Published text:

Genetic Associations With Thalidomide Mediated Venous Thrombotic Events in Myeloma Identified using Targeted Genotyping

Johnson DC¹, Corthals S², Ramos C³, Hoering A⁴, Cocks K⁵, Dickens NJ¹, Haessler J⁴, Goldschmidt H⁶, Child JA⁵, Bell SE⁵, Jackson GH⁷, Baris D⁸, Rajkumar SV⁹, Davies FE¹, Durie BG¹⁰, Crowley J⁴, Sonneveld P², Van Ness B³, Morgan GJ¹.

¹ Institute of Cancer Research, London, UK.
² Erasmus Medical Center, Rotterdam, the Netherlands.
³ University of Minnesota. Minneapolis, USA.
⁴ Cancer Research and Biostatistics (CRAB), Seattle, USA.
⁵ University of Leeds, Leeds, UK.
⁶ University of Heidelberg, Heidelberg, Germany.
⁷ University of Newcastle, Newcastle, UK.
⁸ National Cancer Institute, Bethesda, Maryland, USA.
⁹ Mayo Clinic, Rochester, Minnesota, USA.
¹⁰ Cedars Sinai, Los Angeles, USA.

Keywords
Thalidomide, thrombosis, mechanism, risk, single nucleotide polymorphism.

Author for Correspondence
Professor GJ Morgan
Section of Haematological Oncology
Institute of Cancer Research
Belmont
Sutton
Surrey
United Kingdom
SM2 5NG

This work is supported by the International Myeloma Foundation.
Abstract

A venous thromboembolism (VTE) with the subsequent risk of pulmonary embolism is a major concern in the treatment of multiple myeloma patients with thalidomide. The susceptibility to developing a VTE in response to thalidomide therapy is likely to be influenced by both genetic and environmental factors. To test genetic variation associated with treatment related VTE in patient peripheral blood DNA, we used a custom-built molecular inversion probe (MIP) based single nucleotide polymorphism (SNP) chip containing 3404 SNPs. SNPs on the chip were selected in “functional regions” within 964 genes spanning 67 molecular pathways thought to be involved in the pathogenesis, treatment response and side effects associated with myeloma therapy. Cases and controls were taken from three large clinical trials: MRC Myeloma IX, Hovon-50 and ECOG EA100, which compared conventional treatments with thalidomide in myeloma patients. Our analysis showed that the set of SNPs associated with thalidomide-related VTE were enriched in genes and pathways important in drug transport/metabolism, DNA repair and cytokine balance. The effects of the SNPs associated with thalidomide related VTE may be functional at the level of the tumor cell, the tumor-related microenvironment, and the endothelium. The clinical trials described in this paper have been registered as follows: MRC Myeloma IX: ISRCTN68454111, HOVON50: www.clinicaltrials.gov under identifier NCT00028886, and ECOG EA100: www.clinicaltrials.gov under identifier NCT00033332.
Introduction

The introduction of thalidomide and other immunomodulatory drugs has revolutionised clinical management of patients with myeloma. Thalidomide treatment has achieved response rates of 30% at relapse and even higher rates at presentation.\(^1\) Investigation of the specific effects of thalidomide in myeloma remains an active area of research where up regulation of ICAM-1\(^2\), VCAM-1, IL10,\(^3,4\) IL12,\(^5\) and decreased levels of VEGF,\(^6\) βFGF,\(^7-9\) HGF,\(^10\) TNF\(\alpha\),\(^11\) IL6,\(^12\) sIL6-R,\(^13\) are thought to play a role in the mechanism of action, which suggests that thalidomide effects the myeloma cell directly as well as its microenvironment.\(^14\)

The therapeutic use of thalidomide has focused attention on venous thrombotic events (VTEs). There appears to be a background rate of 5-10% VTE\(^15,16\) in myeloma possibly due to enhanced expression of tissue factor and VEGF,\(^17\) acquired cytokine mediated activated protein C resistance\(^18\) and down-regulation of thrombospondin.\(^19\) In intensively treated patients exposed to thalidomide the rate of VTE increases to 10-15%,\(^16,20,21\) the mechanisms leading to this are uncertain, but it is known that thalidomide regulates the level of COX-2\(^22-25\), a well described pro-thrombotic factor. Thalidomide may also modulate the VTE risk by its effects on cytokine levels acting on the endothelial cell, a mechanism dependent on the differential apoptotic effects of thalidomide in myeloma plasma cells compared to endothelial cells, which are protected from apoptosis by decrease of VEGF by thalidomide,\(^26-28\) In this context, it is known that stressed human umbilical vein endothelial cells (HUVECs) up regulate a number of pro-coagulant factors including PAR-1, P-selectin, E-selectin and tissue factor, with thalidomide protecting these cells from apoptosis potentially enhancing these pro-coagulant effects, there is some clinical evidence for this mechanism in non-myeloma settings.\(^29-33\)

The risk of developing a VTE following thalidomide exposure depends upon a number of factors including: Disease stage, the type of chemotherapy combination and the supportive therapy used. Patient specific variables also contribute to the excess risk of VTE including immobility, poor performance status, and dehydration. An important clinical observation is that VTEs occur early after the initiation of thalidomide treatment.
and VTE rates are increased in patients when used in conjunction with anthracycline and dexamethasone and can decrease following exposure to bortezomib.

The excess risk of thalidomide associated VTE in myeloma has been managed by a number of different strategies, ranging from the identification of high risk patients suitable for prophylaxis to prophylactic anticoagulation for all patients. Aspirin has been suggested to be effective, but its use is controversial because of the lack of a readily applicable mechanism justifying its use. In this work we have examined inherited genetic variation associated with VTE following thalidomide exposure in myeloma patients, using a custom array based SNP detection tool, in an effort to elucidate the molecular mechanisms contributing to increased risk.

Materials and methods

Clinical samples
Peripheral blood DNA samples were obtained from 544 myeloma cases derived from three randomised clinical trials comparing standard induction treatment for presenting patients with thalidomide containing regimens derived from the MRC Myeloma IX, 1966 patients, the ECOG EA100, 900 patients and the Hovon-50 study, 400 patients. The dose of thalidomide (100-200mg daily) was comparable between the three studies but the chemotherapy combination used differed. The samples were used as the basis for two nested case control comparisons examining the inherited genetic contribution to the risk of VTE as a consequence of thalidomide exposure. In a discovery set analysis we compared the genotype results derived from 157 Myeloma IX patients with VTEs of which 104 were related to thalidomide exposure and 53 unrelated, to a control group of 315 age and sex matched myeloma patients also in the trial who did not develop a VTE (198 thalidomide exposed and 117 non thalidomide exposed). To validate the frequency distributions we carried out a second case control comparison using 23 patients with VTE treated with thalidomide and 49 thalidomide treated controls. To ensure homogeneity of allelic frequencies only patients of European descent were included. This study has been approved by The United Kingdom Multicentre Ethics Committee.
Clinical trials
The Myeloma IX study comprises of two randomisations; an intensive pathway for younger fitter patients comparing CVAD: cyclophosphamide (500mg orally weekly), vincristine (0.4mg iv d1-4), doxorubicin (9.0mg/m² d1-4), dexamethasone (40mg d1-4, d12-15), delivered by a central venous access device with oral CTD (cyclophosphamide, thalidomide, dexamethasone) using the same doses of cyclophosphamide and dexamethasone combined with 200mg of thalidomide. The second randomisation, for older less fit patients, compared an attenuated dose of CTD (thalidomide (100-200 mg) to melphalan (7.0mg/m² od p.o. d1-4 every 28 days) and prednisolone (MP). All patients at high risk of VTE, defined by clinical criteria, were identified and prophylactic anti-coagulation was considered by the treating physician, but it was not specified. The ECOG EA100 study randomised patients to either dexamethasone alone 40mg daily d1-4, d12-15 or the same dose in combination with thalidomide 200mg daily. In the study set, from which samples was available; no thromboprophylaxis was used on either arm. The Hovon-50 study randomised patients to either: 3 cycles of VAD: vincristine (0.4 mg, iv rapid infusion on days 1-4), doxorubicin (9 mg/m2, iv rapid infusion on days 1-4) and dexamethasone (40 mg orally, days 1-4, 9-12, 17-20) or the same regimen but with thalidomide replacing the vincristine (TAD). Thalidomide was given daily at a dose of 200 mg, but could be escalated to 400 mg. All patients in the TAD arm, received thromboprophylaxis with Low Molecular Weight Heparin (LMWH). Incident cases of VTE were defined using clinical criteria, and no screening approach was used. The identification of VTE represents current clinical practice with initial clinical identification and subsequent confirmation and definition of the extent of thrombosis using a definitive radiological investigation. Central venous thrombosis and line related thrombosis were defined by clinical criteria and subsequently confirmed by ultrasound.

Genotyping, SNP selection and chip design
DNA was extracted from frozen white blood cell pellets using Qiagen® Flexigene kit and quantified using a Nanodrop® Spectrophotometer. Genotyping was performed using the Affymetrix® Targeted genotyping™ platform which is based on a Molecular
Inversion Probe technology. Patients samples were assayed using a custom built 3.0K panel comprising 3,400 SNPs. SNPs were selected using a hypothesis driven strategy. Pertinent candidate genes were nominated by myeloma groups in the International Myeloma Foundation led 'Bank On A Cure™ (BOAC) consortium. An initial list was supplemented with referencing pathway databases including BioCarta, KEGG\(^46,47\), and Pathway Assist\(^48\) generating a candidate gene list spanning some 67 molecular pathways important in the biology of myeloma, treatment response and side effects to conventional and novel agents, which included important genes within the clotting and pro-thrombotic pathways. Taking the BOAC candidate genes we completed a literature search\(^49\) to identify SNPs that had been previously reported as having a functional consequence or being relevant in prior aetiological or treatment outcomes studies. SNPs with a minor allele frequency (MAF) > 2% were then systematically selected from the candidate gene list using the following criteria: non-synonymous SNPs present in dbSNP/SNP 500\(^50\); promoter variants present in homologous regions between human and mouse, in or adjacent to a transcription binding site utilising the Promolign database;\(^51\) promoter SNPs identified in the Functional Element SNPs Database (FESD).\(^52\) We then included Tag-SNPs in genes considered to be of particular relevance along with population discriminating admixture variants from the X chromosome;\(^53\) Finally we included all non-synonymous SNPs present in the dbSNP database in phosphatase, kinase, and transferase genes with a MAF > 2%. The genes and SNPs comprising this panel with allele frequencies are available as online supplementary data.\(^54\)

**Statistical Analyses**

We carried out a Fisher Exact Hardy-Weinberg equilibrium (HWE) Test at a level p <= 0.001 on all SNPs across the control samples and removed SNPs departing from HWE from the analysis to filter erroneously performing SNPs. We then carried out a ‘test of missing-ness’ on case and control status, to control for any bias in missing data. We performed basic Fisher’s test (allellic) association test for disease trait based on a comparison of cases with controls. We then completed the analysis using three genetic models: additive (Cochran-Armitage trend test), dominant and recessive. To account for multiple testing we carried out label swapping permutation procedures on
each of the SNP assay with their most significant model being used to calculate an empirical p-value for each SNP. The size of the dataset generated on the BOAC panel is much larger than a typical candidate-gene study; we therefore carried out the above analysis in the program Plink, an open-source whole genome association analysis toolset, designed for large dataset analysis. The test for epistasis involved the testing all pair-wise combinations of SNPs. The output consists only of pair wise epistatic results above $P < 0.001$; for each SNP. Combinations were restricted to SNPs more than 1 Mb apart, or on different chromosomes. This test is only an approximation of the extent of epistasis (SNP-SNP interaction), as it is a naive statistic that does not take linkage disequilibrium (LD) into account. We characterised the haplotypes using Haploview 4.0, and completed Haplotype trend regression in Helix-tree®. Meta-analysis was performed in SPSS 14.0, utilising a meta-analysis macro written by Marta Garcia-Granero. Combined odds ratios were calculated using Mantel Haenszel method for fixed events.

**Biological relevance of the associated SNPs**

To examine the possible functionality of the thalidomide related VTE associated SNPs, we used two complementary *in silico* algorithms for prediction of the putative impact of missense variants on protein function, PolyPhen (structural) and the SIFT (conservation), shown in Table S8. We then used a bioinformatics approach to define the pathways potentially deregulated by the associated and validated genes. We used the functional annotation tool on the DAVID Bioinformatics Resources/Database, to characterise which pathways are most represented in associated gene groups from our single point analysis. The gene coverage of the BOAC chip was used to form a template/background set, against which associated genes and validated associated SNPs with VTE were tested, and is presented in Table S9.

**Recursive Partitioning**

In order to develop a predictive model for the identification of patients at high risk of VTE, we first divided the combined data set into a training and validation set. We then applied the method of recursive partitioning to the training set. In this approach a regression tree is built by first finding the SNP which best splits the data into two
groups (VTE, no VTE). This process is repeated over and over again for the individual subsets until the subgroups reach a minimum size or no improvement can be made. The second stage in recursive partitioning consists of cross-validation by trimming back (pruning) the typically complex full tree. The best pruned trees are examined to find which one has the largest classification rate while using the smallest number of SNPs. Sensitivity and specificity are determined for the training and validation set. A receiver operator characteristic (ROC) curve is used to determine the best sensitivity and specificity trade-off.

Results

Clinical Results

The Myeloma IX analysis is based on 1966 randomised patients, on 984 cases treated with CTD, 557 with CVAD and 425 with MP. In the intensive pathway, the overall rate of VTE was identical in both arms (Table S1). However, there was a qualitative difference between the two arms with DVT being predominant in the thalidomide treated group and line related thrombosis being predominant in the CVAD group. In the non-intensive pathway very few VTEs were seen in the MP group, whereas in the CTD group there were 15.0% VTE. The median time to VTE in each of the groups was approximately 12 weeks from treatment initiation. The Hovon-50 study had VTE rates of 12.1% and 11.8% in the thalidomide related and standard arms respectively with median time to first event of 8.9 weeks. In the ECOG EA100 study the rates were 17.0% and 3.0%.58

Panel, Sample and SNP Assay Validation

Affymetrix constructed and validated the SNP panel reagents. 59 DNA samples from the extensively characterised and genotyped Coriell CEPH Hapmap series were assayed to validate the call performance of the BOAC panel. 58 Coriell CEPH HAPMAP samples were also used in a correlation analysis between the BOAC chip and HAPMAP study. We did not obtain Hapmap data for the remaining Coriell sample and did not perform a correlation analysis. 2606 SNPs were present on both the chip and HAPMAP, there was a SNP call correlation of 96.1% at 95% confidence levels, SNPs falling below this were removed from the analysis (132 SNPs). The Coriell
sample genotype validation was replicated in BOAC labs to ensure there was no
differential bias in genotyping scoring between sites. Cases and controls were also
genotyped together throughout the experiment to avoid any differential bias in
genotype scoring. We observed complete agreement between the known gender and
inferred SNP based gender in all samples. We used Eigenstrat,\textsuperscript{59} to highlight
population stratification and removed four population outliers from the analysis. A
number of admixture SNPs were included in the SNP panel,\textsuperscript{53} genotype calls for these
SNPs demonstrated that cases and controls reflected a sample set drawn from a
Caucasian population.

**Genotyping results and validation**

The MRC Myeloma IX study is the largest of the data sets in this study and in order to
capitalise on this, we chose to focus our discovery set on this study and to validate the
results on combined data sets from Hovon-50 and ECOG EA100 trials. A set of SNPs
for validation was defined, by separately determining the distribution of the most
significant SNPs in the MRC myeloma IX study, Hovon50 and ECOG EA100 studies.
Following testing allelic distributions using Fisher’s test to an empirical $p<0.05$, in the
discovery set, 120 SNPs were found to be associated with thalidomide associated
VTE, involving 71 genes (Table S2). Further genetic model association analysis in
Myeloma IX data is listed in Table S3, allelic and genetic model association for non-
thalidomide related VTEs are listed in Table S4 and S5, with allelic and genetic model
association for Hovon-50/EA100 trials shown in Tables S6 and S7. With the aim of
identifying genes modulating the risk of thalidomide related thrombosis we compared
the distribution of the SNPs identified in this analysis between three subgroups,
thalidomide associated VTE MRC IX, non thalidomide associated VTE MRC IX and
thalidomide associated VTE from the combined Hovon50/ECOG EA100 studies. A
Venn diagram, depicting the overlapping associated genes in the three analyses is
shown in Figure 2.

**Validation of thalidomide related VTE associated SNPs**

To validate the genotyping results in the discovery set, we discarded SNPs with
conflicting frequency distributions between studies. This approach may have led to the
removal of a number of true positives from the analysis because of the small sizes of the Hovon-50 and ECOG EA100 data sets. SNPs with small effect sizes were also removed from the analysis. As a result of this process we found 24 SNPs (Table 1) associated with VTE in the “discovery set” with consistent distributions in the two validation datasets. Haplotype analysis showed that 6 SNPs were in linkage with a stronger proxy SNP and as such were discarded leaving 18 validated SNPs associated with thalidomide related venous thrombotic events. A “Forest plot” with Odds ratios (ORs) and Confidence Intervals (CI) for the combined and individual data sets of the validated SNPs was generated. (Figure 3)

Gene-gene interactions
To examine gene-gene interactions we looked for pair-wise combinations mediating risk, the epistatic interactions with a p< 0.001 are shown in Table S10.

Recursive partitioning analysis
To maximise the size of the data set and thus to maximise the ability to identify relevant SNPs we combined all the data sets into one and randomly split it into a two third training set and one third validation set. The data was stratified by trial and VTE cases to insure that the training and validation sets are comparable. This data included 165 subjects without VTE and 84 subjects with VTE in the training set and 82 subjects without VTE and 42 subjects with VTE in the validation set. The training set was used to identify the top associated genes and SNPs by association at the level of p<0.05 listed in Table S2. These SNPs were used in a recursive partitioning analysis carried out on the test set with the aim of finding the combination with the best sensitivity and specificity for the identification of VTE. We pruned the tree to find the tree with highest classification and smallest number of SNPs. The results of this analysis (Figure 4) showed that using 7 SNPs (rs7011 in CINP, rs289747 in CETP, rs610529 in ALDH1A1, rs3829963 in CDKN1A, rs2608555 in GAN, rs699947 in VEGF and rs168351 in ALDH1A) it was possible to identify VTE correctly in 70% of individuals with a specificity of 59% and sensitivity of 81%. This set of SNPs performed well in the validation set being able to correctly classify VTE in 61% of individuals with a specificity of 30% and sensitivity of 77%. (Tables S12 and S13)
Discussion
This study has analysed data from three large randomised clinical studies comprising 3100 patients, comparing induction treatment for newly presenting patients with myeloma with and without thalidomide. The results of this analysis show that the background rate of VTE in MP-treated patients is very low and significantly increases with the addition of thalidomide. In addition we provide further evidence that infusional regimes based on VAD increase VTE rates to around 15%, which is similar to the rates seen with oral thalidomide combinations. The nature of the thrombotic events is qualitatively different between regimes, with all events being either DVT or PE in the oral thalidomide treated patients, whereas in the intravenous treatment 50% of the events are central line related. There is a doubling of non central line related VTE rates in the thalidomide treated patients compared to those receiving infusional induction regimens. The median time to VTE in each of the treatments is approximately 50-60 days after the initiation of treatment, a time reflecting the rapid dissolution of the myeloma clone. We have shown previously that response rate is enhanced in thalidomide containing regimes compared to VAD like regimes and we postulate that this is important in determining the VTE risk. The mechanistic importance of increased response rates with VTE risk, may explain the reduced numbers of VTEs seen in relapse patients, who are frequently drug resistant and show lower response rates. It is also important not to discount increased VTE risk due to changes in the disease biology related pro-coagulant profiles of such relapse-patients.

Using a nested case control design, with readily defined exposure and clinical endpoint, this study has given useful information about inherited genetic variants with a moderate effect size affecting the thrombotic response to thalidomide exposure. We chose to use the MRC Myeloma IX study as our initial discovery set because it was the largest and had the most data available with it. Validation in the combined Hovon50/ECOG study represents a pragmatic decision based on study size, study design and our desire to identify penetrant variants that can be replicated with relevance to different studies and data sets.
Despite a comprehensive analysis of the genetic variation within the coagulation and pro-thrombotic pathways, we could not find evidence for a significant association of genetic variation within these pathways with VTE risk following thalidomide exposure. Although we found Factor 5 Leiden (rs6025) to be associated with an increased risk of VTE in this analysis, the thrombotic risk was not increased in patients treated with thalidomide, similar results were seen for polymorphisms in MTHFR and FGB. We saw no association with thalidomide related VTE in commonly reported VTE risk alleles in F2 -455G/A (rs3136430) splice variant 20210G/A (rs3136431). We did find weak associations with genes known to mediate the coagulation pathway including MTRR, PLAUR, PPARD, PPARGC1A, PPARGC1B, THBS4 and WNK, but the associated risk was not high. We conclude that we can exclude a major contribution of genetic variation within the coagulation and pro thrombotic pathways based on this targeted approach, although smaller contribution to the phenotype may be missed because of the study size and design. Our findings are consistent with previous clinical observations and work by some of the authors, who failed to identify relevant changes in functional assays investigating this pathway. 61-63

The lack of a strong association with variation in the coagulation cascade suggests that VTE risk is mediated via alternative mechanisms. We identified the following SNPs which validated across the three data sets (Figure 3). Using the whole BOAC panel as the background gene set in the DAVID Functional Annotation Clustering tool against the 18 validated genes, generated three major enriched annotation clusters. The annotation clusters consisted of two ‘response to stress’ groups; a response to DNA damage group including: CHEK1, XRCC5, LIG1, ERCC6, DCLRE1B and PARP1; and a cytokine response group containing NFKB1, TNFRSF17, IL12B and LEP; and a third related group of ‘apoptosis’ with CASP3, PPARD and NFKB1. These enrichment groups indicate that genetic variation in response to DNA damage and cytokine mediated apoptosis modulates risk of developing a thalidomide related thrombosis.
High dose dexamethasone enhances haemostasis, increases platelet activation and promotes von Willebrand factor antigen (vWF) dependent thrombosis.\textsuperscript{64} Extremely high levels of factor VIII-coagulant (FVIII:C) activity and vWF have been found in thalidomide exposed patients.\textsuperscript{65} Patients that develop a subsequent VTE had higher VWF-Ag levels but not FVIII:C. High FVIII:C/VWF-Ag levels are found in patients with active myeloma and this is probably a reflection of increased bone marrow angiogenesis in myeloma. These pro-thrombogenic circumstances would contribute to VTE during treatment with a thalidomide-dexamethasone combination.\textsuperscript{65} In line with vWF mediating the pro-thrombotic effects of dexamethasone in thalidomide related VTE, we saw a protective effect of vWF non-synonymous SNP (rs216321) and synonymous SNP (rs216902) in thalidomide treated controls.

Although there is evidence to suggest thalidomide may damage DNA directly,\textsuperscript{66} it is important to note the majority of cases in this analysis were derived from the MRC and Hovon-50 studies that included either cyclophosphamide or doxorubicin/adriamycin in the treatment regime, which may explain an association with DNA repair genes. Variation in DNA repair capacity could readily affect the response of the myeloma clone to treatment, due to the direct relationship between the extent of DNA damage accumulation and the clinical response to alkylating agents.\textsuperscript{67} A rapid response and dissolution of myeloma clones with an impaired double stranded DNA repair pathway would release greater pro-thrombotic factors that could be either micro-particles with surface tissue factor or cytokines and tissue factor. The greater thrombogenesis due to increased dissolution of the myeloma clone may act additively with a dexamethasone-thalidomide interaction on plasma cells,\textsuperscript{68} giving rise to an increased number of VTEs in the MRC and HOVON studies.\textsuperscript{69,70} An alternative mechanism to explain the increase risk of a VTE associated with DNA repair genes could be based on the observation that thalidomide can protect endothelial cells from doxorubicin-induced apoptosis by restoring PAR-1 expression\textsuperscript{71} promoting sub-endothelial tissues factor exposure, endothelial dysfunction, platelet activation and consequently increase the thrombosis risk.\textsuperscript{71-73} Under these conditions decreased DNA repair capacity could promote clot formation at the endothelium.
The enrichment of cytokine mediated apoptosis genes in SNPs associated with thalidomide related thrombosis risk may also give clues to the role bortezomib and aspirin play in VTE management. Low rates of VTE are seen in myeloma patients treated with bortezomib (Velcade™) in thalidomide combinations, possibly through the prevention of the up regulation of pro-thrombotic molecules such as thrombomodulin, cytokines, and E-selectin by bortezomib. A number of clinical studies have suggested that aspirin is effective at preventing the excess of VTE seen in thalidomide exposed individuals. Aspirin is classically thought to inhibit platelet COX-2, reducing platelet adhesiveness, modulating risk of arterial thrombosis. Aspirin can also lead to decrease levels of circulating TNF-α, by inhibiting IKK and therefore NFkB. Higher levels of TNF-α and COX-2 lead to an increased risk of apoptosis in endothelial cells, which also become pro-adhesive to non-activated platelets. In a thalidomide treatment setting, aspirin may be able to inhibit thalidomide VTE mediated events by lowering circulating TNF-α.

Genetic analysis of the multi-factorial phenotype that is thalidomide related venous thrombosis is challenging. To minimise experimental artefacts that can be found in many association studies, we have associated a discrete clinical outcome from a homogenous population of similarly treated patients with high quality genotype data with stringent quality controls. We took a hypothesis driven candidate gene approach rather than a whole genome scan (WGS) based approach, because it was clear that the number of events to be analysed would be small and we were aiming to identify pertinent functional loci variants with moderate to large effect size. We accept that future GWS and sequencing approaches may add relevant variants in unknown pathways. As part of the analysis we took an exploratory approach to defining whether the SNPs identified could be used to identify patients at high risk of VTE and consequently guide clinical intervention. Guidelines have recently been established to govern clinical indicators for intervention but these prognostic factors can be difficult to identify and use clinically. The US Food and Drug Administration (FDA) and EMEA have published warnings suggesting the use of thromboprophylaxis with any IMiD based regimen. The results of this recursive partitioning analysis have identified a limited number of SNPs which when analysed together can predict the risk of VTE.
Testing for these SNPs has the potential for being clinically useful for identifying high risk patients for whom therapeutic intervention is required. For clinically defined high risk patients intervention strategies may not change, but for patients at genetic high risk for whom aspirin was the chosen strategy, intervention with warfarin or LMWH would be more appropriate.

Acknowledgments
This study was supported by International Myeloma Foundation (IMF) as part of the Bank on a Cure (BOAC) Consortium. We are also grateful for support received from Myeloma UK. We would like to thank the staff at the Haematological Malignancy Diagnostic Service, Leeds and members of the MRC Myeloma IX Trial Management Group, as well as consultants and patients entered into the MRC Myeloma IX, ECOG and Hovan-50 trials.

Author contributions
D.C.J wrote the paper, performed research and analyzed data
S.C performed research
C.R. performed research
A.H analyzed data
K.H analyzed data
N.J.D analyzed data
J.H analyzed data
H.G designed the research
J.A.C analyzed data
S.E.B analyzed data
G.H.J analyzed data
D.B designed the research
V.R designed the research
F.E.D designed the research and wrote the paper
B.D designed the research
J.C analyzed data
P.S designed the research
B.V.N designed the research
G.J.M designed the research and wrote the paper

The authors declare no conflict of interest.
References


### Table 1: Allele distributions of cross-trial validated associated thalidomide related VTE SNPs.

<table>
<thead>
<tr>
<th>SNP</th>
<th>CHR</th>
<th>Gene</th>
<th>Functional Class</th>
<th>Minor allele</th>
<th>MAF cases</th>
<th>Major allele</th>
<th>MAF controls</th>
<th>P</th>
<th>OR</th>
<th>L95</th>
<th>US</th>
<th><strong>Log P</strong></th>
<th><strong>P-value</strong></th>
<th>Trial</th>
<th>In Linkage</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2302387</td>
<td>7</td>
<td>ABCB4</td>
<td>coding-synon</td>
<td>A</td>
<td>0.085</td>
<td>A</td>
<td>0.375</td>
<td>0.0425</td>
<td>G</td>
<td>0.5557</td>
<td>0.3133</td>
<td>0.9857</td>
<td>0.04113</td>
<td>Myeloma IX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1049216</td>
<td>4</td>
<td>CASP3</td>
<td>untranscribed</td>
<td>C</td>
<td>0.19</td>
<td>C</td>
<td>0.2778</td>
<td>0.2981</td>
<td>T</td>
<td>0.5769</td>
<td>0.3804</td>
<td>0.875</td>
<td>0.02068</td>
<td>Myeloma IX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs506504</td>
<td>11</td>
<td>CHEK1</td>
<td>coding-non-synon</td>
<td>T</td>
<td>0.00633</td>
<td>C</td>
<td>0.1111</td>
<td>0.03115</td>
<td>C</td>
<td>2.409</td>
<td>1.059</td>
<td>5.481</td>
<td>0.04414</td>
<td>Myeloma IX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7011</td>
<td>14</td>
<td>CINP</td>
<td>coding-synon</td>
<td>T</td>
<td>0.3061</td>
<td>C</td>
<td>0.3899</td>
<td>0.2296</td>
<td>C</td>
<td>1.52</td>
<td>1.034</td>
<td>2.233</td>
<td>0.0361</td>
<td>Myeloma IX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1202378</td>
<td>1</td>
<td>DCURE1B</td>
<td>coding-synon</td>
<td>T</td>
<td>0.2323</td>
<td>C</td>
<td>0.3333</td>
<td>0.1536</td>
<td>T</td>
<td>1.667</td>
<td>1.083</td>
<td>2.665</td>
<td>0.02621</td>
<td>Myeloma IX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs342311</td>
<td>10</td>
<td>ERE6</td>
<td>coding-synon</td>
<td>T</td>
<td>0.285</td>
<td>C</td>
<td>0.1667</td>
<td>0.197</td>
<td>C</td>
<td>2.036</td>
<td>4.2</td>
<td>0.3973</td>
<td>0.44</td>
<td>0.2614</td>
<td>Hovon50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs299295</td>
<td>5</td>
<td>HMMR</td>
<td>coding-synon</td>
<td>T</td>
<td>0.285</td>
<td>C</td>
<td>0.3421</td>
<td>0.1979</td>
<td>C</td>
<td>1.576</td>
<td>1.059</td>
<td>2.346</td>
<td>0.04119</td>
<td>Myeloma IX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs82537</td>
<td>3</td>
<td>IL1A</td>
<td>intron</td>
<td>A</td>
<td>0.515</td>
<td>A</td>
<td>0.5933</td>
<td>0.3996</td>
<td>G</td>
<td>1.132</td>
<td>2.251</td>
<td>0.00488</td>
<td>Myeloma IX</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs12049476</td>
<td>7</td>
<td>LEP</td>
<td>Promoter</td>
<td>T</td>
<td>0.405</td>
<td>G</td>
<td>0.3333</td>
<td>0.3115</td>
<td>G</td>
<td>1.504</td>
<td>1.054</td>
<td>1.047</td>
<td>0.02634</td>
<td>Myeloma IX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs20579</td>
<td>19</td>
<td>LIG1</td>
<td>untranscribed</td>
<td>T</td>
<td>0.065</td>
<td>T</td>
<td>0.05566</td>
<td>0.1263</td>
<td>C</td>
<td>0.2188</td>
<td>0.481</td>
<td>0.2544</td>
<td>0.01666</td>
<td>Myeloma IX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs13815</td>
<td>22</td>
<td>MT</td>
<td>coding-synon</td>
<td>T</td>
<td>0.2576</td>
<td>C</td>
<td>0.5933</td>
<td>0.3679</td>
<td>G</td>
<td>0.5961</td>
<td>0.4078</td>
<td>0.8716</td>
<td>0.01617</td>
<td>Myeloma IX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2410558</td>
<td>8</td>
<td>NAT2</td>
<td>locus</td>
<td>T</td>
<td>0.215</td>
<td>T</td>
<td>0.3333</td>
<td>0.3032</td>
<td>C</td>
<td>0.8295</td>
<td>0.4208</td>
<td>0.9416</td>
<td>0.02383</td>
<td>Myeloma IX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3774968</td>
<td>4</td>
<td>NRP1</td>
<td>intron</td>
<td>A</td>
<td>0.355</td>
<td>A</td>
<td>0.3421</td>
<td>0.4581</td>
<td>G</td>
<td>0.651</td>
<td>0.475</td>
<td>0.9263</td>
<td>0.01757</td>
<td>Myeloma IX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs185414</td>
<td>1</td>
<td>PARP1</td>
<td>coding-synon</td>
<td>C</td>
<td>0.25</td>
<td>C</td>
<td>0.3421</td>
<td>0.3376</td>
<td>T</td>
<td>0.2091</td>
<td>0.6539</td>
<td>0.4459</td>
<td>0.9592</td>
<td>0.03326</td>
<td>Myeloma IX</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2267669</td>
<td>6</td>
<td>PPARD</td>
<td>intron</td>
<td>G</td>
<td>0.198</td>
<td>A</td>
<td>0.0199</td>
<td>0.2319</td>
<td>A</td>
<td>0.861</td>
<td>0.475</td>
<td>0.9263</td>
<td>0.01757</td>
<td>Myeloma IX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs207682</td>
<td>7</td>
<td>SERPINE1</td>
<td>intron</td>
<td>C</td>
<td>0.365</td>
<td>C</td>
<td>0.3333</td>
<td>0.4562</td>
<td>T</td>
<td>0.3628</td>
<td>0.475</td>
<td>0.9263</td>
<td>0.01757</td>
<td>Myeloma IX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs12922317</td>
<td>16</td>
<td>TNSFRF17</td>
<td>intron,TagSNP</td>
<td>G</td>
<td>0.29</td>
<td>G</td>
<td>0.3333</td>
<td>0.3682</td>
<td>T</td>
<td>0.9828</td>
<td>1.986</td>
<td>0.04731</td>
<td>Myeloma IX</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2440</td>
<td>2</td>
<td>XRCDC5</td>
<td>untranscribed</td>
<td>T</td>
<td>0.4286</td>
<td>T</td>
<td>0.4167</td>
<td>0.3486</td>
<td>C</td>
<td>0.6185</td>
<td>1.401</td>
<td>0.9828</td>
<td>0.04731</td>
<td>Myeloma IX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: Simplified treatment arms of ECOG EA100, HOVON-50 and Myeloma IX studies.

Figure 2: A Venn diagram showing overlapping VTE associated genes between thalidomide Myeloma IX, non-thalidomide Myeloma IX and thalidomide Hovon 50/ECOG EA100 analyses

Figure 3: Forest plots showing distribution of validated SNPs associated with thalidomide related VTE across the Myeloma IX, Hovon-50 and ECOG EA100 trials

Figure 4: Predictive tree of thalidomide related thrombosis in myeloma patients following recursive partitioning analysis

Figure 5: Thalidomide treatment in combination with alkylating agents in myeloma promotes pro-thrombotic conditions at the endothelium surface via a combination of mechanisms, including; rapid apoptosis of myeloma cells leading to circulating tissue factor (TF), exposed tissue factor by endothelium cells salvaged from apoptosis, increased circulating cytokines e.g. TNFα, with T cell activation by antigen presenting cells (APCs); and activated platelets in response to increased circulating cytokines.
Figure 1

- Thalidomide + Dexamethasone vs. ECOG EA100 vs. Dexamethasone
- Thalidomide + Dexamethasone + Doxorubicin vs. HOVON-50 vs. Vincristine + Doxorubicin + Dexamethasone
- Thalidomide + Dexamethasone + Cyclophosphamide vs. Myeloma IX intensive vs. Vincristine + Doxorubicin + Dexamethasone + Cyclophosphamide
- Thalidomide + Dexamethasone + Cyclophosphamide vs. Myeloma IX non-intensive vs. Melphalan + Prednisolone
Figure 2: 
Figure 4:

Recursive Partitioning for DVT

<table>
<thead>
<tr>
<th>rs number</th>
<th>Rank*</th>
<th>Univariate Pval*</th>
<th>Functional type</th>
<th>Gene Symbol</th>
<th>Target Allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7011</td>
<td>1</td>
<td>0.001231074</td>
<td>coding-nonsynonymous</td>
<td>CINP</td>
<td>C/T</td>
</tr>
<tr>
<td>rs289747</td>
<td>5</td>
<td>0.007021794</td>
<td>3' UTR</td>
<td>CETP</td>
<td>A/G</td>
</tr>
<tr>
<td>rs610529</td>
<td>7</td>
<td>0.008176454</td>
<td>intron, TagSNP</td>
<td>ALDH1A1</td>
<td>C/T</td>
</tr>
<tr>
<td>rs3629963</td>
<td>17</td>
<td>0.017841244</td>
<td>Promoter</td>
<td>CDKN1A</td>
<td>A/C</td>
</tr>
<tr>
<td>rs2608555</td>
<td>24</td>
<td>0.027787883</td>
<td>coding-synonymous</td>
<td>GAN</td>
<td>C/T</td>
</tr>
<tr>
<td>rs699947</td>
<td>30</td>
<td>0.031169521</td>
<td>Promoter</td>
<td>VEGF</td>
<td>A/C</td>
</tr>
<tr>
<td>rs168351</td>
<td>80</td>
<td>0.143913327</td>
<td>intron, TagSNP</td>
<td>ALDH1A1</td>
<td>A/G</td>
</tr>
</tbody>
</table>

*Using Fisher’s exact test
Figure 5: