BDL-SP: **B**io-inspired **D**eep **L**earning Architecture for the identification of altered **S**ignaling **P**athways

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Outline

- Background
- Motivation
- Description of Global WES data repositories
- Infographic abstract of AI-driven workflow and BDL-SP model
- AI-driven workflow for feature extraction and description of genomic feature matrix
- Post-hoc analysis of BDL-SP model for identification of top genomic attributes
  - Algorithm for estimation of best ShAP score for genomic attributes
  - Qualitative assessment of top-performing model at group level (MM/MGUS)
  - Qualitative assessment of top-performing model at sample level (MM/MGUS)
  - Pathway Enrichment Analysis using Enrichr
- Key-points of bio-inspired AI-driven workflow
- Conclusions
- References

Background

- Multiple Myeloma (MM) is a cancer of plasma cells that is preceded by a premalignant stage of Monoclonal Gammopathy of Undetermined Significance (MGUS).

- In clinical practice, the distinction between MM and MGUS is based on clinical symptoms, disease load (for example, number of aberrant plasma cells, monoclonal protein level secreted by aberrant plasma cells, etc.) which is at time ambiguous.

- Several studies involving exome sequencing data analysis have been performed to understand the genomic abnormalities in MM and revealed that the primary events in MM are either hyperdiploidy (such as trisomy of chromosome) or non-hyperdiploidy (such as translocations) [1].

- Further, the primary events are then followed by the secondary events (such as secondary translocations, loss of heterozygosity, etc.) [1].

- In a landmark study, the analysis of MGUS and MM paired samples reaffirmed the clonal heterogeneity and presence of majority of genetic changes at MGUS stage [2].

- The existence of the majority of genetic abnormalities seen in MM at the MGUS stage poses a challenge in distinguishing MM from MGUS based on the genomic signatures and in defining critical genomic events responsible for the progression of MGUS to MM [3,4,5,6].

Challenges and Motivation

Challenges:

- MGUS patients patients share many genetic aberrations of MM without showing any overt clinical symptoms of MM that makes it challenging to differentiate MGUS from MM.
- The early diagnosis of MM and the identification of relevant differentiating genetic and genomic biomarkers between MGUS and MM present several challenges at the genomic-level and the subject-level.
- The subject-level challenges includes the lack of paired sequencing data as not all MGUS subjects are progressed to MM, limited information about subject's treatment time intervals, etc.
- Similarly, the genomic-level challenges includes the availability of reliable workflow for analysing a pool of large mutational information, identification of biomarkers, altered signaling pathways, etc.

Motivations:

- With increasing complexity of genomics data, the recent advancements in machine/deep learning (ML/DL) methods (such as somatic SNV prediction[7,8], CNV prediction [9,10], survival outcome and treatment-sensitivity in MM [11,12], etc.) can be helpful to identify the meaningful patterns and infer the salient information from multi-omics datasets that can be utilized to halt the disease progression.
- In recent years, geometric deep learning (GDL) has emerged to incorporate graph structures into a deep learning framework that enables the integration of exomic mutational profiles with gene-gene interaction information (PPI network).

Description of Global WES data repositories

- We have considered three datasets in this work (shown Table-1) that consists of whole exome sequencing (WES) samples from three different ethnicities.
- The overall datasets contains 1154 MM samples and 61 MM samples that makes it an imbalanced datasets (95% MM and 5% MGUS cases).

Table-1: Three global WES repositories were used in this work for identification of pivotal biomarkers to differentiate MM and MGUS.

Data Repository	# of MM/# MGUS Patients (including all time points)	Ethnicity	Sources	Available data type
phs0000748	MM patients: 1092	American Population	MMRF [13]	Processed VCF files
EGAD00001001901	MGUS: 33	Caucasian Population	European Genome Archive (EGA) [14]	Processed BAM files
PRJNA685283, PRJNA 694218	MM: 82, MGUS: 28	Asian Population	AIIMS [15]	Raw FASTQ files

Infographic abstract of AI-driven workflow and BDL-SP model

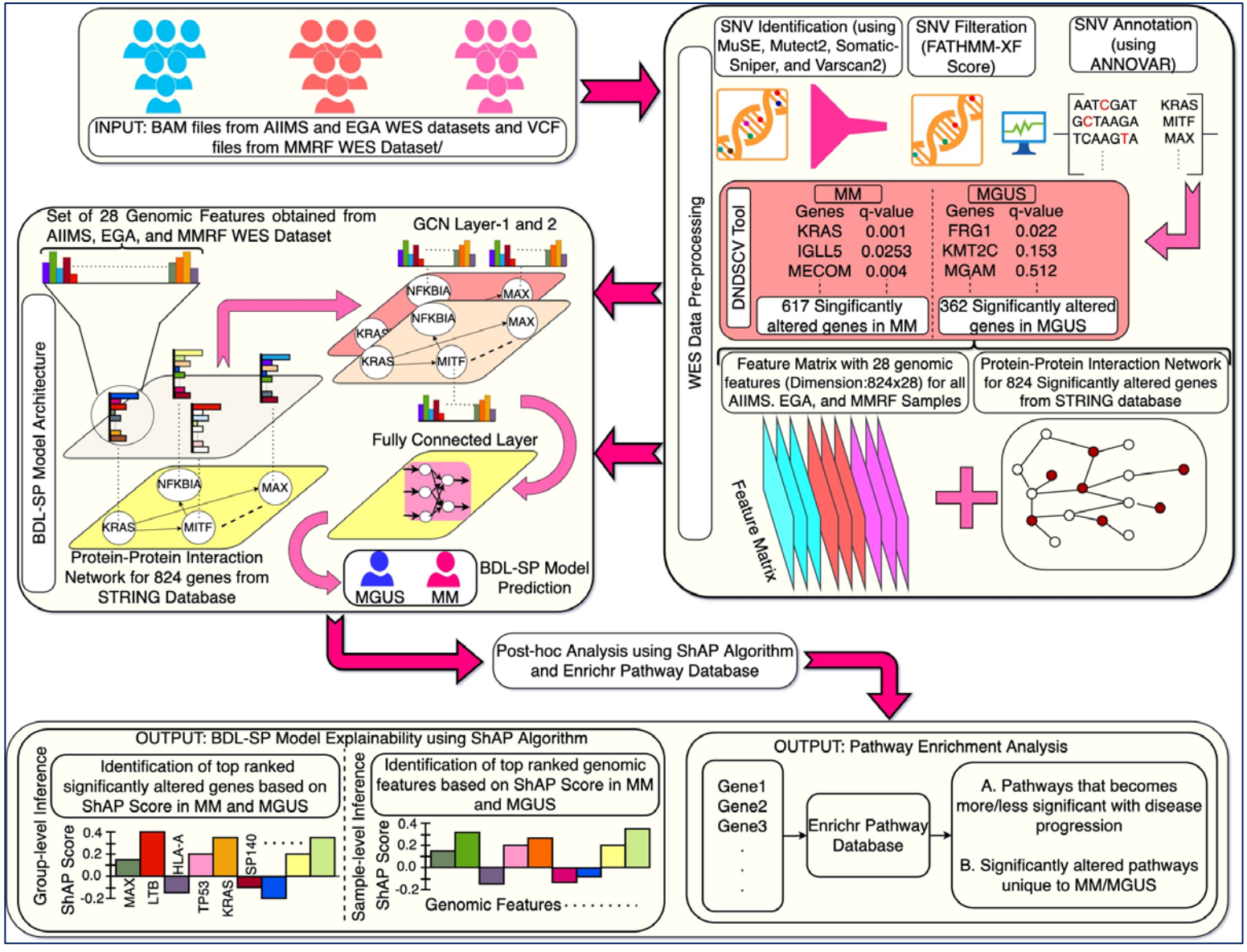


Figure-1: Infographic representation of AI-based workflow with BDL-SP model architecture and post- hoc analysis for the identification of pivotal genetic biomarkers that distinguish MGUS from MM.

AI-driven workflow for feature extraction and description of genomic feature matrix

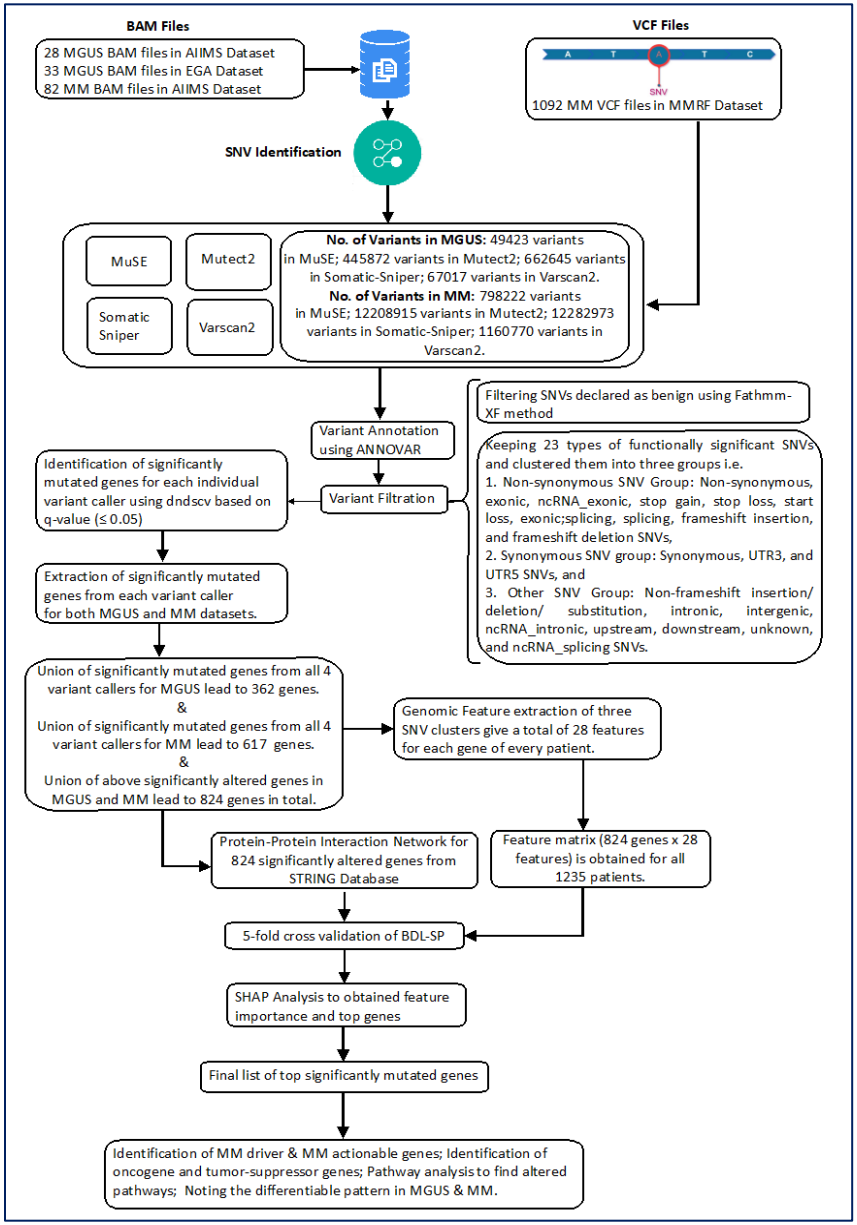
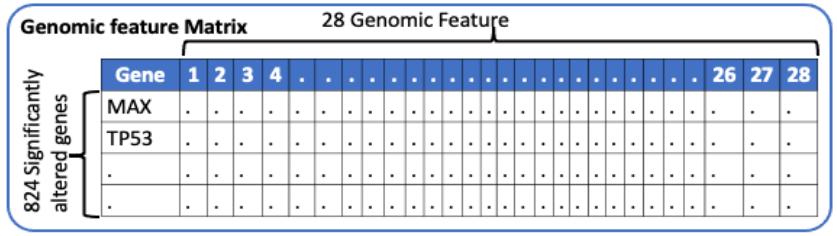


Figure-2: AI-based workflow to infer differentiable genomic biomarkers to identify MGUS and MM using the whole-exome sequencing (WES) data



Description of genomic features used in the feature matrix

Feature Number	Feature Name
1	Total number of the SNVs
2	Total number of the SNVs in synonymous group
3	Maximum VAF of the SNVs in synonymous group
4	Median VAF of the SNVs in synonymous group
5	Mean VAF of the SNVs in synonymous group
6	VAF's standard deviation of the SNVs in synonymous group
7	Maximum AD of the SNVs in synonymous group
8	Median AD of SNVs in synonymous group
9	Mean AD of SNVs in synonymous group
10	AD's standard deviation of the SNVs in synonymous group
11	Total number of SNVs in non-synonymous group
12	Maximum VAF of SNVs in non-synonymous group
13	Median VAF of SNVs in non-synonymous group
14	Mean VAF of SNVs in non-synonymous group
15	VAF's standard deviation of the SNVs in non-synonymous group
16	Maximum AD of SNVs in non-synonymous group
17	Median AD of SNVs in non-synonymous group
18	Mean AD of SNVs in non-synonymous group
19	AD's standard deviation of the SNVs in non-synonymous group
20	Total number of SNVs in other group
21	Maximum VAF of SNVs in other group
22	Median VAF of SNVs in other group
23	Mean VAF of SNVs in other group
24	VAF's standard deviation of the SNVs in other group
25	Maximum AD of SNVs in other group
26	Median AD of SNVs in other group
27	Mean AD of SNVs in other group
28	AD's standard deviation of the SNVs in other group

Figure-3: Schematic layout of genomic feature matrix used for the training of proposed BDL-SP model.

Algorithm for estimation of best ShAP score for genomic attributes

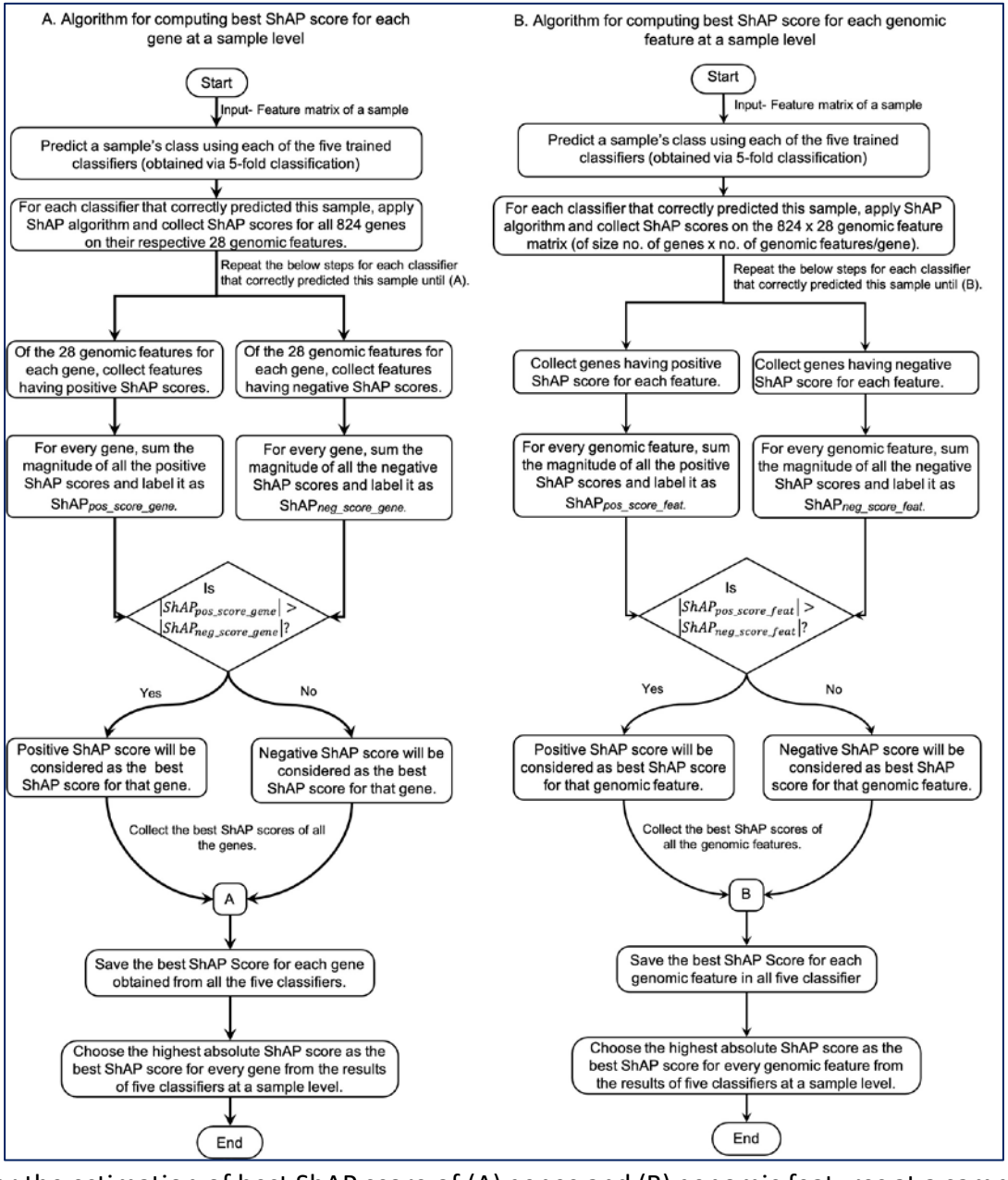


Figure-4: Algorithm for the estimation of best ShAP score of (A) genes and (B) genomic features at a sample level

Benchmarking of proposed BDL-SP model [Quantitative Benchmarking]

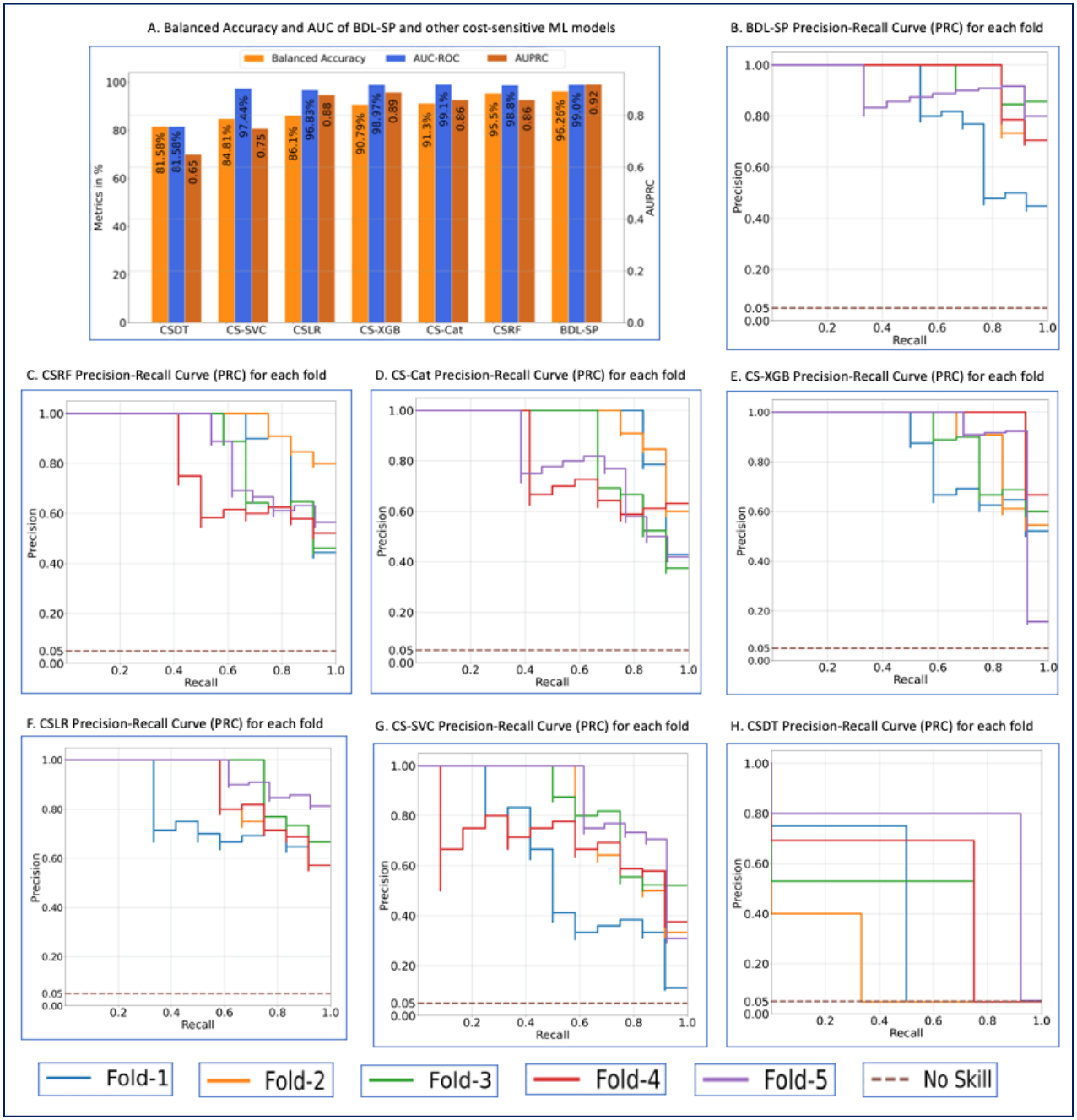


Table-2: Confusion matrix of top-3 performing models

Model Name	Confusion Matrix { TP=MGUS, FP=Not MM, TN=MM, FN=Not MGUS }
BDL-SP	{ TP=60, FP=67, TN=1092, FN=2 }
CS-RF	{ TP=59, FP=68, TN=1086, FN=2 }
CS-Cat	{ TP=52, FP=33, TN=1121, FN=9 }

Figure-5: (A) Benchmarking of proposed BDL-SP model with baseline machine learning models, Precision-Recall curves (PRC) for all five folds of (B). BDL-SP, (C). CSRF, (D). CS-Cat, (E). CS-XGB, (F). CSLR, (G) CS-SVC, and (H) CSDT.

Benchmarking of proposed BDL-SP model

Post-hoc Analysis of top-performing models

Table-3: Types of four different gene categories (OG, TSG, ODG, and AG) and their counts in 824 significantly altered genes

Gene Type	Total number of previously reported genes present in the list of 824 significantly altered genes
Oncogenes (OG)	20
Tumor Suppresor genes (TSG)	35
Both Oncogenes and driver genes (ODG)	5
Actionable genes (AG)	19

Table-4: Counts of four gene categories in top-250 and top-500 genes obtained from the post-hoc analysis of top-3 models (BDL-SP, CS-RF, and CS-Cat).

Top genes	BDL-SP model (Top performing model)				CS-RF model (Top performing model)				CS-Cat model (Top performing model)			
	OG	TSG	ODG	AG	OG	TSG	ODG	AG	OG	TSG	ODG	AG
Top-250	15	20	5	12	7	10	1	4	6	5	1	4
Top-500	19	30	5	16	7	10	1	4	6	5	1	4

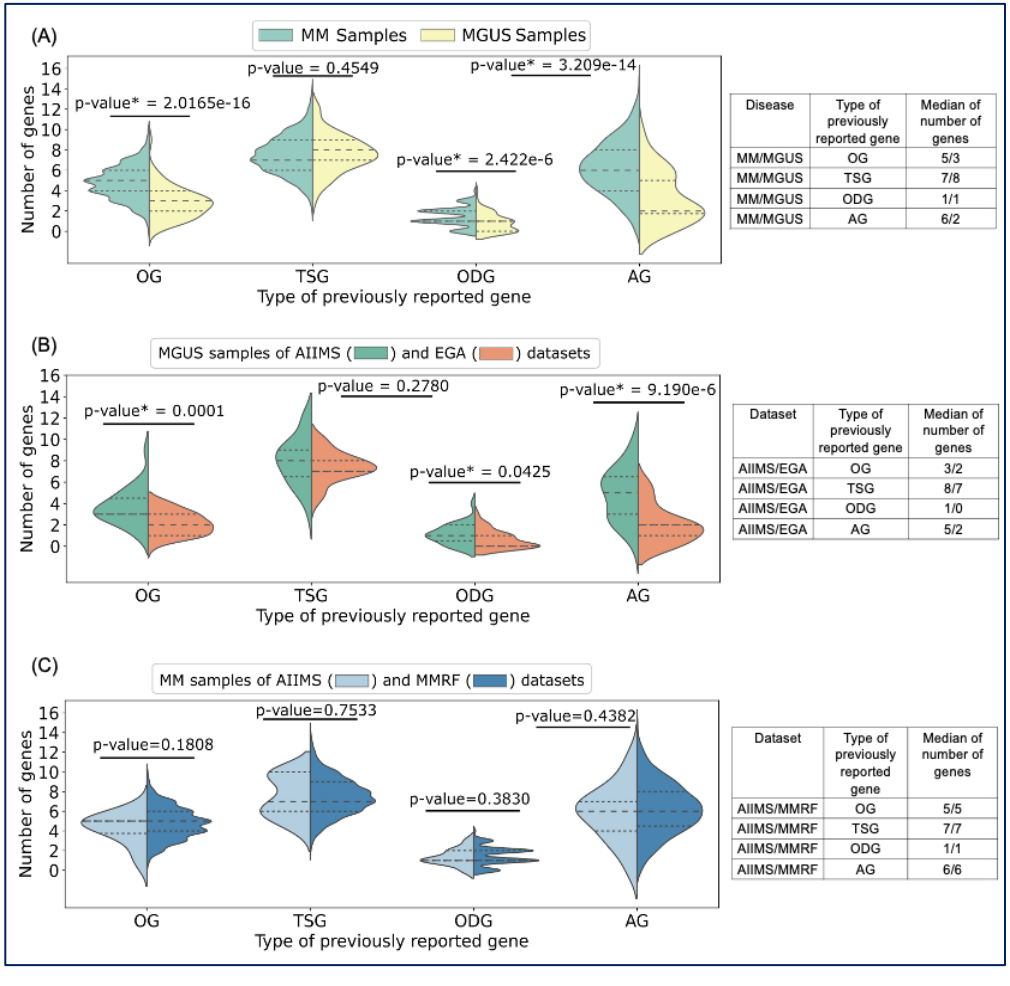
Qualitative assessment of top-performing model at group level (MM/MGUS)

Table-5: (Top) Oncogenes (OG) and actionable genes (AG) (B) Tumor suppressor genes (TSG) and Both oncogenes and driver genes (ODG) reported by top-3 performing models

Top genes in MM and MGUS	Top significantly altered previously genes reported by BDL-SP Model		Top significantly altered previously genes reported by CS-RF Model		Top significantly altered previously genes reported by CS-Cat Model	
	Oncogenes	Actionable genes	Oncogenes	Actionable genes	Oncogenes	Actionable genes
Top-250 genes in MM and MGUs combined SHAP analysis	KRAS, LTB, CARD11, NOTCH1, FGFR3, VAV1, IRS1, MGAM, ABL2, NRAS, BRAF, TCL1A, PGR, MITF, RPTOR	KRAS, NOTCH1, TP53, FGFR3, NFKBIA, NF1, NRAS, BRAF, MITF, ARID1B, ARID2, RPTOR	TCL1A, LTB, RPTOR, ABL2, TAL1, VAV1, NOTCH1	RPTOR, NF1, NFKBIA, NOTCH1	TCL1A, MGAM, ABL2, VAV1, PGR, BRD4	NFKBIA, APC, BRD4, BRAF
Top-500 genes in MM and MGUs combined SHAP analysis	KRAS, LTB, CARD11, NOTCH1, FGFR3, VAV1, IRS1, MGAM, ABL2, NRAS, BRAF, TCL1A, PGR, MITF, RPTOR, TERT, BRD4, MECOM, TAL1	KRAS, NOTCH1, TP53, FGFR3, NFKBIA, NF1, NRAS, BRAF, MITF, ARID1B, ARID2, RPTOR, BRD4, RB1, FANCD2, APC	TCL1A, LTB, RPTOR, ABL2, TAL1, VAV1, NOTCH1	RPTOR, NF1, NFKBIA, NOTCH1	TCL1A, MGAM, ABL2, VAV1, PGR, BRD4	NFKBIA, APC, BRD4, BRAF

Top genes in MM and MGUS	Top significantly altered previously genes reported by BDL-SP Model		Top significantly altered previously genes reported by CS-RF Model		Top significantly altered previously genes reported by CS-Cat Model	
	Tumor suppresor genes	Both oncogene and driver genes	Tumor suppresor genes	Both oncogene and driver genes	Tumor suppresor genes	Both oncogene and driver genes
Top-250 genes in MM and MGUs combined SHAP analysis	HLA-A, HLA-C, HLA-B, LTB, NOTCH1, TRAF3, TP53, EGR1, NFKBIA, SDHA, IRF1, SAMHD1, NF1, DIS3, MITF, ARID1B, ARID2, CYLD, SP140, KMT2C	KRAS, LTB, FGFR3, NRAS, BRAF	NCOR1, LTB, CYLD, NFKBIA, NF1, EGR1, TRAF3, NOTCH1, SDHA, KMT2C	LTB	NCOR1, CYLD, NFKBIA, APC, MAX	BRAF
Top-500 genes in MM and MGUs combined SHAP analysis	HLA-A, HLA-C, HLA-B, LTB, NOTCH1, TRAF3, TP53, EGR1, NFKBIA, SDHA, IRF1, SAMHD1, NF1, DIS3, MITF, ARID1B, ARID2, CYLD, SP140, KMT2C, MAX, RB1, FANCD2, ZFHX3, NCOR1, KMT2D, KMT2B, APC, CMTR2, AMER1	KRAS, LTB, FGFR3, NRAS, BRAF	NCOR1, LTB, CYLD, NFKBIA, NF1, EGR1, TRAF3, NOTCH1, SDHA, KMT2C	LTB	NCOR1, CYLD, NFKBIA, APC, MAX	BRAF

Qualitative assessment of BDL-SP model at sample level (MM/MGUS)



Disease	Type of previously reported gene	Median of number of genes
MM/MGUS	OG	5/3
MM/MGUS	TSG	7/8
MM/MGUS	ODG	1/1
MM/MGUS	AG	6/2

Dataset	Type of previously reported gene	Median of number of genes
AIIMS/EGA	OG	3/2
AIIMS/EGA	TSG	8/7
AIIMS/EGA	ODG	1/0
AIIMS/EGA	AG	5/2

Dataset	Type of previously reported gene	Median of number of genes
AIIMS/MMRF	OG	5/5
AIIMS/MMRF	TSG	7/7
AIIMS/MMRF	ODG	1/1
AIIMS/MMRF	AG	6/6

Figure-6: (A) The distribution of the number of previously reported genes in four gene groups (OGs, TSGs, ODGs, and AGs) found significantly altered and ranked in top-100 across all MM and MGUS samples in AIIMS, EGA, and MMRF datasets, (B) The distribution of the number of previously reported genes found significantly altered and ranked in top-100 across all MGUS samples in AIIMS and EGA datasets. (C) The distribution of the number of previously reported genes found significantly altered and ranked in top-100 across all MM samples in AIIMS and MMRF datasets.

- On post-hoc analysis of BDL-SP model using ShAP algorithm at a sample level, we identified the top-100 significantly altered genes for each sample and compared them with previously reported genes (OG, TSG, ODG, and AG) from MM related studies.
- We compared the number of previously reported genes (OG, TSG. ODG, and AG) samples from different ethnicities (American, Caucasian, and Asian) statistically using unpaired Wilcoxon ranksum test and observed the impact of ethnicity among MM and MGUS samples.
- The number of OG, ODG, and AG were statistically different among MM and MGUS samples (figure-6(A) and 6(B)).
- On the other hand, none of the gene group (OG, TSG, ODG, and AG) were statistically different among MM samples in between Asian and American ethnicity.
- These observation indicates that the ethnicity might be playing a significant role in the disease development and should not be overlooked.

Qualitative assessment of top-performing model at sample level (MM/MGUS)

- On ranking of genomic features in post-hoc analysis of BDL-SP model using ShAP algorithm at a sample level, we observed that the total number of SNVs and genomic features associated to synonymous SNV group (that contains synonymous SNVs, UTR3, and UTR5 SNVs) were the most contributing SNVs in differentiating MM and MGUS.

- Although, the role of synonymous SNVs are unclear in MM but, in recent studies, the synonymous SNVs were observed as significant contributor SNVs in multiple cancer types [16,17,18].

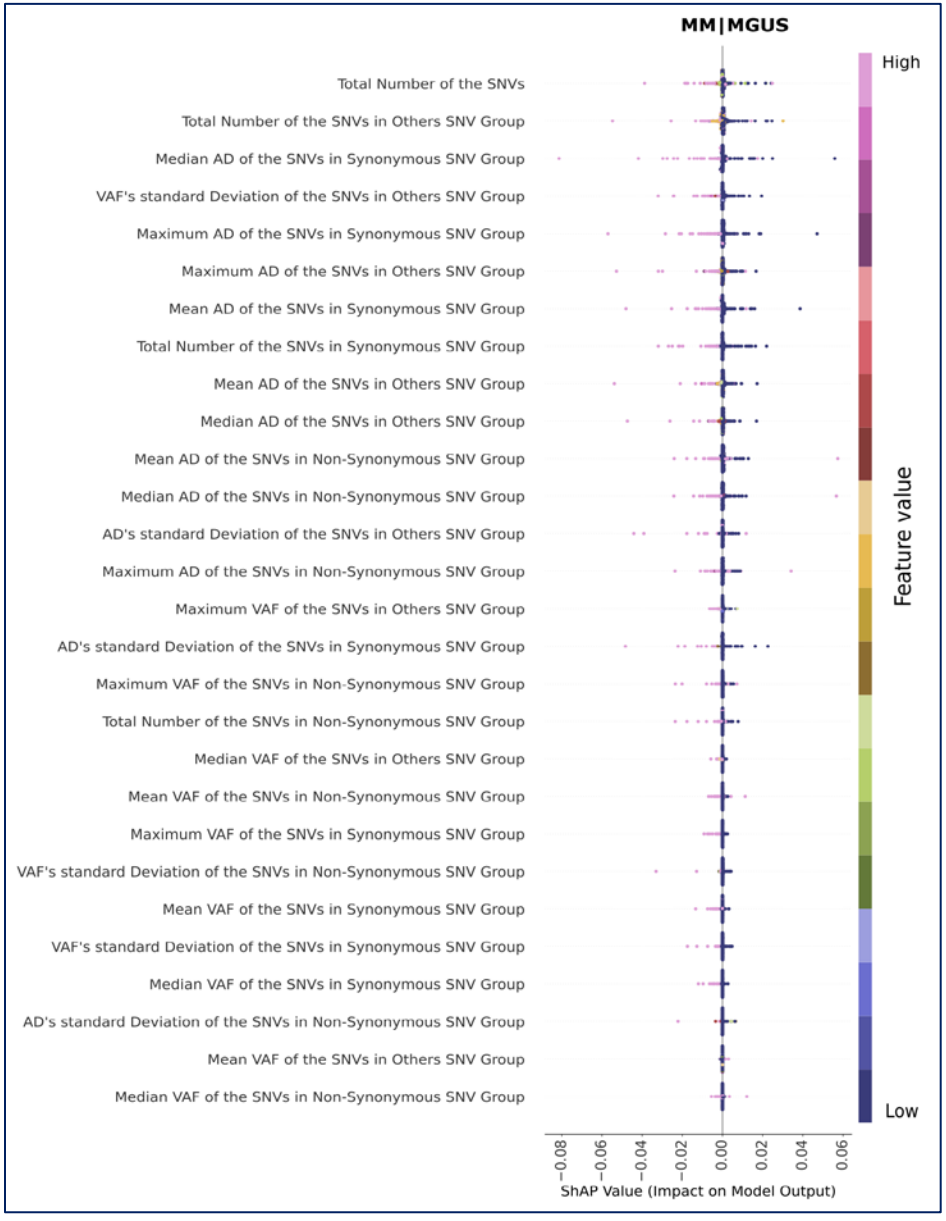


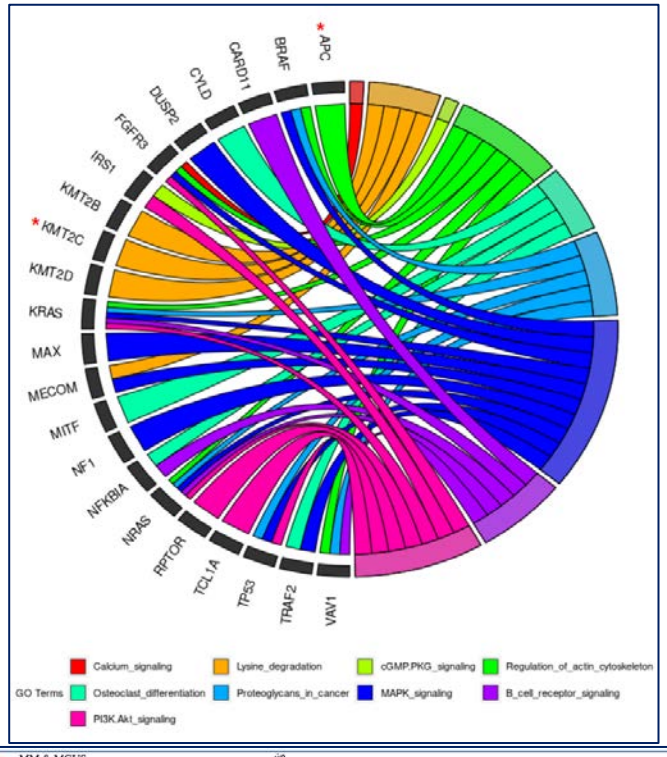
Figure-7: Genomic Feature ranking based on the BDL-SP post-hoc explainability in MM and MGUS using ShAP algorithm

Pathway Enrichment Analysis using Enrichr [13,14,15]

(A)

Adjusted P-value		Adjusted P-value
0.16102	Calcium signaling pathway	0.00002
RYR1; CACNA1B; ITPR1; ITPR2; SLC25A5; RYR3; CACNA1G		RYR1; RYR2; CHRM3; PDE1C; CACNA1B; ITPR1; CACNA1A; ITPR2; ATP2B3; ITPR3; ATP2B2; CACNA1F; RYR3; GRIN2D; CACNA1I; PTK2B; NOS1; SLC25A5; FGFR3
0.73326	B cell receptor signaling pathway	0.00002
LILRB1		NFKBIA; NRAS; INPP5D; LILRB1; KRAS; LILRA1; LILRB2; LILRA2; VAV1; CARD11; LILRA4
0.57833	MAPK signaling cascade	0.00034
CACNA1B; FLNC; PAK2; CACNA1G		MAP4K1; DUSP2; MAX; CACNA1B; CACNA2D2; CACNA1A; BRAF; TRAF2; CACNA1F; CACNA1I; NRAS; MECOM; NF1; FLNA; KRAS; TP53; PAK2; FGFR3
0.42260	Regulation of actin cytoskeleton	0.00040
ENAH; APC; ACTR3B; PAK2		CHRM3; GSN; ITGA2; BRAF; ACTR3B; VAV1; MYL5; NRAS; MYH14; MYH11; KRAS; EZR; PAK2; FGFR3; VCL
0.42260	Osteoclast differentiation	0.00048
FCGR2A; SIRPA; LILRB1		NFKBIA; CYLD; SIRPA; MITF; TRAF2; LILRB1; LILRA1; LILRB2; LILRA2; SIRPB1; LILRA4
0.36247	cGMP-PKG signaling pathway	0.00314
ITPR1; ITPR2; SLC25A5; MYH7		IRS1; ITPR1; ITPR2; ATP2B3; ITPR3; ATP2B2; CACNA1F; SLC25A5; MYH6; MYH7; ADCY5
0.42260	Proteoglycans in cancer	0.00428
GPC1; ITPR1; ITPR2; FLNC		NRAS; ITGA2; ITPR1; FZD8; FLNA; ITPR2; BRAF; ITPR3; KRAS; EZR; TP53; VAV1
0.18224	Lysine degradation	0.00683
KMT2E; KMT2C; ASH1L		KMT2E; KMT2D; MECOM; KMT2C; KMT2B; ASH1L
0.88434	PI3K-Akt signaling pathway	0.01655
HSP90AB1; LAMA3		HSP90AB1; VWF; LAMA2; IRS1; ITGA2; LAMA3; LAMC2; YWHAZ; RPTOR; TCL1A; NRAS; EIF4EBP1; KRAS; TP53; FGFR3

(B)



(C)

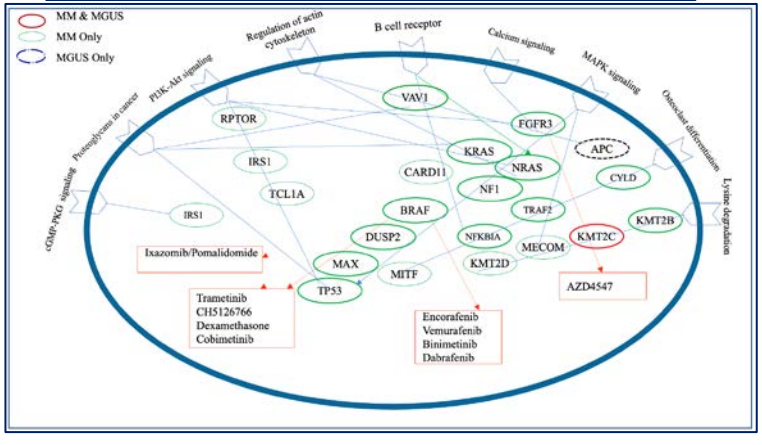


Figure-8: (A) KEGG pathways found to be significantly involved in progression of MGUS to MM. Top significantly altered genes that were identified by post-hoc analysis of BDL-SP using ShAP algorithm as unique either in MGUS or in MM are shown in red colored font. (B) GOChord plot reveals the association of driver/ TSG/ Onco/ Actionable genes associated with important pathways. The gene KMT2C was observed to be significantly mutated in MGUS and MM; and the gene APC was mutated only in MGUS. All other genes are significantly mutated in MM only. © Important pathways significantly altered in MM. Further, drugs used for pathways directed therapies associated with mutation in genes are highlighted.

Key points of bio-inspired AI-driven workflow

- The term bio-inspired refers to the ability of the workflow to comprehend the topological information from PPI network and then imbibe the genomic features to identify the differentiating biomarkers for MM and MGUS.
- The integrated learning using topological information obtained from PPI network and exomic mutational profiles enables the BDL-SP model to rank the genes based on their role in disease progression more efficiently with few layers of graph convolutions network than the traditional machine learning models that were trained only on the exomic mutational profiles.
- When two ML/DL model perform similar quantitatively, it would be better if the model is chosen on the basis of the interpretability with reference to the application domain. Here, The quantitative performance of top-3 models (BDL-SP, CS-RF, and CS-Cat) were similar and on further post-hoc benchmarking revealed the BDL-SP model has identified most number of previously reported genes.
- The post-hoc analysis of the BDL-SP model also revealed the role of ethnicity in the initial phase of MM (that is, MGUS) as the number of significantly altered OG, ODG, and TSG were statistically different among MM and MGUS samples obtained from American, Caucasian and Asian ethnicity.
- Further, total number of SNVs and the genomic features associated to synonymous SNVs were observed to be the most contributing genomic features in differentiating MM from MGUS. Thus, further analysis on the clinical impact of synonymous SNVs is strongly suggested.
- Using this workflow, we have identified the the significantly altered pathways using Enrichr with the help of top-500 significantly altered genes obtained from BDL-SP that become more/less significant with disease progression from MGUS to MM. We observed that several pathways are selectively dysregulated in MM and also, the pathways who lost their significance from MGUS to MM were actually related to other cancer types.

Conclusion

- We designed and implemented an AI-driven work with novel “bio-inspired deep learning architecture for the identification of altered signaling pathways in MM (BDL-SP)” to identify the differentiating biomarkers in MM and MGUS.
- The PPI network can be helpful to infer the genes playing significant role in MM disease pathogenesis. We incorporated the PPI network of 824 significantly altered genes obtained using three global WES data repositories.
- On benchmarking the proposed BDL-SP model with several baseline machine learning models, the BDL-SP models performed almost similar to CS-RF and CS-Cat models in terms of AUC.
- When two model performed similar quantitatively, the model should be chosen on the basis of interpretability with reference to the application domain using proper post-hoc analysis/benchmarking.
- On post-hoc benchmarking of BDL-SP model using ShAP algorithm, the BDL-SP model reported the highest number of previously reported genes (OG, ODG, and AG) as comparison to the traditional ML models.
- The number of previously reported genes among MM and MGUS samples in multiple ethnicities were observed to be statistically different and indicates the significant role of ethnicity during the initial phase of disease and should not be overlooked.
- We further validated our findings by performing pathway analysis on the top mutated genes and observed that several signaling pathways such as Calcium signaling path- way, B-Cell receptor signaling pathway, PI3K-Akt signaling pathway, MAPK signaling pathway, etc. are selectively and more significantly deregulated with disease progression.
- The genomic mutation associated to the synonymous SNV group (synonymous SNVs, UTR3, and UTR5) were found to be the most significantly contributing differentiating MM from MGUS.

References

1. S. Manier, K. Salem, S. V. Glavey, A. M. Roccaro, I. M. Ghobrial, Genomic aberrations in multiple myeloma, Plasma Cell Dyscrasias (2016) 23–34.
2. A. K. Dutta, J. L. Fink, J. P. Grady, G. J. Morgan, C. G. Mullighan, L. B. To, D. R. Hewett, A. C. Zannettino, Subclonal evolution in disease progression from mgus/smm to multiple myeloma is characterised by clonal stability, Leukemia 33 (2) (2019) 457–468.
3. A. Mikulasova, J. Smetana, M. Wayhelova, H. Janyskova, V. Sandecka, Z. Kufova, M. Almasi, J. Jarkovsky, E. Gregora, P. Kessler, et al., Genomewide profiling of copy-number alteration in monoclonal gammopathy of undetermined significance, European journal of haematology 97 (6) (2016) 568–575.
4. B. A. Walker, C. P. Wardell, L. Melchor, A. Brioli, D. C. Johnson, M. F. Kaiser, F. Mirabella, L. Lopez-Corral, S. Humphray, L. Murray, et al., Intraclonal heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms, Leukemia 28 (2) (2014) 384–390.
5. A. Mikulasova, C. P. Wardell, A. Murison, E. M. Boyle, G. H. Jackson, J. Smetana, Z. Kufova, L. Pour, V. Sandecka, M. Almasi, et al., The spectrum of somatic mutations in monoclonal gam- mopathy of undetermined significance indicates a less complex genomic landscape than that in multiple myeloma, Haematologica 102 (9) (2017) 1617.
6. A. Farswan, A. Gupta, L. Jena, V. Ruhela, G. Kaur, R. Gupta, Characterizing the mutational landscape of mm and its precursor mgus, American journal of cancer research 12 (4) (2022) 1919.
7. I. Anzar, A. Sverchkova, R. Stratford, T. Clancy, Neomutate: an ensemble machine learning frame- work for the prediction of somatic mutations in cancer, BMC medical genomics 12 (1) (2019) 1–14.
8. Y.-C. Hsu, Y.-T. Hsiao, T.-Y. Kao, J.-G. Chang, G. S. Shieh, Detection of somatic mutations in exome sequencing of tumor-only samples, Scientific reports 7 (1) (2017) 1–9.
9. V. K. Pounraja, G. Jayakar, M. Jensen, N. Kelkar, S. Girirajan, A machine-learning approach for accurate detection of copy number variants from exome sequencing, Genome research 29 (7) (2019) 1134–1143.

References

10. T. Huang, J. Li, B. Jia, H. Sang, Cnv-meann: A neural network and mind evolutionary algorithm-based detection of copy number variations from next-generation sequencing data, Frontiers in Genetics 12 (2021).
11. A. Mosquera Orgueira, M. S. González Pérez, J. Á. Díaz Arias, B. Antelo Rodríguez, N. Alonso Vence, Á. Bendaña López, A. Abuín Blanco, L. Bao Pérez, A. Peleteiro Raíndo, M. Cid López, et al., Survival prediction and treatment optimization of multiple myeloma pa- tients using machine-learning models based on clinical and gene expression data, Leukemia 35 (10) (2021) 2924–2935.
12. [25] L. Venezian Povoa, C. H. C. Ribeiro, I. T. d. Silva, Machine learning predicts treatment sensitivity in multiple myeloma based on molecular and clinical information coupled with drug response, PloS one 16 (7) (2021) e0254596.
13. M. V. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, A. Lachmann, et al., Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, Nucleic acids research 44 (W1) (2016) W90–W97.
14. [60] Z. Xie, A. Bailey, M. V. Kuleshov, D. J. Clarke, J. E. Evangelista, S. L. Jenkins, A. Lachmann, M. L. Wojciechowicz, E. Kropiwnicki, K. M. Jagodnik, et al., Gene set knowledge discovery with enrichr, Current protocols 1 (3) (2021) e90.
15. [61] E. Y. Chen, C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark, A. Ma'ayan, Enrichr: interactive and collaborative html5 gene list enrichment analysis tool, BMC bioinformatics 14 (1) (2013) 1–14.
16. D. Chu, L. Wei, Nonsynonymous, synonymous and nonsense mutations in human cancer-related genes undergo stronger purifying selections than expectation, BMC cancer 19 (1) (2019) 1–12.
17. Y. Sharma, M. Miladi, S. Dukare, K. Boulay, M. Caudron-Herger, M. Groß, R. Backofen, S. Diederichs, A pan-cancer analysis of synonymous mutations, Nature communications 10 (1) (2019) 1–14.

References

18. T. Soussi, P. E. Taschner, Y. Samuels, Synonymous somatic variants in human cancer are not infamous: a plea for full disclosure in databases and publications, Human mutation 38 (4) (2017) 339–342.