Details of the research work duly signed by the applicant, for which the "Sun Pharma Science Foundation Research Award-Clinical Sciences" is claimed, including references and illustrations (not to exceed 6000 words).

I have contributed to the understanding of the pathogenesis, clonal evolution, immune microenvironment, biomarker discovery, risk prediction models and therapeutic monitoring in Multiple Myeloma (MM).

My research group analyzed the genomic and transcriptomic signatures of MM at diagnosis and progression, and of Monoclonal Gammopathy of Undetermined Significance (MGUS) which is precursor state of MM to deduced clinically relevant information.

The Mutational Signatures and Tumor mutational burden (TMB) have emerged as prognostic biomarkers in cancer genomics. However, the association of TMB with overall survival (OS) is still unknown in newly diagnosed multiple myeloma (NDMM) patients. Further, the change in the mutational spectrum involving both synonymous and non-synonymous mutations as MGUS progresses to MM is unexplored. To address these questions, we carried out an extensive evaluation of the WES data of 1018 NDMM patients and 61 MGUS patients from three different global regions and inferred single base substitutions, mutational signatures and TMB. This study finds a change in the mutational spectrum with a statistically significant increase from MGUS to MM in the frequency of all the three categories of variants, non-synonymous (NS), synonymous (SYN), and others (OTH) ($P<0.05$: **Figure 1A**). We also reported that 3' and 5'UTR mutations were more frequent in MM and might be responsible for driving MGUS to MM via regulatory binding sites (**Figure 1B**). High TMB is associated with poor progression-free survival (PFS) and OS in NDMM patients (**Figure 1C**). A statistically significant association between the APOBEC activity and poor overall survival in MM was discovered (**Figure 1**).

This work (**Am J Cancer Res 2022;12(4):1919-1933. PMID: 35530275)** demonstrates the role of TMB and regulatory binding sites in pathogenesis of MM. These findings have potential clinical relevance and can assist in designing therapies to inhibit the progression of MGUS to MM and prolong the overall survival in high-risk MM patients.

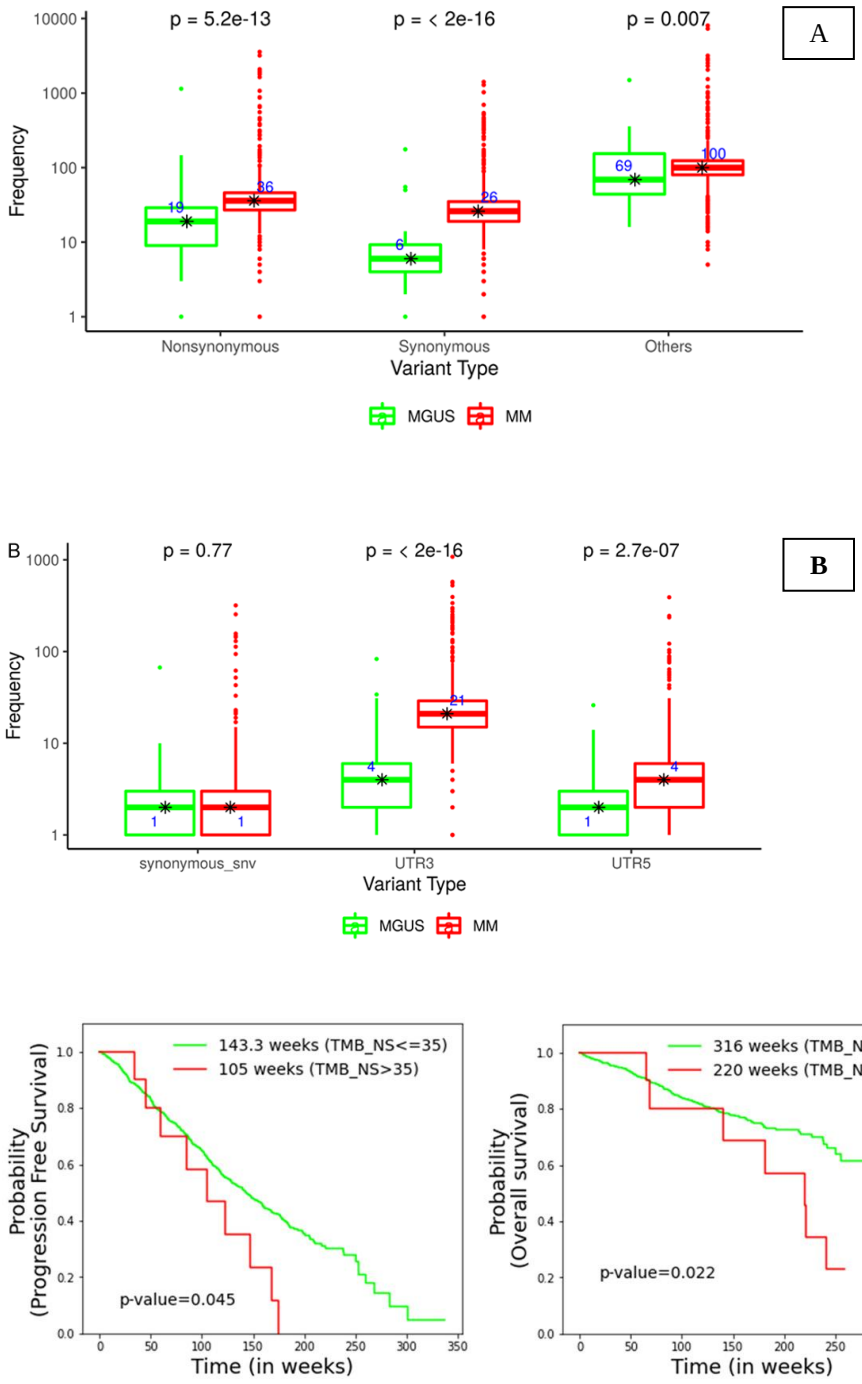


Figure 1: A change in the mutational spectrum with a statistically significant increase from MGUS to MM in the frequency of all the three categories of variants, non-synonymous (NS), synonymous (SYN), and others (OTH) (P<0.05: **Figure 1A**). 3' and 5'UTR mutations are more frequent in MM (**Figure 1B**). High TMB is associated with poor progression-free survival (PFS) and OS in NDMM patients (**Figure 1C**).

The MM arises from its premalignant precursor stages through a complex cascade of interactions between clonal mutations and co-evolving microenvironment and the temporo-spatial evolutionary trajectories of MM are established early during myelomatogenesis in precursor stages and retained in MM. Such molecular events impact subsequent disease progression and clinical outcomes. Identification of clonal sweeps of actionable gene targets in MM could reveal potential vulnerabilities that may exist in early stages and thus potentiate prognostication and customization of early therapeutic interventions. To address this, we evaluated clonal evolution at multiple time points in MM patients. The major findings of this study (**Transl Oncol. 2022 Jun 28; 23:101472. PMID: 35777247**) are (a) MM progresses predominantly through branching evolution, (b) there is a heterogeneous spectrum of mutational landscapes that include unique actionable gene targets at diagnosis compared to progression, (c) unique clonal gains/ losses of mutant driver genes can be identified in patients with different cytogenetic aberrations, (d) there is a significant correlation between co-occurring oncogenic mutations/ co-occurring subclones e.g., with mutated TP53+SYNE1, NRAS+MAGI3, and anticorrelative dependencies between FAT3+FCGBP gene pairs (**Figure 2**). Such co-trajectories synchronize molecular events of drug response, myelomatogenesis and hold potential for early prognostication and development of risk stratified personalized therapies in MM.

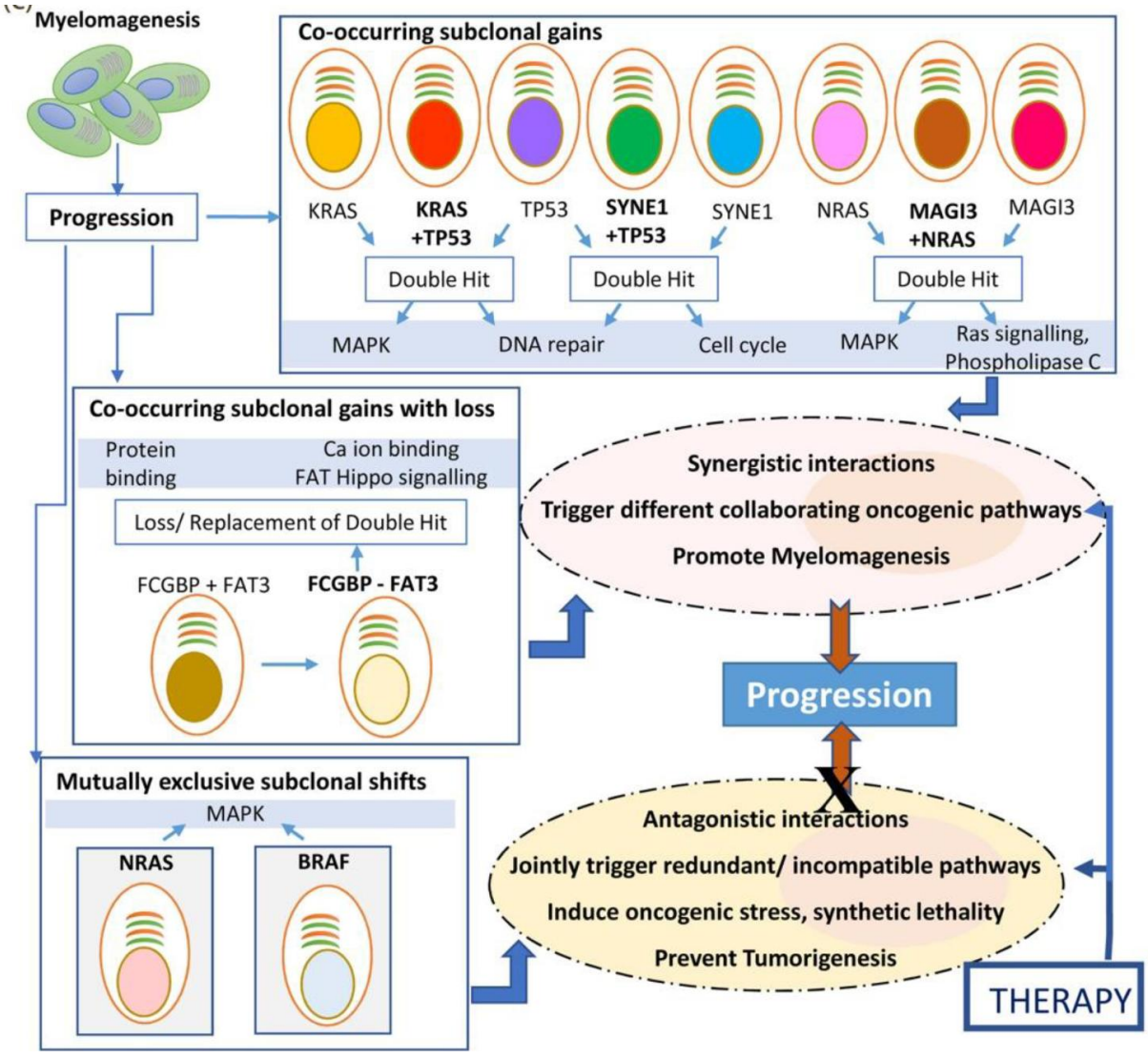


Figure 2: Cartoon showing three types of subclonal shifts (Co-occurring subclonal gains, Co-occurring subclonal gains with loss and mutually exclusive) that can be observed either within same clone or separate clones.. The co-occurring clonal shifts (Co-occurring subclonal gains or Co-occurring subclonal gains with loss) act like 'Double hits' and when present in same clone may promote progression by altering multiple pathways synergistically whereas mutually exclusive clonal shifts tend to result in antagonist synthetic lethal interactions that prevent tumorigenesis. The pathways triggered by such clonal shifts could be explored for development of novel and risk appropriate combinatorial therapies.

Multiple Myeloma (MM) arises from malignant transformation and deregulated proliferation of clonal plasma cells (PCs) harbouring heterogeneous molecular anomalies. The effect of evolving mutations on clone fitness and their cellular prevalence shapes the progressing myeloma genome and impacts clinical outcomes. Although clonal heterogeneity in MM is well established, which subclonal mutations emerge/persist/perish with progression in MM and which of these can be targeted therapeutically remains an open question. To address this, we sequenced pairwise whole exomes of 62 MM patients collected at diagnosis and on progression and demonstrated marked intraclonal heterogeneity and identified actionable/druggable gene targets that varied at different time points within the same patient **(Am J Cancer Res 2021; 11(11):5659-5679. PMID: 34873486).** Branching evolution was predominant pattern of clonal evolution and associated with had low TMBs (<10) and 2 or more founder clones. A distinct temporal fall in subclonal driver mutations was identified from diagnosis to progression e.g., in *PABPC1, BRAF, KRAS, CR1, DIS3* and *ATM* genes suggesting such patients could be treated early with target specific drugs like Vemurafenib/Cobimetinib. An analogous rise in driver mutations was observed in *KMT2C, FOXD4L1, SP140, NRAS* whereas a few drivers such as *FAT4, IGLL5* and *CDKN1A* retained consistent distribution patterns at two time points **(Figure 3)**. These findings are clinically relevant and provide evidence for multi time point evaluation of subclonal mutational landscapes for time to time risk stratification and tailoring time-adapted combination therapies in MM.

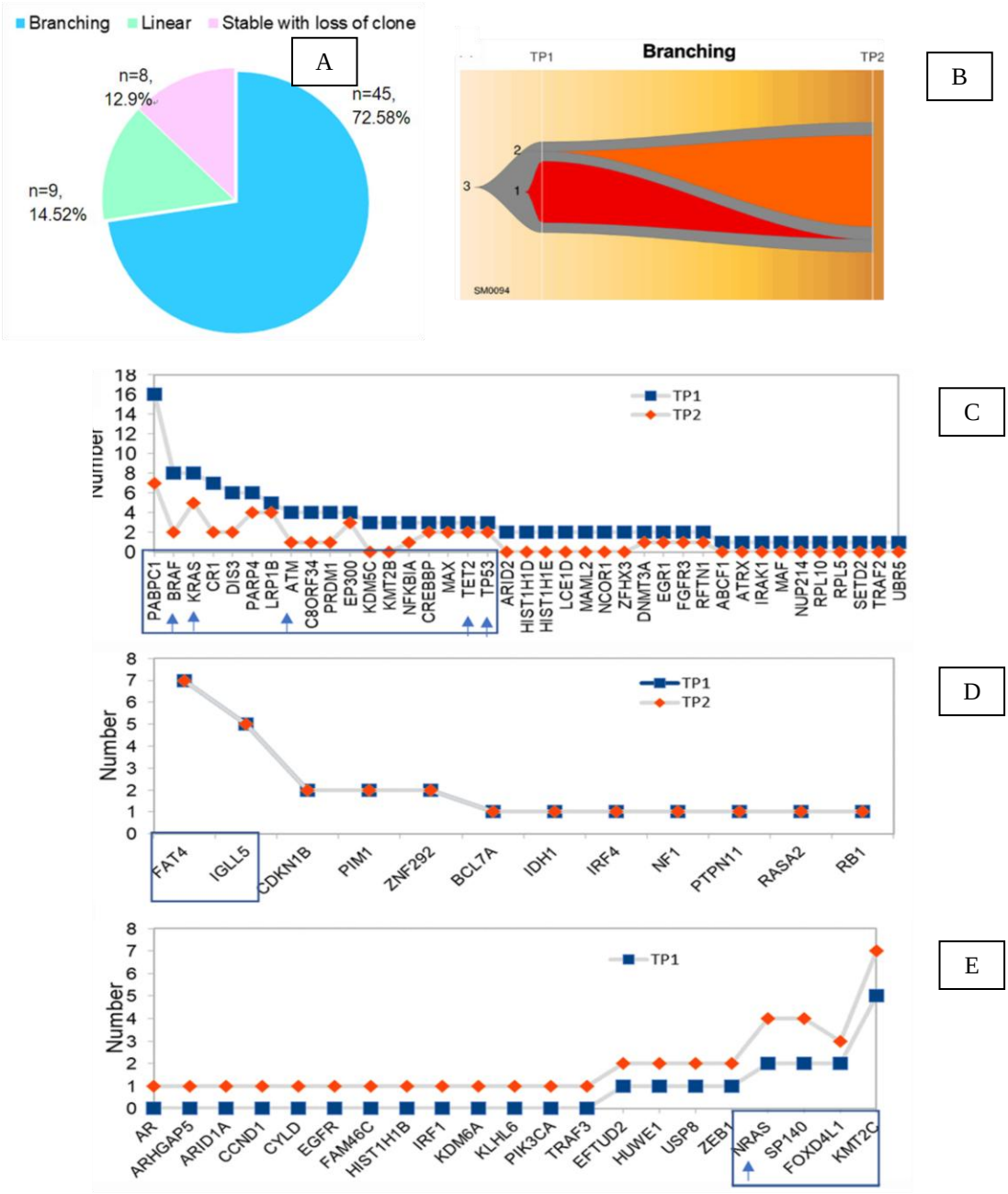


Figure 3: Branching evolution is the predominant pattern of clonal evolution in multiple myeloma (A&B). Temporal changes in distribution of driver genes on progression. Distribution of mutated driver genes in MM patients at TP1 and compared to TP2. (C) Falling mutated drivers whose frequencies decreased in TP2, (D) Drivers that are maintained at constant frequencies throughout the disease, and (E) Rising mutated drivers whose preponderance increased in patients at TP2. Driver mutation profiles observed in atleast 3 or more patients are shown inside boxed frames. Actionable genes are indicated by arrows on X axis.

As is evident from the previous works detailed above, the identification of genomic features responsible for the progression of MM from its precancerous stage MGUS can improve the understanding of the disease pathogenesis and, in devising suitable preventive and treatment measures. We set out to design an innovative AI-based model, namely, the Bio-inspired Deep Learning architecture for the identification of altered Signaling Pathways (BDL-SP) to discover pivotal genomic biomarkers that can potentially distinguish MM from MGUS **(Am J Cancer Res. 2023 Apr 15;13(4):1155-1187. PMID: 37168334).** The proposed BDL-SP model comprehends gene-gene interactions using the PPI network and analyzes genomic features using a deep learning (DL) architecture to identify significantly altered genes and signaling pathways in MM and MGUS. In the quantitative benchmarking with the other popular machine learning models, BDL-SP performed almost similar to the two other best performing predictive ML models of Random Forest and CatBoost. However, an extensive post-hoc explainability analysis, capturing the application specific nuances, clearly established the significance of the BDL-SP model. This analysis revealed that BDL-SP identified a maximum number of previously reported oncogenes, tumor-suppressor genes, and actionable genes of high relevance in MM as the top significantly altered genes. Further, the post-hoc analysis revealed a significant contribution of the total number of single nucleotide variants (SNVs) and genomic features associated with synonymous SNVs in disease stage classification. Finally, the pathway enrichment analysis of the top significantly altered genes showed that many cancer pathways are selectively and significantly dysregulated in MM compared to its precursor stage of MGUS and can thus, be targeted using inhibitors and/or modulators of these specific pathways (**Figure 4).** These observations may pave the way for appropriate therapeutic interventions to halt the progression to overt MM in the future.

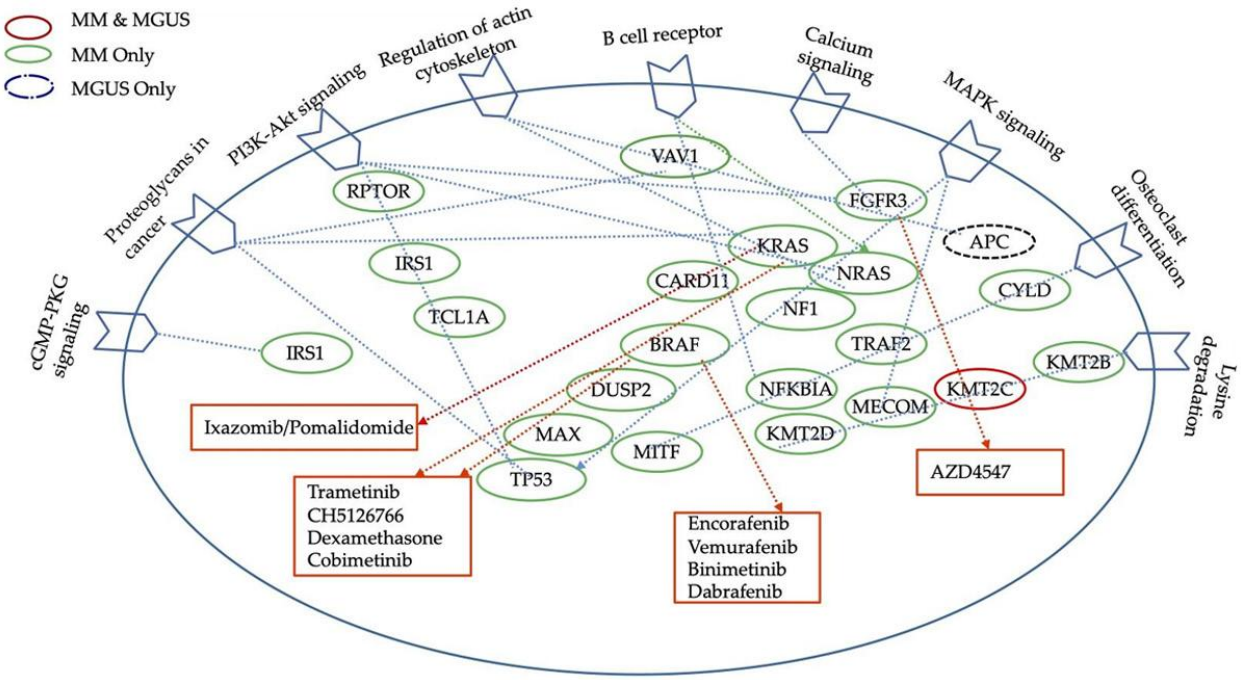
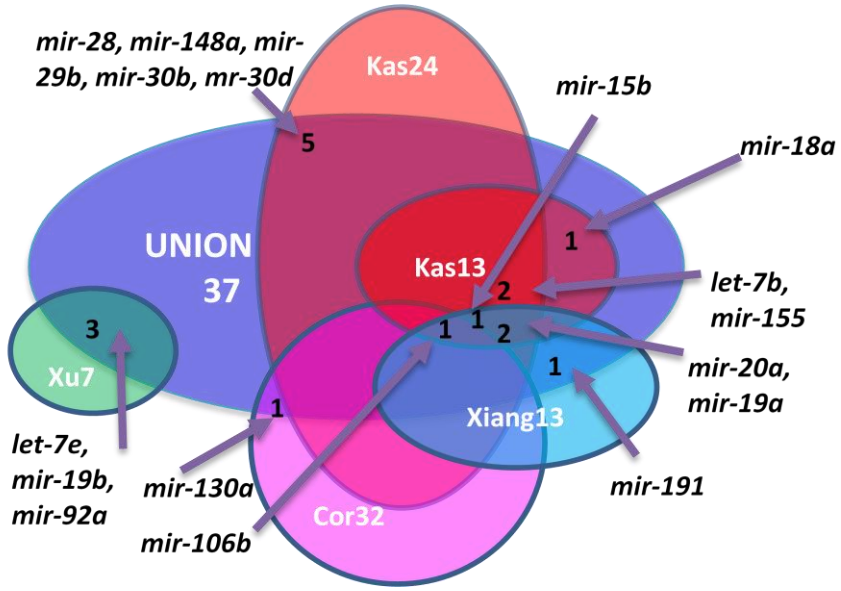


Figure 4: Important pathways significantly altered in MM. Drugs used for pathway-directed therapies associated with mutations in genes are also shown with red colored text-boxes and arrows.

As the differentially expressed genes (DEGs) and miRNAs (DEMs) in MM may influence disease pathogenesis, clinical presentation and drug sensitivities, we conducted genome-wide meta-analysis of DEGs/DEMs in MM versus Normal Plasma Cells (NPCs) and derived unified putative signatures for MM **(Sci Rep. 2021; 11(1):10957. PMID: 34040057).** 100 DEMs and 1,362 DEGs were found deranged between MM and NPCs. Signatures of 37 DEMs ('Union 37') and 154 DEGs ('Union 154') were deduced that shared 17 DEMs and 22 DEGs with published prognostic signatures, respectively (**Figure 5**). Two miRs (miR-16-2-3p, 30d-2-3p) correlated with survival outcomes. PPI analysis identified 5 topmost functionally connected hub genes (UBC, ITGA4, HSP90AB1, VCAM1, VCP). Transcription factor regulatory networks were determined for five seed DEGs with ≥ 4 biomarker applications (CDKN1A, CDKN2A, MMP9, IGF1, MKI67) and three topmost up/ down regulated DEMs (miR-23b, 195, let7b/ miR-20a, 155, 92a). This study identified the functionally connected hub genes and transcription factor regulatory networks with miR-16–2-3p and 30d-2-3p as prognostically relevant miRs that correlated with survival outcomes in MM (**Figure 6**).



UNION 37			
hsa-mir-130a	Cor 32		
hsa-mir-106b	Kas 13	Xiang 13	Cor 32
hsa-mir-19a	Kas 13	Xiang 13	
hsa-mir-20a	Kas 13	Xiang 13	
hsa-mir-18a	Kas 13		
hsa-mir-15b	Kas 24	Kas 13	Xiang 13
hsa-let-7b	Kas 24	Kas 13	
hsa-mir-155	Kas 24	Kas 13	
hsa-mir-28	Kas 24		
hsa-mir-148a	Kas 24		
hsa-mir-29b	Kas 24		
hsa-mir-30b	Kas 24		
hsa-mir-30d	Kas 24		
hsa-mir-191	Xiang 13		
hsa-let-7e	Xu 7		
hsa-mir-19b	Xu 7		
hsa-mir-92a	Xu 7		

Figure 5: Union 37 DEM signatures: A comparison of commonality with analogous published signatures showing an overlap of 22 DEGs and 17 DEMs found in this study

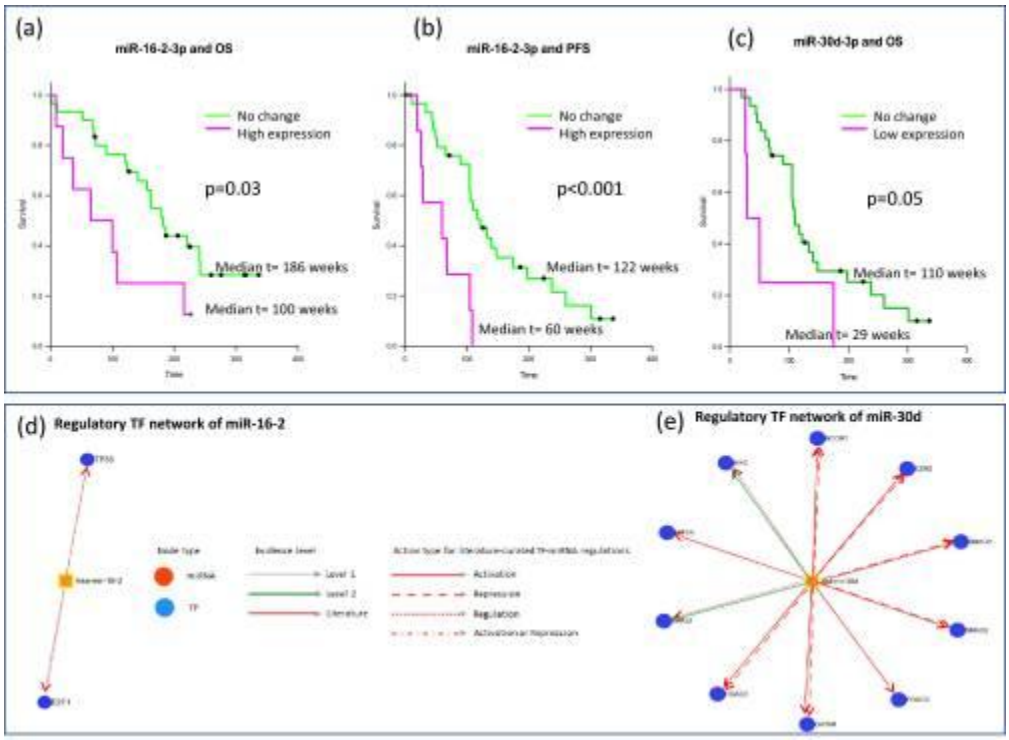
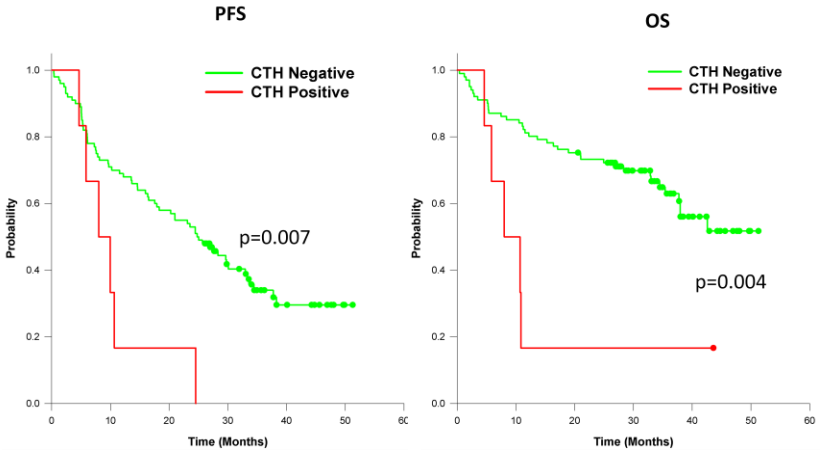


Figure 6: Kaplan Meier plots showing associations of (**a**) miR-16–2-p with OS, (**b**) miR-16–2-3p with PFS and (**c**) miR-30d-3p with OS. Regulatory transcription factor networks of miR-16–2 and of miR-30d are shown in (**d**) and (**e**) respectively.

Complex Chromosomal Rearrangements (CCRs) are increasingly being reported as genetic risk factors of clinical significance in cancer owing to their identification using high resolution whole genome profiling technologies. This study employed high resolution CGH + SNP microarrays for whole genome copy number variations (CNV) profiling and identified CCRs in 10% of newly diagnosed MM patients. Multivariable Cox regression model demonstrated a significant association of CTH with poor PFS (HR = 3.09, p = 0.010) and OS (HR = 3.31, p = 0.024; **Figure 7**) which suggests that CTH is an additional independent prognostic marker in multiple myeloma. Addition of CTH in risk stratification models in clinical setting in multiple myeloma is helpful in upfront identification of high risk patients for suitable customized therapy **(Leuk Res. 2019; 76:58-64. PMID: 30576858)**.



.

Figure 7: Kaplan Meier survival curve analysis of (a) progression free survival and (b) overall survival in multiple myeloma patients with (CTH positive) *Vs* without chromothripsis (CTH negative)

The existing risk predictions models in MM are based on western cohorts. In a systematic evaluation of 2000 patients of MM, we evaluated the risk predictions in the context of ethnicity-specific information and validated robust AI-based models for Indian patients that outperform the existing Revised international staging system. These risk prediction calculators are available online. This work is a value addition as it establishes novel and robust risk-staging models that can be employed in India irrespective of the existing diversity and disparity of our health care infrastructure.

These models include

1) Modified risk staging (MRS) when genomic data on high risk cytogenetic aberrations is not available, a common scenario in a resource constraint setting **(Transl Oncol 2021; 14(9):101157. PMID: 34247136).**

An efficient readily employable risk prognostication method is desirable for MM in settings where genomics tests cannot be performed owing to geographical/economical constraints. In this work, a new Modified Risk Staging (MRS) has been proposed for newly diagnosed Multiple Myeloma (NDMM) that exploits six easy-to-acquire clinical parameters i.e. age, albumin, β2-microglobulin (β2M), calcium, estimated glomerular filtration rate (eGFR) and hemoglobin. MRS was designed using a training set of 716 NDMM patients of our inhouse MM Indian (MMIn) cohort and validated on MMIn (n=354) cohort and MMRF (n=900) cohort. Risk staging rules, obtained via training a J48 classifier, were used to build MRS. New thresholds were identified for albumin (3.6 g/dL), β2M (4.8 mg/L), calcium (11.13 mg/dL), eGFR (48.1 mL/min), and hemoglobin (12.3 g/dL) using KAP on the MMIn dataset (**Figure 8A & B**). On the MMIn dataset, MRS outperformed ISS for OS prediction in terms of C-index, hazard ratios, and its corresponding p-values, but performs comparable in prediction of PFS. On both MMIn and MMRF datasets, MRS performed better than RISS in terms of C-index and p-values. A simple online tool was also designed to allow automated calculation of MRS based on the values of the parameters. Our ML-derived yet simple staging system, MRS, although does not employ genetic features, outperforms RISS as confirmed by better separability in KM survival curves and higher values of C-index and can be implemented in India with limited resources for genomic testing.

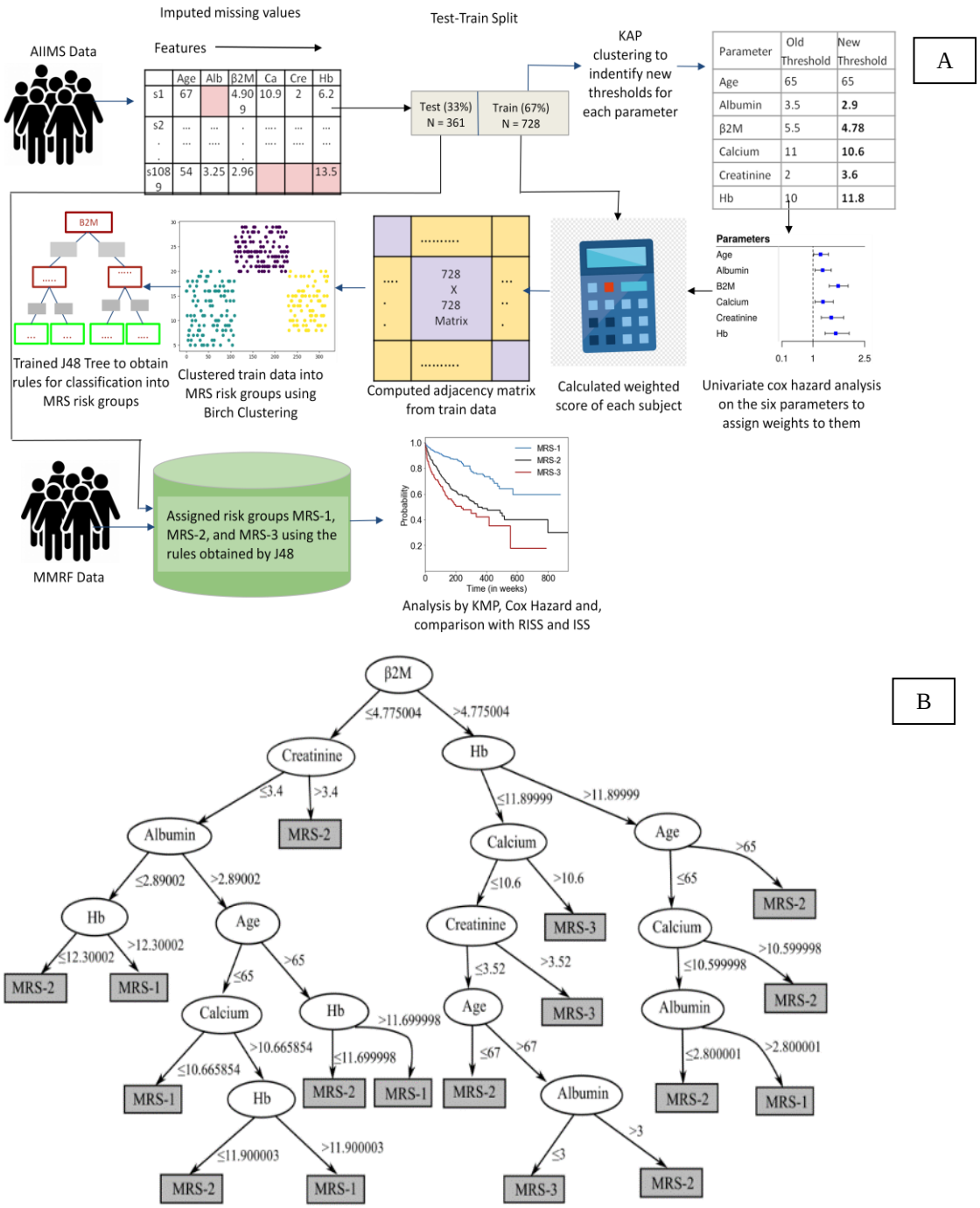


Figure 8: (A) Workflow for the development of Modified Risk staging (MRS) system for Multiple Myeloma.(B) Hierarchical rule based tree structure to assign data samples to MRS-1, MRS-2 and MRS-3. Parameters: Age: Age; Alb: Albumin; β2M: beta2- macroglobulin; Ca: Calcium; eGFR: estimated glomerular filtration rate and Hb: hemoglobin.

2) Consensus based risk-stratification system (CRSS) which integrates genomic data with a SHapley Additive explanation to deduce the biological relevance of the risk predictions (**Front Oncol. 2021; 11:720932. PMID: 34858811)**

Current risk predictors of multiple myeloma do not integrate ethnicity-specific information. However, the impact of ethnicity on disease biology cannot be overlooked. In this study, we have investigated the impact of ethnicity in multiple myeloma risk prediction. In addition, an efficient and robust artificial intelligence (AI)-enabled risk-stratification system is developed for newly diagnosed multiple myeloma (NDMM) patients that utilizes ethnicity-specific cutoffs of key prognostic parameters. K-adaptive partitioning is used to propose new cutoffs of parameters for two different datasets-the MMIn (MM Indian dataset) dataset and the MMRF (Multiple Myeloma Research Foundation) dataset belonging to two different ethnicities. The Consensus-based Risk-Stratification System (CRSS) is designed using the Gaussian mixture model (GMM) and agglomerative clustering. CRSS is validated via Cox hazard proportional methods, Kaplan-Meier analysis, and log-rank tests on PFS and OS. SHAP (SHapley Additive exPlanations) is utilized to establish the biological relevance of the risk prediction by CRSS (**Figure 9**). There is a significant variation in the key prognostic parameters of the two datasets belonging to two different ethnicities. CRSS demonstrates superior performance as compared with the R-ISS in terms of C-index and hazard ratios on both the MMIn and MMRF datasets. An online calculator has been built that can predict the risk stage of a multiple myeloma (MM) patient based on the values of parameters and ethnicity. Our methodology discovers changes in the cutoffs with ethnicities from the established cutoffs of prognostic features and the best predictor model for both cohorts was obtained with the new ethnicity-specific cutoffs of clinical parameters. Our study revealed the efficacy of AI in building a deployable risk prediction system for Indian patients with MM which is now available to Indian medical community.

Link to online calculators:
http://sbilab.iiitd.edu.in/pub_files/CRRScalculator_edit.html

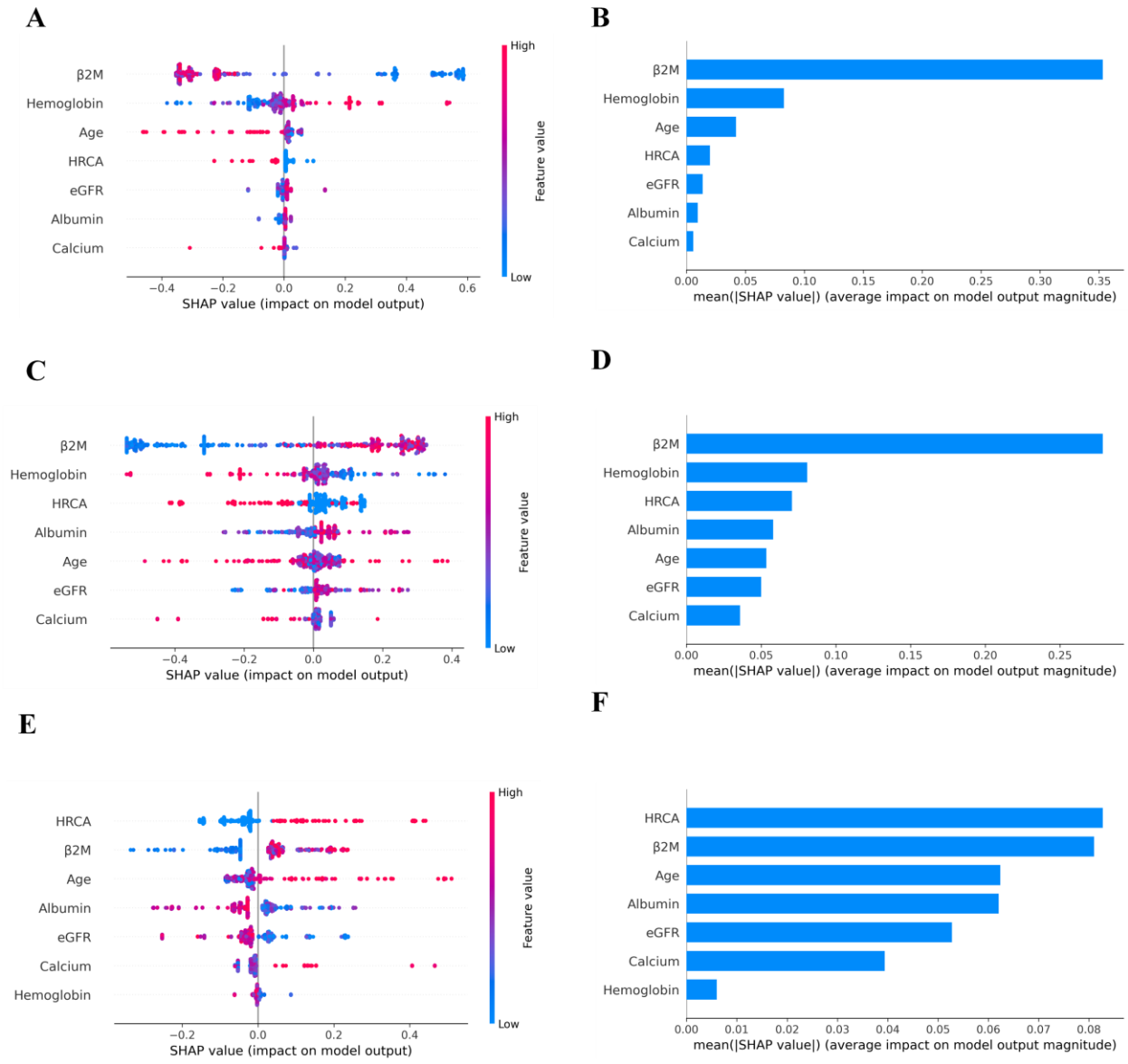


Figure 9: Model interpretation using SHAP (Shapley Additive Explanations). SHAP summary plots for different risk stages inferred in MMIn data showing the relative impact of different parameters (top to bottom) contributing to a particular risk stage prediction.
A, B - CRSS-1: Normal levels of β2M & hemoglobin are key contributors to the low risk stage prediction. Further, high values of age on the left side of the summary plot are pushing the model away from the low risk prediction and are indicative of either intermediate or high risk. Overall β2M has the highest impact and calcium has the lowest impact on low risk stage prediction.
C, D - CRSS-2: Elevated levels of β2M and lower levels of hemoglobin are the key contributors to the intermediate risk stage.
E, F - CRSS-3: Presence of HRCA is contributing the most to the high risk stage. Elevated values of β2M, calcium and lower levels of albumin, hemoglobin, and eGFR are contributing towards the high risk stage prediction.

3) Graph Convolution Risk Staging (GCRS), a deep neural network based model which further improves risk stratification **(Comput Biol Med. 2022; 149:106048. doi: 10.1016/j.compbiomed.2022.106048. PMID: 36113255).**

In this study, we present an efficient Graph Convolutional Network based Risk Stratification system (GCRS) for cancer risk-stage prediction of newly diagnosed multiple myeloma (NDMM) patients. GCRS is a hybrid graph convolutional network consisting of a fusion of multiple connectivity graphs that are used to learn the latent representation of topological structures among patients. This proposed risk stratification system integrates these connectivity graphs prepared from the clinical and laboratory characteristics of NDMM cancer patients for partitioning them into three cancer risk groups: low, intermediate, and high. Extensive experiments demonstrate that GCRS outperforms the existing state-of-the-art methods in terms of C-index and hazard ratio on two publicly available datasets of NDMM patients. We have statistically validated our results using the Cox Proportional-Hazards model, Kaplan-Meier analysis, and log-rank test on progression-free survival (PFS) and overall survival (OS). We have also evaluated the contribution of various clinical parameters as utilized by the GCRS risk stratification system using the SHapley Additive exPlanations (SHAP) analysis, an interpretability algorithm for validating AI methods. Our study reveals the utility of the deep learning approach in building a robust system for cancer risk stage prediction.

With a substantial number of patients with multiple myeloma (MM) experiencing disease relapse, the quest for more sensitive methods to assess deeper responses indicative of cure continues. In this prospective analysis of patients with MM after autologous stem cell transplant, we evaluated the predictive value of conventional response, measurable residual disease, and neoplastic plasma cell index**.**

We, thus, established the Next Generation Flow cytometry (NGF) assay for Measurable Residual Disease (MRD) and defined cut-offs for dynamic risk monitoring and, introduced a novel statistic of neoplastic plasma cell index (NPCI) to overcome the caveat of hemodilution that mars the utility of NGF in clinical practice **(Am J Clin Pathol 2023; IF=5.4).** Significantly better progression-free survival (PFS) and overall survival (OS) were observed with deepening conventional response. Conventional response-based stratification within the MRD-positive and MRD-negative subgroups showed a significantly higher PFS and OS in the conventional response-positive/MRD-positive group compared with the conventional response-negative/MRD-positive group. Using K-adaptive partitioning to find the optimum threshold for MRD, patients achieving less than 0.001% MRD had superior PFS and OS. The C index and Akaike information criterion metrics demonstrated the superior performance of the NPCI compared with MRD in predicting treatment outcome. Progressive deepening of response, conventional as well as MRD correlates with superior survival outcomes. The NPCI proved to be a superior determinant of survival and can be explored as a better statistic than MRD, especially in hemodiluted aspirates.

We also described the pattern-based recognition of neoplastic plasma cells and immunomodulation after chemotherapy that impacts residual disease assessment in MM (**Clinical Cytometry 2022; Am J Clin Pathol 2009**). This work has enhanced the disease monitoring in the clinical stem cell transplantation program at AIIMS.