only independent SNPs ($r^2 < 0.2$) on the autosomes were kept for PCA. We then performed PCA using EIGENSTRAT [23]. MDS analysis was also performed with the independent SNPs using PLINK.

Association analyses

For genotyped SNPs, association was tested by basic allelic test (chi-square test) and the odds ratio was calculated with respect to the minor allele using PLINK [22]. Logistic regression analyses were additionally performed, including the first 10 coordinates from the MDS analysis as covariates. Similar analyses were also performed including the first 10 principal components from the PCA analysis as covariates. Conditional association analysis was performed by including the genotype of the most associated SNP rs953387 as a covariate.

Imputation

Imputation was carried out using software MACH 1.0 [24], with the reference panel of the HapMap CEU samples (HapMap release 22, NCBI build 36). The default two-step procedure was adopted for imputation. Imputed SNPs with MAF <0.01 in either cases or controls and SNPs with poor imputation quality ($r^2 < 0.3$) were excluded from further analysis. We also zeroed out imputed genotypes with a posterior probability of <0.9.

Meta-analysis

We performed meta-analysis using the inverse variance based method implemented in software METAL [25], which accounts for the direction of association relative to a consistent reference allele and adopts a fixed effect model. In this method, the effect size estimate of each cohort is weighted by its corresponding standard error. All meta-analyses comply with MOOSE guidelines (URLs).

Targeted resequencing

We selected 480 patients with JIA and 480 controls without history of any autoimmune or inflammatory diseases, all of European ancestry based on the above MDS analysis. Samples were pooled in batches of 8 cases or 8 controls. One control pool was excluded for final analysis because it failed QC. Library preparation for

targeted resequencing was performed according to the TruSeq (Illumina) sample-preparation protocol. DNA libraries were then hybridized to customized probes for capturing *CXCR4* with NimbleGen SeqCap EZ Choice Library (Roche NimbleGen). The captured region is chr2:136871907–136895725, including introns. *CXCR4*-enriched libraries were sequenced on the HiSeq 2000 (Illumina).

We performed sequence alignment using BWA against the reference human genome (UCSC hg19). We achieved ~320X coverage per pool or ~40X per individual. We performed variant calling using SNVer [26], a statistical tool designed for pooled sequencing data. We used ANNOVAR software [27] to annotate variants. Each pool has $2 \times 8 = 16$ haplotypes, so we estimated allele frequency by rounding $X/K \times 16$, where X is the number of reads carrying the alternative allele, and K is the total coverage.

Statistical analysis for targeted sequencing

For the sequencing data, we employed the SUM test [28] for testing the association of the identified multiple non-synonymous variants. We computed the SUM association p value using the R package Assotester. Specifically, we used a permutation version of SUM, in order to prevent an inflated Type I error.

Sanger sequencing

We performed Sanger sequencing to validate the rare non-synonymous and stop-gain sequence variants identified by targeted resequencing. Primers for all the five variants were designed using software Primer3 [29, 30] (URLs). Purification of PCR products was conducted using ExoSAP-it (USB, Affymetrix), and Sanger sequencing on both strands was performed using Big Dye Terminator Cycle Sequencing Kit v3.1 Kit (Applied Biosystems) with ABI 3130xl Genetic Analyzer (Applied Biosystems).

eQTL analyses

To test association between SNP genotypes and gene expression quantified in immortalized B-lymphocytes and T-cells, we performed in silico analysis using publicly available data from genome-wide expression analysis of quantitative trait loci (eQTL) of the 270 individuals genotyped in the HapMap Project (including 30 Caucasian trios of Northern and Western European origin [CEU]) [31–34] and the 85 individuals of the GenCord project (a collection of cell lines from umbilical cords of individuals of Western European origin) [35]. Linear regression was used to test the association between gene expression and SNP genotypes under additive model [32]. SNP genotype was coded as 0, 1, and 2, corresponding to the counts of the minor allele in each genotype.

Results

Replication of known JIA susceptibility loci

We performed a genome-wide association study in the discovery cohort (TSRHC + CMHC) including 388 children with JIA and 2500 genetically matched controls of European ancestry with high-quality SNP array data (Table 1, Additional file 1: Table S1). Then three independent JIA cohorts (CHOP, MCRI, OUH) were analyzed for replication and followed by meta-analysis of all the cohorts in a total of 1166 cases and 9500 controls (Table 1, Additional file 1: Table S1). The age at onset, gender, and subtype distributions of all case cohorts are similar to those reported previously for JIA [1].

We focused on the phenotypic commonality amongst our patients, that is, chronic inflammation of the joints. We observed a strong genome-wide significant association at the *HLA* locus on 6p21 (Additional file 1: Table S2, Additional file 1: Figure S1), A total of 24 SNPs surpassed genome-wide significance threshold at this region in the discovery cohort. Furthermore, three loci which previously have been implicated in JIA and several other autoimmune diseases [10, 36, 37] – the *PTPN22* locus on 1p13, the *IL2RA* locus on 10p15, and the *ANTXR2* locus on 4q21.21 – were nominally associated with JIA (Additional file 1: Table S3). Replication of these known JIA susceptibility loci demonstrated the validity of our study.

Novel association of *CXCR4* common variants

In addition to these known loci, we found a novel association signal at 2q22.1, with the most significant marker being rs953387 ($p = 2.07 \times 10^{-10}$; OR 0.59, 95 % CI 0.50–0.69). Six genotyped SNPs and ten additional imputed SNPs located on 2q22.1 of p value $< 5 \times 10^{-8}$ (Additional file 1: Table S4–S5, Additional file 1: Figure S1–S2) in the discovery cohort. Conditional on the top SNP rs953387, the significant association of the other SNPs were ablated, suggesting a single association signal at this locus. These SNPs were all nominally significant in each of our three independent replication cohorts. (Additional file 1: Table S4). Combined meta-analysis of all four cohorts indicated that four SNPs at the 2q22.1 locus were of p values $< 5 \times 10^{-8}$ (Additional file 1: Table S4).

However, as *CXCR4* is near a known stratified locus (nearby the lactase gene) [38], logistic regression analysis including the first ten coordinates from the MDS analysis as covariates brought the locus p value below genome-wide significance (top associated SNP rs1016269, $P = 4.70 \times 10^{-5}$). Similarly, using the first 10 principal components as the covariates in the logistic regression association, we obtained similar results (top associated SNP rs1016269, $P = 2.80 \times 10^{-5}$). Thus, our original association results were inflated due to subpopulation stratification within the subjects of European ancestry.

The association of *CXCR4* was nominally significant with the best p -value <0.0001 . When we evaluated the association by another method, permutation tests with the inclusion of the first ten coordinates from the MDS analysis as covariates in each association analysis, we obtained similar results of nominally significant results based on 10,000 permutations. *CXCR4* is an interesting gene because of its known role in immune regulation in a variety of immune cell types as a chemokine receptor. Gene *CXCR4*, is expressed on the surface of T-cells, B-cells, monocytes, neutrophils and dendritic cells (Additional file 1: Figure S3). The important function of *CXCR4* during B cell development and activation [39, 40], and the key role of B cells in pathogenesis of JIA [41] make *CXCR4* an intriguing target worth of further examination. Because of sub-population stratification, additional replication is still necessary for this locus to be considered a true risk locus for JIA. Its association with JIA need to be further investigated by stratified analysis in more refined homogeneous sub-populations of European ancestry, with a larger sample size. Further functional studies would also be required to demonstrate its potential contribution to JIA etiology.

Correlation between SNP genotype and *CXCR4* expression

Additionally, rs953387 was significantly associated with the mRNA levels of *CXCR4* ($p = 0.014$; Additional file 1: Figure S4) among HapMap dataset of 30 CEU children in the Sanger Institute GENE Expression VARIation (Genevar) public database (URLs), which profiles gene expression in immortalized B-lymphocyte samples [31]. Analysis of an independent dataset of T-cell lines from umbilical cords of 75 individuals of Western European origin [35] also demonstrated association between genotypes of another significant SNP rs1016269 at *CXCR4* locus (Additional file 1: Table S5) and levels of *CXCR4* expression

($p = 0.0054$; Additional file 1: Figure S3). These data suggest that common variants at *CXCR4* locus may regulate mRNA expression of this chemokine receptor.

Novel association of *CXCR4* rare variants

In addition to common variants, rare variants also contribute to disease susceptibility and usually confer a larger effect than common variants. We further examined rare variants at this locus by conducting deep resequencing in 480 JIA cases and 480 controls. Table 2 shows the variants identified in the coding region of *CXCR4*. Of interest are 4 non-synonymous and 1 stop-gain SNVs. These are all rare variants that are not catalogued in current variant databases (1000G and ESP6500) or with an extreme low frequency (0.0003). Four of these five rare variants are present only in JIA patients, not in controls. The SUM test produced a significance p value of 0.015 when testing the association of these 5 variants, suggesting that rare variants in the *CXCR4* gene may be contributing to the pathogenesis of JIA.

Then we performed Sanger sequencing to validate their genotype status. The primer sequences for each rare variant are shown in Additional file 1: Table S6. Among 40 samples of the 5 pools from which the rare variants were detected, 33 samples were left with enough DNA for Sanger sequencing. No genomic DNA was left for 4 samples in pool from which variant chr2:136872970 was identified and 3 samples in pool from which variant chr2:136873341 was found. Among the 33 samples with enough genomic DNA, 4 samples were found carrying the rare variant identified in deep sequencing (Fig. 1).

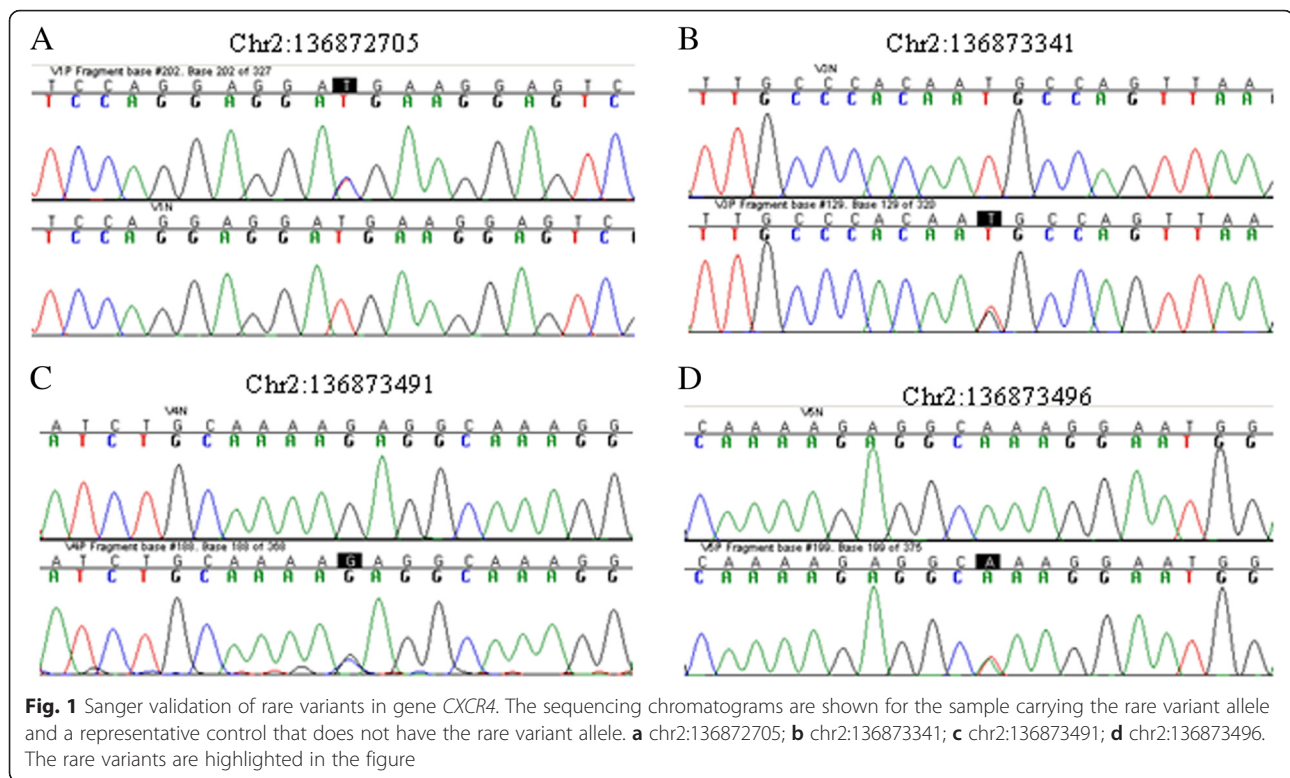
Discussion

In a combined sample set of more than 1100 JIA subjects of European ancestry, we replicated common

Table 2 Summary of variants in *CXCR4* coding-region by deep resequencing

Chr	Pos	Ref	Obs	Exonic Function	ESP6500	1000G	dbSNP135	CASE Freq.	CTRL Freq.
chr2	136872553	G	A	synonymous SNV	0.000461	0.0014	rs144110709	0	1
chr2	136872565	G	C	synonymous SNV				0	1
chr2	136872705	T	C	nonsynonymous SNV				0	1
chr2	136872715	G	A	synonymous SNV	0.003306	0.0018	rs148279552	3	7
chr2	136872727	A	G	synonymous SNV	0.000077		rs146627075	1	0
chr2	136872970	G	T	nonsynonymous SNV				1	0
chr2	136873084	G	A	synonymous SNV	0.036983	0.06	rs2228014	40	41
chr2	136873341	T	G	nonsynonymous SNV	0.000308		rs56400844	2	0
chr2	136873491	G	C	nonsynonymous SNV				1	0
chr2	136873496	A	T	stopgain SNV				1	0

Chr chromosome, Pos position on human genome build hg19, Ref reference allele, Obs the variant allele observed in our JIA dataset, ESP6500 variant allele frequency in Exome Sequencing Project, 1000G variant allele frequency in Thousand Genome Project, dbSNP 135 SNP rs ID in Single Nucleotide Polymorphism database of human genome build 135, CASE Freq variant allele frequency among cases, CTRL freq variant allele frequency among controls



genetic variants at loci of *HLA*, *PTPN22*, *IL2RA* and *ANTXR2* that have been reported for association with susceptibility to JIA. We also found a nominally significant association at *CXCR4* which has been implicated in immune regulation, demonstrated the correlation between *CXCR4* variants genotype and its gene expression level. We further showed that rare non-synonymous and stop-gain variants in *CXCR4* are enriched in JIA cases. Our data support a role for altered expression of *CXCR4* in JIA pathogenesis, and present the first genetic demonstration of a potential role for the chemokine receptor, *CXCR4*, in the pathogenesis of autoimmune disease. Because this locus is subjected to population stratification within the subjects of European ancestry, additional replication is still necessary for this locus to be considered a true risk locus for JIA.

Few genetic studies of JIA have been carried out at genome-wide scale. Previously reported GWASs of JIA were all limited by small size [42, 43]. A more recent study using Immunochip, took a hypothesis-driven approach, targeting immune-specific genes. Several significant new loci were identified from the study [13], however the Immunochip does not capture variants across a large proportion of the genome. Our study was composed of a large dataset with comprehensive genome SNP coverage, and the candidate gene identified was examined using targeted resequencing with a large number of samples and high coverage.

Chemokine receptors impact immune system development and function in part via regulation of cell migration. The G protein-coupled chemokine receptor, *CXCR4*, is expressed on the surface of T-cells, B-cells, monocytes, neutrophils and dendritic cells (Additional file 1: Figure S3), and is activated exclusively by CXCL12 (also known as stromal-derived-factor-1, SDF-1), a small peptide mediator and potent chemoattractant for leukocytes, including B- and T-cells. *CXCR4* and its ligand, CXCL12, have been shown to play a role in B-cell production, myelopoiesis, integrin activation, angiogenesis, and chemotaxis [25]. Intriguingly, the human immunodeficiency virus (HIV-1) has usurped *CXCR4*'s unique CXCL12 binding site, exploiting *CXCR4* as a co-receptor in later stages of HIV-1 infection, and *CXCR4* antagonists have been explored as treatments for HIV infection. Binding of *CXCR4* to CXCL12 is also proposed to play a role in cancer metastases, and *CXCR4* antagonists are under study in human clinical trials for solid and non-solid tumors [44]. Available therapeutic agents targeting the *CXCR4*-CXCL12 axis for activation or inhibition include plerixafor (AMD3100), recombinant CXCL12, and high-affinity *CXCR4* and CXCL12 monoclonal antibodies, some of which are already in use in the clinic but not approved for use in children. The recent report of crystal structures of *CXCR4* with small-molecule and cyclic peptide inhibitors [45] provide new opportunities for drug discovery efforts targeting this receptor.

CXCR4 and CXCL12 have been implicated in the pathogenesis of autoimmune diseases [44, 46]. In mouse models of autoimmune disease, modulation of CXCR4 alters trafficking of leukocytes to peripheral organs and polarization of regulatory T cells, and accelerates onset of disease [47, 48]. This is consistent with our data showing that a risk variant of *CXCR4* correlates with decreased expression of CXCR4. An alternate hypothesis is that the effect of low CXCR4 expression is indirect and leads to a compensatory increase of the CXCR4 ligand, CXCL12. Our preliminary data suggest that the risk variant of *CXCR4* correlates with increased expression of CXCL12 (data not shown). This is consistent with models of collagen-induced arthritis in which CXCL12 acts as a pro-inflammatory factor in the pathogenesis of inflammatory arthritis [49, 50], and with human studies in which CXCL12 enhances cellular proliferation and cytokine expression by peripheral blood T cells, upregulates expression of cytokines and chemokines by fibroblast-like synoviocytes from patients with RA [51], and mediates lymphocyte ingress into RA synovial tissue, synovial neovascularisation, and osteoclastogenesis [52]. Translation of our discovery into therapeutic benefit for a specific group of JIA patients will require additional genetic replication studies, proper functional validation and greater insight into the pathogenic role of CXCR4 in JIA.

Conclusions

We uncovered common SNPs at the *CXCR4* locus associated with JIA and further found rare functional variants at *CXCR4* enriched in JIA cases. Our results suggest that genetic variants at the *CXCR4* locus may predispose to the development of JIA. Because of sub-population stratification, additional replication is still necessary for this locus to be considered a true risk locus for JIA. The extensive literature surrounding the biology of CXCR4 and its binding partner CXCL12 in health and disease, and the ready availability of targeted therapeutic agents make *CXCR4* a particularly attractive candidate for further genetic replication and functional investigation of its pathogenic role and therapeutic potential in JIA and other autoimmune diseases affecting children.

Additional file

Additional file 1: Supplemental Data: Table S1. The clinical characteristics of samples in each JIA cohort. Table S2. Genome-wide significant associations at the *HLA* locus ($p < 5 \times 10^{-8}$ in the discovery cohort). Table S3. Association results for top SNPs in known JIA associated genes *PTPN22*, *IL2RA*, *ANTXR2*. Table S4. The most significantly associated SNPs at *CXCR4* locus on chromosome 2q22.1. Table S5. Genomewide association results for imputed SNPs ($p < 1 \times 10^{-4}$ in combined analysis) in the vicinity of *CXCR4* in our JIA cohort. Table S6. Primers used in Sanger sequencing validation of rare variants at *CXCR4* locus.

Figure S1. Genome-wide association results for JIA. Figure S2. Regional association plot for the 2q22.1 region. Figure S3. *CXCR4* tissue-specific gene expression levels. Figure S4. *CXCR4* expression levels stratified by SNP genotype. (DOC 466 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contribution

THF, HH conceived and designed the experiments; JG, CK, EF, DS, KT, RC, CH, KX, JS, HQ, FM and KW performed the experiments; WW, ZW, JL and HZ analyzed the data; EB, ER, SL, CW, MP, MB, JM, BF, OF, ST, CL, BL, JE and HH contributed reagents/materials/analysis tools; THF, JL, ZW, JE and HH wrote the paper. All authors read and approved the final manuscript.

Acknowledgments

We gratefully thank all the children with JIA and their families at the participating study sites who were enrolled in this study and all the control subjects who donated blood samples to the Children's Hospital of Philadelphia (CHOP) for genetic research purposes. We thank the technical staff at the Center for Applied Genomics at CHOP for producing the genotypes used for analyses, and the nursing, medical assistant and medical staff for their help with recruitment of patient and control subjects for the study. This study was supported by Institutional Development Funds from The Children's Hospital of Philadelphia, RC1AR058606 and U01HG006830. The expression work reported in this project was funded in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN26120080001E, and by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

Web resources (URLs)

The URLs for data presented herein are as follows:
 JIA Calculator, <http://www.jra-research.org/JIAcalc/index.php>
 PLINK, <http://pngu.mgh.harvard.edu/~purcell/plink/>
 MACH, <http://www.sph.umich.edu/csg/abecasis/MaCH/index.html>
 Metal, <http://www.sph.umich.edu/csg/abecasis/Metal/index.html>
 MOOSE, <http://jama.ama-assn.org/cgi/content/full/283/15/2008>
 Genevar, <http://www.sanger.ac.uk/science/tools/genevar-gene-expression-variation-archive>
 HapMap, <http://hapmap.ncbi.nlm.nih.gov/>
 BWA, <http://bio-bwa.sourceforge.net/>
 ANNOVAR, <http://annovar.openbioinformatics.org/en/latest/>
 SNVer, <http://snver.sourceforge.net/>
 AssotesteR, <http://cran.r-project.org/web/packages/AssotesteR/index.html>
 Primer3, <http://bioinfo.ut.ee/primer3-0.4.0/>

Author details

¹Division of Rheumatology, The Children's Hospital of Philadelphia, 19104 Philadelphia, PA, USA. ²Department of Pediatrics, University of Pennsylvania School of Medicine, 19104 Philadelphia, PA, USA. ³The Center for Applied Genomics, The Children's Hospital of Philadelphia, 19104 Philadelphia, PA, USA. ⁴Department of Computer Science, New Jersey Institute of Technology, 07102 New Jersey, NJ, USA. ⁵Basic Research Laboratory, Center for Cancer Research, National Cancer Institute, Leidos Biomedical Research Inc., Frederick National Laboratory, 21702 Frederick, MD, USA. ⁶Sarah M. and Charles E. Seay Center for Musculoskeletal Research, Texas Scottish Rite Hospital for Children, 75219 Dallas, TX, USA. ⁷Division of Rheumatology, Texas Scottish Rite Hospital, 75219 Dallas, TX, USA. ⁸Division of Rheumatology, Children's Mercy-Kansas City, 64108 Kansas City, MO, USA. ⁹Murdoch Children's Research Institute, 3052 Parkville, VIC, Australia. ¹⁰Paediatric Rheumatology Unit, Royal Children's Hospital, 3052 Parkville, VIC, Australia. ¹¹Department of Rheumatology, Oslo University Hospital, Rikshospitalet, Norway. ¹²Cincinnati Children's Hospital Medical Center, 45229 Cincinnati, OH, USA. ¹³Center for Public Health Genomics and Department of Biostatistical Sciences, Wake Forest School of Medicine, 27157 Winston-Salem, NC, USA. ¹⁴Department of Psychiatry and Behavioral Sciences, UUniversity of Southern California, 90089 Los Angeles, CA, USA. ¹⁵Institute of Immunology, Oslo University Hospital, 0027 Rikshospitalet, Oslo, Norway. ¹⁶Department of Paediatrics, University of Melbourne, 3010 Melbourne, Australia. ¹⁷Division of Human Genetics, The Children's Hospital of Philadelphia, 19104 Philadelphia, PA, USA. ¹⁸Present

Address: Department of Pediatrics, Nemours Research Institute, Nemours Children's Hospital, 32827 Orlando, FL, USA.

Received: 20 August 2014 Accepted: 4 March 2016

Published online: 22 March 2016

References

- Ravelli A, Martini A. Juvenile idiopathic arthritis. *Lancet*. 2007;369(9563):767–78.
- Prakken B, Albani S, Martini A. Juvenile idiopathic arthritis. *Lancet*. 2011;377(9783):2138–49.
- Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, Birnbaum NS, Burmester GR, Bykerk VP, Cohen MD. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum*. 2010;62(9):2569–81.
- Petty RE, Southwood TR, Manners P, Baum J, Glass DN, Goldenberg J, He X, Maldonado-Cocco J, Orozco-Alcala J, Prieur A-M. International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. *J Rheumatol*. 2004;31(2):390–2.
- Helmick CG, Felson DT, Lawrence RC, Gabriel S, Hirsch R, Kwoh CK, Liang MH, Kremers HM, Mayes MD, Merkel PA. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. *Arthritis Rheum*. 2008;58(1):15–25.
- Yelin E, Murphy L, Cisternas MG, Foreman AJ, Pasta DJ, Helmick CG. Medical care expenditures and earnings losses among persons with arthritis and other rheumatic conditions in 2003, and comparisons with 1997. *Arthritis Rheum*. 2007;56(5):1397–407.
- Centers for Disease C, Prevention. National and state medical expenditures and lost earnings attributable to arthritis and other rheumatic conditions—United States. *MMWR Morbidity and mortality weekly report* 2007. 2003;56(1):4–7.
- Foster HE, Marshall N, Myers A, Dunkley P, Griffiths ID. Outcome in adults with juvenile idiopathic arthritis: a quality of life study. *Arthritis Rheum*. 2003;48(3):767–75.
- Ellis JA, Munro JE, Ponsonby A-L. Possible environmental determinants of juvenile idiopathic arthritis. *Rheumatology (Oxford)*. 2010;49(3):411–25.
- Woo P, Colbert RA. An overview of genetics of paediatric rheumatic diseases. *Best Pract Res Clin Rheumatol*. 2009;23(5):589–97.
- Hollenbach JA, Thompson SD, Bugawan TL, Ryan M, Sudman M, Marion M, Langefeld CD, Thomson G, Erlich HA, Glass DN. Juvenile idiopathic arthritis and HLA class I and class II interactions and age-at-onset effects. *Arthritis Rheum*. 2010;62(6):1781–91.
- Prahalad S, Shear ES, Thompson SD, Giannini EH, Glass DN. Increased prevalence of familial autoimmunity in simplex and multiplex families with juvenile rheumatoid arthritis. *Arthritis Rheum*. 2002;46(7):1851–6.
- Hinks A, Cobb J, Marion MC, Prahalad S, Sudman M, Bowes J, Martin P, Comeau ME, Sajuthi S, Andrews R. Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. *Nat Genet*. 2013;45(6):664–9.
- Cortes A, Brown MA. Promise and pitfalls of the Immunochip. *Arthritis Res Ther*. 2011;13(1):101.
- Behrens EM, Finkel TH, Bradford JP, Kim CE, Linton L, Casalunovo T, Frackelton EC, Santa E, Otieno FG, Glessner JT. Association of the TRAF1-C5 locus on chromosome 9 with juvenile idiopathic arthritis. *Arthritis Rheum*. 2008;58(7):2206–7.
- Wise CA, Bennett LB, Pascual V, Gillum JD, Bowcock AM. Localization of a gene for familial recurrent arthritis. *Arthritis Rheum*. 2000;43(9):2041–5.
- Becker ML, Gaedigk R, van Haandel L, Thomas B, Lasky A, Hoeltzel M, Dai H, Stobaugh J, Leeder JS. The effect of genotype on methotrexate polyglutamate variability in juvenile idiopathic arthritis and association with drug response. *Arthritis Rheum*. 2011;63(1):276–85.
- Eike MC, Nordang GBN, Karlsen TH, Boberg KM, Vatn MH, Dahl-Jørgensen K, Rønningen KS, Joner G, Flatø B, Bergquist A. The FCRL3-169T > C polymorphism is associated with rheumatoid arthritis and shows suggestive evidence of involvement with juvenile idiopathic arthritis in a Scandinavian panel of autoimmune diseases. *Ann Rheum Dis*. 2008;67(9):1287–91.
- Ellis JA, Ponsonby A-L, Pezic A, Chavez RA, Allen RC, Akikusa JD, Munro JE. CLARITY - Childhood Arthritis Risk factor Identification sTudy. *Pediatric Rheumatology Online J*. 2012;10(1):37.
- Behrens EM, Beukelman T, Gallo L, Spangler J, Rosenkranz M, Arkachaisri T, Ayala R, Groh B, Finkel TH, Cron RQ. Evaluation of the presentation of systemic onset juvenile rheumatoid arthritis: data from the Pennsylvania Systemic Onset Juvenile Arthritis Registry (PASOJAR). *J Rheumatol*. 2008;35(2):343–8.
- Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661–78.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, Bakker PIW, Daly MJ. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904–9.
- Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet*. 2009;10:387–406.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England)*. 2010;26(17):2190–1.
- Wei Z, Wang W, Hu P, Lyon GJ, Hakonarson H. SNVer: a statistical tool for variant calling in analysis of pooled or individual next-generation sequencing data. *Nucleic Acids Res*. 2011;39(19):e132.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
- Pan W. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genet Epidemiol*. 2009;33(6):497–507.
- Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. *Bioinformatics*. 2007;23(10):1289–91.
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3—new capabilities and interfaces. *Nucleic Acids Res*. 2012;40(15):e115.
- Yang T-P, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, Deloukas P, Dermitzakis ET. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics (Oxford, England)*. 2010;26(19):2474–6.
- Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis SE, Tavaré S. Genome-wide associations of gene expression variation in humans. *PLoS Genet*. 2005;1(6):e78.
- Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, Grassi A, Lee C. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science (New York, NY)*. 2007;315(5813):848–53.
- Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D. Population genomics of human gene expression. *Nat Genet*. 2007;39(10):1217–24.
- Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science (New York, NY)*. 2009;325(5945):1246–50.
- Thompson SD, Sudman M, Ramos PS, Marion MC, Ryan M, Tsoras M, Weiler T, Wagner M, Keddache M, Haas JP. The susceptibility loci juvenile idiopathic arthritis shares with other autoimmune diseases extend to PTPN2, COG6, and ANGPT1. *Arthritis Rheum*. 2010;62(11):3265–76.
- Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat Rev Genet*. 2009;10(1):43–55.
- Campbell CD, Ogburn EL, Lunetta KL, Lyon HN, Freedman ML, Groop LC, Altshuler D, Ardlie KG, Hirschhorn JN. Demonstrating stratification in a European American population. *Nat Genet*. 2005;37(8):868–72.
- Ma Q, Jones D, Springer TA. The chemokine receptor CXCR4 is required for the retention of B lineage and granulocytic precursors within the bone marrow microenvironment. *Immunity*. 1999;10(4):463–71.
- Nie Y, Waite J, Brewer F, Sunshine MJ, Littman DR, Zou YR. The role of CXCR4 in maintaining peripheral B cell compartments and humoral immunity. *J Exp Med*. 2004;200(9):1145–56.
- Wiegering V, Girschick HJ, Morbach H. B-cell pathology in juvenile idiopathic arthritis. *Arthritis*. 2010;2010:759868.
- Hinks A, Barton A, Shephard N, Eyre S, Bowes J, Cargill M, Wang E, Ke X, Kennedy GC, John S. Identification of a novel susceptibility locus for juvenile idiopathic arthritis by genome-wide association analysis. *Arthritis Rheum*. 2009;60(1):258–63.

43. Thompson SD, Marion MC, Sudman M, Ryan M, Tsoras M, Howard TD, Barnes MG, Ramos PS, Thomson W, Hinks A. Genome-wide association analysis of juvenile idiopathic arthritis identifies a new susceptibility locus at chromosomal region 3q13. *Arthritis Rheum.* 2012;64(8):2781–91.
44. Chong BF, Mohan C. Targeting the CXCR4/CXCL12 axis in systemic lupus erythematosus. *Expert Opin Ther Targets.* 2009;13(10):1147–53.
45. Wu B, Chien EYT, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science (New York, NY).* 2010;330(6007):1066–71.
46. Koch AE. Chemokines and their receptors in rheumatoid arthritis: future targets? *Arthritis Rheum.* 2005;52(3):710–21.
47. Aboumrad E, Madec AM, Thivolet C. The CXCR4/CXCL12 (SDF-1) signalling pathway protects non-obese diabetic mouse from autoimmune diabetes. *Clin Exp Immunol.* 2007;148(3):432–9.
48. Meiron M, Zohar Y, Anunu R, Wildbaum G, Karin N. CXCL12 (SDF-1 α) suppresses ongoing experimental autoimmune encephalomyelitis by selecting antigen-specific regulatory T cells. *J Exp Med.* 2008;205(11):2643–55.
49. De Klerck B, Geboes L, Hatse S, Kelchtermans H, Meyvis Y, Vermeire K, Bridger G, Billiau A, Schols D, Matthys P. Pro-inflammatory properties of stromal cell-derived factor-1 (CXCL12) in collagen-induced arthritis. *Arthritis Res Ther.* 2005;7(6):R1208–20.
50. Bradfield PF, Amft N, Vernon-Wilson E, Exley AE, Parsonage G, Rainger GE, Nash GB, Thomas AMC, Simmons DL, Salmon M. Rheumatoid fibroblast-like synoviocytes overexpress the chemokine stromal cell-derived factor 1 (CXCL12), which supports distinct patterns and rates of CD4+ and CD8+ T cell migration within synovial tissue. *Arthritis Rheum.* 2003;48(9):2472–82.
51. Nanki T, Nagasaka K, Hayashida K, Saita Y, Miyasaka N. Chemokines regulate IL-6 and IL-8 production by fibroblast-like synoviocytes from patients with rheumatoid arthritis. *J Immunol.* 2001;167(9):5381–5.
52. Nanki T, Hayashida K, El-Gabalawy HS, Suson S, Shi K, Girschick HJ, Yavuz S, Lipsky PE. Stromal cell-derived factor-1-CXC chemokine receptor 4 interactions play a central role in CD4+ T cell accumulation in rheumatoid arthritis synovium. *J Immunol.* 2000;165(11):6590–8.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

