

PRIMARY RESEARCH

Open Access



# Transcriptome-wide association study of multiple myeloma identifies candidate susceptibility genes

Molly Went<sup>1,2\*</sup> , Ben Kinnersley<sup>1</sup>, Amit Sud<sup>1</sup>, David C. Johnson<sup>2</sup>, Niels Weinhold<sup>3</sup>, Asta Försti<sup>4</sup>, Mark van Duin<sup>5</sup>, Giulia Orlando<sup>1</sup>, Jonathan S. Mitchell<sup>1</sup>, Rowan Kuiper<sup>5</sup>, Brian A. Walker<sup>6</sup>, Walter M. Gregory<sup>7</sup>, Per Hoffmann<sup>8,9</sup>, Graham H. Jackson<sup>10</sup>, Markus M. Nöthen<sup>8,11</sup>, Miguel Inacio da Silva Filho<sup>4</sup>, Hauke Thomsen<sup>4</sup>, Annemiek Broyl<sup>5</sup>, Faith E. Davies<sup>6</sup>, Unnur Thorsteinsdottir<sup>12</sup>, Markus Hansson<sup>13,14</sup>, Martin Kaiser<sup>2</sup>, Pieter Sonneveld<sup>6</sup>, Hartmut Goldschmidt<sup>3,8</sup>, Kari Stefansson<sup>12</sup>, Kari Hemminki<sup>4</sup>, Björn Nilsson<sup>14,15</sup>, Gareth J. Morgan<sup>6</sup> and Richard S. Houlston<sup>1</sup>

## Abstract

**Background:** While genome-wide association studies (GWAS) of multiple myeloma (MM) have identified variants at 23 regions influencing risk, the genes underlying these associations are largely unknown. To identify candidate causal genes at these regions and search for novel risk regions, we performed a multi-tissue transcriptome-wide association study (TWAS).

**Results:** GWAS data on 7319 MM cases and 234,385 controls was integrated with Genotype-Tissue Expression Project (GTEx) data assayed in 48 tissues (sample sizes,  $N = 80-491$ ), including lymphocyte cell lines and whole blood, to predict gene expression. We identified 108 genes at 13 independent regions associated with MM risk, all of which were in 1 Mb of known MM GWAS risk variants. Of these, 94 genes, located in eight regions, had not previously been considered as a candidate gene for that locus.

**Conclusions:** Our findings highlight the value of leveraging expression data from multiple tissues to identify candidate genes responsible for GWAS associations which provide insight into MM tumorigenesis. Among the genes identified, a number have plausible roles in MM biology, notably *APOBEC3C*, *APOBEC3H*, *APOBEC3D*, *APOBEC3F*, *APOBEC3G*, or have been previously implicated in other malignancies. The genes identified in this TWAS can be explored for follow-up and validation to further understand their role in MM biology.

**Keywords:** Genome-wide association study, Gene expression, Multiple myeloma, Transcriptome-wide association study

## Background

Multiple myeloma (MM) is the second most common hematologic malignancy in economically developed countries, and despite improvements in therapy, the disease essentially remains incurable. The aetiology of MM is poorly understood; however, the two- to four-fold increased risk of MM in relatives of patients has provided evidence for

an inherited basis [1]. Direct evidence for inherited genetic susceptibility is provided by genome-wide association studies (GWAS), which have so far discovered 23 genomic regions harbouring risk variants for MM [2].

Consistent with findings from many different cancer GWAS, bar a few notable exceptions, the functional variants and target susceptibility genes at the MM risk regions are yet to be identified. Knowledge of the causal genes responsible for defining disease predisposition is important in furthering our understanding of MM tumorigenesis and has the potential to inform the development of novel therapeutic strategies [3]. While most GWAS risk variants map to non-coding regions of the

\* Correspondence: [molly.went@icr.ac.uk](mailto:molly.went@icr.ac.uk)

<sup>1</sup>Division of Genetics and Epidemiology, The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK

<sup>2</sup>Division of Molecular Pathology, The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK

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



genome, they are enriched for variants correlated with gene expression levels [4, 5]. Exploiting this characteristic, the integration of GWAS signals with expression quantitative trait loci (eQTLs) has implicated *ELL2* and *CDCA7L* as the risk genes likely to be responsible for the 5q15 and 7p15.3 MM associations, respectively [6–9]. The high frequency of eQTLs coupled with linkage disequilibrium (LD) across regions can, however, make disentangling the risk genes from spurious co-localization at the same region problematic.

Transcriptome-wide association studies (TWAS) have been proposed as a strategy to identify risk genes underlying complex traits [10]. This approach imputes genetic data from GWAS using reference sets of weights generated from eQTL data, before correlating this genetic component of gene expression with the phenotype of interest. Since TWAS aggregates the effects of multiple variants into a single testing unit, and facilitates prioritisation of genes at known risk regions for functional validation, it potentially also affords increased study power to identify new risk regions.

While MM is caused by the clonal expansion of malignant plasma cells, if a TWAS is to be based on expression data from a single cell deciding on the most appropriate source is inherently problematic [11]. Utilising eQTL data from tumours is complicated by copy number alterations and essentially represents terminal stage in disease progression. Moreover, the effect of any risk allele may be acting at the level of the tumour micro-environment [12]. Studies have shown that eQTLs strongly enriched in GWAS signals are not necessarily specific to the eQTL discovery tissue [5]. Taking advantage of this principle allows a multi-tissue TWAS to be conducted integrating expression across multiple tissues, thereby leveraging information on shared eQTLs for candidate gene discovery [13].

Herein, we report a multi-tissue TWAS to prioritise candidate causal genes at known risk regions for MM and search for new risk regions. Specifically, we have analysed gene expression data from 48 tissue panels measured in 8756 individuals in conjunction with summary association statistics on 7319 MM cases and 234,385 controls of European descent. We identify 108 genes at 13 loci associated with MM risk and provide additional evidence of a potential role for a number of genes dysregulated in MM tumorigenesis.

## Results

We evaluated the association between predicted gene expression levels and MM risk using MetaXcan with summary statistics for GWAS SNPs in 7319 MM cases and 234,385 controls. In total, the expression levels of 25,520 genes across 48 tissues were tested for an association with MM risk. Quantile-quantile plots of TWAS

association statistics did not show evidence of systematic inflation (Additional file 1: Figure S1). Figure 1 shows Manhattan plots for respective GWAS and TWAS associations.

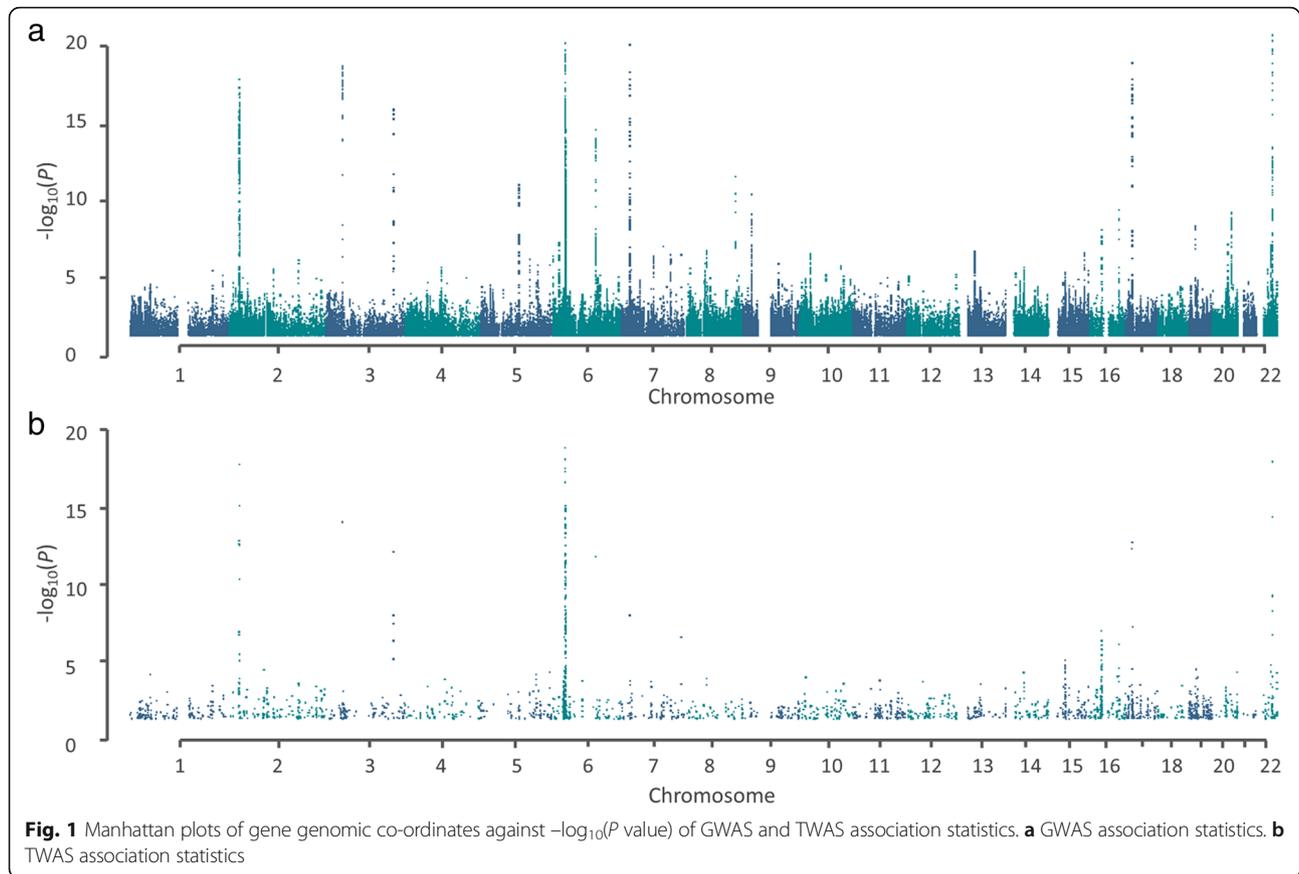
Applying a Bonferroni threshold, we identified 108 genes at 13 independent regions associated with MM (Table 1, Additional file 1: Table S1). All identified genes except those localising to the HLA region on chromosome 6p21 were within 1 Mb of previously reported MM risk SNPs. For all loci, except those in the HLA region, association signals were abrogated after adjusting for the top risk SNP, consistent with variation in expression of the identified gene being functionally related to the MM risk association. The complex LD patterns within the HLA region make deconvolution of significant results within the region difficult [14, 15]; therefore, our principal focus was confined to 31 genes at 12 loci outside 6p21.

For many loci, our TWAS findings support the involvement of a number of genes that have previously been implicated in defining MM [2, 16–19]. Specifically, single-gene associations were identified at 3p22.1 (*ULK4*), 6q21 (*ATG5*), 7p15.3 (*CDCA7L*), 7q36.1 (*CHPF2*) and 16q23.1 (*RFWD3*). However, at a number of regions, our analysis identified multiple significant genes, notably, 2p23.3 (*KIF3C*, *EPT1*, *CENPO*, *DTNB*, *DNM3TA*, *PTGES3P2*, *DNAJC27*), 3q26.2 (*MYNN*, *LRR34*, *LRR1Q4*, *ACTRT3*), 16p11.2 (*QPRT*, *RNF40*, *PRR14*, *C16orf93*, *RP11-2C24.5*, *PRSS53*) and 17p11.2 (*TBC1D27*, *USP32P1*, *PEMT*). A complete list of novel genes identified at known GWAS risk loci is provided in Additional file 1: Table S2.

Interestingly, several of the *APOBEC* genes were identified at 22q13.1. These genes localise within a distinct LD block adjacent to the one to which the sentinel GWAS risk SNPs maps (Fig. 2). We sought to gain insight into the potential for genome-wide significant SNPs in 22q13.1 in to influence regulation via a *cis*-regulatory enhancer, by mapping looping interaction and histone modifications in the lymphoblastoid cell line GM12878, which was chosen as a model for early B cell differentiation, with negligible genetic and phenotypic abnormalities [20]. We found evidence of enhancer marks and looping interactions from SNPs in 22q13.1 to *APOBEC* genes (Fig. 2), highlighting active chromatin and spatial proximity present in this region, necessary to mediate gene expression [21]. No significant genes were identified at 12 reported MM risk regions (2q31.1, 5q15, 5q23.2, 6p22.3, 7q22.3, 7q31.33, 8q24.21, 9p21.3, 10p12.1, 17p11.2, 19p13.1, 20q13.1).

## Discussion

In this large TWAS involving 7319 MM cases of European ancestry, we identified genetically predicted expression levels in 108 genes associated with MM risk. Of these,



there were 94 genes located in eight regions that, although mapping within 1 Mb of a MM risk locus had not previously been considered as a candidate gene for that locus.

Our findings provide further support for a number of the genes previously implicated by GWAS whose expression influences the risk of developing MM, including *CDCA7L* at 7p15.3, which has been functionally validated. At 7p15.3, rs4487645 resides in an enhancer of *c-Myc*-interacting *CDCA7L* and increases *IRF4* binding, affecting MM proliferation [7]. Furthermore, *ULK4* at 3p22.1, *ATG5* at 6q21 and *RFWD3* at 16q23 have been identified here and implicated previously. Additionally, our TWAS implicates new genes at known risk regions, notably *APOBEC3C*, *APOBEC3D*, *APOBEC3F*, *APOBEC3G* and *APOBEC3H* at 22q13.1 as playing a role in defining MM predisposition. Aberrant *APOBEC* cytidine deaminase activity has been shown to correlate with an increased mutational burden and is a recognised feature of MM, caused by triggering DNA mutation through dC deamination [22–24]. Furthermore, *KIF3C*, identified at 2p23.3, is a gene which regulates microtubule dynamics and has been previously implicated in breast cancer [25, 26]. Also at 2p23.3, this analysis identified *CENPO*, a gene involved in cell cycle progression via regulation of kinetochore assembly [27]. At 16p11.2, *RNF40* is a promising candidate for

MM susceptibility due to its role in double-strand break repair during homologous recombination (HR) and class switch recombination [28, 29]. This gene has also been implicated in colorectal cancer [30]. A further candidate at this locus, *QPRT* has been demonstrated to confer resistance to chemotherapy and radiotherapy when studied in glioma and leukaemia [31, 32]. As such, genes identified within this TWAS build upon previously suggested candidate disease mechanisms which may confer MM predisposition [2], including anti-apoptotic effects, roles in DNA double-strand break repair and cell cycle regulation. Furthermore, many of the genes identified have been previously investigated in vitro for their roles in cancer and this adds further support as plausible candidate genes for MM predisposition [24, 26, 30–32].

6p21.33, which encodes much of the major histocompatibility complex, is an especially gene rich region. As well as the class I HLA-A and class II genes HLA-DQA1 and HLA-DRB1/5, multiple genes localise to the region including *TCF19* which encodes the cell cycle progression and proliferation transcription factor 19 [33, 34]. Complex LD patterns within this region make deconvolution of significant results within the region inherently problematic [14]. Additional work is required to reveal the contribution of genes in this region to MM development.

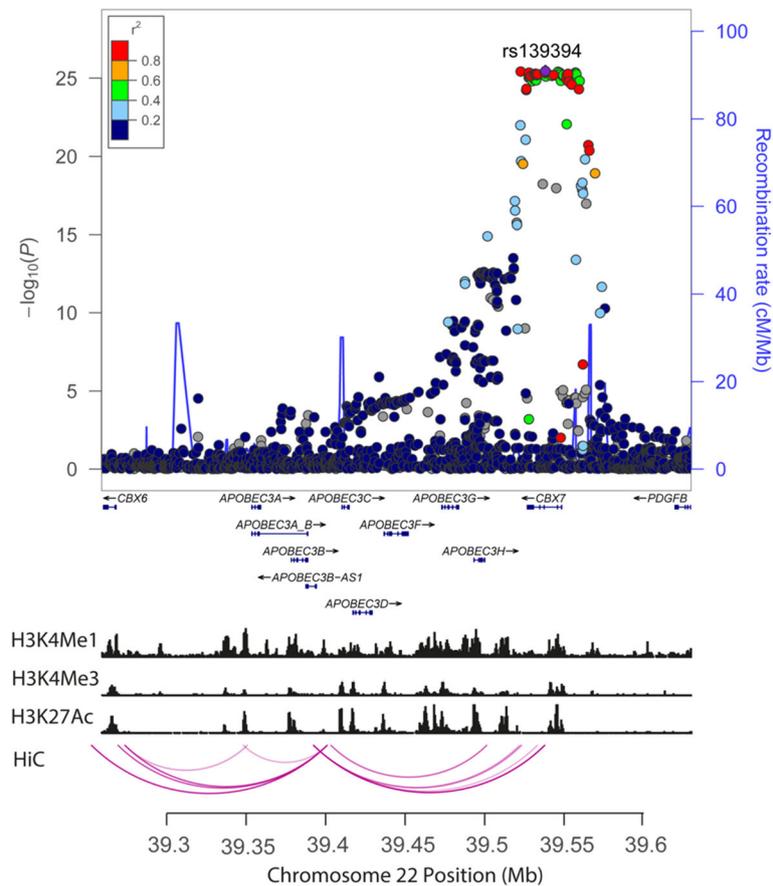
**Table 1** Genes significantly associated with risk of multiple myeloma

Locus	Gene	<i>P</i> value	<i>N</i> / <i>N</i> <sub>indep</sub>	Z-score min	Z-score max	Z-score mean	Z-score s.d.	SNP adjusting for	<i>P</i> value after SNP adjustment
16p11.2	<i>QPRT</i>	$1.01 \times 10^{-7}$	17/8	-2.73	3.04	-0.59	1.63	rs13338946	0.15
16p11.2	<i>RNF40</i>	$4.02 \times 10^{-7}$	24/3	0.05	5.68	4.67	1.48	rs13338946	0.89
16p11.2	<i>PRR14</i>	$4.28 \times 10^{-7}$	2/2	-5.38	-0.20	-2.79	3.66	rs13338946	0.34
16p11.2	<i>C16orf93</i>	$8.07 \times 10^{-7}$	13/5	-5.74	-0.34	-4.59	1.73	rs13338946	0.24
16p11.2	<i>RP11-2C24.5</i>	$1.54 \times 10^{-6}$	5/5	-5.64	4.43	-0.58	3.80	rs13338946	0.73
16p11.2	<i>PRSS53</i>	$1.71 \times 10^{-6}$	16/8	-5.19	3.68	-1.04	2.71	rs13338946	0.79
16q23.1	<i>RFWD3</i>	$7.71 \times 10^{-7}$	34/7	-3.41	6.35	2.51	3.26	rs7193541	0.47
17p11.2	<i>TBC1D27</i>	$1.95 \times 10^{-13}$	6/6	-1.91	4.19	0.51	2.16	rs34562254	0.89
17p11.2	<i>USP32P1</i>	$4.88 \times 10^{-13}$	3/3	-7.29	2.80	-1.36	5.27	rs34562254	0.01
17p11.2	<i>PEMT</i>	$5.65 \times 10^{-8}$	14/7	-1.74	5.43	1.36	1.93	rs34562254	0.01
22q13.1	<i>APOBEC3C</i>	$1.10 \times 10^{-18}$	21/8	-8.93	0.24	-4.09	2.21	rs139402	0.13
22q13.1	<i>APOBEC3H</i>	$4.28 \times 10^{-15}$	7/5	-5.45	7.92	-0.95	4.38	rs139402	0.76
22q13.1	<i>FAM83F</i>	$4.65 \times 10^{-10}$	11/8	-4.25	2.56	-0.48	2.01	rs139402	$1.1 \times 10^{-4}$
22q13.1	<i>APOBEC3D</i>	$6.2 \times 10^{-10}$	29/7	-8.38	-0.85	-4.15	1.56	rs139402	0.04
22q13.1	<i>APOBEC3F</i>	$5.15 \times 10^{-9}$	5/4	-6.34	6.15	1.09	5.07	rs139402	0.13
22q13.1	<i>APOBEC3G</i>	$1.81 \times 10^{-7}$	43/2	0.36	6.57	4.94	1.17	rs139402	0.17
2p23.3	<i>KIF3C</i>	$1.65 \times 10^{-18}$	6/6	-9.40	4.35	-1.19	4.50	rs7577599	$1.4 \times 10^{-9}$
2p23.3	<i>EPT1</i>	$8.37 \times 10^{-16}$	9/9	-1.76	6.00	1.30	2.72	rs7577599	$2.1 \times 10^{-5}$
2p23.3	<i>CENPO</i>	$1.48 \times 10^{-13}$	12/8	-6.60	2.22	-0.05	2.57	rs7577599	$6.1 \times 10^{-8}$
2p23.3	<i>DNMT3A</i>	$2.44 \times 10^{-13}$	8/8	-2.89	7.96	1.94	3.07	rs7577599	0.01
2p23.3	<i>AC010150.1</i>	$2.90 \times 10^{-13}$	4/4	-0.88	7.89	1.61	4.20	rs7577599	$8.9 \times 10^{-10}$
2p23.3	<i>PTGES3P2</i>	$4.46 \times 10^{-11}$	7/5	-4.23	2.03	-2.46	2.08	rs7577599	$1.1 \times 10^{-4}$
2p23.3	<i>DTNB</i>	$1.16 \times 10^{-7}$	11/10	-3.88	5.78	0.36	2.38	rs7577599	$3.1 \times 10^{-3}$
2p23.3	<i>DNAJC27</i>	$1.74 \times 10^{-7}$	8/8	-0.74	4.52	1.95	1.58	rs7577599	0.11
3p22.1	<i>ULK4</i>	$9.01 \times 10^{-15}$	43/6	0.90	8.89	6.60	2.24	rs6599192	0.85
3q26.2	<i>MYNN</i>	$7.84 \times 10^{-13}$	6/6	-7.91	1.58	-1.66	3.32	rs10936600	0.17
3q26.2	<i>LRRIQ4</i>	$9.63 \times 10^{-9}$	3/2	-5.94	-0.88	-4.25	2.92	rs10936600	0.03
3q26.2	<i>LRRC34</i>	$3.35 \times 10^{-8}$	21/2	3.97	6.47	5.12	0.66	rs10936600	0.82
3q26.2	<i>ACTRT3</i>	$4.28 \times 10^{-7}$	4/4	-0.94	5.80	1.56	2.94	rs10936600	0.48
6q21	<i>ATG5</i>	$1.55 \times 10^{-12}$	4/4	0.93	5.89	3.72	2.41	rs9372120	0.07
7p15.3	<i>CDCA7L</i>	$9.61 \times 10^{-9}$	8/8	-3.11	4.61	1.12	2.42	rs75341503	0.23
7q36.1	<i>CHPF2</i>	$2.53 \times 10^{-7}$	6/6	-2.01	2.13	0.40	1.49	rs7781265	0.06

Excludes associations found in the HLA region. s.d., standard deviation. Detailed are the S-MultiXcan *P* values for association between gene expression MM, and the corresponding Z-scores quantifying this relationship (e.g. a positive score indicates increased gene expression increases risk). *N* and *N*<sub>indep</sub> indicate the total number of single-tissue results used for S-MultiXcan analysis and the number of independent components after singular value decomposition, respectively

A number of previously reported MM risk regions were not implicated in our TWAS. At some regions such as 5q15, the high tissue specificity associated with the causal gene *ELL2* [6] may not be best modelled herein. At other loci, it is less obvious why an association was not detected. Speculatively, models at earlier developmental stages may yield greater insights at these loci, especially if they are influencing differentiation along B cell lineages. Additionally, other mechanistic effects may explain the functional basis of such loci, including methylation and splicing.

The increasing appreciation that regulation of gene expression forms the mechanistic basis of many GWAS risk regions makes the TWAS an attractive approach to identify causal genes. Traditionally, studies have only tended to consider an eQTL and risk SNP to overlap if they are in linkage at a specified threshold. This is, however, conservative as multiple local SNPs may independently contribute to risk. Furthermore, stipulating genome-wide significance thresholds for the GWAS signal (i.e.  $P < 5 \times 10^{-8}$ ) and linkage strength (i.e. LD > 0.5) between pairs of



**Fig. 2** Regional plot of association results at 22q13 in MM alongside recombination rates and histone marks in GM12878. Plot shows discovery association results of both genotyped and imputed SNPs in the GWAS samples and recombination rates.  $-\log_{10} P$  values (y axes) of the SNPs are shown according to their chromosomal positions (x axes). The colour of each symbol reflects the extent of LD with the top genotyped SNP. Genetic recombination rates, estimated using HapMap samples from Utah residents of western and northern European ancestry (CEU), are shown with a blue line. Physical positions are based on NCBI build 37 of the human genome. Also shown are the relative positions of GENCODE v19 genes mapping to the region of association. Below the association plot are the relative positions of GENCODE v19 genes mapping to the region of association and the histone marks and chromatin loops for lymphoblastoid cell line, GM12878

SNPs for evidence of expression influencing risk, constrains study power. The TWAS approach is essentially agnostic as it jointly considers all SNPs in the region, regardless of reported GWAS association strength. There are, however, limitations to TWAS. Firstly, TWAS is based on fitting predictive linear models of gene expression based on local genotype data, followed by prediction into large cohorts and subsequent association testing; therefore, it does not capture total expression which includes environmental and technical components [35]. Secondly, TWAS will also lose power if gene expression is a nonlinear function of local SNPs, or when trans (or distal) regulation is a major determinant of expression levels.

All conclusions from our TWAS come with several caveats. While TWAS associations are consistent with models of gene expression level influencing MM risk, we acknowledge the possibility of confounding. Imputed gene expression levels are generated from weighted

linear combinations of SNPs, and many of which may tag non-regulatory mechanisms driving risk and result in inflated association statistics. Inevitably, despite addressing LD, since genes with eQTLs are common, associations may be the result of chance co-localization between eQTLs and MM risk.

Our ability to identify gene expression significantly associated with MM risk in this TWAS may be affected by tissue specificity. On the basis of the power calculation, our TWAS analysis had only 80% power to detect an odds ratio of  $\sim 1.1$  for MM risk per one standard deviation increase (or decrease) in the expression level of a gene whose *cis*-heritability is 60% respectively in EBV-transformed lymphocytes (Additional file 1: Figure S2), which we used as a proxy for plasma cells. In light of abundant shared *cis*-regulation of expression across tissues, by combining data, we would expect any model to yield greater power as the number of tissues increases in

which a variant is functional. Hence, we aimed to robustly capture genetically regulated gene expression using a large sample size.

## Conclusions

Our findings highlight the value of integrating expression with GWAS to prioritise candidate causal genes. A number of identified genes have plausible roles in MM tumorigenesis (e.g. *APOBEC*, *RNF40*) or have been previously implicated in other malignancies (e.g. *QPRT*). The genes identified in this TWAS can be explored for follow-up and validation to further understand their role in MM biology.

## Methods

### GWAS data

MM genotyping data were derived from the most recent meta-analysis of 7 GWAS datasets totalling 7319 cases and 234,385 controls of European descent. After imputation, these related > 3.5 million genetic variants to MM. Comprehensive details of the genotyping and quality control of these GWAS have been previously reported [2, 16–19] and are summarised in Additional file 1: Tables S3 and S4.

### Association analysis of predicted gene expression with myeloma risk

Associations between predicted gene expression and MM risk were examined using MetaXcan [10], which combines GWAS and eQTL data, accounting for LD-confounded associations. Briefly, genes likely to be disease-causing were prioritised using S-PrediXcan [10] which uses GWAS summary statistics and pre-specified weights to predict gene expression, given co-variances of SNPs. SNP weights and their respective covariance in 48 tissues from 80 to 491 individuals were obtained from predict.db (<http://predictdb.org/>), which is based on GTEx version 7 eQTL data [36]. A full list of the sample count by tissue can be found at <https://gtexportal.org/home/tissueSummaryPage>. To combine S-PrediXcan data across the different tissues taking into account tissue-tissue correlations, we used S-MultiXcan [13].

To determine if associations between genetically predicted gene expression and MM risk were influenced by variants previously identified by GWAS, we performed conditional analyses adjusting for sentinel GWAS risk SNPs (Additional file 1: Table S5) using GCTA-COJO [37]. Adjusted output files were provided as the input GWAS summary statistics for S-PrediXcan analyses as above. To account for multiple comparisons, we considered a Bonferroni-corrected  $p$  value threshold of  $1.96 \times 10^{-6}$  (i.e. 0.05/25,520 genes) as being statistically significant.

### Regulatory annotation

To map risk SNPs to interactions involving promoter contacts and identify genes involved in MM susceptibility at the 22q13.1 locus, we analysed previously published promoter capture Hi-C data on the GM12878 downloaded from the ArrayExpress database, accession code E-MTAB-2323 cell line as a model B cell [38]. Reads from technical replicates were combined before processing and valid pairs were identified using HICUP [39]. Two biological replicates were analysed to assure reproducibility and significant interactions were determined using CHiCAGO [40]. ChIP-Seq on H3K4Me1, H3K4Me3, and H3K27Ac in GM12878 were from the ENCODE project (ENCODE Project Consortium, 2012).

### Statistical power for association tests

To estimate the power of our TWAS to identify associations, we performed a simulation analysis adopting a similar methodology to Wu et al. [41]. We set the number of cases and controls as 7319 and 234,385, respectively. An estimate of the population prevalence of MM was obtained from Cancer Research UK (<https://www.cancerresearchuk.org>). We generated the gene expression levels from the empirical distribution of gene expression levels in GTEx normalised expression dataset for each tissue. We calculated statistical power at  $P < 1.96 \times 10^{-6}$ , corresponding to the TWAS genome-wide significance level, according to various *cis*-heritability ( $h^2$ ) thresholds that are assumed to be equivalent to gene expression prediction models ( $R^2$ ). The results are based on 1000 replicates.

## Additional file

**Additional file 1: Table S1.** Genes significantly associated with risk of multiple myeloma. **Table S2.** New and previously implicated<sup>1-5</sup> genes at each genome wide significant multiple myeloma locus. **Table S3.** Quality control filters applied to samples from the seven published GWAS. **Table S4.** Quality control filters applied to SNPs from each GWAS. **Table S5.** MM GWAS risk SNPs. **Figure S1.** Quantile-Quantile Plots of  $-\log_{10}(P\text{-value})$  associations. **Figure S2.** TWAS power plot in EBV-transformed lymphocytes. (DOCX 1515 kb)

### Authors' contributions

MW and RSH conceived and designed the study. MW performed bioinformatics with contribution from BK. MW and RSH wrote the manuscript with contributions from BK. All authors reviewed the final manuscript. All authors read and approved the final manuscript.

### Funding

This work was supported by grants from Myeloma UK, Bloodwise and Cancer Research UK (C1298/A8362). 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.

### Availability of data and materials

Details and availability of SNP genotyping data that support the findings of this study have been previously published [2, 16–19].

**Ethics approval and consent to participate**

The TWAS was undertaken using previously reported GWAS data. Hence, ethical approval was not sought for this project because all data came from the summary statistics from the published GWAS, and no individual-level data were used [2, 16–19].

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Division of Genetics and Epidemiology, The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK. <sup>2</sup>Division of Molecular Pathology, The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK. <sup>3</sup>Department of Internal Medicine V, University of Heidelberg, 69117 Heidelberg, Germany. <sup>4</sup>German Cancer Research Center, 69120 Heidelberg, Germany. <sup>5</sup>Department of Hematology, Erasmus MC Cancer Institute, 3075, EA, Rotterdam, The Netherlands. <sup>6</sup>Myeloma Institute for Research and Therapy, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA. <sup>7</sup>Clinical Trials Research Unit, University of Leeds, Leeds LS2 9PH, UK. <sup>8</sup>Institute of Human Genetics, University of Bonn, D-53127 Bonn, Germany. <sup>9</sup>Division of Medical Genetics, Department of Biomedicine, University of Basel, 4003 Basel, Switzerland. <sup>10</sup>Royal Victoria Infirmary, Newcastle upon Tyne NE1 4LP, UK. <sup>11</sup>Department of Genomics, Life & Brain Center, University of Bonn, D-53127 Bonn, Germany. <sup>12</sup>deCODE Genetics, Sturlugata 8, IS-101, Reykjavik, Iceland. <sup>13</sup>Hematology Clinic, Skåne University Hospital, SE-221 85 Lund, Sweden. <sup>14</sup>Hematology and Transfusion Medicine, Department of Laboratory Medicine, BMC B13, SE-221 84 Lund, Sweden. <sup>15</sup>Broad Institute, 7 Cambridge Center, Cambridge, MA 02142, USA.

Received: 30 June 2019 Accepted: 12 August 2019

Published online: 20 August 2019

**References**

- Altieri A, Chen B, Bermejo JL, Castro F, Hemminki K. Familial risks and temporal incidence trends of multiple myeloma. *Eur J Cancer (Oxford, England: 1990)*. 2006;42(11):1661–70.
- Went M, Sud A, et al. Identification of multiple risk loci and regulatory mechanisms influencing susceptibility to multiple myeloma. *Nat Commun*. 2018;9(1):3707.
- Sud A, Kinnersley B, Houlston RS. Genome-wide association studies of cancer: current insights and future perspectives. *Nat Rev Cancer*. 2017; 17(11):692–704.
- Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, et al. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet*. 2011;43(6):513–8.
- Ip HF, Jansen R, Abdellaoui A, Bartels M, Boomsma DI, Nivard MG. Characterizing the relation between expression QTLs and complex traits: exploring the role of tissue specificity. *Behav Genet*. 2018;48:374.
- Li N, Johnson DC, Weinhold N, Kimber S, Dobbins SE, Mitchell JS, et al. Genetic predisposition to multiple myeloma at 5q15 is mediated by an ELL2 enhancer polymorphism. *Cell Rep*. 2017;20(11):2556–64.
- Li N, Johnson DC, Weinhold N, Studd JB, Orlando G, Mirabella F, et al. Multiple myeloma risk variant at 7p15.3 creates an IRF4-binding site and interferes with CDCA7L expression. *Nat Commun*. 2016;7:13656.
- Weinhold N, Meissner T, Johnson DC, Seckinger A, Moreaux J, Försti A, et al. The 7p15.3 (rs4487645) association for multiple myeloma shows strong allele-specific regulation of the MYC-interacting gene CDCA7L in malignant plasma cells. *Haematologica*. 2015;100(3):e110.
- Ali M, Ajore R, Wihlborg A-K, Niroula A, Swaminathan B, Johnsson E, et al. The multiple myeloma risk allele at 5q15 lowers ELL2 expression and increases ribosomal gene expression. *Nat Commun*. 2018;9(1):1649.
- Gamazon ERA-Ohoo, Wheeler HEA-Ohoo, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet*. 2015;47:1091–98.
- Landgren O, Rajkumar SV. New developments in diagnosis, prognosis, and assessment of response in multiple myeloma. <https://doi.org/10.1158/1078-0432.CCR-16-0866>.
- Sud A, Thomsen H, Orlando G, Försti A, Law PJ, Broderick P, et al. Genome-wide association study implicates immune dysfunction in the development of Hodgkin lymphoma. *Blood*. 2018;132(19):2040.
- Barbeira AN, Pividori M, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet*. 2019;15(1):e1007889.
- Moutsianas L, Gutierrez-Achury J. Genetic Association in the HLA Region. *Methods Mol Biol (Clifton, NJ)*. 2018;1793:111–34.
- Beksac M, Gragert L, Fingerson S, Maiers M, Zhang MJ, Albrecht M, et al. HLA polymorphism and risk of multiple myeloma. *Leukemia*. 2016;30:2260.
- Mitchell JS, Li N, Weinhold N, Forsti A, Ali M, van Duin M, et al. Genome-wide association study identifies multiple susceptibility loci for multiple myeloma. *Nat Commun*. 2016;7:12050.
- Broderick P, Chubb D, Johnson DC, Weinhold N, Forsti A, Lloyd A, et al. Common variation at 3p22.1 and 7p15.3 influences multiple myeloma risk. *Nat Genet*. 2011;44(1):58–61.
- Chubb D, Weinhold N, Broderick P, Chen B, Johnson DC, Forsti A, et al. Common variation at 3q26.2, 6p21.33, 17p11.2 and 22q13.1 influences multiple myeloma risk. *Nat Genet*. 2013;45(10):1221–5.
- Swaminathan B, Thorleifsson G, Joud M, Ali M, Johnsson E, Ajore R, et al. Variants in ELL2 influencing immunoglobulin levels associate with multiple myeloma. *Nat Commun*. 2015;6:7213.
- Hussain T, Mulherkar R. Lymphoblastoid cell lines: a continuous in vitro source of cells to study carcinogen sensitivity and DNA repair. *Int J Mol Cell Med*. 2012;1(2):75–87.
- Risca VI, Greenleaf WJ. Unraveling the 3D genome: genomics tools for multiscale exploration. *Trends Genet*. 2015;31(7):357–72.
- Bolli N, Maura F, Minvielle S, Gloznik D, Szalat R, Fullam A. Genomic patterns of progression in smoldering multiple myeloma. *Nat Commun*. 2018;9(1):3363.
- Maura F, Petljak M, Lionetti M, Cifola I, Liang W, Pinatel E, et al. Biological and prognostic impact of APOBEC-induced mutations in the spectrum of plasma cell dyscrasias and multiple myeloma cell lines. *Leukemia*. 2018;32(4):1044–8.
- Walker BA, Wardell CP, Murison A, Boyle EM, Begum DB, Dahir NM, et al. APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. *Nat Commun*. 2015;6:6997.
- Guzik-Lendrum S, Rayment I, Gilbert SP. Homodimeric kinesin-2 KIF3CC promotes microtubule dynamics. *Biophys J*. 2017;113(8):1845–57.
- Wang C, Wang C, Wei Z, Li Y, Wang W, Li X, et al. Suppression of motor protein KIF3C expression inhibits tumor growth and metastasis in breast cancer by inhibiting TGF-beta signaling. *Cancer Lett*. 2015;368(1):105–14.
- Eskat A, Deng W, Fau, Hofmeister A, Hofmeister A, Fau, Rudolph S, Rudolph S, Fau, Emmerth S, Emmerth S, Hellwig D, Hellwig D, Fau, Ulbricht T, et al. Step-wise assembly, maturation and dynamic behavior of the human CENP-P/O/R/Q/U kinetochore sub-complex. *PLoS One*. 2012;7(9):e44717.
- So CC, Ramachandran S, Martin A. E3 Ubiquitin Ligases RNF20 and RNF40 are required for Double-Stranded Break (DSB) repair: evidence for Monoubiquitination of Histone H2B Lysine 120 as a Novel Axis of DSB signaling and repair. *Mol Cell Biol*. 2019;39(8):e00488–18.
- Shiloh Y, Shema E, Moyal L, Oren M. RNF20-RNF40: a ubiquitin-driven link between gene expression and the DNA damage response. *FEBS Lett*. 2011; 585(18):2795–802.
- Schneider D, Chua RL, Molitor N, Hamdan FH, Rettenmeier EM, Prokakis E, et al. The E3 ubiquitin ligase RNF40 suppresses apoptosis in colorectal cancer cells. *Clin Epigenetics*. 2019;11(1):98.
- Ullmark T, Montano G, Jarvstrat L, Jernmark Nilsson H, Hakansson E, Drott K, et al. Anti-apoptotic quinolinate phosphoribosyltransferase (QPRT) is a target gene of Wilms' tumor gene 1 (WT1) protein in leukemic cells. *Biochem Biophys Res Commun*. 2017;482(4):802–7.
- Sahm F, Oezen I, Opitz CA, Radlwimmer B, von Deimling A, Ahrendt T, et al. The endogenous tryptophan metabolite and NAD+ precursor quinolinic acid confers resistance of gliomas to oxidative stress. *Cancer Res*. 2013; 73(11):3225–34.
- Krautkramer KA, Linnemann AK, Fontaine DA, Whillock AL, Harris TW, Schleis GJ, et al. Tcf19 is a novel islet factor necessary for proliferation and survival in the INS-1 beta-cell line. *Am J Physiol Endocrinol Metab*. 2013;305(5):E600–10.
- Sen S, Sanyal S, Srivastava DK, Dasgupta D, Roy S, Das C. Transcription factor 19 interacts with histone 3 lysine 4 trimethylation and controls gluconeogenesis via the nucleosome-remodeling-deacetylase complex. *J Biol Chem*. 2017;292(50):20362–78.

35. Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, et al. Opportunities and challenges for transcriptome-wide association studies. *Nat Genet.* 2019;51(4):592–9.
36. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580–5.
37. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88(1):76–82.
38. Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet.* 2015;47:598.
39. Wingett S, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research.* 2015;4:1310.
40. Cairns J, Freire-Pritchett P, Wingett SW, Varnai C, Dimond A, Plagnol V, et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* 2016;17(1):127.
41. Wu L, Shi W, Long J, Guo XA-Ohoo, Michailidou K, Beesley J, et al. A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. (1546–1718 (Electronic)).

### Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

