Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis

Anne Hinks1,2,22, Joanna Cobb1,2,22, Miranda C Marion3,4,22, Sampath Prahalad5,6, Marc Sudman7, John Bowes1,2, Paul Martin1,2, Mary E Comeau3,4, Satria Sajuthi3,4, Robert Andrews8, Milton Brown5, Wei-Min Chen9, Patrick Concannon9, Panos Deloukas8, Sarah Edkins8, Stephen Eyre1,2, Patrick M Gaffney10, Stephen L Guthery11,12, Joel M Guthridge10, Sarah E Hunt8, Judith A James10, Mehdí Keddache13, Kathy L Moser10, Peter A Nigrovic14,15, Suna Onengut-Gumuscu9, Mitchell L Onslow7, Carlos D Rosé14,15, Stephen S Rich9, Kathryn J A Steel11,12, Edward K Wakeland16, Carol A Wallace17, Lucy R Wedderburn18, Patricia Woo19, Boston Children’s JIA Registry20, British Society of Paediatric and Adolescent Rheumatology (BSPAR) Study Group20, Childhood Arthritis Response to Medication Study (CHARMS)20, German Society for Pediatric Rheumatology (GKRI)20, JIA Gene Expression Study20, NIAMS JIA Genetic Registry20, TREAT Study20, United Kingdom Juvenile Idiopathic Arthritis Genetics Consortium (UKJIAGC)20, John F Bohnsack11,12, Johannes Peter Haas21, David N Glass7, Carl D Langefeld3,4, Wendy Thomson1,2,23 & Susan D Thompson7,23

We used the Immunochip array to analyze 2,816 individuals with juvenile idiopathic arthritis (JIA), comprising the most common subtypes (oligoarticular and rheumatoid factor-negative polyarticular JIA), and 13,056 controls. We confirmed association of 3 known JIA risk loci (the human leukocyte antigen (HLA) region, PTPN22 and PTPN2) and identified 14 loci reaching genome-wide significance ($P < 5 \times 10^{-8}$) for the first time. Eleven additional new regions showed suggestive evidence of association with JIA ($P < 1 \times 10^{-6}$). Dense mapping of loci along with bioinformatics analysis refined the associations to one gene in each of eight regions, highlighting crucial pathways, including the interleukin (IL)-2 pathway, in JIA disease pathogenesis. The entire Immunochip content, the HLA region and the top 27 loci ($P < 1 \times 10^{-6}$) explain an estimated 18, 13 and 6% of the risk of JIA, respectively. In summary, this is the largest collection of JIA cases investigated so far and provides new insight into the genetic basis of this childhood autoimmune disease.

JIA is the most common chronic rheumatic disease of childhood and describes a group of clinically heterogeneous arthritides that begin before the age of 16 years, persist for at least 6 weeks and have an unknown cause. It has been established that there is a strong genetic contribution to the risk of JIA, with a sibling risk ratio of ~11.6 (ref. 2) and higher risk for other autoimmune diseases in the families of individuals with JIA. Using International League of Associations for Rheumatology (ILAR) criteria, JIA can be divided into subtypes on the basis of clinical features. A recent genome-wide association study (GWAS) identified a number of JIA susceptibility regions. Additional loci have been identified through candidate gene association studies and confirmed in multiple independent studies. However, until now, only three loci have reached genome-wide significance (the HLA region, PTPN22 and PTPN2).

Many confirmed and nominally associated JIA susceptibility loci show association with other autoimmune diseases. This marked overlap of autoimmune disease susceptibility loci may occur when...
the same variants contribute to multiple diseases or when different variants in the same gene confer susceptibility to different autoimmune diseases. Thus, dense mapping of as many susceptibility loci as possible will be important to understand how individual variants contribute to risk of disease. To facilitate these efforts, a custom Illumina Infinium genotyping array has been designed by the Immunochip Consortium on the basis of confirmed risk loci for 12 autoimmune diseases\textsuperscript{15}, not including JIA. The chip includes dense coverage of the extended HLA region and 186 non-HLA loci\textsuperscript{15}. In this study, we report an Immunochip analysis of 2,816 individuals with oligoarticular or rheumatoid factor (RF)-negative polyarticular juvenile idiopathic arthritis (JIA)\textsuperscript{5–12,14} (for further details, see Online Methods).

There were 11 additional regions showing suggestive evidence (5 × 10\textsuperscript{−8} < P < 1 × 10\textsuperscript{−6}) of association with oligoarticular and RF-negative polyarticular JIA\textsuperscript{18}. We also confirmed association with PTPN2 at rs2847293 (OR = 1.31; P = 1.44 × 10\textsuperscript{−12}), which lies in the intergenic region of PTPN2 and is in LD (r\textsuperscript{2} = 0.94) with rs1893217, a SNP previously associated with oligoarticular and RF-negative polyarticular JIA\textsuperscript{2}. Stepwise logistic regression including the most significant SNP in the PTPN2 region as a covariate suggested that there is an uncommon variant, rs149850873 (MAF\textsubscript{controls} = 2%), that confers an independent secondary effect in the region (Supplementary Table 4). The most significant SNP, rs7775055, tagged the HLA-DRB1*0801–HLA-DQA1*0401–HLA-DQB1*0402 haplotype, which has been consistently implicated in conferring risk of JIA\textsuperscript{16,17}; however, other HLA haplotypes have also been associated with JIA. The HLA region SNP rs7775055 showed a highly significant difference in SNP allele frequency between the two subtypes (Supplementary Table 5). Association was stronger in the oligoarticular subtype compared to the RF-negative polyarticular subtype, consistent with previous studies showing differences in HLA region associations between the two subtypes\textsuperscript{16,17}. Further analysis at the amino-acid level is necessary to fully understand this complex region in JIA and its subtypes. The most significant association outside the MHC region was with rs6679677 (OR = 1.59; P = 3.19 × 10\textsuperscript{−25}) at 1p13.2, which contains the PTPN22 gene; rs6679677 is in linkage disequilibrium (LD; r\textsuperscript{2} = 1) with rs2476601, the SNP previously associated with JIA\textsuperscript{5,7} and implicated as the PTPN22 causal variant\textsuperscript{18}. Of the 14 loci newly confirmed as JIA susceptibility loci in this study at the genome-wide significance level (P < 5 × 10\textsuperscript{−8}) (Fig. 1, Table 1 and Supplementary Fig. 1), 5 (STAT4, ANKRD55, IL2-IL21, IL2RA and SH2B3-ATXN2) have supporting evidence of a role in JIA susceptibility from previous studies. The most significant SNP in the STAT4 region (rs10174238) is in high LD with a SNP (rs7574865) previously reported in JIA\textsuperscript{5,8,10} and other autoimmune diseases\textsuperscript{19}. However, stepwise logistic regression analysis suggested the presence of two additional independent effects (rs45539732 and rs13029532), which are located within the adjacent STAT1 gene (Supplementary Fig. 3 and Supplementary Table 6). Notably, rs45539732 is an uncommon SNP (MAF\textsubscript{controls} = 3%).

There were 11 additional regions showing suggestive evidence (5 × 10\textsuperscript{−8} < P < 1 × 10\textsuperscript{−6}) of association with oligoarticular and RF-negative
polyarticular JIA (Table 2), of which 4 have supportive evidence from previous studies (COG6, CCR1-CCR3, TIMMDC1 (also known as C3orf56)-CD80 and AFF3-LONRF2).

We imputed across the non-HLA JIA risk loci identified in this study using 1000 Genomes Project data (Tables 1 and 2, Online Methods and Supplementary Fig. 1). We found only modest differences between the P values of the top genotyped SNP and the top imputed SNP at each locus. We note two regions that are minor exceptions: PRM1-RMI2 (also known as C16orf75) and C5orf56-IRF1 (Supplementary Fig. 1). In the latter region, the top imputed SNP lies within the C5orf56 gene. The lack of a substantial gain of information from imputation of the regions is consistent with other reports on the performance of Immunochip imputation20,21. This is likely owing to the dense fine mapping of most regions represented on the Immunochip.

Of the top 17 regions that reached genome-wide significance, 13 were densely mapped on the Immunochip. LD patterns and functional annotation provided strong evidence that the signal localized to a single gene in eight loci (PTPN2, IL2RA, STAT4, IL21, TYK2, SH2B3-ATXN2 and LTB1) based on LD patterns and PTPN2, SH2B3-ATXN2 and TYK2 based on the most significant SNP being a nonsynonymous coding variant) (Table 3, Supplementary Fig. 1 and Supplementary Table 7); however, further functional analysis is required for confirmation.

All but one of the variants that reached genome-wide significance were common (MAF > 5%). One variant, a nonsynonymous coding variant within TYK2, had a low allele frequency (MAF controls = 5%). In addition, a couple of the secondary effects in PTPN2 and STAT4 were uncommon.

In three regions (TYK2, SH2B3-ATXN2 and LTBR), the most significant SNP (or a SNP with r² > 0.9) was located within a coding region and is therefore a strong candidate for the causal variant. For SH2B3-ATXN2, the same variant has also been associated with celiac disease22, vitiligo23, rheumatoid arthritis24, type 1 diabetes25 and multiple sclerosis26. The TYK2 SNP (rs34536443) was also the lead SNP in the region in rheumatoid arthritis27, primary biliary cirrhosis20 and psoriasis28. Other regions (IL6R, ZEP36L1, IL2-IL21, UBE2L3, LTBR and TIMMDC1-CD80) contain SNPs that show evidence of high mammalian conservation (17-way vertebrate conservation)29 or have a high regulatory potential score (Table 3), calculated using alignments of 7 mammalian genomes30. There is evidence that the associated SNPs in LTBR, UBE2L3 and LNPEP are expression quantitative trait loci (eQTL) (Table 3). The SNP in LNPEP, rs27290,
Table 2  Regions showing suggestive association with oligoarticular and RF-negative polyarticular JIA (5 × 10^{-8} < P < 1 × 10^{-6})

| Gene region | Chr. | Positiona | Most significant SNP | Minor allele | MAF controbs (n = 13,056) | MAF cases (n = 2,816) | Best P value | Model | OR (95% CI) | SNP position |
|-------------|------|-----------|----------------------|--------------|--------------------------|-----------------------|--------------|-------|-------------|--------------|
| LTBR        | 12   | 6495275   | rs2364480            | C            | 0.25                     | 0.28                  | 5.10 × 10^{-7} | Additive | 1.20 (1.12–1.28) Coding (NS) |
| IL6         | 7    | 22798080  | rs7808122            | A            | 0.44                     | 0.48                  | 5.80 × 10^{-8} | Additive | 1.19 (1.11–1.25) Intergenic |
| COG6        | 13   | 40350912  | rs7993214            | A            | 0.35                     | 0.31                  | 1.61 × 10^{-7} | Additive | 0.84 (0.79-0.90) Intergenic |
| PRR5L       | 11   | 36336575  | rs4755450            | A            | 0.35                     | 0.31                  | 3.35 × 10^{-7} | Dominant | 0.80 (0.74–0.87) Intergenic |

*Coordinates are based on the NCBI37 assembly. *Imputed SNP results are included for SNPs that had a better P value than the most significant directly genotyped SNP in the region.

is also in LD (r^2 = 0.78) with rs2248374, a SNP that lies within a splice site for ERAP2 (ref. 31). The rs2248374[G] allele results in a spliced ERAP2 mRNA that encodes a truncated protein. In JIA, the rs2248374[G] minor allele showed a protective association (OR = 0.76; P = 1.8 × 10^{-7}).

**IL2RA**, the IL2-IL21 region and IL2RB are now all considered to be confirmed susceptibility loci for JIA and suggest an important role for the IL-2 pathway in JIA disease pathogenesis. This pathway has a vital role in T cell activation and development, as well as a key role in maintaining immune tolerance through the dependence of regulatory T cells on IL-2. Other confirmed JIA loci identified here are related to this pathway: SH2B3 encodes an adaptor protein involved in T cell activation, and STAT4 encodes a transcription factor important in T cell differentiation.

We next considered the top non-HLA SNP associations separately for each JIA subtype (oligoarticular and RF-negative polyarticular JIA). Only one region (C5orf56-IRF1) showed evidence of differential association, with the association limited to the oligoarticular subtype of JIA. All other regions showed associations with similar effect sizes and directions of effect in the two subtypes (Supplementary Table 5).

Table 3  Potential causal SNPs within the JIA risk regions

| Lead SNP | SNP in strong LD (r^2 > 0.9) with the lead SNP | Chr. | Positionb | r^2 with lead SNP | Location | Regulatory potential | Conservation | Functional predictionb | eQTLc |
|----------|-----------------------------------------------|------|-----------|------------------|----------|----------------------|--------------|-----------------------|-------|
| Genome-wide significant SNPs |
| rs6679677 | rs2476601                                      | 1    | 114377568 | 1                | Exon of PTPN22 | 0.14          | 0.999          | Benign; tolerated |
| rs11265608| rs1205591                                      | 1    | 154298374 | 1                | Intron of ATP8B2 | 0.89          | 0              |                      |
| rs1479924 | rs13144509                                    | 4    | 123473487 | 0.94             | Intergenic between IL2 and IL21 | 0.17          | 1              |                      |
| rs27290  | rs27290                                       | 5    | 96350088  | –                | Intron of LNPEP | 0.21          | 0              |                      |
| rs3184504 | rs3184504                                      | 12   | 111884608 | –                | Exon of SH2B3 | 0.29          | 0.005          | Benign; tolerated |
| rs12434551| rs3825568                                      | 14   | 68096588  | 0.98             | 5’ UTR of ZFP36L1 | 0.55          | 0.002         |                      |
| rs34536443| rs34536443                                     | 19   | 10463118  | –                | Exon of TYK2 | 0.40          | 0.19           | Probably damaging; deleterious |
| rs4536443 | rs74956615                                     | 19   | 10472721  | 1                | Intron of RAVER1 | 0            | 0.998         |                      |
| rs2266959 | rs2266959                                      | 22   | 21922904  | –                | Intron of UBE2L3 | 0.47          | 0.003         |                      |
| rs2266959 | rs2298282                                      | 22   | 21982892  | 1                | Exon of YDUC | 0.37          | 1              | Benign; tolerated |
| rs2266959 | rs4820091                                      | 22   | 21940189  | 1                | Intron of UBE2L3 | 0            | 0              |                      |
| Suggestive SNPs |
| rs4688013 | rs17203104                                     | 3    | 119139575 | 0.92             | Intergenic between CDGAP and TMEM39A | 0            | 0.998         |                      |
| rs2364480 | rs2364481                                      | 12   | 6497260   | –                | Intron of LTBR | 0.36          | 0.002         |                      |
| rs2364480 | rs2364480                                      | 12   | 6495275   | –                | Exon of LTBR | 0.34          | 0.005         |                      |

*Coordinates are based on the NCBI37 assembly. *Functional prediction is based on PolyPhen38. *Data from three studies were considered: lymphoblastoid cell lines (LCLs) from HapMap 3 (Stranger et al.35), fibroblast, LCLs and T cells from the umbilical cords of 75 Geneva Gencord individuals (Dimas et al.36) and adipose, LCLs and skin from 856 healthy female twins of the MuTHER Resource (Grundberg et al.37). Yes indicates evidence of an eQTL (P < 1 × 10^{-5}).
As expected, many of the JIA-associated regions shown in Tables 1 and 2 are also associated with other autoimmune diseases (Supplementary Table 8), with the same SNP or a highly correlated SNP associated with the same direction of effect (assessed by comparing our results with information from the Catalog of Published GWAS and recent studies using the Immunochip in other autoimmune diseases20–22,27,32,33). We found a strong overlap with rheumatoid arthritis risk loci, which is not unexpected owing to the clinical similarities with JIA, which is consistent with previous studies8,10,34. In addition, there was notable overlap with type 1 diabetes and celiac disease. Some regions (IL2–IL21, Casrf56–IRF1, IL2RB, ATP8B2–ILS6, 13q14, CCR1–CCR3, RUNX3 and TIMMDC1–CD80) showed association with other autoimmune diseases, but the top SNPs for the latter were not highly correlated with the most strongly associated SNPs for JIA. Some regions have not previously been associated with autoimmune diseases by GWAS or Immunochip at genome-wide significance. In-depth analysis of the results across all Immunochip studies will be of great value in understanding the contributions of the individual loci to the various diseases.

This study of 2,816 JIA cases is the largest collaborative cohort study of JIA so far and includes samples from across the United States, UK and Germany. The power derived from this cohort plus the large control sample size, combined with the comprehensive coverage of SNPs in regions implicated in autoimmune disease on the Immunochip, has substantially increased our ability to detect associations. In setting the statistical threshold at a stringent genome-wide significance level (P < 5 × 10−8), we report 14 new loci. In addition, a second tier of 11 regions with suggestive evidence of association (P < 1 × 10−6) has been identified that includes plausible candidate risk factors but requires validation. This study substantially increases the number of confirmed susceptibility loci for JIA, but additional genetic risk factors likely remain to be discovered, an idea supported by the quantile-quantile plot (Supplementary Fig. 4) that suggests that there are residual associations after removing those of the above implicated regions. In addition, we calculated that the entire Immunochip content, the HLA region and the top 27 loci explain an estimated 18, 13 and 6% of risk of JIA, respectively. These estimates also suggest that there must be other regions of the genome harboring additional JIA risk loci. In summary, this analysis using the Immunochip has substantially enhanced understanding of the genetic component of JIA, increasing the number of confirmed JIA susceptibility loci from 3 to 17. The dense mapping of confirmed regions has narrowed down the regions to be taken forward into future functional studies. Notably, these studies allow us to begin to understand where JIA fits within the spectrum of autoimmune diseases and identifies a number of new genes and pathways as potential targets for future therapeutic intervention.

ACKNOWLEDGMENTS

We thank P. Gilbert for preparing UK JIA case samples for genotyping and M. Rye for preparing US JIA case samples and the Cincinnati local control samples. Genotyping of the US JIA, German JIA and respective control collections was supported by US National Institutes of Health (NIH) grants R01-AR-058587 and U01-AI-067150S1. In addition, subject recruitment and DNA preparation in the United States was largely funded by US NIH grants N01-AR-42272, P01-AR-048929 and P30-AR-473639, with contributions from the Arthritis Foundation, The Val A. Browning Charitable Foundation in Salt Lake City, Utah, and the Marcus Foundation, Inc., in Atlanta, Georgia, as well as US NIH grants K23-AR-50177 and R01-AR-068093. The Federal Ministry of Education and Research, Germany (BMBF grants 01GM09097 and 01 ZZ 0403) supported subject recruitment and sample preparation in Germany. Genotyping of the UK JIA case samples was supported by Arthritis Research UK (grant 17552). Sparks Childhood Arthritis Response to Medication Study was funded by Sparks, UK (0816CH09) and the Big Lottery Fund, UK (RG/1/010135231). The study is on the UK Medicines for Children Research Network (MCRN) portfolio. We acknowledge support from the Wake Forest School of Medicine Center for Public Health Genomics and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS; R01-AR-057106) for computing resources and data analysis.

Control sample recruitment and genotyping originating at the Oklahoma Medical Research Foundation (OMRF) was supported in part by NIH grants N01-AR-42272, P01-AR-048929 and P30-AR-473639 from NIH, the National Institute of General Medicine Sciences (NIGMS) and the National Institute of Allergy and Infectious Diseases (NIAID). The contents are solely the responsibility of the authors and do not necessarily represent the official views of these institutes or the US NIH.

We thank J. Barrett and C. Wallace for the SNP selection. We thank the Wellcome Trust Sanger Institute Genotyping Facility and, in particular, E. Gray, S. Bumpstead, D. Simpkin and H. Blackburn for typing the UK samples.

We acknowledge use of DNA from the UK Blood Services collection of common controls (UKBS-CC collection), which is funded by Wellcome Trust grant 076113/C/04/Z and by a National Institute for Health Research program grant to National Health Service Blood and Transplant (RP-PG-0310-1002). We acknowledge the use of DNA from the British 1958 Birth Cohort Collection, which is funded by UK Medical Research Council grant G0000934 and Wellcome Trust grant 086854/Z/01. Genotyping of control samples was supported, in part, by grants from Juvenile Diabetes Research Foundation International (JDRF) and the US NIH (U01 DK062418).

We thank P.K. Gregersen at the Feinstein Institute for providing US control genotyping from the Genotype and Phenotype Registry supported by US NIH grant RC2AR059092. We thank the National Institute of Diabetes, Digestive and Kidney Diseases Inflammatory Bowel Disease (NIDDK-IBD) Genetics Consortium for providing North American control genotyping supported by US NIH grants DK062431, DK062422, DK062420, DK062432, DK062434 and DK062423 and DK062429. We gratefully acknowledge contributions from physicians at CCHMC and collaborating clinics. We also acknowledge the assistance of S. Kramer, B. Clifford and L. Ponder in subject recruitment and coordination of clinical information at Cincinnati Children’s Hospital Medical Center, the University of Utah and Emory University, respectively. The Cincinnati normal control DNA collection was supported and made available by Cincinnati Children’s Hospital Medical Center.

AUTHOR CONTRIBUTIONS

S.D.T., W.T., C.D.L., S.P., M.C.M., J.C. and A.H. led the study. A.H., J.C., M.C.M., C.D.L., S.P., W.T. and S.D.T. wrote the manuscript. A.H., J.C., C.D.L., M.C.M., M.S., S.P., J.B., M.E.C. and S.S. performed the data and statistical analyses. A.H. and P.M. performed the bioinformatics analysis. D.N.G., L.P.H., J.F.B., R.A., M.B., W.-M.C., P.D., S. Edkins, S. Eyre, P.M.G., S.L.G., J.M.G., S.E.H., J.A.J., L.R.W. and P.W. contributed primarily to subject ascertainment, sample collection and/or genotyping. All authors reviewed the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Ravelli, A. & Martini, A. Juvenile idiopathic arthritis. Lancet 369, 767–778 (2007).
2. Prahalad, S. et al. Quantification of the familial contribution to juvenile idiopathic arthritis. Arthritis Rheum. 62, 2525–2529 (2010).
3. Prahalad, S., Shear, E.S., Thompson, S.D., Giannini, E.H. & Glass, D.N. Increased prevalence of familial autoimmunity in simplex and multiplex families with juvenile rheumatoid arthritis. Arthritis Rheum. 46, 1851–1856 (2002).
Nature Ge Netics

669

© 2013 Nature America, Inc. All rights reserved.
ONLINE METHODS

Subjects. All cohorts comprised individuals from populations of European descent from the United States, UK and Germany.

The US cohorts after quality control comprised 1,596 US oligoarticular and RF-negative polyarticular JIA cases and 4,048 US controls. Less than half of these cases have already been included in a GWAS and previously described5,6. Notably, 95 of these cases were from multiplex pedigrees, such that, for each pedigree, 1 RF-negative polyarticular or oligoarticular JIA case was randomly selected for genotyping. Clinics enrolling the JIA cases for Cincinnati-based studies (listed in order of number contributed) were located in Cincinnati, Ohio; Atlanta, Georgia; Columbus, Ohio; Little Rock, Arkansas; Long Island, New York; Chicago, Illinois; Dover, Delaware; Salt Lake City, Utah; Cleveland, Ohio; Philadelphia, Pennsylvania; Toledo, Ohio; Nashville, Tennessee; Milwaukee, Wisconsin; and Charleston, South Carolina. Additional DNA samples from JIA cases collected independently by investigators in Salt Lake City, Utah (314 cases; approximately 75% overlap with the replication cohorts in previous GWAS5,6) and Boston, Massachusetts (13 cases) or enrolled as part of the Trial of Early Aggressive Therapy in Juvenile Idiopathic Arthritis (TREAT) Study (clinicaltrials identifier NCT00443430) (22 cases) were made available for genotyping in Cincinnati.

The US controls were derived from 4 sources, including (i) 793 healthy children without known major health conditions recruited from the geographic area served by Cincinnati Children's Hospital Medical Center (CCHMC) and 119 healthy adults collected at CCHMC (previous JIA GWAS have included approximately 75% of only the pediatric controls); (ii) 484 healthy adult controls from Utah screened for autoimmune diseases (all were included in the replication cohorts of previous GWAS5,6); (iii) 848 healthy adult controls collected at the Oklahoma Medical Research Foundation; and (iv) 1,804 healthy US adult controls from the Genotype and Phenotype Registry and the NIDDK IBD Genetics Consortium. Healthy controls from the Oklahoma Medical Research Foundation (OMRF) were provided by the Lupus Family Registry and Repository (LFFR)39 and the Oklahoma Immune Cohort (OIC). Each individual completed the Connective Tissue Disease Screening Questionnaire (CSQ)40, and individuals with ‘probable’ systemic rheumatic disease were excluded. Each individual was enrolled into these studies after appropriate written consent and institutional review board (IRB) approval by the OMRF and the University of Oklahoma Health Sciences Center. Healthy controls were also provided from the University of Minnesota systemic lupus erythematosus (SLE) sibling collection41, and these subjects were enrolled after appropriate written consent and IRB approval by the University of Minnesota.

The US collections and their use in genetic studies have been approved by the IRB of CCHMC and each collaborating center.

The UK cohort after quality control comprised 772 UK oligoarticular and RF-negative polyarticular JIA cases from 5 sources: (i) The British Society for Paediatric and Adolescent Rheumatology (BSPAR) National Repository of JIA; (ii) a group of UK cases with long-standing JIA described previously42; (iii) a cohort collected as part of the Childhood Arthritis Prospective Study (CAPS), a prospective inception cohort study of JIA cases from 5 centers across the UK43; (iv) a cohort of children recruited for the SPARKS-CHARM (Childhood Arthritis Response to Medication) study who fulfilled ILAR criteria for JIA and are about to start new disease-modifying medication for active arthritis44; and (v) an ongoing collection of UK cases from the UK JIA Genetics Consortium (UKJAGC). There was overlap between the JIA cases used in this study and those in previous UK candidate gene studies of JIA5,9-12. JIA cases were classified according to ILAR criteria4, All UK JIA cases were recruited with ethical approval and provided informed consent, including from the North-West Multi-Centre Research Ethics Committee (MREC 99/8/84), the University of Manchester Committee on the Ethics of Research on Human Beings and the National Research Ethics Service (NRES 02/8/104). The 8,530 UK controls comprised the shared UK 1958 Birth Cohort and UK Blood Services Common Controls. The collection was established as part of the Wellcome Trust Case Control Consortium (WTCCC)45.

The German cohort after quality control comprised 448 German oligoarticular and RF-negative polyarticular JIA cases and 478 controls. These cases have already been included as a replication cohort in a GWAS and were previously described5,6. They were recruited from the German Center for Rheumatology in Children and Adolescents, Garmisch-Partenkirchen, Germany; the Department of Pediatrics at the University of Tübingen, Tübingen, Germany; the Children's Rheumatology Unit, St Josef-Stift, Sendenhorst, Germany; and the Department of Pediatrics at the University of Prague, Prague, Czech Republic. JIA was determined retrospectively by chart review. German population-based control samples were prepared from cord blood obtained from healthy newborns in the Survey of Neonates in Pomerania (SNIP) Consortium46. Respectively IRBs approved the collection of these samples and participation in this study.

Demographic breakdown of the cohorts is shown in Supplementary Table 1.

Genotyping and quality controls. Samples were genotyped using the Immunochip, a custom Illumina Infinium array, described previously22. Genotyping was performed according to Illumina's protocols at laboratories in Hinxton, UK; Manchester, UK; Cincinnati, USA; Utah, USA; Charlottesville, USA; and New York, USA. The Illumina GenomeStudio GenTrain2.0 algorithm was used to recluse all 15,872 samples.

SNPs were excluded if they had call rate < 98% or cluster separation score < 0.4. Samples were then excluded for call rate < 98% across 178,203 markers or if there were inconsistencies between recorded and genotype-inferred sex. Duplicates and first- or second-degree relatives were also removed. Principal-component analysis (PCA) was performed using EIGENSOFT v4.2 (refs. 47,48) on the samples merged with HapMap phase 2 individuals (CEU, YRI and CHB) as reference populations to identify ancestry outliers. PCA was performed on a subset of SNPs, removing SNPs in known regions of high LD, with MAF < 0.05 and pruned for LD between markers using a sliding window approach based on r2 = 0.2. To maximize genetic homogeneity within the samples, the initial PCA was followed by five subsequent PCAs, with individuals > 6 s.d. from the mean removed after each iteration. The principal components from the fifth iteration were used as covariates in logistic regression analysis. SNPs were removed from the primary analysis if they had significant differential missingness in cases and controls (P < 0.05), had significant departure from Hardy-Weinberg equilibrium (P < 0.001 in controls) or had MAF < 0.01.

Statistical analysis. To test for an association between a SNP and case-control status, a logistic regression analysis was computed using the five principal components as covariates. Primary inference was based on an additive genetic model, unless there was significant lack of fit to this model (P < 0.05). If there was evidence of departure from an additive model, then inference was based on the most significant of the dominant, additive and recessive genetic models. The additive and recessive models were computed only if there were at least 10 and 20 individuals homozygous for the minor allele, respectively. For analysis of the X chromosome, data analysis was first stratified by sex, and a meta-analysis was performed. The genomic control inflation factor ( λGC) was calculated using a set of SNPs included on the Immunochip for a study investigating the genetic basis for reading and writing ability27. We visually inspected the cluster plots for the most associated SNPs in the regions to confirm genotyping quality. Additionally, we determined the concordance of genotyping data with those previously generated on other platforms. A subset of cases had high-resolution HLA region genotyping. These data were used to investigate whether the SNPs with the strongest statistical associations with JIA were in high LD with classical HLA alleles or haplotypes. To investigate subtype-specific effects, the two main subtypes (oligoarticular JIA and RF-negative polyarticular JIA) were compared separately against the same controls. Disease association heterogeneity was tested by looking for significant differences in SNP allele frequency in the two subtypes. To determine how many independent associations were located within a genomic region, a manual stepwise procedure (forward selection with backward elimination, entry and exit criteria of P < 0.0001) was computed49. Specifically, for each region that reached genome-wide significance, the top SNP was included as a covariate, and the association statistics were recalculated. SNPs were allowed to enter stepwise models in this stage until no additional SNPs met the significance threshold of P < 0.0001. The stepwise procedure was modified slightly in the greater MHC region to have entry and exit criteria of P < 0.0001. These statistical analyses were performed using PLINK v1.07 (ref. 50) and SNPGWAS version 4.0.

The cumulative variance explained by common SNP variation was estimated using a variance component model and restricted maximum-likelihood estimation as implemented in the program GCTA51, adjusting for the principal
components as covariates and using Yang’s correction factor ($c = 0$ from formula 9)\textsuperscript{52} for imperfect LD with causal variants. Estimates are based on SNPs that had $<1\%$ missing genotypes and a stringent relatedness threshold of 0.025.

We computed SNP genotype imputation across the regions of the Immunochip. We used the program SHAPEIT to prephase our Immunochip data and IMPUTE2 with the 1000 Genomes Phase 1 integrated reference panel to impute the SNP genotypes. To account for phase uncertainty, we tested for association using SNPTEST. Only genotyped SNPs of high quality were used to inform imputation. Imputed SNP quality was assessed using the information score ($>0.5$) and the confidence score ($>0.9$).

Regional plots of association, and after adjustment for the strongest SNP association, at each locus, were computed using LocusZoom\textsuperscript{53}.

39. Rasmussen, A. et al. The lupus family registry and repository. *Rheumatology (Oxford)* 50, 47–59 (2011).
40. Karlson, E.W. et al. A connective tissue disease screening questionnaire for population studies. *Ann. Epidemiol.* 5, 297–302 (1995).
41. Gaffney, P.M. et al. Genome screening in human systemic lupus erythematosus: results from a second Minnesota Cohort and combined analyses of 187 sib-pair families. *Am. J. Hum. Genet.* 66, 547–556 (2000).
42. Packham, J.C. & Hall, M.A. Long-term follow-up of 246 adults with juvenile idiopathic arthritis: functional outcome. *Rheumatology (Oxford)* 41, 1428–1435 (2002).
43. Adib, N. et al. Association between duration of symptoms and severity of disease at first presentation to paediatric rheumatology: results from the Childhood Arthritis Prospective Study. *Rheumatology (Oxford)* 47, 991–995 (2008).
44. Moncrieffe, H. et al. Generation of novel pharmacogenomic candidates in response to methotrexate in juvenile idiopathic arthritis: correlation between gene expression and genotype. *Pharmacogenet. Genomics* 20, 665–676 (2010).
45. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–678 (2007).
46. Beyerdorff, A. et al. Survey of Neonates in Pomesania (SniP): a population based analysis of the mothers’ quality of life after delivery with special relations to their social integration. *Int. J. Public Health* 53, 87–95 (2008).
47. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* 2, e190 (2006).
48. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909 (2006).
49. Hosmer, D.W. & Lemeshow, S. *Applied Logistic Regression* (Wiley, New York, 1989).
50. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575 (2007).
51. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82 (2011).
52. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42, 565–569 (2010).
53. Pruim, R.J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26, 2336–2337 (2010).