

RESEARCH ARTICLE

# Comprehensive molecular profiling of 718 Multiple Myelomas reveals significant differences in mutation frequencies between African and European descent cases

Zarko Manojlovic<sup>1,2\*</sup>, Austin Christofferson<sup>2</sup>, Winnie S. Liang<sup>2</sup>, Jessica Aldrich<sup>2</sup>, Megan Washington<sup>2</sup>, Shukmei Wong<sup>2</sup>, Daniel Rohrer<sup>3</sup>, Scott Jewell<sup>3</sup>, Rick A. Kittles<sup>4</sup>, Mary Derome<sup>5</sup>, Daniel Auclair<sup>5</sup>, David Wesley Craig<sup>1</sup>, Jonathan Keats<sup>2</sup>, John D. Carpten<sup>1,2</sup>

**1** Department of Translational Genomics, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States of America, **2** Translational Genomics Research Institute, Phoenix, AZ, United States of America, **3** Van Andel Research Institute, Grand Rapids, MI, United States of America, **4** Department of Surgery, Division of Population Genetics, University of Arizona, Tucson, AZ, United States of America, **5** Multiple Myeloma Research Foundation, Norwalk, CT, United States of America

\* [Zarko.manojlovic@med.usc.edu](mailto:Zarko.manojlovic@med.usc.edu)



**OPEN ACCESS**

**Citation:** Manojlovic Z, Christofferson A, Liang WS, Aldrich J, Washington M, Wong S, et al. (2017) Comprehensive molecular profiling of 718 Multiple Myelomas reveals significant differences in mutation frequencies between African and European descent cases. *PLoS Genet* 13(11): e1007087. <https://doi.org/10.1371/journal.pgen.1007087>

**Editor:** Petar Stojanov, UNITED STATES

**Received:** May 30, 2017

**Accepted:** October 23, 2017

**Published:** November 22, 2017

**Copyright:** © 2017 Manojlovic et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The Multiple Myeloma Research Foundation (MMRF) CoMMpass (Relating Clinical Outcomes in MM to Personal Assessment of Genetic Profile) trial (NCT01454297) is a longitudinal observation study of 1000 newly diagnosed myeloma patients receiving various standard approved treatments that aim at collecting tissue samples, genetic information, Quality of Life (QoL) and various disease and clinical outcomes over 10 years. Study Weblink: <https://urldefense.proofpoint.com/v2/url?>

## Abstract

Multiple Myeloma (MM) is a plasma cell malignancy with significantly greater incidence and mortality rates among African Americans (AA) compared to Caucasians (CA). The overall goal of this study is to elucidate differences in molecular alterations in MM as a function of self-reported race and genetic ancestry. Our study utilized somatic whole exome, RNA-sequencing, and correlated clinical data from 718 MM patients from the Multiple Myeloma Research Foundation CoMMpass study Interim Analysis 9. Somatic mutational analyses based upon self-reported race corrected for ancestry revealed significant differences in mutation frequency between groups. Of interest, *BCL7A*, *BRWD3*, and *AUTS2* demonstrate significantly higher mutation frequencies among AA cases. These genes are all involved in translocations in B-cell malignancies. Moreover, we detected a significant difference in mutation frequency of *TP53* and *IRF4* with frequencies higher among CA cases. Our study provides rationale for interrogating diverse tumor cohorts to best understand tumor genomics across populations.

## Author summary

This study represents the largest comprehensive molecular analysis of ethnically defined newly diagnosed treatment naïve Multiple Myeloma (MM). We revealed significant differences in mutation frequencies for important cancer genes between AA and CA MM. This study provides support for interrogating diverse tumor cohorts to best understand tumor genomics across populations.

u=https-3A\_www.themmr.org\_&d=DwlCAw&c=cIK7kQUTWtAVEOVIgvi0NU5BOUHhpN0H8p7CSfnc\_gl&r=rvfVnh4sABPo4MYhXxHnN6nyBTCIX6CKfeq-EEXJQfQ&m=aHWkANVSUoNo8KRzmN8nffZUUiSrG1bTUr5aciciswk&s=KgM\_JmmlPfg5auOrKOd0v42b5eVTiZMSCyTD9KiCErc&e= Study Type: "CoMMpass Longitudinal. The study has been registered with dbGAP (dbGaP Study Accession: phs000748).

**Funding:** This work was supported from Multiple Myeloma Research Foundation (CoMMpass) MMRF-TGen Carpten and USC-Carpten and Start-up Fund, University of Southern California to JDC. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Multiple Myeloma (MM) is a malignancy of plasma cells provoked by immunoglobulin gene rearrangements, accounting for slightly more than 10% of all hematologic cancer diagnoses in the US [1–3]. Pathogenesis evolves from an asymptomatic premalignant stage of clonal plasma cell proliferation termed “monoclonal gammopathy of undetermined significance” (MGUS) [4, 5]. MGUS is present in more than 3% of the population above the age of 50 and progresses to MM, or related malignancy at a rate of 1% per year [2, 6]. High-levels of pathogenic heterogeneity in MM among distinct racial/ethnic distributions have been outlined by epidemiology [7]. For instance, African American (AA) patients matched for socioeconomics, age, and gender are three times more likely to be diagnosed with MM, and with death rates that double those observed among Caucasians (CA) [4, 8, 9]. In addition, reports have shown that AA have decreased frequency of IgM monoclonal gammopathy, and have an earlier age of onset compared to CA patients [4, 8, 10]. However, over the past decade there were effective improvements in treatments and disease management that contributed to an astonishing increase in overall survival for MM, but these improvements were observed predominantly in CA patients [10, 11]. Therefore, a deeper understanding of oncogenic processes driving MM pathogenesis in statistically powered multi-ethnic cohorts is still needed to addressing disparities in incidence and outcomes observed among AA or otherwise African descent patients.

Several previous profiling studies have provided a view of the somatic landscape of MM [12, 13]. However, the representation of tumors from AA has been critically limited. To date, only one study has been reported comparing the frequencies of molecular alterations in MM between AA and CA cases [14]. This study revealed lower frequencies of IgH translocations by Fluorescence In Situ Hybridization among AA [14]. Although seminal, the small sample size of tumors from AA, lack of coding mutation data, and incomplete access to clinical data represented limitations of this study [14].

Recently, a comprehensive longitudinal study (CoMMpass) was initiated with the overall goal to prospectively observe the natural history of MM through comprehensively profiling 1,000 MM cases at diagnosis, with multiple biological and clinical follow-up points throughout the course of clinical management. Genomic profiling includes whole exome sequencing of germline and MM tumors, low pass whole genome sequencing, and RNA sequencing of tumors. Data are publicly distributed as Interim Analyses throughout the course of the study. Interim Analysis 9 (IA9) includes whole exome sequencing (WES) data from bone marrow tumor extracts with matching normal from 796 newly diagnosed MM cases. Patient ethnicity was one of the demographic parameters collected for each patient that included self-identification categories for African American, Caucasian, Asian, Hispanic, Middle Eastern, Other, Declined, and Unknown. This enabled us to perform genomic analyses to assess potential somatic differences from tumors based upon self-reported race. Alternatively, genetic ancestry is characterized by population genetic informative markers derived from allele frequencies of single nucleotide variances across biogeographical distributions [15] and is another way to characterize populations and individuals. The strength of genetic ancestry is that it offers information on ancestral genetic contributions based upon percent admixture within a given individual. While genetic ancestry provides molecular information with direct biological implications within the context of a disease [16, 17], we cannot disregard the importance of self-reporting that is influenced by socio-environmental behaviors that also play a critical role in disease risk. Most MM studies to date have been based upon self-reported race information. However, the availability of self-reported race and WES data from CoMMpass IA9 provided the unprecedented opportunity to search for novel and statistically significant somatic alterations relative to ancestrally-defined population differences in MM cases.

## Results

### CoMMpass IA9 data metrics

For this study, we utilized a subset of the CoMMpass IA9 dataset comprised of high quality WES and RNA-seq data from 721 newly diagnosed MM patients that were self-reported as either AA (N = 128) or CA (N = 593) with longitudinal clinical follow-up (S1 Table). For the CoMMpass study, MM tumor specimens were enriched from bone marrow aspirates by CD138 antibody conjugation yielding on average 99% CD138+ plasma tumor cell purity. Genomic data utilized for the downstream analysis for this study includes matching germline-tumor WES data on 718 MM cases (S1 Table). Sequencing statistics are provided in (S2 Table). For WES data, we achieved a mean coverage of 124X for MM tumor samples and 126X for germline samples. The average percentage of reads mapping to the WES target regions at 30X was 93% for tumor samples and 94% for germline samples. For RNA-seq, we generated an average of 200 million read pairs per sample with 88% mapping on average to annotated gene regions.

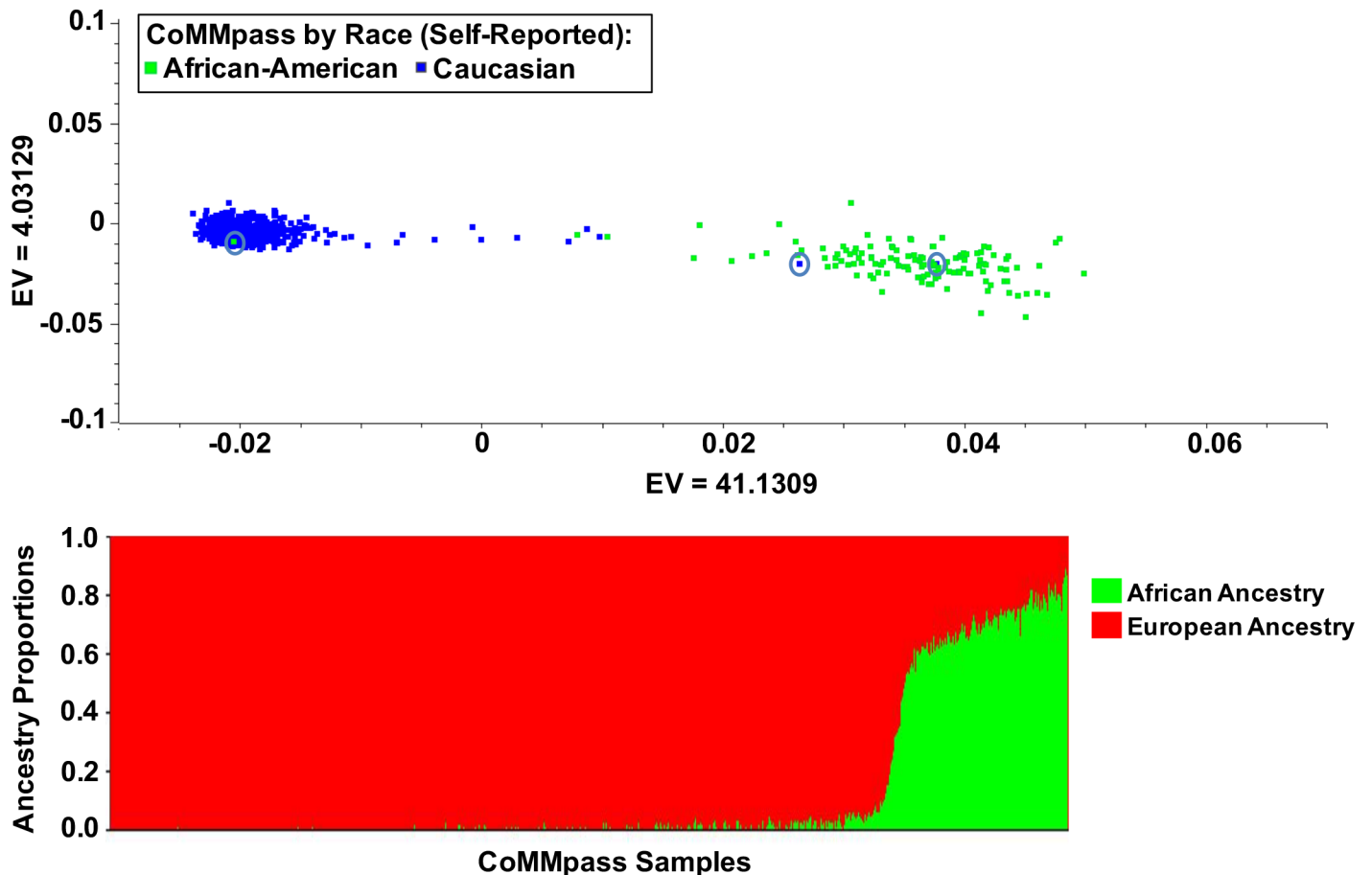
### Characterization of CoMMpass IA9 cases by genetic ancestry

To delineate genetic ancestry, we used a population stratification principal component analysis (PCA) to cluster MM patients by extracting 4,761 Ancestry Informative Markers (AIMs) SNP genotypes derived from germline WES CoMMpass IA9 data (Fig 1A). This allowed us to assess the distribution of genetic ancestry for CoMMpass cases [18]. We also utilized STRUCTURE [19] to determine individual percent ancestry for each CoMMpass case (Fig 1B, S1 Table). Analysis of individual ancestry data revealed that two self-reported CA had greater than 55% African ancestry and one self-reported African American had 99.9% European ancestry (Fig 1A, red circles, S1 Table). These three cases were excluded from our analyses. This resulted in a total of 127 African American and 591 Caucasians that were used for all downstream analyses. The mean European admixture among self-reported AA was 31% (range; 11%–67.8%). The mean west African admixture among self-reported CA was 0.1% (range; 0–34.3).

Analysis of demographics data stratified by race confirmed the previously reported finding by Waxman et al. [10] of a significant ( $p = 0.004$ ; Fisher's) two-fold increase in early age of onset (40–49 years) of MM among AA cases (11%) compared to CA cases (4.6%) (Fig 2A, S3 Table). In addition, there was a reverse effect in later ages of onset (70–79 years) with significantly higher frequency ( $p = 0.04$ ; Fisher's) in CA (22%) compared to AA (14%) (Fig 2A, S3 Table). Interestingly, our data showed no significant difference in overall survival based upon, race, age of onset, and MM karyotype in this cohort of similarly treated MM cases (Fig 2A and 2B, S1 Fig).

### Comparison of somatic mutation profiles between AA and CA MM cases

Somatic mutational analysis of the WES data was performed using a modified Mutation Significance (MutSig CV) algorithm [20] with custom scripts designed to detect differentially mutated genes between AA and CA MM tumors (Fig 3). We did not detect a statistically significant difference in mean nonsynonymous mutation frequencies between AA (mean = 63) and CA (mean = 68) MM cases ( $p = 0.574$ ) (S3A Fig). Furthermore, there was no difference in the mutational signature between AA and CA MM cases (S3B Fig). Somatic mutational analysis across the entire cohort confirmed common mutated MM driver genes such as *KRAS*, *NRAS*, *BRAF*, *TP53*, *DIS3*, and *FAM46C* (Fig 3, S4 Table) [12, 13, 21]. Our comparison of mutated genes between tumors from AA and CA cases identified *RYR1*, *RPL10*, *PTCHD3*, *BCL7A*, *SPEF2*, *MYH13*, *ABI3BP*, *BRWD3*, *GRM7*, *AUTS2*, *PARP4*, *PLD1*, *ANKRD26*, *DDX17* and

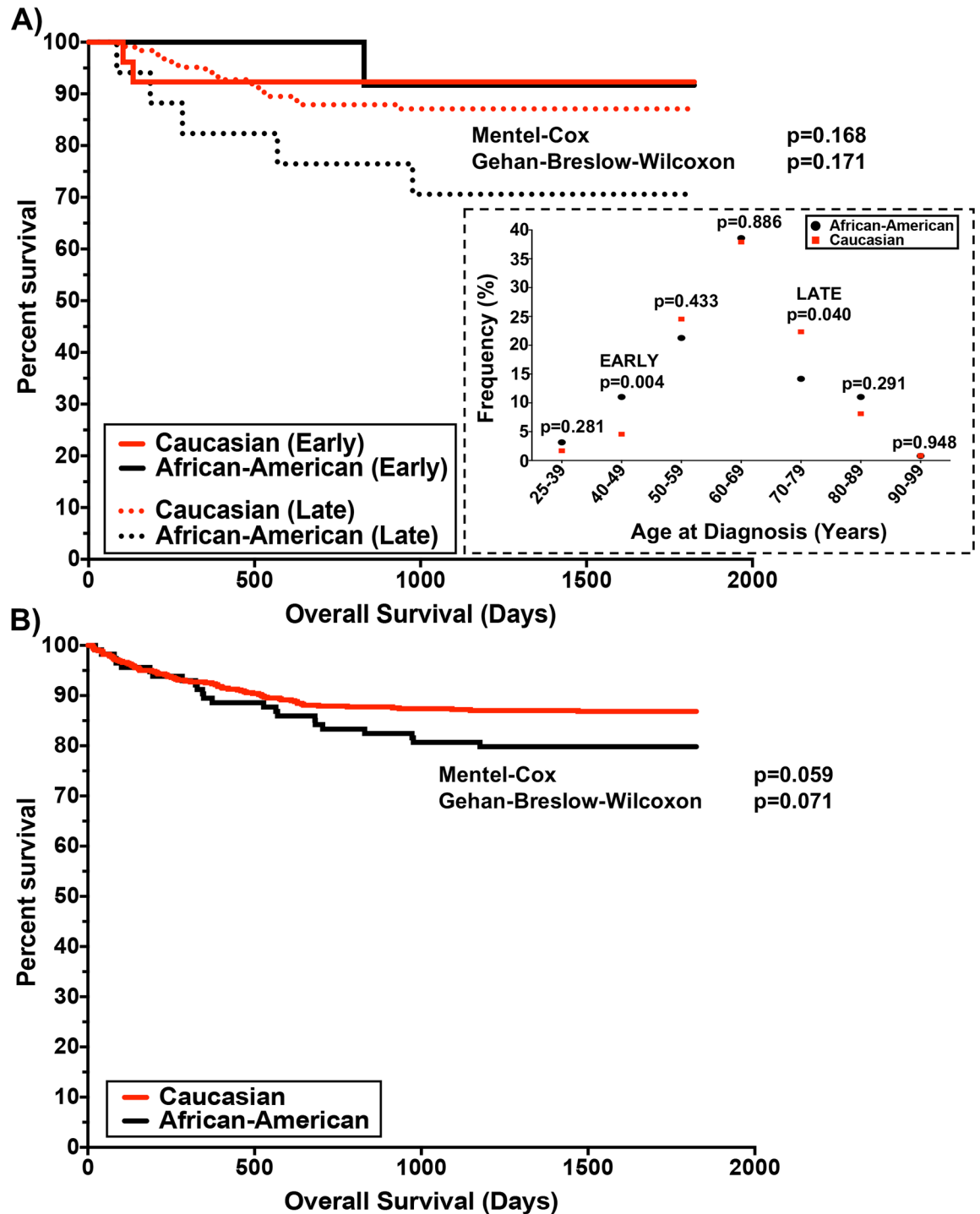


**Fig 1. Population stratification by principal component analysis and STRUCTURE plot defining propositions of calculated genetic ancestry.** (A) Principal component plot across all samples by self-reported race with Caucasian (blue dots) and African American (green dots) using AIMs derived from the whole-exome deep sequencing. The red circles indicate samples that have been removed from the analysis due to misclustering. PCA is calculated using SNP & Variation Suite v8.4.1 (Golden Helix, Inc.) PCA tool by eigenvalue (EV) implementation technique (Methodology). These samples were identified as the self-reported race. (B) STRUCTURE plot (K = 2; 50,000 Burnin period and 100,000 MCMC repeats) used to interfere genetic clusters and percent admixture: European Ancestry (red), African Ancestry (green).

<https://doi.org/10.1371/journal.pgen.1007087.g001>

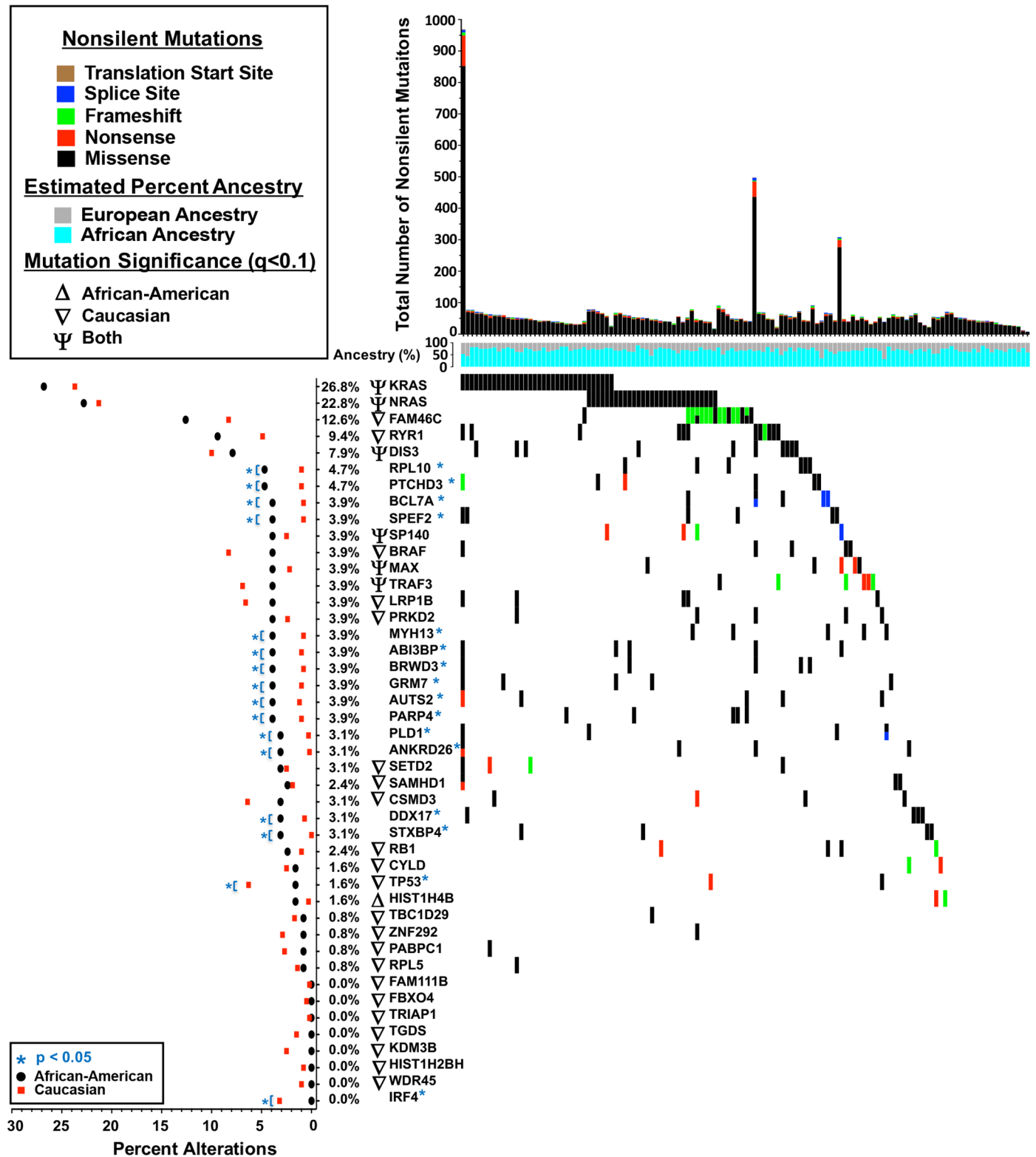
*STXBP4* as genes with significantly higher mutation frequencies in AA MM cases (Fig 3, S4 Table). *FAM46C*, which is commonly mutated in MM, also exhibited a trend toward higher frequency in AA (12.6%) versus CA (8.3%) MM, however the difference did not reach statistical significance ( $p = 0.09$ ). In addition, we further examined differences in BRAF mutation frequency between AA and CA cases. Although we did not detect a difference in overall BRAF mutation frequency, we did observe a difference in BRAF<sup>V600E</sup> mutation between AA (0.8%) versus CA (4.34%), but this difference did not reach nonimnal significance ( $p = 0.053$ ; Fisher’s (S1 Table).

Among the most interesting observations of our mutational analysis was the significantly higher frequency of *IRF4* ( $p = 0.041$ ) and *TP53* ( $p = 0.035$ ) mutations in CA MM cases (Fig 3, S4 Table). Specifically, this analysis revealed a *TP53* somatic mutation frequency of 6.3% in the CA MM cases compared to 1.6% in AA MM cases ( $p = 0.035$ ) (Fig 3, Table 1, S2 Fig, S4 Table). To verify our observation, we used an independent publically available MM somatic whole exome sequencing dataset consisting of 205 MM cases as a validation cohort published by Lohr et al (S5 Table) [13]. This dataset consisted of a mix of newly diagnosed and relapse MM specimens. Although there were only 14 self-reported AA MM cases within this validation



**Fig 2. Overall survival of patients based upon age of MM onset and race.** (A) Kaplan-Meier analysis with long-rank test for overall survival data from CoMMpass IA9 cases stratified by the impact of the early or late onset of MM. The data in the black box demonstrate the distribution of onset across the IA9 data set. The early (term used to label onset between 40–49 years), and late (term used to label onset between 70–79 years) that were significantly different using Fisher’s exact test between the stratified populations. Samples size also summarized in supplemental table (S3 Table) is as following: 25–39 [AA (4/127) vs CA (10/591)], 40–49 [AA (14/127) vs CA (27/591)], 50–59 [AA (27/127) vs CA (145/591)], 60–69 [AA (49/127) vs CA (224/591)], 70–79 [AA (118/127) vs CA (132/591)], 80–89 [AA (14/127) vs CA (48/591)], 90–99 [AA (1/127) vs CA (5/591)]. (B) Kaplan-Meier analysis with long-rank test for overall survival data from CoMMpass IA9 impact of incidents of MM by race-matched-ancestry.

<https://doi.org/10.1371/journal.pgen.1007087.g002>



**Fig 3. Frequencies of recurrently mutated genes in human MM.** (Top) The total number of acquired nonsilent somatic mutations across 127 tumors from African American patients. Percent ancestry track is indicating the distribution of genetic ancestry among each sample. (Center) mutations across recurrent genes among African American patients colored by mutation type. (Left) Mutation significance analysis was performed using MutSigCV (Methods) on two cohorts African American ( $n = 127$ ) and Caucasian ( $n = 591$ ) independently. For each analysis, significance was determined using

false discovery rate ( $q < 0.1$ ) with significant genes labeled as following: exclusive to African American ( $\Delta$ ) and Caucasian ( $\nabla$ ), or if the same gene is identified by both analysis “Both” ( $\psi$ ). (Left) Histogram depicting percent of alterations in each gene between African American (black) and Caucasian (red) cohorts using the Fisher’s exact test with significance set at (blue; \* =  $p < 0.05$ ) between the two stratified populations.

<https://doi.org/10.1371/journal.pgen.1007087.g003>

cohort, we observed differences in *TP53* coding mutations compared between CA (14/157; 8.9%) and AA (0/14; 0%), providing an independent validation, albeit with limited power. Given the significantly higher *TP53* mutation rate among CA MM cases prompted us to assess the distribution of *TP53* mutation status as a function of European ancestry. The analysis demonstrated that *TP53* mutations were strongly associated with MM cases that have high European ancestry ( $>95\%$ ) ( $p = 0.01$ ; Wilcoxon rank-sum test) (Fig 4).

### Comparison of somatic copy number changes between AA and CA MM cases

To uncover potential racial or ancestry differences in copy number events across CoMMpass IA9 WES data, we utilized GISTIC 2.0 analysis [20, 22]. This analysis identified several regions of the genome associated with common copy number gain and loss (S4 Fig, S6 Table). However, we did not detect any statistically significant differences in specific focal copy number events between AA and CA cases. However, to further expand upon our understanding of *TP53* loss based upon mutational analysis data, an integrated somatic copy number and mutational analysis of the *TP53* locus was performed. This analysis revealed a predominance of bi-allelic *TP53* events among CA MM cases (Table 1), however the difference was not statistically significant. Furthermore, integration of genomic data with clinical outcomes demonstrated that CoMMpass cases with tumors harboring bi-allelic perturbation in *TP53* have significantly ( $p = 0.027$ ; Mental-Cox Log-rank test) poorer overall survival (Fig 5). However, there is no difference in overall survival in MM patients with tumors demonstrating mono-allelic events (loss of copy, or mutation only) and wild type *TP53* (Fig 5).

### Assessment of MM transcriptional signatures among MM derived from AA and CA cases

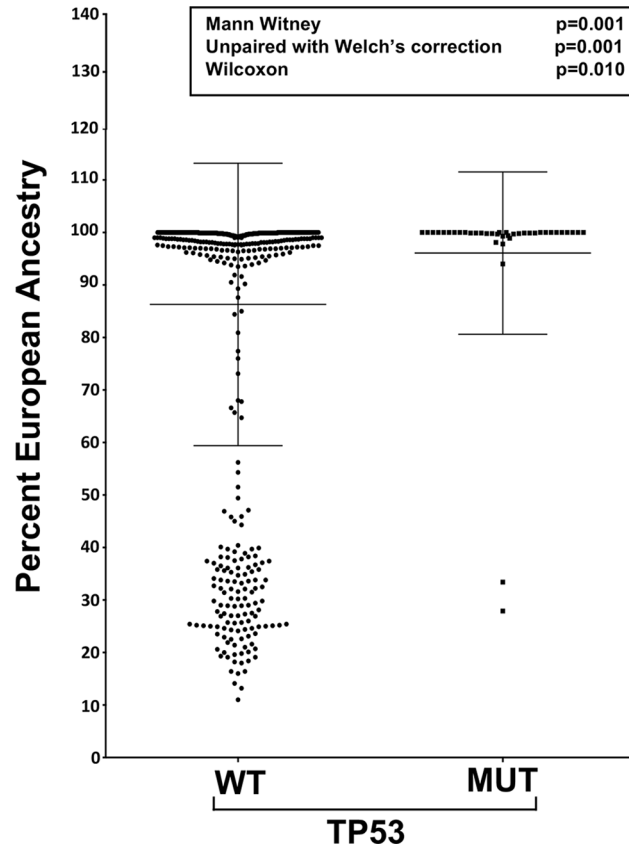
We performed gene expression profile analysis of RNA-seq data to compare the frequency of the University of Arkansas UAMC 70 gene high-risk signature between AA and CA MM cases [23]. Comparison between African and European ancestry and by self-reported race further sub-stratified by MM karyotype did not reveal a significant difference ( $p = 0.662$ ) of the UAMC high-risk signature consistent with our previous report from an independent data set (Table 2) [14]. Further analysis of Ki67 proliferation index showed no significant difference ( $p = 0.560$ ) of Ki67 profile among patients stratified by race or ancestry (S5 Fig)

**Table 1. TP53 mutation profile.**

TP53 Aberration	African-American	Caucasian	p-value (Fisher’s)
Wildtype	84.3%	83.1%	0.75
Mono-Alleic	9.4%	9.1%	0.91
Bi-Alleic	0.8%	4.1%	0.07
LOH	6.3%	8.6%	0.39
Mutation	1.6%	6.6%	0.03
LOH+Mutation	7.9%	15.2%	0.03

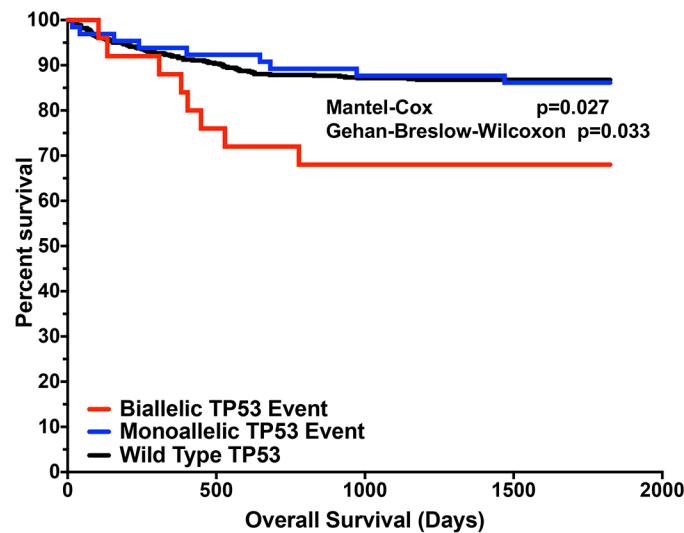
Wildtype (both *TP53* alleles are normal); Mono-Allelic (somatic mutation or copy loss event); Bi-Allelic (somatic *TP53* event in both alleles); LOH (loss of heterozygosity only), Mutation (somatic mutation only); LOH+Mutation (combined loss of heterozygosity and mutations)

<https://doi.org/10.1371/journal.pgen.1007087.t001>



**Fig 4. Alteration of TP-53 across ancestry.** Analysis of TP53 state across the percent European ancestry. Each dot represents individual sample. To validate statistical power, we performed Mann-Witney, Wilcoxon, and Unpaired t-test with Welch's correction with statistical significance set at  $p < 0.05$ .

<https://doi.org/10.1371/journal.pgen.1007087.g004>



**Fig 5. Overall survival of patients based upon TP53 locus alterations.** (A) Kaplan-Meier analysis with long-rank test for overall survival data from CoMMpass IA9 cases stratified by mono-allelic (blue) versus bi-allelic (red) alteration as well as wildtype (black) of the TP53 locus.

<https://doi.org/10.1371/journal.pgen.1007087.g005>



**Table 2. Frequency analysis of Arkansas 70 high-risk genes expression profile.**

Risk Groups	Frequency AA (%)	Frequency E A (%)	p value (Fisher's Exact)
Total	28	26	0.662
Hyperdiploid	20	20	0.999
Nonhyperdiploid	33	33	0.935

<https://doi.org/10.1371/journal.pgen.1007087.t002>

## Discussion

Not surprisingly, the vast majority of our current understanding of MM biology has been derived from data collected from largely European descent cohorts, even given significant disparities in disease incidence and outcomes seen among African American patients [10]. Through the analysis of CoMMpass data, we were able to analyze the largest multi-ethnic MM cohort at diagnosis to date, in an attempt to elucidate in-depth molecular differences in tumors derived from AA and CA cases to further understand the biological determinants of MM as a function of tumors derived from an ancestrally defined dataset. First, the nature of the CoMMpass dataset has allowed us to confirm previously reported results of published clinical and molecular studies of MM. Our study results were similar to previously reported data demonstrating earlier disease onset in MM among African Americans [10]. Furthermore, several MM genomics sequencing studies [12, 24, 25] have been reported in both newly diagnosed and relapse MM and we have validated commonly mutated genes in MM such as *KRAS*, *NRAS*, *FAM46C*, *DIS3* and *TP53*.

Differences in mutation frequencies in genes such as *RYR1*, *RPL10*, *PTCHD3*, *BCL7A*, *SPEF2*, *MYH13*, *ABI3BP*, *BRWD3*, *GRM7*, *AUTS2*, *PARP4*, *PLD1*, *ANKRD26*, *DDX17* and *STXBP4* that were more common in AA MM cases could possibly reflect differences of myelomagenesis by race and/or ancestry. *FAM46C* is among the most commonly mutated genes in MM based on several previously reported WES studies [12]. The reported *FAM46C* mutation frequency in MM ranges from 5%-11% [13, 26], and although the frequency was 8.3% in Caucasian cases in our study, we observed an increased frequency of 12.6% in African American cases in our study. Although the functional role of *FAM46C* in myelomagenesis is still yet to be determined, there seems to be a potential enrichment of its role in MM biology among tumors derived from patients of African ancestry.

BRAF mutations have also been reported in MM, generally at frequencies of 2.8%-5% [27–29]. Although overall BRAF mutation frequencies were not different between AA and CA, we did detect differences when stratifying specifically by BRAF<sup>V600E</sup>, with higher frequencies seen in CA (4%) as compared to AA (0.8%), although this difference did not exceed nominal significance due to limited power. Furthermore, we did not observe a difference in overall survival in primary MM cases harboring any BRAF mutation, nor specifically BRAF<sup>V600E</sup> mutation compared to BRAF wildtype cases ( $p = 0.439$ ,  $p = 0.579$ ; Long-rank, Gehan-Breslow-Wilcoxon). These results are of clinical consequence as BRAF mutations, particularly BRAF<sup>V600E</sup>, can be targeted with select BRAF inhibitors, which has shown effect in mutation positive MM cases [27, 29].

Three of the genes with significantly higher mutation frequency in AA are involved in other B-cell malignancies. *BCL7A* has been shown to be directly involved in a three-way gene translocation with *Myc* and *IgH* in high-grade B-cell non-Hodgkin lymphoma cell lines [30]. As a result of the gene translocation, the N-terminal region of the gene product is disrupted, which is thought to be related to the pathogenesis of a subset of high-grade B cell non-Hodgkin lymphoma [30]. The protein encoded by *BRWD3* is a bromodomain and WD repeat containing protein that is thought to have chromatin-modifying function, and may play a role JAK/STAT pathway activity [31]. Importantly, this gene is involved in translocations in B-cell chronic lymphocytic leukemia [32]. Finally, *AUTS2* is involved in translocations with *PAX5* in B-cell

precursor acute lymphoblastic leukemia and other cancers [33]. These findings are of considerable interest, and further validation and functional characterization of these genes in appropriate myeloma model systems could shed light on their potential role in myelomagenesis particularly in patients of African descent.

Another significant observation of our study was the difference in *TP53* mutation frequency among patients with higher European admixture in our dataset. Integrated mutation and copy number analysis demonstrated a trend towards higher *TP53* bi-allelic inactivation in tumors derived from CA cases. This could have translational significance as bi-allelic *TP53* inactivation is a universally validated predictor of poor outcome [34]. We further detected significant differences in *IRF4* mutation frequency, which was also higher among CA MM cases. *IRF4* is a credentialed MM oncogene, which is known to be an oncogenic fusion partner in MM, and is also among the known significantly mutated genes in MM [35, 36]. Moreover, *IRF4* activity is associated with poor outcome, with potential therapeutic implications for immunomodulatory agents in MM [37, 38]. These data suggest that the significant enrichment of *IRF4* mutations in CA cases could have strong clinical translational implications. Although the collection of longitudinal clinical data is ongoing for CoMMpass, one current limitation of this study is the limited long-term follow-up data from IA9. However, the CoMMpass study design will allow us to assess outcomes longitudinally, along with molecular profiles of both newly diagnosed and relapse disease in upwards of 1,000 MM cases.

Ultimately, our study represents among the most comprehensive (WES and RNAseq) genomics studies of a tumor type in patients of African descent, and sheds light on potential ancestry-related differences in biological mechanisms of myelomagenesis. Three genes that are known to be involved in B-cell malignancy translocations represent new candidate myeloma genes that may have been overlooked because of the lack of AA cases in most large genomic efforts. It is clear that there are molecular differences between MM tumors from AA and CA cases, and that it is absolutely critical to continue to delineate these observations to better improve clinical management of the disease. As the CoMMpass study matures, it will allow scientists to validate these findings as well as expand on studies such as recurrence, better elucidation of driving mutations, and clonal evolution during the course of treatment.

## Materials and methods

### Ethics statement

Samples were obtained under Multiple Myeloma Research Foundation (MMRF) CoMMpass Study Network Institutional Review Board approved Informed Consent; Copernicus IRB (IRB # QUI1-11-217).

### CoMMPass study synopsis

The design of the study is to prospectively profile newly diagnosed, treatment naive MM from 1000 patients with longitudinal clinical follow up. Tumor specimens collected at diagnosis and relapse are interrogated by whole-exome, modified low pass whole-genome, and or RNA sequencing. The longitudinal component requires clinical follow-up of each patient with collection of clinical parameters four times annually over a period of 10 years. Furthermore, each patient participating into the study underwent an IMiD and/or Proteasome inhibitor based treatment regimen at diagnosis determined by the treating oncologist.

CoMMpass data are systematically analyzed and periodically released in the form of Interim Analyses on a biannual basis. Interim Analysis 9 (IA9) is comprised of 796 unique baseline newly diagnosed bone-marrow samples with high quality WES data of which, 75 have confirmed progression with comprehensive clinical annotation. In addition, IA9 is comprised of

520 bone marrow baseline samples that were analyzed by both whole-exome and RNAseq platforms. The data is publically available at dbGAP accession number phs000748.

In this study, we performed our analysis on whole-exome data from 741 of treatment-naïve bone marrow derived MM matched normal samples from IA9, from those cases who self-reported race as either African American or Caucasian. To ensure high quality WES data for downstream copy number analysis, we removed samples that had maximum segmentation count above 2,500 to be in concordance with GISTIC 2.0 recommendation of maximum segment counts [22]. This resulted in final 721 samples that passed the quality threshold and were used to determine genetic ancestry across the samples.

## Sample preparation

Bone marrow aspirates and peripheral blood samples were collected from each patient. Bone marrow aspirates from each patient were subjected to immunomagnetic bead separation using the Miltenyi MACS Cell Separation System (Miltenyi, San Diego, CA) to enrich for CD138-positive malignant MM plasma cells. Only clinically eligible samples with greater than 250,000 cells recovered after CD138 enrichment, which are greater than 80% monoclonal light chain restricted plasma cells move forward for nucleic acid extraction. Genomic DNA was extracted from purified CD138-positive plasma cells (tumor) and matched peripheral blood samples (constitutional) using QIAamp DNA Mini Kit (Qiagen). Total RNA was extracted from CD138-positive plasma cells the using QiaAmp RNeasy Mini Kit (Qiagen). Nucleic acids were quality assessed using the Qubit 2.0 (Thermo Fisher) and Agilent Tape Station to determine quantity and integrity. Samples were stored at  $-80^{\circ}\text{C}$  for subsequent molecular analyses.

## Massively parallel sequencing of DNA and RNA from CoMMpass specimens

DNA samples were used for two different assays including whole exome sequencing (WES) and low pass long insert whole genome sequencing (WGS). WES was prioritized if material was limiting. For WES, 50ng-250ng of genomic DNA was fragmented to an average size of 180bp in length using a Covaris focused-ultrasonicator (Covaris). An Illumina sequencing technology compatible whole genome library was created using Kapa Biosystems Hyper Prep Kits. These libraries were then subjected to whole exome target enrichment using Agilent SureSelect V5+UTR hybrid capture kits.

For RNA-sequencing, either 150ng or 500ng of total RNA was used to enrich for poly-adenylated RNA molecules, which were subsequently fragmented to a target size of 180bp by heat fragmentation. Fragmented molecules were then converted to cDNA using random primers with Superscript II (Invitrogen). After second strand synthesis, the resulting molecules were used for library prep using the Illumina TruSeqRNA library kit.

## Massively parallel sequencing

Parallel sequencing of libraries was performed on Illumina HiSeq2000 or HiSeq2500 systems using version 3 or version 4 chemistry. WES was sequenced using paired-end 83x83bp reads while long-insert whole genome libraries were sequenced using paired-end 86x86bp reads. All sequencing reads were converted to industry standard FASTQ files using BCL2FASTQ v1.8.4.

## Bioinformatics analysis of massively parallel sequencing data alignments

FASTQ files are processed using a custom semi-automated pipeline based upon industry standard software packages and programs. Sequencing reads are initially aligned to the GRCh37

human genome reference using v0.7.8 BWA-MEM aligner [39] to generate BAM files. SAMTOOLS v0.1.19 [40] was used to sort BAM files and PICARD v1.111 (<http://broadinstitute.github.io/picard/>) to mark duplicate read pairs. Post alignment joint INDEL realignment and base quality scores recalibration was performed on the BAM files using GATK v3.1–1 [41].

## Variant detection and mutational analysis

For WES data, final BAM files for each patient's constitutional and tumor data were used for germline and somatic variant detection, respectively. Variants were called from germline and tumor BAM files individually using GATK Haplotype Caller v3.1–1 [42] and SAMTOOLS v0.1.19 [43]. Somatic mutations including single nucleotide variants SNVs and INDELS were called using each patient's germline and tumor BAMs by three independent software packages including MUTECT [44], STRELKA [45], and SEURAT [46]. To make the final mutation list, a mutation had to be detected by at least two out of three independent callers used.

Somatic copy number analysis was performed on WES germline-tumor pairs. For these studies, we utilized the DNACopy segmentation module in BioConductor [47]. We also utilized a comparative germline-tumor copy number approach where by raw data was normalized to physical coverage using circular binary segmentation as well as filtered to remove repetitive regions prior to calculating log<sub>2</sub> comparison across germline-tumor exome data.

Somatic events were assembled in VCF and MAF formats and further annotated using SNPeff [48] to provide additional information on gene states and variant effects.

For RNA-sequencing analysis, we employed the TopHat v2.0.11 for alignment of RNA-seq reads, CuffDiff v2.2.1 for differential expression analysis, and Salmon 0.7.2 for isoform quantification [49].

## Secondary analysis of somatic alterations

Genotypes for germline variants with >98% detection across all samples with exome data, and used for somatic analysis, were deduced using the SNP & Variation Suite v8.4.1 (Golden Helix, Inc., Bozeman, MT, [www.goldenhelix.com](http://www.goldenhelix.com)) genotype tool. Exome specific genome-wide Ancestry Informative Markers (AIMs) were derived from Kosoy et al. [50], Price et al. [51], Tandon et al. [52], and using informativeness estimation established by Rosenber et al. [53]. Population stratification principal component analysis (PCA) was calculated using the SNP & Variation Suite v8.4.1 (Golden Helix, Inc., Bozeman, MT, [www.goldenhelix.com](http://www.goldenhelix.com)) Genotype PCA tool that implements eigenvalue technique described by Patterson et al. 2006 [54] and Price et al. 2006 [51]. The genotype file containing AIMs was further formatted for STRUCTURE analysis using plink and PGDSpider v2.1.0.3 [55]. STRUCTURE was performed as described by Pritchard et al. [19] with set Burning period for each replicate at 50,000 with consecutive 100,000 iterations of MCMC repetitions. Each genetic cluster was run with 3 independent replicates and the number of populations (K) was estimated by implementing both L(K) and ΔK [19]. The reference populations used for the putative ancestral populations were derived from publically available 1000G Population Exome Phase1\_v3 Genotypes [18].

The analysis to identify significant driver mutations from WES somatic mutation data was performed using MUTSIG CV (Mutation Significance) algorithm [22] with adjusted covariates file using myeloma specific expression profile with significance set at  $q < 0.1$  and  $p < 0.05$ . GISTIC 2.0 (The Genomic Identification of Significant Targets in Cancer) was applied to define significantly altered somatic copy number focal events with  $q$  value cut off set at 0.25 [20, 22]. Mutation signatures were deduced using an industry standard publically available analysis tool at <https://bitbucket.org> "analysis-of-mutational-signatures".

## Genetic MM subgroups

Myeloma samples were further stratified by two subtypes: hyperdiploid or nonhyperdiploid. Hyperdiploidy was defined by presence of trisomy of at least three odd-numbered chromosomes [14]. The rest of the samples were identified as nonhyperdiploid subtype.

## High-Risk gene-expression signature

Gene expression profile was performed on RNA extracted from CD138+ plasma cells as described above using HiSeq2500 sequencer (Illumina, Inc.). RNA-seq Sailfish TPM values were log<sub>2</sub> transformed prior to the analysis. To determine the high-risk score, we utilized the UAMC 70-gene expression profile [23]. The high-risk expression signature was calculated and reported as percent frequency across each ancestry and MM subtype.

## Mutation signature

Mutation signature was deduced using all somatic point mutations except INDELs using a publically available tool (<https://bitbucket.org/jtr4v/analysis-of-mutational-signatures>).

## Statistical analysis

Each event between the stratified populations was analyzed appropriately either using the Fisher's exact test [56], non-parametric Mann-Whitney-Wilcoxon when assumption of normality is not maintained, and unpaired t-test with Welch's correction when normal distribution is assumed with unequal variances. Benjamin-Hochberg method [57] was used to adjust for multiple testing. Overall survival was inferred from the clinical follow-up data collection over the 4 years' span with estimations using Kaplan-Meier methods. The p-value of < 0.05 was set for statistical significance. Statistical software packages used throughout the study were R v3.1.1. (<https://www.r-project.org>) and GraphPad Prism 7 (GraphPad Software, Inc.).

## Supporting information

**S1 Fig. Overall survival by multiple myeloma karyotypes.** Analysis was performed using Kaplan-Meier method with long-rank test for group comparisons.  
(TIF)

**S2 Fig. Mutation distribution across TP53 protein structure.** cBioPortal Mutation Mapper tool as described by [Gao et al. \*Sci. Signal.\* 2013](#) & [Cerami et al. \*Cancer Discov.\* 2012](#) was applied to generate the mutation profile across the TP53 domains with top representing mutation profile among Caucasian, and bottom representing African American respectively.  
(TIF)

**S3 Fig. Somatic mutation frequency of significantly mutated genes in tumors from African-American and Caucasian.** (A) Comparison of nonsilent mutation frequency between ancestry and self-reporting using Wilcoxon, Mann-Whitney, and Unpaired t-test to determine statistical significance. (B) Mutation signature associated with African and European ancestry.  
(TIF)

**S4 Fig. GISTIC 2.0 plot.** Differentially altered events indicated by arrows. The red graph indicate copy number gains and blue is deletions for each stratified group.  
(TIF)

**S5 Fig. Expression profile.** Ki67 expression profile in TPM across patients with MM.  
(TIF)

**S1 Table. Genomic data utilized, gender, reported race, and inferred genetic ancestry utilized in this study.**

(XLSX)

**S2 Table. Sequencing statistics matrix for Whole exome Sequencing (WES), and RNA sequencing.** (RNAseq) used in this study.

(XLSX)

**S3 Table. Clinical outline, treatment summary, and karyotype summary.**

(PDF)

**S4 Table. Summary of significantly mutated genes.**

(PDF)

**S5 Table. Data downloaded from Lohr, et al., cancer cell, 2014. 25(1).**

(XLSX)

**S6 Table. Significant focal copy number perturbations called by GISTIC 2.0.**

(XLSX)

## Acknowledgments

The authors would like to thank MMRF CoMMpass Network (Abella, Eugenia; Anderson, Larry; Ascensao, Joao; Azaceta, Gemma; Bahlis, Nizar; Balaraman, Rama; Bar, Michael; Bargay, Joan; Belani, Rajesh; Berdeja, Jesus; China, Anabelle; Conde, Miguel; Costello, Caitlin; Dakhil, Shaker; Fernández de Larrea, Carlos; Gasparetto, Cristina; Giever, Thomas; Graham, Mark; Granell, Miquel; Grossbard, Michael; Gutiérrez, Norma; Harroff, Allyson; Hassoun, Hani; Hernández, Miguel; Hofmeister, Craig; Hsu, Gerald; Jagannath, Sundar; Jakubowiak, Andrzej; Kambhampati, Suman; Kaya, Hakan; Klein, Leonard; Kolibaba, Kathryn; Krsnik, Isabel; Kumar, Shaji; Kuzma, Charles; Levy, Moshe; Lewis, DeQuincy; Liles, Darla; Lonial, Sagar; Lunning, Matthew; Martinez-Chamorro, Carmen; Masso Asensio, Pilar; Matkiwsky, May; Meehan, Kenneth; Menter, Alex; Mikhael, Joseph; Milner, Carter; Min, Frederick; Ngaiza, Justinian; Niesvizky, Ruben; Onitilo, Adedayo; Orloff, Gregory; Patel, Dilip; Posada, Juan; Richards, Donald; Rifkin, Robert; Ríos, Pablo; Robles, Robert; Rodríguez, Paula; Roy, Vivek; Sampol, Antonia; Scott, Emma; Sebag, Michael; Siegel, David; Solomon, William; Srinivas, Shanti; Trudel, Suzanne; Usmani, Saad; Venner, Christopher; Vij, Ravi; Volterra, Fabio; Wachsmann, William; Whittenberger, Brock; Wolf, Jeffrey; Xia, Chunzhi; Yalamanchili, Madhuri; Yost, Kathleen; Zonder, Jeffrey) for their contributions to the CoMMpass patient and data collection.

## Author Contributions

**Conceptualization:** Zarko Manojlovic, Jonathan Keats, John D. Carpten.

**Data curation:** Zarko Manojlovic, Austin Christofferson, Jessica Aldrich, Megan Washington, Shukmei Wong, David Wesley Craig, Jonathan Keats, John D. Carpten.

**Formal analysis:** Zarko Manojlovic, Austin Christofferson, Jessica Aldrich, Jonathan Keats, John D. Carpten.

**Funding acquisition:** Daniel Auclair, Jonathan Keats, John D. Carpten.

**Investigation:** Jonathan Keats, John D. Carpten.

**Methodology:** Zarko Manojlovic, Austin Christofferson, Winnie S. Liang, Jessica Aldrich, Daniel Rohrer, Rick A. Kittles, David Wesley Craig, Jonathan Keats, John D. Carpten.

**Project administration:** Winnie S. Liang, Daniel Rohrer, Scott Jewell, Mary Derome, Daniel Auclair, Jonathan Keats, John D. Carpten.

**Resources:** Winnie S. Liang, Daniel Rohrer, Scott Jewell, Mary Derome, Daniel Auclair, David Wesley Craig, Jonathan Keats, John D. Carpten.

**Software:** Zarko Manojlovic, Austin Christofferson, Jessica Aldrich, David Wesley Craig, Jonathan Keats.

**Supervision:** Daniel Auclair, David Wesley Craig, Jonathan Keats, John D. Carpten.

**Validation:** Zarko Manojlovic, David Wesley Craig, Jonathan Keats, John D. Carpten.

**Visualization:** Zarko Manojlovic, Rick A. Kittles, David Wesley Craig, Jonathan Keats, John D. Carpten.

**Writing – original draft:** Zarko Manojlovic, John D. Carpten.

**Writing – review & editing:** Zarko Manojlovic, Rick A. Kittles, Daniel Auclair, David Wesley Craig, John D. Carpten.

## References

- Greenberg AJ, Vachon CM, Rajkumar SV. Disparities in the prevalence, pathogenesis and progression of monoclonal gammopathy of undetermined significance and multiple myeloma between blacks and whites. *Leukemia*. 2012; 26(4):609–14. <https://doi.org/10.1038/leu.2011.368> PMID: 22193966; PubMed Central PMCID: PMC3629947.
- Hideshima T, Bergsagel PL, Kuehl WM, Anderson KC. Advances in biology of multiple myeloma: clinical applications. *Blood*. 2004; 104(3):607–18. <https://doi.org/10.1182/blood-2004-01-0037> PMID: 15090448.
- Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. *CA Cancer J Clin*. 2009; 59(4):225–49. <https://doi.org/10.3322/caac.20006> PMID: 19474385.
- Landgren O, Kyle RA, Pfeiffer RM, Katzmann JA, Caporaso NE, Hayes RB, et al. Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study. *Blood*. 2009; 113(22):5412–7. <https://doi.org/10.1182/blood-2008-12-194241> PMID: 19179464; PubMed Central PMCID: PMC362689042.
- Weiss BM, Abadie J, Verma P, Howard RS, Kuehl WM. A monoclonal gammopathy precedes multiple myeloma in most patients. *Blood*. 2009; 113(22):5418–22. <https://doi.org/10.1182/blood-2008-12-195008> PMID: 19234139; PubMed Central PMCID: PMC362689043.
- Benjamin M, Reddy S, Brawley OW. Myeloma and race: a review of the literature. *Cancer Metastasis Rev*. 2003; 22(1):87–93. PMID: 12716040.
- Becker N. Epidemiology of multiple myeloma. *Recent Results Cancer Res*. 2011; 183:25–35. [https://doi.org/10.1007/978-3-540-85772-3\\_2](https://doi.org/10.1007/978-3-540-85772-3_2) PMID: 21509679.
- Landgren O, Graubard BI, Katzmann JA, Kyle RA, Ahmadzadeh I, Clark R, et al. Racial disparities in the prevalence of monoclonal gammopathies: a population-based study of 12,482 persons from the National Health and Nutritional Examination Survey. *Leukemia*. 2014; 28(7):1537–42. <https://doi.org/10.1038/leu.2014.34> PMID: 24441287; PubMed Central PMCID: PMC362689042.
- StatBite. StatBite: Multiple myeloma and African Americans: higher incidence but fewer autologous stem cell transplants. *J Natl Cancer Inst*. 2009; 101(23):1610. <https://doi.org/10.1093/jnci/djp435> PMID: 19910557.
- Waxman AJ, Mink PJ, Devesa SS, Anderson WF, Weiss BM, Kristinsson SY, et al. Racial disparities in incidence and outcome in multiple myeloma: a population-based study. *Blood*. 2010; 116(25):5501–6. <https://doi.org/10.1182/blood-2010-07-298760> PMID: 20823456; PubMed Central PMCID: PMC362689042.
- Srivastava G, Rana V, Lacy MQ, Buadi FK, Hayman SR, Dispenzieri A, et al. Long-term outcome with lenalidomide and dexamethasone therapy for newly diagnosed multiple myeloma. *Leukemia*. 2013; 27(10):2062–6. <https://doi.org/10.1038/leu.2013.143> PMID: 23648667; PubMed Central PMCID: PMC362689042.
- Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, et al. Initial genome sequencing and analysis of multiple myeloma. *Nature*. 2011; 471(7339):467–72. <https://doi.org/10.1038/nature09837> PMID: 21430775; PubMed Central PMCID: PMC362689042.

13. Lohr JG, Stojanov P, Carter SL, Cruz-Gordillo P, Lawrence MS, Auclair D, et al. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell*. 2014; 25(1):91–101. <https://doi.org/10.1016/j.ccr.2013.12.015> PMID: 24434212; PubMed Central PMCID: PMC4241387.
14. Baker A, Braggio E, Jacobus S, Jung S, Larson D, Therneau T, et al. Uncovering the biology of multiple myeloma among African Americans: a comprehensive genomics approach. *Blood*. 2013; 121(16):3147–52. <https://doi.org/10.1182/blood-2012-07-443606> PMID: 23422747; PubMed Central PMCID: PMC3630830.
15. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. Genetic structure of human populations. *Science*. 2002; 298(5602):2381–5. <https://doi.org/10.1126/science.1078311> PMID: 12493913.
16. Kumar R, Seibold MA, Aldrich MC, Williams LK, Reiner AP, Colangelo L, et al. Genetic ancestry in lung-function predictions. *N Engl J Med*. 2010; 363(4):321–30. <https://doi.org/10.1056/NEJMoa0907897> PMID: 20647190; PubMed Central PMCID: PMC2922981.
17. Yang JJ, Cheng C, Devidas M, Cao X, Fan Y, Campana D, et al. Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. *Nat Genet*. 2011; 43(3):237–41. <https://doi.org/10.1038/ng.763> PMID: 21297632; PubMed Central PMCID: PMC3104508.
18. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015; 526(7571):68–74. <https://doi.org/10.1038/nature15393> PMID: 26432245; PubMed Central PMCID: PMC4750478.
19. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155(2):945–59. PMID: 10835412; PubMed Central PMCID: PMC1461096.
20. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013; 499(7457):214–8. <https://doi.org/10.1038/nature12213> PMID: 23770567; PubMed Central PMCID: PMC3919509.
21. Egan JB, Shi CX, Tembe W, Christoforides A, Kurdoglu A, Sinari S, et al. Whole-genome sequencing of multiple myeloma from diagnosis to plasma cell leukemia reveals genomic initiating events, evolution, and clonal tides. *Blood*. 2012; 120(5):1060–6. <https://doi.org/10.1182/blood-2012-01-405977> PMID: 22529291; PubMed Central PMCID: PMC3412329.
22. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*. 2011; 12(4):R41. <https://doi.org/10.1186/gb-2011-12-4-r41> PMID: 21527027; PubMed Central PMCID: PMC3218867.
23. Shaughnessy JD Jr., Zhan F, Burington BE, Huang Y, Colla S, Hanamura I, et al. A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood*. 2007; 109(6):2276–84. <https://doi.org/10.1182/blood-2006-07-038430> PMID: 17105813.
24. Bergsagel PL, Kuehl WM. Molecular pathogenesis and a consequent classification of multiple myeloma. *J Clin Oncol*. 2005; 23(26):6333–8. <https://doi.org/10.1200/JCO.2005.05.021> PMID: 16155016.
25. Chng WJ, Kumar S, Vanwier S, Ahmann G, Price-Troska T, Henderson K, et al. Molecular dissection of hyperdiploid multiple myeloma by gene expression profiling. *Cancer Res*. 2007; 67(7):2982–9. <https://doi.org/10.1158/0008-5472.CAN-06-4046> PMID: 17409404.
26. Walker BA, Boyle EM, Wardell CP, Murison A, Begum DB, Dahir NM, et al. Mutational Spectrum, Copy Number Changes, and Outcome: Results of a Sequencing Study of Patients With Newly Diagnosed Myeloma. *J Clin Oncol*. 2015; 33(33):3911–20. <https://doi.org/10.1200/JCO.2014.59.1503> PMID: 26282654.
27. Andrulis M, Lehnert N, Capper D, Penzel R, Heining C, Huellein J, et al. Targeting the BRAF V600E mutation in multiple myeloma. *Cancer Discov*. 2013; 3(8):862–9. Epub 2013/04/25. <https://doi.org/10.1158/2159-8290.CD-13-0014> PMID: 23612012.
28. O'Donnell E, Raje NS. Targeting BRAF in multiple myeloma. *Cancer Discov*. 2013; 3(8):840–2. Epub 2013/08/10. <https://doi.org/10.1158/2159-8290.CD-13-0297> PMID: 23928771.
29. Rustad EH, Dai HY, Hov H, Coward E, Beisvag V, Myklebost O, et al. BRAF V600E mutation in early-stage multiple myeloma: good response to broad acting drugs and no relation to prognosis. *Blood Cancer J*. 2015; 5:e299. Epub 2015/03/21. <https://doi.org/10.1038/bcj.2015.24> PMID: 25794135; PubMed Central PMCID: PMC4382665.
30. Zani VJ, Asou N, Jadayel D, Heward JM, Shipley J, Nacheva E, et al. Molecular cloning of complex chromosomal translocation t(8;14;12)(q24.1;q32.3;q24.1) in a Burkitt lymphoma cell line defines a new gene (BCL7A) with homology to caldesmon. *Blood*. 1996; 87(8):3124–34. PMID: 8605326.
31. Muller P, Kuttenukeuler D, Gesellchen V, Zeidler MP, Boutros M. Identification of JAK/STAT signalling components by genome-wide RNA interference. *Nature*. 2005; 436(7052):871–5. <https://doi.org/10.1038/nature03869> PMID: 16094372.



32. Kalla C, Nentwich H, Schlotter M, Mertens D, Wildenberger K, Dohner H, et al. Translocation t(X;11) (q13;q23) in B-cell chronic lymphocytic leukemia disrupts two novel genes. *Genes Chromosomes Cancer*. 2005; 42(2):128–43. <https://doi.org/10.1002/gcc.20131> PMID: 15543602.
33. Denk D, Nebral K, Bradtke J, Pass G, Moricke A, Attarbaschi A, et al. PAX5-AUTS2: a recurrent fusion gene in childhood B-cell precursor acute lymphoblastic leukemia. *Leuk Res*. 2012; 36(8):e178–81. Epub 2012/05/15. <https://doi.org/10.1016/j.leukres.2012.04.015> PMID: 22578776; PubMed Central PMCID: PMC3389344.
34. Weinhold N, Ashby C, Rasche L, Chavan SS, Stein C, Stephens OW, et al. Clonal selection and double-hit events involving tumor suppressor genes underlie relapse in myeloma. *Blood*. 2016; 128(13):1735–44. <https://doi.org/10.1182/blood-2016-06-723007> PMID: 27516441.
35. Iida S, Rao PH, Butler M, Corradini P, Boccadoro M, Klein B, et al. Deregulation of MUM1/IRF4 by chromosomal translocation in multiple myeloma. *Nat Genet*. 1997; 17(2):226–30. <https://doi.org/10.1038/ng1097-226> PMID: 9326949.
36. Shaffer AL, Emre NC, Lamy L, Ngo VN, Wright G, Xiao W, et al. IRF4 addiction in multiple myeloma. *Nature*. 2008; 454(7201):226–31. <https://doi.org/10.1038/nature07064> PMID: 18568025; PubMed Central PMCID: PMC2542904.
37. Fionda C, Abruzzese MP, Zingoni A, Cecere F, Vulpis E, Peruzzi G, et al. The IMiDs targets IKZF-1/3 and IRF4 as novel negative regulators of NK cell-activating ligands expression in multiple myeloma. *Oncotarget*. 2015; 6(27):23609–30. <https://doi.org/10.18632/oncotarget.4603> PMID: 26269456; PubMed Central PMCID: PMC4695140.
38. Lopez-Girona A, Heintel D, Zhang LH, Mendy D, Gaidarova S, Brady H, et al. Lenalidomide downregulates the cell survival factor, interferon regulatory factor-4, providing a potential mechanistic link for predicting response. *Br J Haematol*. 2011; 154(3):325–36. <https://doi.org/10.1111/j.1365-2141.2011.08689.x> PMID: 21707574.
39. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168; PubMed Central PMCID: PMC2705234.
40. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943; PubMed Central PMCID: PMC2723002.
41. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010; 20(9):1297–303. <https://doi.org/10.1101/gr.107524.110> PMID: 20644199; PubMed Central PMCID: PMC2928508.
42. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011; 43(5):491–8. <https://doi.org/10.1038/ng.806> PMID: 21478889; PubMed Central PMCID: PMC3083463.
43. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011; 27(21):2987–93. <https://doi.org/10.1093/bioinformatics/btr509> PMID: 21903627; PubMed Central PMCID: PMC3198575.
44. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013; 31(3):213–9. <https://doi.org/10.1038/nbt.2514> PMID: 23396013; PubMed Central PMCID: PMC3833702.
45. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK, et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012; 28(14):1811–7. <https://doi.org/10.1093/bioinformatics/bts271> PMID: 22581179.
46. Christoforides A, Carpten JD, Weiss GJ, Demeure MJ, Von Hoff DD, Craig DW. Identification of somatic mutations in cancer through Bayesian-based analysis of sequenced genome pairs. *BMC Genomics*. 2013; 14:302. <https://doi.org/10.1186/1471-2164-14-302> PMID: 23642077; PubMed Central PMCID: PMC3751438.
47. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods*. 2015; 12(2):115–21. <https://doi.org/10.1038/nmeth.3252> PMID: 25633503; PubMed Central PMCID: PMC4509590.
48. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012; 6(2):80–92. <https://doi.org/10.4161/fly.19695> PMID: 22728672; PubMed Central PMCID: PMC3679285.
49. Patro R, Mount SM, Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol*. 2014; 32(5):462–4. <https://doi.org/10.1038/nbt.2862> PMID: 24752080; PubMed Central PMCID: PMC4077321.

50. Kosoy R, Nassir R, Tian C, White PA, Butler LM, Silva G, et al. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Hum Mutat.* 2009; 30(1):69–78. <https://doi.org/10.1002/humu.20822> PMID: 18683858; PubMed Central PMCID: PMCPMC3073397.
51. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006; 38(8):904–9. <https://doi.org/10.1038/ng1847> PMID: 16862161.
52. Tandon A, Patterson N, Reich D. Ancestry informative marker panels for African Americans based on subsets of commercially available SNP arrays. *Genet Epidemiol.* 2011; 35(1):80–3. <https://doi.org/10.1002/gepi.20550> PMID: 21181899; PubMed Central PMCID: PMCPMC4386999.
53. Rosenberg NA, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers for inference of ancestry. *Am J Hum Genet.* 2003; 73(6):1402–22. <https://doi.org/10.1086/380416> PMID: 14631557; PubMed Central PMCID: PMCPMC1180403.
54. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006; 2(12):e190. <https://doi.org/10.1371/journal.pgen.0020190> PMID: 17194218; PubMed Central PMCID: PMCPMC1713260.
55. Lischer HE, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics.* 2012; 28(2):298–9. <https://doi.org/10.1093/bioinformatics/btr642> PMID: 22110245.
56. Agresti A. A survey of models for repeated ordered categorical response data. *Stat Med.* 1989; 8(10):1209–24. PMID: 2814070.
57. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res.* 2001; 125(1–2):279–84. PMID: 11682119.