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

Molly Went[1], Amit Sud[1], Helen Speedy[1], Nicola J Sunter[2], Asta Försti[3,4], Philip J. Law[1], David C. Johnson[5], Fabio Mirabella[5,] Amy Holroyd[1], Ni Li[1], Giulia Orlando[1], , Niels Weinhold[6,7], Mark van Duin[8] , Bowang Chen[3], Jonathan S. Mitchell[1], Larry Mansouri[9], Gunnar Juliusson[10], Karin E Smedby[11], Sandrine Jayne[12], Aneela Majid[12], Claire Dearden[5], David J Allsup[13], James R Bailey[14], Guy Pratt[15], Chris Pepper[16], Chris Fegan[17], Richard Rosenquist[9], Rowan Kuiper[8], Owen W. Stephens[6], Uta Bertsch[3,18], Peter Broderick[1], Hermann Einsele[19], Walter M. Gregory[20], Jens Hillengass[7], Per Hoffmann[21,22], Graham H. Jackson[23], Karl-Heinz Jöckel[24], Jolanta Nickel[7], Markus M. Nöthen[25,26], Miguel Inacio da Silva Filho[3], Hauke Thomsen[3], Brian A. Walker[6], Annemiek Broyl[8], Faith E. Davies[6], Markus Hansson[4,26], Hartmut Goldschmidt[7,18], Martin J. S. Dyer, Martin Kaiser[5], Pieter Sonneveld[8], Gareth J. Morgan[6], Kari Hemminki[3,4], Björn Nilsson[27,28], Daniel Catovsky[5], James M. Allan[2] and Richard S. Houlston[1,5]

[1]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, SW7 3RP

[2]Northern Institute for Cancer Research, Newcastle University, Newcastle upon Tyne NE2 4HH, UK

[3]German Cancer Research Center, 69120 Heidelberg, Germany

[4]Center for Primary Health Care Research, Lund University, SE-205 02 Malmo, Sweden

[5]Division of Molecular Pathology, The Institute of Cancer Research, London, SW7 3RP

[6]Myeloma Institute for Research and Therapy, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205, USA

[7]Department of Internal Medicine V, University of Heidelberg, 69117 Heidelberg, Germany

[8]Department of Hematology, Erasmus MC Cancer Institute, 3075 EA Rotterdam, The Netherlands

[9]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, 75105 Uppsala, Sweden

[10] Lund Strategic Research Center for Stem Cell Biology and Cell Therapy, Hematology and Transplantation, Lund University, Lund, Sweden.

[11] Unit of Clinical Epidemiology, Department of Medicine, Karolinska Institutet, Stockholm, Sweden.

[12]Medical Research Council Toxicology Unit, Leicester University, Leicester, UK.
[13]Department of Haematology, Hull Royal Infirmary, Hull, UK.
[14]Hull York Medical School and University of Hull, Hull, UK.
[15]Department of Haematology, Birmingham Heartlands Hospital, Birmingham, UK.
[16]Department of Haematology, School of Medicine, Cardiff University, Cardiff, UK.
[17]Cardiff and Vale National Health Service Trust, Heath Park, Cardiff, UK.
[18]National Centre of Tumor Diseases, 69120 Heidelberg, Germany
[19]University Clinic of Würzburg, 97080 Würzburg, Germany
[20]Clinical Trials Research Unit, University of Leeds, Leeds LS2 9PH, UK
[21]Institute of Human Genetics, University of Bonn, D-53127 Bonn, Germany
[22]Division of Medical Genetics, Department of Biomedicine, University of Basel, 4003 Basel, Switzerland
[23]Royal Victoria Infirmary, Newcastle upon Tyne, NE1 4LP, UK
[24]Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, University of Duisburg–Essen, Essen, Germany
[25]Institute of Human Genetics, University of Bonn, D-53127 Bonn, Germany
[26]Department of Genomics, Life & Brain Center, University of Bonn, D-53127 Bonn, Germany
[27]Hematology and Transfusion Medicine, Department of Laboratory Medicine, BMC B13, SE-221 84 Lund University, Sweden
[28]Broad Institute, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA

Correspondence to: Molly Went, Tel: +44 (0) 208 722 4431; E-mail: molly.went@icr.ac.uk

Running title: Genetic correlation between MM and CLL provides evidence for shared genetic aetiology

Word Count: 5244

Abstract: 140

Figures: 2

Tables: 1

References: 61

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) datasets, 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 maturity (1-3). Evidence for inherited predisposition to CLL and MM comes from the 6 and 2-fold increased risk of respective diseases seen in relatives of patients(4).

Recent genome-wide association studies (GWAS) have transformed our understanding of genetic susceptibility to the B-cell malignancies, identifying 45 CLL (5-8) and 23 MM risk loci (9-12). Furthermore, statistical modelling of GWAS data indicates that common genetic variation is likely to account for 34% of CLL and 15% of MM heritability (6, 13). Epidemiological observations on familial cancer risks across the different B-cell malignancies suggest an element of shared inherited susceptibility, especially between CLL and MM (4).

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 (14). Conventional LD score regression regresses trait χ^2 statistics against the LD score for a given SNP, with the coefficient of the regression line providing an estimate of trait heritability. 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 covariance between the two traits (15). This method can be applied to summary statistics, is not biased by sample overlap and does not require multiple traits to be measured for each individual.

By analysis of GWAS data for MM and CLL and applying cross-trait LD score regression we have been able to demonstrate a positive genetic correlation between CLL and MM. We find evidence of shared 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 datasets

The data from six previously reported MM GWAS (9-12) are summarized in **Supplementary Table 1.** All these studies were based on individuals of European ancestry and comprised: Oncoarray-GWAS (878 cases 7,054 controls) UK-GWAS (2,282 cases, 5,197 controls), Swedish-GWAS (1,714 cases, 10,391 controls), German-GWAS (1,508 cases, 2,107 controls), Netherlands-GWAS (555 cases, 2,669 controls) and US-GWAS (780 cases, 1,857 controls).

The data from three previously reported CLL GWAS (8-12) are summarized in **Supplementary Table 2.** All these studies were based on individuals of European ancestry and comprised: CLL UK1 (505 cases and 2,698 controls), CLL UK2 (1,236 cases and 2,501 controls) and CLL US (2,174 cases and 2,682 controls).

Ethics

Collection of patient samples and associated clinico-pathological 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-IX trial by the Medical Research Council (MRC) Leukaemia Data Monitoring and Ethics committee (MREC 02/8/95, ISRCTN68454111), 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, AFmu-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, PLCO: (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 UCSF2: 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 Leukaemia guidelines.

Quality control

Standard quality-control measures were applied to the GWAS (16). 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 (17) (phase 1 integrated release 3 March 2012) and UK10K (18). 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(19). 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 λ was based on the 90% least-significant SNPs and assessment of λ_{1000}. Details of SNP QC are provided in 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 dataset**.** Eigenvectors for the GWAS data sets were inferred using smartpca (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 (20). 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* (15). Using summary statistics from the GWAS meta-analysis we implemented filters as recommended by the authors (15). 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 (15).

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* (21). 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*. (22). 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 naïve B-cells was downloaded from Blueprint Epigenome Project (23). 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 median of the null

distribution. Thus, the enrichment score is the number of standard deviations of the AVS mapping tally from the median of the null distribution tallies. An enrichment plot for naïve B cells is shown in **Supplementary Figure 2**.

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* (24).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 project (23). In addition, we included in our analysis 4 MM cell lines-KMS11, JJN3, MM1-S and L363 processed as previously described (25). A heat map of results is shown in **Supplementary Figure 3**.

eQTL

eQTL analyses were performed using publicly available whole blood data downloaded from GTeX (26). The relationship between SNP genotype and gene expression we carried out using Summary-data-based Mendelian Randomization (SMR) analysis as per Zhu *et al* (27). 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_{zx} is the effect of *z* on *x*, and b_{zy} be the effect of *z* on *y*, b_{xy} (b_{zy}/b_{zx}) 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_{xy} 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_{xy} 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 $P_{SMR} < 2.5 \times 10^{-5}$ 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 Figure 4-5** and a summary of results are shown in **Supplementary Table 5.**

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, GSE2658 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 phs000207.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 Blueprint Epigenome Project [https://osf.io/u8tzp/]. ChIP-seq data for H3K27ac, H3K4Me1, H3K27Me3, H3K9Me3, H3K36Me3 and H3K27Me3 from naïve B-cells is publicly available and was obtained from Blueprint Epigenome Project [http://www.blueprint-epigenome.eu/].

RESULTS

Genetic correlation and heritability

We performed cross trait LD-score regression using summary statistics from two recent GWAS meta-analyses based on 7,717 MM cases and 21,587 controls and 4,017 CLL cases and 7,881 controls (**Figure 1, Supplementary Table 1-4**). While these datasets have been previously subject to quality control (QC) (5-7, 9-11) for the current analysis we implemented additional filtering steps as per Bulik-Sullivan *et al* (15), resulting in 1,055,728 harmonized SNPs between the two datasets. Heritability estimates from cross-trait LD score regression of 9.2 (±1.8%) and 22 (±5.9%) were comparable with previous estimates for MM (13) and CLL (6). LD-score regression revealed a significant positive genetic correlation between MM and CLL with an R_g value of 0.44 ($P = 4.6\times10^{-3}$).

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.* $P_{adj} < 0.05/45$). We then repeated the process, examining MM susceptibility SNPs in CLL, applying a significance level of $P_{adj} < 0.05/23$. Of the 45 CLL risk loci, four were associated with MM ($P_{adj} < 0.0011$) while, of 23 MM risk loci, five were significantly associated in CLL ($P_{adj} < 0.0022$) (**Table 1, Figure 2**). Correlated SNPs ($r^2 > 0.8$) at 3q26.2 are associated with both CLL and MM at genome-wide significance (**Figure 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 (24). Since 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 naïve 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 naïve B-cells (**Supplementary Figure 3**).

Most GWAS signals map to non-coding regions of the genome (28, 29) and influence gene expression through chromatin looping interactions (30, 31). 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 naïve 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 naïve B-cells from Blueprint (23). We also sought to gain insight into the possible biological mechanisms for associations by performing an expression quantitative trait locus (eQTL) analysis using mRNA 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 naïve 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 (32) and *FAS,* a member of the TNF-receptor superfamily. *FAS,* has a central role in regulating the immune response through apoptosis of B-cells (33, 34). At 2q31.1, looping interactions implicated transcription factor *SP3,* which has been shown to influence expression of

germinal centre genes, (35, 36). Variants at 6p25.3 reside in the 3'UTR of *IRF4*, which has an established role in B-cell regulation (37, 38) and MM oncogenesis (39, 40).

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 (41) and is necessary for replication fork restart (42). Variants in this locus demonstrated enrichment of H3K4me3 marks in two samples of naïve 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 (43, 44). Variant rs1317082 at 3q26.2 is located proximal to TERC, a gene which has been shown to influence telomere length (45). Additionally, we observed looping interactions to a number of genes at 3q26.2 including *SEC62*, which has been proposed as a cancer biomarker (46-49). Intriguingly, variants at 3q26.2 this locus have been implicated in colorectal (50), thyroid (51) and bladder (52) cancer.

Several genes were implicated at 22q13.33 by looping interactions for *SCO2*, *LMF2, ODF3B, TYMP/ECGF1, NCAPH2*, *SYCE3* and *ARSA,* with *TYMP/ECGF1* and *SCO2* demonstrating evidence of correlation in GWAS and eQTL effect size, albeit not significant after multiple testing (P_{SMR} = 2.38×10^{-4} and 3.19×10^{-4}). 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 (53, 54). 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(55).* Furthermore, *SCO2* (synthesis of cytochrome c oxidase), also mapping to this locus, has been implicated in the development of breast (56, 57), gastric (58) and leukemia (59), through glucose metabolism reprogramming (60), a hallmark of cancer (61). Tumour suppressor, p53, regulates metabolic pathways, p53-transactivated

381 TP53-induced glycolysis (TIGAR), and regulation of apoptosis in part through SCO2 (57, 58,
382 60).
383
384 Finally, while these data were indifferent to decipher 8q24.21, this locus has also been
385 shown to harbour risk SNPs for other cancers which localize within distinct LD blocks and
386 likely reflect tissue-specific effects on cancer risk through regulation of MYC (29).
387
388

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 naïve 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 (4). 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 aetiology 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.

ACKNOWLEDGMENTS

In the United Kingdom, Myeloma UK and Bloodwise (#05001, #06002 and #13044) provided principal funding. Additional funding was provided by Cancer Research UK (C1298/A8362 supported by the Bobby Moore Fund) and The Rosetrees Trust. M.W. is supported by funding from Mr Ralph Stockwell. A.S. is supported by a clinical fellowship from Cancer Research UK and the Royal Marsden Haematology Research Fund. This study made use of genotyping data on the 1958 Birth Cohort generated by the Wellcome Trust Sanger Institute (http://www.wtccc.org.uk). We are grateful to all investigators who contributed to NSCCG and GELCAPS, from which controls in the replication were drawn. We also thank the staff of the CTRU University of Leeds and the NCRI Haematology Clinical Studies Group. We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z) for the generation of UK myeloma Oncoarray data. The BCAC study would not have been possible without the contributions of the following: Manjeet K. Bolla, Qin Wang, Kyriaki Michailidou and Joe Dennis. BCAC is funded by Cancer Research UK (C1287/A10118, C1287/A16563). For the BBCS study, we thank Eileen Williams, Elaine Ryder-Mills and Kara Sargus. The BBCS is funded by Cancer Research UK and Breast Cancer Now and acknowledges NHS funding to the National Institute of Health Research (NIHR) Biomedical Research Centre (BRC), and the National Cancer Research Network (NCRN). We thank the participants and the investigators of EPIC (European Prospective Investigation into Cancer and Nutrition). The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by: Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF) (Germany); the Hellenic Health Foundation, the Stavros Niarchos Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), PI13/00061 to Granada, PI13/01162 to EPIC-

445 Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra,
446 ISCIII RETIC (RD06/0020) (Spain); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491
447 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk,
448 MR/M012190/1 to EPIC-Oxford) (UK). We thank the SEARCH and EPIC teams which were
449 funded by a programme grant from Cancer Research UK [A16561] and supported by the UK
450 NIHR BRC at the University of Cambridge. We thank Breast Cancer Now and The Institute of
451 Cancer Research (ICR) for support and funding of the UKBGS, and the study participants,
452 study staff, and the doctors, nurses and other health care providers and health information
453 sources who have contributed to the study. We acknowledge NHS funding to the Royal
454 Marsden/ICR NIHR BRC. UKGPCS would like to thank The Institute of Cancer Research and
455 The Everyman Campaign for funding support. The UKGPCS acknowledges The Prostate
456 Cancer Research Foundation, Prostate Action, The Orchid Cancer Appeal, The National
457 Cancer Research Network UK, The National Cancer Research Institute (NCRI), the NIHR
458 funding to the NIHR Biomedical Research data managers and consultants for their work in
459 the UKGPCS study and urologists and other persons involved in the planning, and data
460 collection of the CAPS study. Genotyping of the OncoArray was funded by the US National
461 Institutes of Health (NIH) [U19 CA 148537 for ELucidating Loci Involved in Prostate cancer
462 SuscEptibility (ELLIPSE) project and X01HG007492 to the Center for Inherited Disease
463 Research (CIDR) under contract number HHSN268201200008I]. Additional analytic support
464 was provided by NIH NCI U01 CA188392 (PI: Schumacher). The PRACTICAL consortium was
465 supported by Cancer Research UK Grants C5047/A7357, C1287/A10118, C1287/A16563,
466 C5047/A3354, C5047/A10692, C16913/A6135, European Commission's Seventh Framework
467 Programme grant agreement n° 223175 (HEALTH-F2-2009-223175), and The National
468 Institute of Health (NIH) Cancer Post-Cancer GWAS initiative grant: No. 1 U19 CA 148537-01
469 (the GAME-ON initiative). We would also like to thank the following for funding support: The
470 Institute of Cancer Research and The Everyman Campaign, The Prostate Cancer Research
471 Foundation, Prostate Research Campaign UK (now Prostate Action), The Orchid Cancer
472 Appeal, The National Cancer Research Network UK, The National Cancer Research Institute
473 (NCRI) UK. We are grateful for support of NIHR funding to the NIHR Biomedical Research
474 Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust.
475 The APBC BioResource, which forms part of the PRACTICAL consortium, consists of the
476 following members: Wayne Tilley, Gail Risbridger, Renea Taylor, Judith A Clements, Lisa

Horvath, Vanessa Hayes, Lisa Butler, Trina Yeadon, Allison Eckert, Pamela Saunders, Anne-Maree 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 (110,131), the Federal Ministry of Education and Research (CLIOMMICS 01ZX1309), The German Research Council (DFG; Project SI 236/81, SI 236/)-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 GMMG and DSMM studies. The German GWAS made use of genotyping data from the population-based HNR study, which is supported by the Heinz Nixdorf Foundation (Germany). The genotyping of the Illumina HumanOmni-1 Quad BeadChips of the HNR subjects was financed 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 (KF10-0009), 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, ALF grants to the University and Regional Laboratories (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 Jörgen Adolfsson, Tomas Axelsson, Anna Collin, Ildikó Frigyesi, Patrik Magnusson, Bertil Johansson, Jan Westin and Helga Ögmundsdóttir for their assistance. This work was supported by Center for Translational Molecular Medicine (BioCHIP), a clinical research grant from the European Hematology Association, an EMCR Translational Research Grant, a BMBF grant from CLIOMMICS (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 P20GM125503. We thank the staff of the HOVON, 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 and other individuals who

participated in the project. The Genome-Wide Association Study (GWAS) of Non-Hodgkin Lymphoma (NHL) project, from which US-CLL samples were obtained, was supported by the intramural program of the Division of Cancer Epidemiology and Genetics (DCEG), National Cancer Institute (NCI), National Institutes of Health (NIH). The datasets have been accessed through the NIH database for Genotypes and Phenotypes (dbGaP) under accession # phs000801. A full list of acknowledgements can be found in the supplementary note (Berndt SI et al., Nature Genet., 2013, PMID: 23770605).

AUTHOR CONTRIBUTIONS

M.W., A.S. and R.S.H. designed the study. M.W. and R.S.H. drafted the manuscript with contributions from A.S. In the UK MM study: M.W. performed principal statistical and bioinformatics analyses. A.S., N.L., J.S.M. and G.O. performed additional bioinformatics analyses. P.B. coordinated UK laboratory analyses. A.H. performed sequencing of UK MM and CLL samples. F.M maintained and prepared MM cell lines for ChIP-seq. D.C.J. managed and prepared UK and Oncoarray MM Case Study DNA samples. M.K., G.J.M., F.E.D., W.A.G. and G.H.J. performed ascertainment and collection of Case Study samples in the US and UK. In the German MM study: H.G., U.B., J.H., J.N., and N.W. coordinated and managed Heidelberg samples. C.L. and H.E. coordinated and managed Ulm/Wurzburg samples. A.F. coordinated German genotyping. C.C. and O.R.B performed German genotyping. P.H. and M.M.N. performed GWAS of German cases and controls. H.T., B.C. and M.I.d.S.F. carried out statistical analysis. K.H. coordinated the German part of the project. M.M.N. generated genotype data from the Heinz-Nixdorf recall study. K-H.J., contributed towards the Heinz-Nixdorf control data set. N.N. from Bonn and K-H.J. provided samples for the German GWAS. In the Swedish MM study: M.H. and B.N. coordinated the Swedish/Norwegian part of the project. M.A., E.J., A.-K.W., U.-H.M., H.N., A.V., N.F.A., A.W., I.T. and U.G. performed sample acquisition, sample preparation, clinical data acquisition and data analyses of Sweden/Norway samples. In the Dutch MM study: M.v.D., P.S., A.B. and R.K. coordinated and prepared HOVON65/GMMG-HD4, HOVON87/NMSG18 and HOVON95/EMN02 studies for participating in this study, and coordinated genotyping and pre-processing. In the American MM study: N.W. coordinated the US part of the project and performed statistical analyses. O.W.S. and N.W managed Case Study samples. G.J.M. and F.E.D. performed

ascertainment and collection of Case Study samples. In the UK CLL studies: At the ICR, P.J.L., G.P.S. and H.E.S. performed bioinformatic and statistical analyses. H.E.S. performed CLL project management and supervised genotyping of CLL samples. C.D. and D.C. performed recruitment of samples. In Newcastle, J.M.A. and D.J.A. conceived of the NCLLC; J.M.A. obtained financial support, supervised laboratory management and oversaw genotyping of cases with NCLLC; N.J.S. performed sample management of cases. D.J.A., J.R.B., G.P., C.P. and C.F. developed protocols for recruitment of individuals with CLL and sample acquisition and performed sample collection of cases. In Leicester, M.J.S.D. performed overall management, collection and processing of samples; and S.J. and A.M. performed DNA extractions. In Sweden, L.M. and R.R. performed collection of CLL cases. G.J. and K.E.S. performed sample collection in the Scandinavian Lymphoma Etiology (SCALE) study.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

559 **REFERENCES**

560

561 1. Barlogie B, Gale RP. Multiple myeloma and chronic lymphocytic leukemia: parallels
562 and contrasts. The American journal of medicine. 1992;93(4):443-50.
563 2. Shaffer AL, Rosenwald A, Staudt LM. Lymphoid malignancies: the dark side of B-cell
564 differentiation. Nature reviews Immunology. 2002;2(12):920-32.
565 3. Kuppers R. Mechanisms of B-cell lymphoma pathogenesis. Nature reviews Cancer.
566 2005;5(4):251-62.
567 4. Sud A, Chattopadhyay S, Thomsen H, Sundquist K, Sundquist J, Houlston RS, et al.
568 The landscape of familial risk of hematological malignancies: an analysis of 153,115 cases.
569 (Under review).
570 5. Crowther-Swanepoel D, Broderick P, Di Bernardo MC, Dobbins SE, Torres M,
571 Mansouri M, et al. Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence
572 chronic lymphocytic leukemia risk. Nature genetics. 2010;42(2):132-6.
573 6. Law PJ, Berndt SI, Speedy HE, Camp NJ, Sava GP, Skibola CF, et al. Genome-wide
574 association analysis implicates dysregulation of immunity genes in chronic lymphocytic
575 leukaemia. 2017;8:14175.
576 7. Speedy HE, Di Bernardo MC, Sava GP, Dyer MJ, Holroyd A, Wang Y, et al. A genome-
577 wide association study identifies multiple susceptibility loci for chronic lymphocytic
578 leukemia. Nature genetics. 2014;46(1):56-60.
579 8. Berndt SI, Skibola Cf Fau - Joseph V, Joseph V Fau - Camp NJ, Camp Nj Fau - Nieters
580 A, Nieters A Fau - Wang Z, Wang Z Fau - Cozen W, et al. Genome-wide association study
581 identifies multiple risk loci for chronic lymphocytic leukemia. (1546-1718 (Electronic)).
582 9. Mitchell JS, Li N, Weinhold N, Forsti A, Ali M, van Duin M, et al. Genome-wide
583 association study identifies multiple susceptibility loci for multiple myeloma. Nature
584 communications. 2016;7:12050.
585 10. Broderick P, Chubb D, Johnson DC, Weinhold N, Forsti A, Lloyd A, et al. Common
586 variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. Nature genetics.
587 2011;44(1):58-61.
588 11. Chubb D, Weinhold N, Broderick P, Chen B, Johnson DC, Forsti A, et al. Common
589 variation at 3q26.2, 6p21.33, 17p11.2 and 22q13.1 influences multiple myeloma risk. Nature
590 genetics. 2013;45(10):1221-5.
591 12. Swaminathan B, Thorleifsson G, Joud M, Ali M, Johnsson E, Ajore R, et al. Variants in
592 ELL2 influencing immunoglobulin levels associate with multiple myeloma. Nature
593 communications. 2015;6:7213.
594 13. Mitchell JS, Johnson DC, Litchfield K, Broderick P, Weinhold N, Davies FE, et al.
595 Implementation of genome-wide complex trait analysis to quantify the heritability in
596 multiple myeloma. Scientific reports. 2015;5:12473.
597 14. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J. LD Score regression
598 distinguishes confounding from polygenicity in genome-wide association studies.
599 2015;47(3):291-5.
600 15. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR. An atlas of genetic
601 correlations across human diseases and traits. 2015;47(11):1236-41.
602 16. Turner S, Armstrong LL, Bradford Y, Carlson CS, Crawford DC, Crenshaw AT, et al.
603 Quality control procedures for genome-wide association studies. Current protocols in
604 human genetics. 2011;Chapter 1:Unit1.19.

605 17. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A
606 map of human genome variation from population-scale sequencing. Nature.
607 2010;467(7319):1061-73.
608 18. Huang J, Howie B, McCarthy S, Memari Y, Walter K, Min JL, et al. Improved
609 imputation of low-frequency and rare variants using the UK10K haplotype reference panel.
610 Nature communications. 2015;6:8111.
611 19. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for
612 genome-wide association studies by imputation of genotypes. Nature genetics.
613 2007;39(7):906-13.
614 20. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, et al. Meta-analysis
615 and imputation refines the association of 15q25 with smoking quantity. Nature genetics.
616 2010;42(5):436-40.
617 21. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, et al. Partitioning
618 heritability by functional annotation using genome-wide association summary statistics.
619 2015;47(11):1228-35.
620 22. Cowper-Sal lari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eeckhoute J, et al. Breast
621 cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene
622 expression. Nature genetics. 2012;44(11):1191-8.
623 23. Fernandez JM, de la Torre V, Richardson D, Royo R, Puiggros M, Moncunill V, et al.
624 The BLUEPRINT Data Analysis Portal. Cell systems. 2016;3(5):491-5.e5.
625 24. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, et al. Chromatin marks identify
626 critical cell types for fine mapping complex trait variants. Nature genetics. 2013;45(2):124-
627 30.
628 25. Li N, Johnson DC, Weinhold N, Kimber S, Dobbins SE, Mitchell JS, et al. Genetic
629 Predisposition to Multiple Myeloma at 5q15 Is Mediated by an ELL2 Enhancer
630 Polymorphism. Cell reports. 2017;20(11):2556-64.
631 26. The Genotype-Tissue Expression (GTEx) project. Nature genetics. 2013;45(6):580-5.
632 27. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of
633 summary data from GWAS and eQTL studies predicts complex trait gene targets. Nature
634 genetics. 2016;48(5):481-7.
635 28. Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, et al.
636 Principles for the post-GWAS functional characterization of cancer risk loci. Nature genetics.
637 2011;43(6):513-8.
638 29. Sud A, Kinnersley B, Houlston RS. Genome-wide association studies of cancer:
639 current insights and future perspectives. Nature reviews Cancer. 2017;17(11):692-704.
640 30. Orlando G, Kinnersley B, Houlston RS. Capture Hi-C Library Generation and Analysis
641 to Detect Chromatin Interactions. Current protocols in human genetics. 2018:e63.
642 31. Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, et al.
643 Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C.
644 Nature genetics. 2015;47:598.
645 32. Guo DC, Pannu H, Tran-Fadulu V, Papke CL, Yu RK, Avidan N, et al. Mutations in
646 smooth muscle alpha-actin (ACTA2) lead to thoracic aortic aneurysms and dissections.
647 Nature genetics. 2007;39(12):1488-93.
648 33. Akagi T, Yoshino T, Kondo E. The Fas antigen and Fas-mediated apoptosis in B-cell
649 differentiation. Leukemia & lymphoma. 1998;28(5-6):483-9.

650 34. Figgett WA, Fairfax K, Vincent FB, Le Page MA, Katik I, Deliyanti D, et al. The TACI
651 receptor regulates T-cell-independent marginal zone B cell responses through innate
652 activation-induced cell death. Immunity. 2013;39(3):573-83.
653 35. Park SR, Zan H, Pal Z, Zhang J, Al-Qahtani A, Pone EJ, et al. HoxC4 binds to the
654 promoter of the cytidine deaminase AID gene to induce AID expression, class-switch DNA
655 recombination and somatic hypermutation. Nature immunology. 2009;10(5):540-50.
656 36. Steinke JW, Hodsdon W, Parenti S, Ostraat R, Lutz R, Borish L, et al. Identification of
657 an Sp factor-dependent promoter in GCET, a gene expressed at high levels in germinal
658 center B cells. Molecular immunology. 2004;41(12):1145-53.
659 37. Willis SN, Good-Jacobson KL, Curtis J, Light A, Tellier J, Shi W, et al. Transcription
660 factor IRF4 regulates germinal center cell formation through a B cell-intrinsic mechanism.
661 Journal of immunology (Baltimore, Md : 1950). 2014;192(7):3200-6.
662 38. Nutt SL, Hodgkin PD, Tarlinton DM, Corcoran LM. The generation of antibody-
663 secreting plasma cells. Nature reviews Immunology. 2015;15(3):160-71.
664 39. Zhang S, Xu J, Wu S, Wang R, Qu X, Yu W, et al. IRF4 promotes cell proliferation by
665 JNK pathway in multiple myeloma. Medical oncology (Northwood, London, England).
666 2013;30(2):594.
667 40. Ohguchi H, Hideshima T, Bhasin MK, Gorgun GT, Santo L, Cea M. The KDM3A-KLF2-
668 IRF4 axis maintains myeloma cell survival. 2016;7:10258.
669 41. Inano S, Sato K, Katsuki Y, Kobayashi W, Tanaka H, Nakajima K, et al. RFWD3-
670 Mediated Ubiquitination Promotes Timely Removal of Both RPA and RAD51 from DNA
671 Damage Sites to Facilitate Homologous Recombination. Molecular cell. 2017;66(5):622-
672 34.e8.
673 42. Elia AE, Wang DC, Willis NA, Boardman AP, Hajdu I, Adeyemi RO, et al. RFWD3-
674 Dependent Ubiquitination of RPA Regulates Repair at Stalled Replication Forks. Molecular
675 cell. 2015;60(2):280-93.
676 43. Rice C, Shastrula PK, Kossenkov AV, Hills R, Baird DM, Showe LC, et al. Structural and
677 functional analysis of the human POT1-TPP1 telomeric complex. Nature communications.
678 2017;8:14928.
679 44. Pinzaru AM, Hom RA, Beal A, Phillips AF, Ni E, Cardozo T, et al. Telomere Replication
680 Stress Induced by POT1 Inactivation Accelerates Tumorigenesis. Cell reports.
681 2016;15(10):2170-84.
682 45. Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL, et al. Identification
683 of seven loci affecting mean telomere length and their association with disease. Nature
684 genetics. 2013;45:422.
685 46. Jung V, Kindich R, Kamradt J, Jung M, Muller M, Schulz WA, et al. Genomic and
686 expression analysis of the 3q25-q26 amplification unit reveals TLOC1/SEC62 as a probable
687 target gene in prostate cancer. Molecular cancer research : MCR. 2006;4(3):169-76.
688 47. Linxweiler M, Schick B, Zimmermann R. Let's talk about Secs: Sec61, Sec62 and Sec63
689 in signal transduction, oncology and personalized medicine. Signal transduction and
690 targeted therapy. 2017;2:17002.
691 48. Bergmann TJ, Fumagalli F, Loi M, Molinari M. Role of SEC62 in ER maintenance: A link
692 with ER stress tolerance in SEC62-overexpressing tumors? Molecular & cellular oncology.
693 2017;4(2):e1264351.
694 49. Greiner M, Kreutzer B, Lang S, Jung V, Cavalie A, Unteregger G, et al. Sec62 protein
695 level is crucial for the ER stress tolerance of prostate cancer. The Prostate.
696 2011;71(10):1074-83.

50. Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, Howarth K, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. Nature genetics. 2010;42(11):973-7.
51. Gudmundsson J, Thorleifsson G, Sigurdsson JK, Stefansdottir L, Jonasson JG, Gudjonsson SA, et al. A genome-wide association study yields five novel thyroid cancer risk loci. 2017;8:14517.
52. Figueroa JD, Ye Y, Siddiq A, Garcia-Closas M, Chatterjee N, Prokunina-Olsson L, et al. Genome-wide association study identifies multiple loci associated with bladder cancer risk. Human molecular genetics. 2014;23(5):1387-98.
53. Deves C, Rostirolla DC, Martinelli LK, Bizarro CV, Santos DS, Basso LA. The kinetic mechanism of Human Thymidine Phosphorylase - a molecular target for cancer drug development. Molecular bioSystems. 2014;10(3):592-604.
54. Bijnsdorp IV, Capriotti F, Kruyt FA, Losekoot N, Fukushima M, Griffioen AW, et al. Thymidine phosphorylase in cancer cells stimulates human endothelial cell migration and invasion by the secretion of angiogenic factors. British journal of cancer. 2011;104(7):1185-92.
55. Liu H, Liu Z, Du J, He J, Lin P, Amini B, et al. Thymidine phosphorylase exerts complex effects on bone resorption and formation in myeloma. Science translational medicine. 2016;8(353):353ra113.
56. Liu F, Zhang W, You X, Liu Y, Li Y, Wang Z, et al. The oncoprotein HBXIP promotes glucose metabolism reprogramming via downregulating SCO2 and PDHA1 in breast cancer. Oncotarget. 2015;6(29):27199-213.
57. Won KY, Lim SJ, Kim GY, Kim YW, Han SA, Song JY, et al. Regulatory role of p53 in cancer metabolism via SCO2 and TIGAR in human breast cancer. Human pathology. 2012;43(2):221-8.
58. Kim SH, Choi SI, Won KY, Lim SJ. Distinctive interrelation of p53 with SCO2, COX, and TIGAR in human gastric cancer. Pathology, research and practice. 2016;212(10):904-10.
59. Papadopoulou LC, Kyriazou AV, Bonovolias ID, Tsiftsoglou AS. Imatinib inhibits the expression of SCO2 and FRATAXIN genes that encode mitochondrial proteins in human Bcr-Abl(+) leukemia cells. Blood cells, molecules & diseases. 2014;53(1-2):84-90.
60. Nath A, Chan C. Genetic alterations in fatty acid transport and metabolism genes are associated with metastatic progression and poor prognosis of human cancers. Scientific reports. 2016;6:18669.
61. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.

TABLE AND FIGURE LEGENDS

Table 1: Risk loci demonstrating association of alleles at respective loci in both chronic lymphocytic leukaemia (CLL) and multiple myeloma (MM). – indicates SNP not present in filtered data.

Figure 1: Schematic outlining the processing of datasets used in the genetic correlation.

Figure 2: Overlap of loci in multiple myeloma and chronic lymphocytic leukaemia. *correlated variants at 3q26.2 had been previously published as genome wide significant in each dataset prior to this analysis.

749 **TABLES**

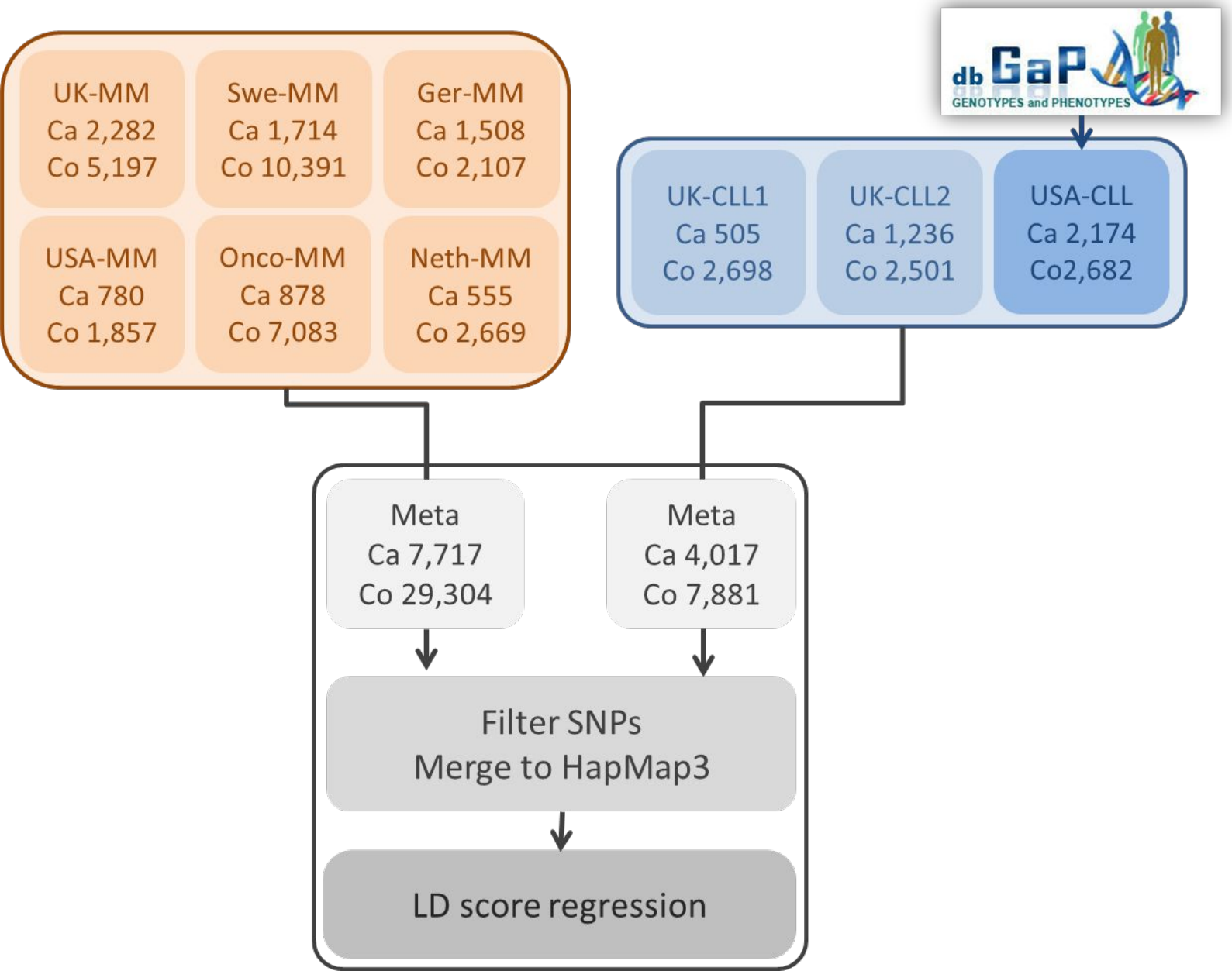
750

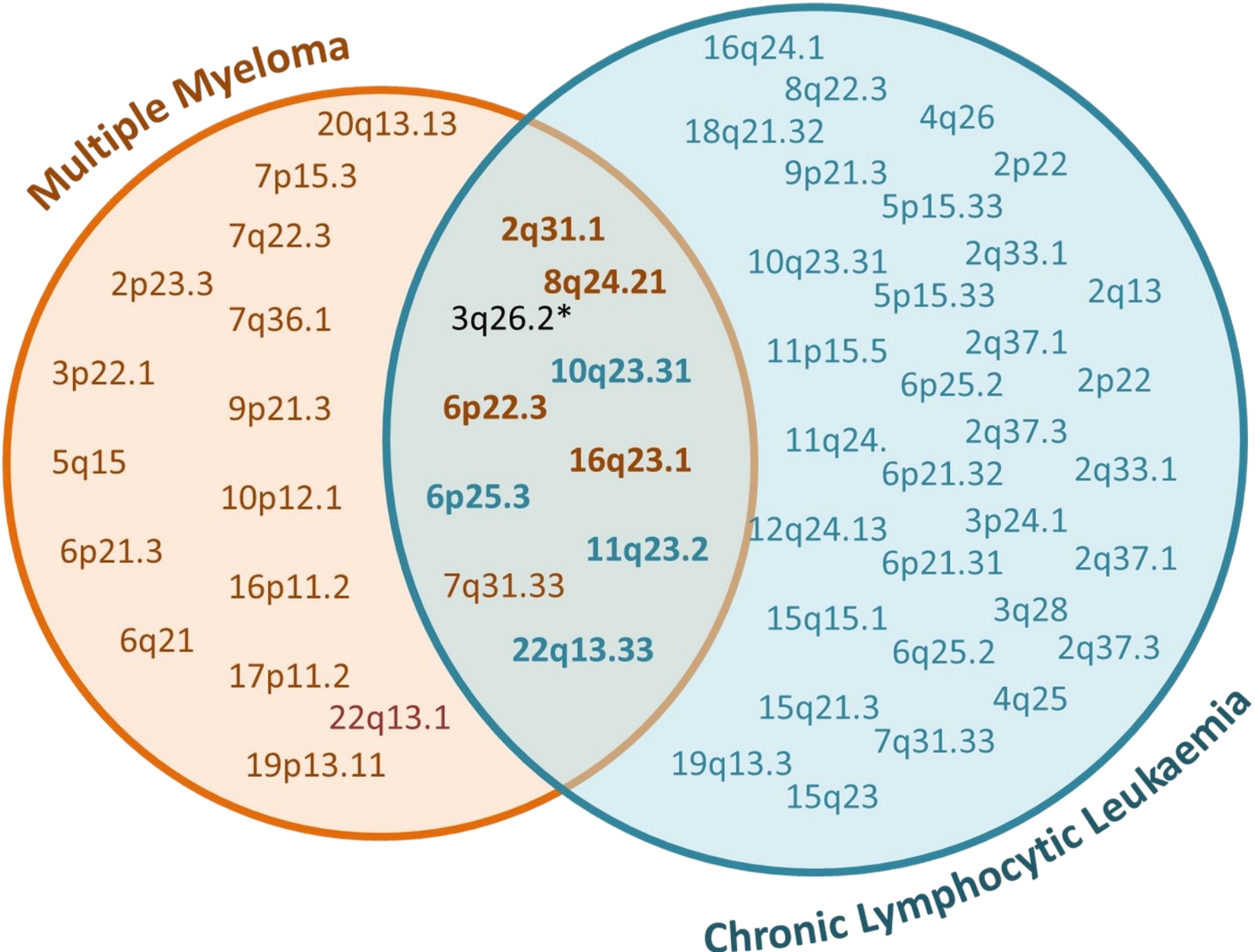
Locus	Discovery GWAS	Sentinel Variant	Correlated Variant	Position (hg19)	Risk allele		Odds Ratio		*P*	
					CLL	MM	CLL	MM	CLL	MM
2q31.1	MM	rs4325816		174,808,899	T	T	1.11	1.12	2.0×10^{-3}	6.4×10^{-7}
			rs72919402	174,750,200	T	-	1.13	-	4.6×10^{-4}	-
3q26.2	MM & CLL	rs1317082		169,497,585	A	A	1.20	1.19	7.1×10^{-8}	2.2×10^{-16}
			rs3821383	169,489,946	A	A	1.20	1.18	4.2×10^{-8}	4.5×10^{-15}
6p25.3	CLL	rs872071		411,064	G	G	1.37	1.10	2.8×10^{-27}	7.5×10^{-7}
			rs1050976	408,079	T	T	1.37	1.10	1.9×10^{-27}	3.7×10^{-7}
6p22.3	MM	rs34229995		15,244,018	G	G	1.37	1.36	8.5×10^{-3}	5.6×10^{-8}
			rs13197919	15,282,334	T	T	1.35	1.32	1.3×10^{-3}	3.42×10^{-7}
7q31.33	MM	rs58618031		124,583,896	T	T	1.15	1.11	3.2×10^{-5}	1.7×10^{-7}
			rs59294613	124,554,267	C	-	1.16	-	4.4×10^{-6}	-
8q24.21	MM	rs1948915		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		rs140522		50,971,266	T	T	1.17	1.08	3.7×10^{-7}	1.2×10^{-4}
			-	-	-	-	-	-	-	-

751

752 **Table 1: Risk loci demonstrating enrichment of alleles at respective loci in both chronic**
753 **lymphocytic leukaemia (CLL) and multiple myeloma (MM). – indicates SNP not present in**
754 **filtered data.**

755


UK-MM
Ca 2,282
Co 5,197
Swe-MM
Ca 1,714
Co 10,391
Ger-MM
Ca 1,508
Co 2,107
USA-MM
Ca 780
Co 1,857
Onco-MM
Ca 878
Co 7,083
Neth-MM
Ca 555
Co 2,669
db GaP
GENOTYPES and PHENOTYPES
UK-CLL1
Ca 505
Co 2,698
UK-CLL2
Ca 1,236
Co 2,501
USA-CLL
Ca 2,174
Co2,682
Meta
Ca 7,717
Co 29,304
Meta
Ca 4,017
Co 7,881
Filter SNPs
Merge to HapMap3
LD score regression


Multiple Myeloma
20q13.13
7p15.3
7q22.3
2p23.3
7q36.1
3p22.1
9p21.3
5q15
10p12.1
6p21.3
16p11.2
6q21
17p11.2
22q13.1
19p13.11
2q31.1
8q24.21
3q26.2*
10q23.31
6p22.3
16q23.1
6p25.3
11q23.2
7q31.33
22q13.33
16q24.1
8q22.3
4q26
18q21.32
9p21.3
2p22
5p15.33
10q23.31
2q33.1
5p15.33
2q13
11p15.5
2q37.1
6p25.2
2p22
11q24.
2q37.3
6p21.32
2q33.1
12q24.13
3p24.1
6p21.31
2q37.1
15q15.1
3q28
6q25.2
2q37.3
15q21.3
4q25
7q31.33
19q13.3
15q23
Chronic Lymphocytic Leukaemia