Genetic correlation between multiple myeloma and chronic lymphocytic leukaemia provides evidence for shared aetiology

Molly Went1, Amit Sud1, Helen Speedy1, Nicola J. Sunter2, Asta Först3,4, Philip J. Law5, David C. Johnson6, Fabio Mirabella5, Amy Holroyd1, Ni Li5, Giulia Orlando1, Niels Weinhold6,7, Mark van Duin8, Bowang Chen3, Jonathan S. Mitchell1, Larry Mansouri9, Gunnar JULIUSSON10, Karin E Smedby11, Sandrine Jayne12, Aneela Majid12, Claire Dearden5, David J. Allsup13, James R. Bailey14, Guy Pratt15, Chris Pepper16, Chris Fegan17, Richard Rosenquist9, Rowan Kuiper8, Owen W. Stephens6, Uta Bertsch18, Peter Broderick1, Hermann Einsele19, Walter M. Gregory20, Jens Hillengass7, Per Hoffmann21,22, Graham H. Jackson23, Karl-Heinz Jöckel24, Jolanta Nickel7, Markus M. Nöthen25,26, Miguel Inacio da Silva Filho3, Haucke Thomsen5, Brian A. Walker6, Annemie Broyle8, Faith E. Davies6, Markus Hansson27,28, Hartmut Goldschmidt7,18, Martin J. S. Dyer12, Martin Kaiser2, Pieter Sonneveld8, Gareth J. Morgan9, Kari Hemminki14, Björn Nilsson27,28, Daniel Catovsky5, James M. Allan2 and Richard S. Houlston1,5

Abstract

The clustering of different types of B-cell malignancies in families raises the possibility of shared aetiology. To examine this, we performed cross-trait linkage disequilibrium (LD)-score regression of multiple myeloma (MM) and chronic lymphocytic leukaemia (CLL) genome-wide association study (GWAS) data sets, totalling 11,734 cases and 29,468 controls. A significant genetic correlation between these two B-cell malignancies was shown ($R_g = 0.4$, $P = 0.0046$). Furthermore, four of the 45 known CLL risk loci were shown to associate with MM risk and five of the 23 known MM risk loci associate with CLL risk. By integrating eQTL, Hi-C and ChIP-seq data, we show that these pleiotropic risk loci are enriched for B-cell regulatory elements and implicate B-cell developmental genes. These data identify shared biological pathways influencing the development of CLL and MM and further our understanding of the aetiological basis of these B-cell malignancies.

Introduction

Chronic lymphocytic leukaemia (CLL) and multiple myeloma (MM) are both B-cell malignancies, which arise from the clonal expansion of progenitor cells at different stages of B-cell maturity1–3. Evidence for inherited predisposition to CLL and MM comes from the six- and two-fold increased risk of the respective diseases seen in relatives of patients4.

Recent genome-wide association studies (GWAS) have transformed our understanding of genetic susceptibility to the B-cell malignancies, identifying 45 CLL5–8 and 23 MM risk loci9–13. Furthermore, statistical modelling of GWAS data indicates that common genetic variation is likely to account for 34% of CLL and 15% of MM heritability6,14.
Epidemiological observations on familial cancer risks across the different B-cell malignancies suggest an element of shared inherited susceptibility, especially between CLL and MM.

Linkage disequilibrium (LD) score regression is a method which exploits the feature of a test statistic for a given single nucleotide polymorphism (SNP), whereby that test statistic will incorporate the effects of correlated SNPs. Conventional LD score regression regresses trait $\chi^2$ statistics against the SNP LD score, with the slope providing an estimate of genetic covariance between the two traits. This method can be modified by instead regressing the product of SNP Z-scores from two traits against the SNP LD score, with the slope providing an estimate of genetic susceptibility at 10 known risk loci and by integrating promoter capture Hi-C (PCHi-C) data, ChIP-seq and expression data we provide insight into the shared biological basis of CLL and MM.

Methods

GWAS data sets

The data from six previously reported MM GWAS are summarized in Supplementary Table 1. All these studies were based on individuals of European ancestry and comprised: Oncoarray-GWAS (878 cases 7054 controls) UK-GWAS (2282 cases, 5197 controls), Swedish-GWAS (1714 cases, 10,391 controls), German-GWAS (1508 cases, 2107 controls), Netherlands-GWAS (555 cases, 2699 controls) and US-GWAS (780 cases, 1857 controls).

The data from three previously reported CLL GWAS are summarized in Supplementary Table 2. All these studies were based on individuals of European ancestry and comprised: CLL UK1 (505 cases and 2698 controls), CLL UK2 (1236 cases and 2501 controls) and CLL US (2174 cases and 2682 controls).

Ethics

Collection of patient samples and associated clinicopathological information was undertaken with written informed consent and relevant ethical review board approval at respective study centres in accordance with the tenets of the Declaration of Helsinki.

Specifically for the Myeloma-XI trial by the Oxfordshire Research Ethics Committee (MREC 17/09/09, ISRCTN49407852), HOVON65/GMMG-HD4 (ISRCTN 644552890; METC 13/01/2015), HOVON87/NMSG18 (EudraCTnr 2007-004007-34, METC 20/11/2008), HOVON95/EMN02 (EudraCTnr 2009-017903-28, METC 04/11/10), University of Heidelberg Ethical Commission (229/2003, S-337/2009, AFinu-119/2010), University of Arkansas for Medical Sciences Institutional Review Board (IRB 202077), Lund University Ethical Review Board (2013/54, the Norwegian REK 2014/97, and the Danish Ethical Review Board (no: H-16032570).

Specifically, the centres for UK-CLL1 and UK-CLL2 are: UK Multi-Research Ethics Committee (MREC 99/1/082); GEC: Mayo Clinic Institutional Review Board, Duke University Institutional Review Board, University of Utah, University of Texas MD Anderson Cancer Center Institutional Review Board, National Cancer Institute, ATBC: NCI Special Studies Institutional Review Board, BCCA: UBC BC Cancer Agency Research Ethics Board, CPS-II: American Cancer Society, ENGELA: IRB00003888—Comite d’ Evaluation Ethique de l’Inserm IRB #1, EPIC: Imperial College London, Epilymph: International Agency for Research on Cancer, HPFS: Harvard School of Public Health (HSPH) Institutional Review Board, Iowa-Mayo SPORE: University of Iowa Institutional Review Board, Italian GxE: Comitato Etico Azienda Ospedaliero Universitaria di Cagliari, Mayo Clinic Case-Control: Mayo Clinic Institutional Review Board, MCCS: Cancer Council Victoria’s Human Research Ethics Committee, MSKCC: Memorial Sloan-Kettering Cancer Center Institutional Review Board, NCI-SEER (NCI Special Studies Institutional Review Board), NHS: Partners Human Research Committee, Brigham and Women’s Hospital, NSW: NSW Cancer Council Ethics Committee, NYU-WHS: New York University School of Medicine Institutional Review Board, PLOCO: (NCI Special Studies Institutional Review Board), SCALE: Scientific Ethics Committee for the Capital Region of Denmark, SCALE: Regional Ethical Review Board in Stockholm (Section 4) IRB#5, Utah: University of Utah Institutional Review Board, UCSF and UCSD: University of California San Francisco Committee on Human Research, Women’s Health Initiative (WHI): Fred Hutchinson Cancer Research Center and Yale: Human Investigation Committee, Yale University School of Medicine. Informed consent was obtained from all participants.

The diagnosis of MM (ICD-10 C90.0) in all cases was established in accordance with World Health Organization guidelines. All samples from patients for genotyping were obtained before treatment or at presentation. The diagnosis of CLL (ICD-10-CM C91.10, ICD-O M9823/3 and 9670/3) was established in accordance with the International Workshop on Chronic Lymphocytic Leukemia guidelines.
Quality control

Standard quality-control measures were applied to the GWAS\textsuperscript{17}. Specifically, individuals with low SNP call rate (<95\%) as well as individuals evaluated to be of non-European ancestry (using the HapMap version 2 CEU, JPT/CHB and YRI populations as a reference) were excluded. For apparent first-degree relative pairs, we excluded the control from a case-control pair; otherwise, we excluded the individual with the lower call rate. SNPs with a call rate <95\% were excluded as were those with a MAF <0.01 or displaying significant deviation from Hardy–Weinberg equilibrium ($P < 10^{-5}$). GWAS data were imputed to >10 million SNPs using IMPUTE2 v4 (for CLL) and IMPUTE2 v2.3 (for MM) software in conjunction with a merged reference panel consisting of data from 1000 Genomes Project\textsuperscript{18} (phase 1 integrated release 3 March 2012) and UK10K\textsuperscript{19}. Genotypes were aligned to the positive strand in both imputation and genotyping. We imposed predefined thresholds for imputation quality to retain potential risk variants with MAF >0.01 for validation. Poorly imputed SNPs with an information measure <0.80 were excluded. Tests of association between imputed SNPs and MM were performed under an additive model in SNPTESTv2.5\textsuperscript{20}. The adequacy of the case-control matching and possibility of differential genotyping of cases and controls was evaluated using a Q–Q plot of test statistics. The inflation $\lambda$ was based on the 90\% least-significant SNPs and assessment of $\lambda_{1000}$. Details of SNP QC are provided in Supplementary Table 3 and 4. Four principal components, generated using common SNPs, were included to limit the effects of cryptic population stratification in the US-CLL data set. Eigenvectors for the GWAS data sets were inferred using smartpc (part of EIGENSOFT) by merging cases and controls with phase II HapMap samples.

Meta-analysis

Meta-analyses were performed using the fixed-effects inverse-variance method using META v1.6\textsuperscript{21}. Cochran’s Q-statistic to test for heterogeneity and the $I^2$ statistic to quantify the proportion of the total variation due to heterogeneity was calculated.

LD score regression

To investigate genetic correlation between MM and CLL, we implemented cross-trait LD score regression by Bulik-Sullivan et al.\textsuperscript{16}. Using summary statistics from the GWAS meta-analysis we implemented filters as recommended by the authors\textsuperscript{16}. Specifically, filtering SNPs to INFO >0.9, MAF >0.01, and harmonizing to Hap Map3 SNPs with 1000 Genomes EUR MAF >0.05, removing indels and structural variants, removing strand-ambiguous SNPs and removing SNPs where alleles did not match those in 1000 Genomes. This was performed by running the munge-sumstats.pr script included with ldsc. We ran ldsc.py, part of the ldsc package, excluding the HLA region. We report heritability estimates on the observed scale. There is no distinction between observed and liability scale genetic correlation for case/control traits\textsuperscript{16}.

Shared risk loci

To identify pleiotropic risk loci, that is genetic loci that influence two traits, we identified SNPs previously reported to be associated with each disease at genome-wide significance ($P < 5 \times 10^{-8}$), as well as highly correlated variants ($r^2 > 0.8$) at the 45 and 23 known risk loci for CLL and MM, respectively. Within these correlated variant sets at each locus, we determined how many of the CLL susceptibility loci were associated with MM at region-wide significance after Bonferroni correction for multiple testing (i.e. $P_{adj} < 0.05/45$). We then repeated the process, examining MM susceptibility SNPs in CLL, applying a significance level of $P_{adj} < 0.05/23$. A full list of results is summarized in Supplementary Data File 1 and 2.

Partitioned heritability

A variation of LD score regression, namely stratified LD score regression, can be used to partition heritability according to different genomic categories. For both MM and CLL we applied stratified LD score regression across the baseline model used in Finucane et al.\textsuperscript{22}. We plotted the enrichment of functional categories for each disease—this is defined as proportion heritability divided by the total heritability. We excluded from our plot additional flanking regions around each functional category, which authors designed to allow observation of enrichment of SNP heritability in intermediary regions. A plot of the results is found in Supplementary Figure 1.

Variant set enrichment

To examine enrichment in specific histone mark binding across shared risk loci, we adapted the method of Cowper-Sal lari et al.\textsuperscript{23}. Briefly, for each risk locus, a region of strong LD (defined as $r^2 > 0.8$ and $D’ > 0.8$) was determined, and these SNPs were considered the associated variant set (AVS). Publically available ChIP-seq data for 6 histone marks from naive B cells was downloaded from Blueprint Epigenome Project\textsuperscript{24}. For each mark, the overlap of the SNPs in the AVS and the binding sites was assessed to generate a mapping tally. A null distribution was produced by randomly selecting SNPs with the same characteristics as the risk-associated SNPs, and the null mapping tally calculated. This process was repeated 10,000 times, and $P$-values calculated as the proportion of permutations where null mapping tally was greater or equal to the AVS mapping tally. An enrichment score was calculated by normalizing the tallies to the
Under the hypothesis of pleiotropy (HEIDI), using multiple SNPs in each tested for heterogeneity in dependent instruments one affecting gene expression the other affecting trait we top associated cis-eQTL is in LD with two causal variants, results are shown in Supplementary Table 5. 

Cell-type-specific analyses
We considered chromatin mark overlap enrichment for genome-wide significant loci in different cell types using the methodology of Trynka et al.25. This approach scores GWAS SNPs based on proximity to chromatin mark and fold-enrichment of respective chromatin mark, assessing significance using a tissue-specific permutation method. We obtained chip-seq data for H3K4me3 from primary blood cells and CLL samples downloaded from Blueprint Epigenome project24. In addition, we included in our analysis 4 MM cell lines- KMS11, JJN3, MM1-S and L363 processed as previously described26. A heat map of results is shown in Supplementary Figure 3.

eQTL
eQTL analyses were performed using publicly available whole-blood data downloaded from GTEx27. The relationship between SNP genotype and gene expression we carried out using Summary-data-based Mendelian Randomization (SMR) analysis as per Zhu et al.28. Briefly, if \( b_{xy} \) is the effect size of \( x \) (gene expression) on \( y \) (slope of \( y \) regressed on the genetic value of \( x \)), \( b_{xz} \) is the effect of \( x \) on \( z \), and \( b_{zy} \) be the effect of \( z \) on \( y \), \( b_{xy} (b_{xz}/b_{zy}) \) is the effect of \( x \) on \( y \). To distinguish pleiotropy from linkage where the top associated cis-eQTL is in LD with two causal variants, one affecting gene expression the other affecting trait we tested for heterogeneity in dependent instruments (HEIDI), using multiple SNPs in each cis-eQTL region. Under the hypothesis of pleiotropy \( b_{zy} \) values for SNPs in LD with the causal variant should be identical. For each probe that passed significance threshold for the SMR test, we tested the heterogeneity in the \( b_{zy} \) values estimated for multiple SNPs in the cis-eQTL region using HEIDI.

GWAS summary statistics files were generated from the meta-analysis. For the disease discovery GWAS, we set a threshold for the SMR test of \( P_{\text{SMR}} < 2.5 \times 10^{-5} \) corresponding to a Bonferroni correction for the number of probes which demonstrated an association in the SMR test. For all genes passing this threshold we generated plots of the eQTL and GWAS associations at the locus, as well as plots of GWAS and eQTL effect sizes (i.e. input for the HEIDI heterogeneity test). HEIDI test \( P \)-values <0.05 were considered as reflective of heterogeneity. This threshold is, however, conservative for gene discovery because it retains fewer genes than when correcting for multiple testing. SMR plots for significant eQTLs are shown in Supplementary Figures 4, 5 and a summary of results are shown in Supplementary Table 5.

Results
Genetic correlation and heritability
We performed cross-trait LD-score regression using summary statistics from two recent GWAS meta-analyses based on 7717 MM cases and 21,587 controls and 4017 CLL cases and 7881 controls (Fig. 1, Supplementary Table 1-4). While these data sets have been previously subject to quality control (QC)5–7,9–12 for the current analysis we implemented additional filtering steps as per Bulik-Sullivan et al.16, resulting in 1,055,728 harmonized SNPs between the two data sets. Heritability estimates from cross-trait LD score regression of 9.2 (±1.8%) and 22 (±5.9%) were comparable with previous estimates for MM14 and CLL5. LD-score regression revealed a significant-positive genetic correlation between MM and CLL with an \( R_g \) value of 0.44 (\( P = 4.6 \times 10^{-3} \)).

Identification of pleiotropic risk loci
We identified SNPs previously reported to be associated with each disease at genome-wide significance (\( P < 5 \times 10^{-8} \)), as well as highly correlated variants (\( r^2 > 0.8 \)) at the 45 and 23 known risk loci for CLL and MM, respectively. To identify pleiotropic risk loci, that is genetic loci that influence two traits, we determined how many of the CLL susceptibility loci were associated with MM at region-wide significance after Bonferroni correction for multiple testing (i.e. \( P_{\text{adj}} < 0.05/45 \)). We then repeated the process, examining MM susceptibility SNPs in CLL, applying a significance level of \( P_{\text{adj}} < 0.05/23 \). Of the 45 CLL risk loci, four were associated with MM (\( P_{\text{adj}} < 0.0011 \)) while, of 23 MM risk loci, five were significantly associated in CLL (\( P_{\text{adj}} < 0.0022 \)) (Table 1, Fig. 2). Correlated SNPs (\( r^2 > 0.8 \))
at 3q26.2 are associated with both CLL and MM at genome-wide significance (Fig. 2), bringing the total number of pleiotropic loci to 10.

**Biological inference**

Trynka et al. have recently shown that chromatin marks highlighting active regulatory regions overlap with phenotype-associated variants in a cell-type-specific manner. As H3K4me3 was shown to be the most phenotypically cell-type-specific chromatin mark, we examined cell-type specificity of the 10 pleiotropic risk loci by analysing H3K4me3 chromatin marks in normal haematopoietic cells and CLL patient samples from Blueprint, and de novo data on KMS11, MM1S, JJN3 and L363 MM cell lines. Cell types showing the strongest enrichment of risk SNPs at H3K4me3 marks included naive B cells and CD38-B cells. Notably, variants at 2q31.1, 6p25.3, 8q24.21, 16q23.1 and 22q13.33 were enriched for H3K4me3 in naive B cells (Supplementary Figure 3).

Most GWAS signals map to non-coding regions of the genome and influence gene expression through chromatin looping interactions. Application of partitioned heritability analysis, stratifying across 53 genomic categories demonstrated enrichment of CLL and MM heritability in functional elements of the genome, in particular FANTOM5 enhancers (CLL and MM) transcription start sites (CLL) and 5′ untranslated region and coding regions (MM) (Supplementary Figure 1). Furthermore, we found significant enrichment of SNPs in the shared loci within regions of active chromatin, as indicated by the presence of H3K27ac and H3K4Me3 marks in naive B cells, supporting the principle that SNPs in shared loci influence risk through regulatory effects (Supplementary Figure 2). To identify target genes we analysed PCHi-C data on naive B cells from Blueprint.

**Table 1 Risk loci demonstrating association of alleles at respective loci in both chronic lymphocytic leukaemia (CLL) and multiple myeloma (MM)**

| Locus  | Discovery GWAS | Sentinel variant | Correlated variant | Position (hg19) | Risk allele | Odds Ratio | P-value |
|--------|----------------|-----------------|-------------------|----------------|-------------|------------|---------|
|        |                |                 |                   |                | CLL MM      | CLL MM     | CLL MM  |
| 2q31.1 | MM            | rs4325816       |                   | 174,808,899    | T T         | 1.11       | 1.12    | 2.0 × 10^{-3}  | 6.4 × 10^{-7} |  |
|        |                | rs72919402      |                   | 174,750,200    | T -         | 1.13       | -      | 4.6 × 10^{-4}  | -             |  |
| 3q26.2 | MM & CLL      | rs1317082       |                   | 169,497,585    | A A         | 1.20       | 1.19    | 7.1 × 10^{-6}  | 2.2 × 10^{-16} |  |
|        |                | rs3821383       |                   | 169,489,946    | A A         | 1.20       | 1.18    | 4.2 × 10^{-6}  | 4.5 × 10^{-15} |  |
| 6p25.3 | CLL           | rs872071        |                   | 411,064        | G G         | 1.37       | 1.10    | 2.8 × 10^{-7}  | 7.5 × 10^{-7}  |  |
|        |                | rs1050976       |                   | 408,079        | T T         | 1.37       | 1.10    | 1.9 × 10^{-7}  | 3.7 × 10^{-7}  |  |
| 6p22.3 | MM            | rs34229995      |                   | 15,244,018     | G G         | 1.37       | 1.36    | 8.5 × 10^{-7}  | 5.6 × 10^{-8}  |  |
|        |                | rs13197919      |                   | 15,282,334     | T T         | 1.35       | 1.32    | 1.3 × 10^{-4}  | 3.42 × 10^{-7} |  |
| 7q31.33| MM            | rs58618031      |                   | 124,583,896    | T T         | 1.15       | 1.11    | 3.2 × 10^{-6}  | 1.7 × 10^{-7}  |  |
|        |                | rs59294613      |                   | 124,554,267    | C -         | 1.16       | -      | 4.4 × 10^{-6}  | -             |  |
| 8q24.21| MM            | rs1949815       |                   | 128,222,421    | C C         | 1.17       | 1.15    | 7.6 × 10^{-7}  | 2.5 × 10^{-12} |  |
|        |                | -               |                   | -              | -           | -         | -      | -              | -             |  |
| 10q23.31| CLL          | rs6586163       |                   | 90,752,018     | A A         | 1.28       | 1.06    | 1.1 × 10^{-16} | 1.8 × 10^{-3}  |  |
|        |                | rs7082101       |                   | 90,741,615     | - C         | 1.06       | -      | 8.2 × 10^{-4}  | -             |  |
| 11q23.2| CLL           | rs11601504      |                   | 113,526,853    | C C         | 1.20       | 1.09    | 2.3 × 10^{-5}  | 8.5 × 10^{-4}  |  |
|        |                | -               |                   | -              | -           | -         | -      | -              | -             |  |
| 16q23.1| MM            | rs7193541       |                   | 74,664,743     | T T         | 1.12       | 1.12    | 1.0 × 10^{-4}  | 3.7 × 10^{-10} |  |
|        |                | CLL             |                   | -              | -           | -         | -      | -              | -             |  |
| 22q13.33| MM           | rs140522        |                   | 50,971,266     | T T         | 1.17       | 1.08    | 3.7 × 10^{-7}  | 1.2 × 10^{-4}  |  |
|        |                | -               |                   | -              | -           | -         | -      | -              | -             |  |

- indicates SNP not present in filtered data
expression data on blood from GTEx. Applying Summary data-based Mendelian Randomization (SMR) methodology, we tested for pleiotropy between GWAS signal and cis-eQTL for genes to identify a causal relationship. Broadly, our analysis of the shared loci groups them into those which act on a B-cell regulation and differentiation and those which underpin the distinctive biology of cancer; specifically, loci relating to genome instability, angiogenesis and dysregulated apoptosis (Supplementary Table 6).

Of the shared loci, three were related to B-cell regulation. This included composite evidence at 10q23.31, from looping interactions in naive B cells and correlation in GWAS effect size and expression, which provide evidence for two candidate genes ACTA2, encoding smooth muscle (α)-2 actin, a protein involved in cell movement and contraction of muscles and FAS, a member of the TNF-receptor superfamily. FAS, has a central role in regulating the immune response through apoptosis of B cells. At 2q31.1, looping interactions implicated transcription factor SP3, which has been shown to influence expression of germinal centre genes. Variants at 6p25.3 reside in the 3'-UTR of IRF4, which has an established role in B-cell regulation and MM oncogenesis.

Three of the 10 loci contain genes with roles in maintenance of genomic stability. Specifically, evidence from expression and PCHi-C data implicated RFWD3 at 16q23.1. This gene encodes an E3 ubiquitin–protein ligase, which has been shown to promote progression to late stage homologous recombination through ubiquitination and timely removal of RAD51 and RPA at sites of DNA damage and is necessary for replication fork restart. Variants in this locus demonstrated enrichment of H3K4me3 marks in two samples of naive B cells, which represents a plausible cell of disease origin. rs58618031 (7q31.33) maps 5' of POT1, the protection of telomeres 1 gene, which is part of the shelterin complex and functions to maintain chromosomal stability. Variant rs1317082 at 3q26.2 is located proximal to TERC, a gene which has been shown to influence telomere length. Additionally, we observed looping interactions to a number of genes at 3q26.2 including SEC62, which has been proposed as a cancer biomarker. Intriguingly, variants at 3q26.2 this locus have been implicated in colorectal, thyroid and bladder cancer.

Several genes were implicated at 22q13.33 by looping interactions for SOC2, LMF2, ODF3B, TYMP/ECGF1, NCAPH2, SYCE3 and ARSA, with TYMP/ECGF1 and SOC2 demonstrating evidence of correlation in GWAS and eQTL effect size, albeit not significant after multiple testing (P_{SMR} = 2.38 \times 10^{-4} and 3.19 \times 10^{-5}). Variants within this locus were enriched in H3K4me3 chromatin marks in both CD38- B cells and inflammatory macrophages. TYMP (alias ECGF1) encodes thymidine phosphorylase, which is often overexpressed in tumours and has been linked to angiogenesis. A detailed study on this gene has implicated TYMP in the development of lytic bone lesions in MM, via a mechanism involving activation of PI3K/Akt signalling and increased DNMT3A expression resulting in hypermethylation of RUNX2, osterix, and IRF8. Furthermore, SOC2 (synthesis of cytochrome c oxidase), also mapping to this locus, has been implicated in the development of breast, gastric, and leukaemia, through glucose metabolism reprogramming, a hallmark of cancer. Tumour suppressor, p53, regulates metabolic pathways, p53-transactivated TP53-induced glycolysis (TIGAR), and regulation of apoptosis in part through SOC2.

Finally, whereas these data were indifferent to decipher 8q24.21, this locus has also been shown to harbour risk SNPs for other cancers, which localize within distinct LD blocks and likely reflect tissue-specific effects on cancer risk through regulation of MYC.

**Discussion**

Our analysis provides evidence of a genetic correlation between MM and CLL. Furthermore, we have identified shared genetic susceptibility at 10 known risk loci. While requiring biological validation, integration of data from PCHi-C, chromatin mark enrichment and eQTL at shared loci has provided insight into how these loci may confer susceptibility to both CLL and MM. Applying a working hypothesis that the loci may act in pleiotropic fashion, we selected relevant cells representing a common tissue of disease origin; namely naive B cells.

A significant genetic correlation between MM and CLL, as well as the discovery of risk loci shared between them,
supports epidemiological data demonstrating elevated familial risks between these B-cell malignancies. Furthermore, the shared loci we identified could be broadly grouped into those containing genes related to B-cell regulation and differentiation and those containing genes involved in angiogenesis, genome stability and apoptosis, supporting the tenet that these alleles can influence etiology of either disease. With the expansion of GWAS of the B-cell malignancies, more detailed characterisation of common underlying risk alleles and affected pathways can inform the biology of B-cell oncogenesis.

Data availability
SNP genotyping data that support the findings of this study have been deposited in Gene Expression Omnibus with accession codes GSE21349, GSE19784, GSE24080, GSE26568 and GSE15695, in the European Genome-phenome Archive (EGA) with accession code EGAS00000000001, in the European Bioinformatics Institute (Part of the European Molecular Biology Laboratory) (EMBL-EBI) with accession code E-MTAB-362 and E-TABM-1138; and in the database of Genotypes and Phenotypes (dbGaP) with accession code phs00207.v1.p1. The remaining data are contained within the paper and Supplementary Files or available from the author upon request. Naïve B-cell HiC data used in this work is publicly available from Blueprint Epigenome Project [https://osf.io/ultzpq/]. ChiP-seq data for H3K27ac, H3K4Me1, H3K27Me3, H3K36Me3 and H3K40Me3 from naïve B cells are publicly available and was obtained from Blueprint Epigenome Project [http://www.blueprint-epigenome.eu/].

Acknowledgements
In the United Kingdom, Myeloma UK and Bloodwise (#05001, #06002 and #13044) provided principal funding. Additional funding was provided by Cancer Research UK (C1287/A10692, C1287/A16563, C5047/A3354, C5047/A10692, C16913/A6135, European Commission’s Seventh Framework Programme grant n° 221175 (HEALTH-F2-2009-223175), and The National Institute of Health (NIH) Cancer Post-Cancer GWAS initiative grant. No. 1 L19 L18537-01 (the GAME-ON initiative). We would also like to thank the following for funding support: The Institute of Cancer Research and The Everman Campaign, The Prostate Cancer Foundation, Prostate Research Campaign UK (now Prostate Action), The Orchid Cancer Appeal, The National Cancer Research Network UK, The National Cancer Research Institute (NCRI) UK. We are grateful for support of NIH funding to the NIH Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. The APRC BioResource, which forms part of the PRACTICAL consortium, consists of the following members: Wayne Tilley, Gail Ribisnig, Renee Taylor, Judith A Clements, Lisa Horvath, Vanessa Hayes, Lisa Butler, Trina Yeadon, Allison Eckert, Pamela Saunders, Anne-Marie Haynes, Melissa Papargiris. The US GWAS was supported by a grant from the National Institutes of Health (P01CA055819). The German study was supported by the Dietmar-Hopp-Stiftung, Germany, the German Cancer Aid (1103131), the Federal Ministry of Education and Research (CILOMMCs 01ZX1309), The German Research Council (DFG, Project SI 236/81, SI 236/1-1, ER 155/6-1) and the DFG CRI 216), the Harald Huppert Foundation and the Multiple Myeloma Research Foundation. The patients were collected by the GWMG and DSWM studies. The German GWAS made use of genotyping data from the population-based HNR study, which is supported by the Heinz Nixdorff Foundation (Germany). The genotyping of the Illumina HumanOmni-Quad BeadChips of the HNR subjects was performed by the German Center for Neurodegenerative Disorders (DZNE), Bonn. We are grateful to all investigators who contributed to the generation of this data set. The German replication controls were collected by Peter Bugert, Institute of Transfusion Medicine and Immunology, Medical Faculty Mannheim, Heidelberg University, German Red Cross Blood Service of Baden-Württemberg-Hessen, Mannheim, Germany. This work was supported by research grants from the Swedish Foundation for Strategic Research (KF01-0009), the Marianne and Marcus Wallenberg Foundation (2010.0112), the Knut and Alice Wallenberg Foundation (2012.0193), the Swedish Research Council (2012-1753), the Royal Swedish Academy of Science, ALE grants provided to the University of Skåne, (Labmedicin Skåne), the Siv-Inger and Per-Erik Andersson Foundation, the Medical Faculty at Lund University, the Borås foundation for cancer research, and the Swedish Society of Medicine. We thank Jorgen Adolfsson, Tomas Axelsson, Anna Collin, Ildiko Frigyes, Patrik Magnusson, Bert Johansson, Jan Westin and Helga Omgundsdottir for their assistance. This work was supported by Center for Translational Molecular Medicine (BioCHP), a clinical research grant from the European Hematology Association, an EMCR Translational Research Grant, a BMBF grant from CILOMMCs (01ZX1309A) and FP7 grant MSCNET (LSHC-Ct-2006-037602). N.W. was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R25GM125303. We thank the staff of the HoxOn, as well as patients and physicians at participating sites. In addition, we also thank Jasper Koenders, Michael Vermeulen, André Uitterlinden and Nathalie van der Velde for their assistance. We are indebted to the clinicians who contributed samples to Swedish, Norwegian and Danish biobanks. We are indebted to the patients
Conflict of interest

The authors declare that they have no conflict of interest.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information

accompanies this paper at (https://doi.org/10.1038/s41408-018-0162-8).

Received: 26 August 2018 Accepted: 19 November 2018
Published online: 21 December 2018

References

1. Barlogie, B. & Gale, R. P. Multiple myeloma and chronic lymphocytic leukemia: parallels and contrasts. Am. J. Med. 93, 448–450 (1992).
2. Shaffer, A. L., Rosenwald, A. & Staudt, L. M. Lymphoid malignancies: the dark side of B-cell differentiation. Nat. Rev. Immunol. 2, 930–952 (2002).
3. Kuppers, R. Mechanisms of B-cell lymphoma pathogenesis. Nat. Rev. Cancer 5, 251–262 (2005).
4. Sud, A., et al. The landscape of familial risk of hematological malignancies: an analysis of 153,115 cases. (Under review).
5. Growther-Swanepoel, D. et al. Common variants at 2q37.3, 8q24.2, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. Nat. Genet. 42, 132–136 (2010).
6. Law, P. J. et al. Genome-wide association analysis implicates dysregulation of immunity genes in chronic lymphocytic leukaemia. Nat. Commun. 8, 14175 (2017).
7. Speedy, H. E. et al. A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. Nat. Genet. 46, 56–60 (2014).
8. Went, M. et al. Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukaemia (1546–1718 (Electronic)).
9. Went, M. et al. Identification of multiple risk loci and regulatory mechanisms influencing susceptibility to multiple myeloma. Nat. Commun. 9, 3707 (2018).
10. Mitchell, J. S. et al. Genome-wide association study identifies multiple susceptibility loci for multiple myeloma. Nat. Commun. 7, 12050 (2016).
11. Bredereck, P. et al. Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. Nat. Genet. 44, 58–61 (2011).
12. Chubb, D. et al. Common variation at 3q26.2, 6p21.33, 17p11.2 and 22q13.1 influences multiple myeloma risk. Nat. Genet. 45, 1221–1225 (2013).
13. Swaminathan, B. et al. Variants in ELL2 influencing immunglobulin levels associate with multiple myeloma. Nat. Commun. 6, 7213 (2015).
14. Mitchell, J. S. et al. Implementation of genome-wide complex trait analysis to quantify the heritability in multiple myeloma. Sci. Rep. 5, 12473 (2015).
15. Bulk-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S. & Yang, J. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47, 291–295 (2015).
16. Bulk-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R. An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241 (2015).
