Whole Genome Sequencing Reveals the Extent of Structural Variants in Multiple Myeloma and Identifies Recurrent Mutational Hotspots Within the Non-Coding Regions

Cody Ashby, Michael A. Bauer, Owen W. Stephens, Christopher P. Wardell, Ruslana G. Tytarenko, Aneta Mikulasova, Purvi Patel, Shayu Deshpande, Frits van Rhee, Maurizio Zangari, Sharmilan Thanendrarajan, Carolina D. Schinke, Niels Weinhold, Leo Rasche, Faith E Davies, Gareth J. Morgan and Brian A Walker

UAMS Myeloma Institute, University of Arkansas for Medical Sciences, Little Rock, AR, USA

INTRODUCTION

- Genetic analysis of Multiple Myeloma (MM) has focused on the coding region, comprising 2% of the total genome. This has left important driver events and mutational mechanisms undefined within the much larger non-coding regions.
- Recent advances in sequencing have allowed phasing of barcoded reads, permitting a more faithful representation of DNA sequence and an ability to better identify structural variants (SVs), the sites at which they occur, and the underlying mechanisms leading to their generation.
- We have used 10X Genomics Chromium Whole Genome Sequencing (WGS) to study the genome of a series of newly diagnosed MM (NDMM) patient samples with a focus on the non-coding regions, structural events and mutational mechanisms.

MATERIALS AND METHODS

- We analyzed 70 pairs of tumor and germline control NDMM samples using WGS.
- The samples consisted of the following cytogenetic groups:
  - t(4;14) (n=12), t(11;14) (n=11), t(14;16) (n=18) , t(14;20) (n=10)
  - hyperdiploidy (n=23; some with co-occurring translocations).
  - gene expression derived high risk (n=30) and low risk (n=40)
- Analysis pipeline is shown in **Figure 1** and alignment metrics are shown in **Table 1.**

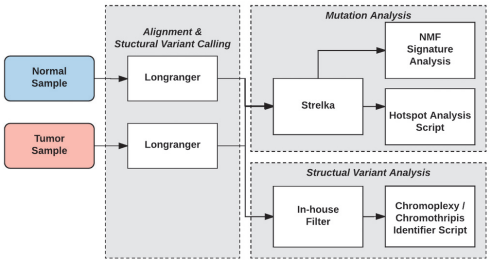

Figure 1 – Pipeline used for analysis. Longranger was used on alignment of tumor and normal samples which were used in mutation and structural variant calling.

Table 1: Quality control metrics for samples aligned using Longranger.

metric	median	min	max
Depth	51x	31x	78x
Mean Molecule Length	28 kb	6.7 kb	65 kb
Phase block N50	761 kb	49 kb	6320 kb

RESULTS

WGS Identifies Multiple Types of Structural Variants

Samples exhibited multiple types of structural variants including:

- **Chromothripsis** - chromosomal shattering and recombination involving up to hundreds of segments usually affecting 1 or 2 chromosomes (**Figure 1A**). Chromothripsis was seen in 19% of samples.
- **Chromoplexy** – complex structural events involving 5 or more chromosomes (**Figure 1B**). Chromoplexy was seen in 10% of samples.

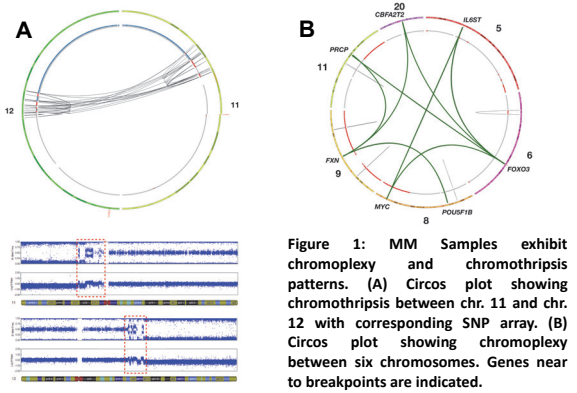

Figure 1: MM Samples exhibit chromoplexy and chromothripsis patterns. (A) Circos plot showing chromothripsis between chr. 11 and chr. 12 with corresponding SNP array. (B) Circos plot showing chromoplexy between six chromosomes. Genes near to breakpoints are indicated.

Number of Structural Variants Differs by Chromosome

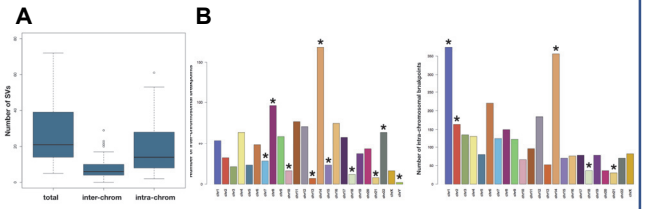

Figure 2: Number of SVs across samples varies, but certain chromosomes have significantly more or less breakpoints. (A) boxplot of total, inter-chromosomal and intra-chromosomal SVs. (B) Number of inter-chromosomal (left) and intra-chromosomal (right) breakpoints per chromosome (asterisk indicates significance by Wilcoxon $p < 0.05$).

RESULTS

Samples with ABOBEC Signature Have More Mutations and SVs

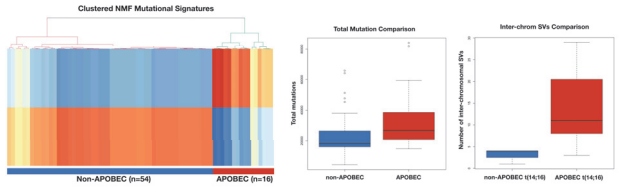

Figure 3: Samples with APOBEC signatures have more mutations and in the case of t(14;16) more inter-chromosomal SVs.

Mutational Analysis Identifies Hotspots in the Non-coding Regions

Table 2: Mutational counts for samples using Strelka2 (after filtering).

metric	median	min	max
Total Mutations	19521	4240	84041
Genic Mutations	10131	2192	45249
Intergenic Mutations	9343	2048	38792

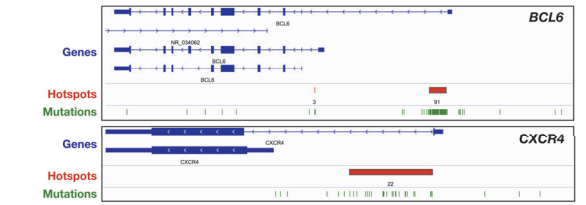

Figure 4: Integrative Genome Viewer showing mutational hotspots around the TSS of *BCL6* (top) and *CXCR4* (bottom).

CONCLUSIONS

- Using 10X Genomics WGS we show there is significant somatic variation in the non-coding regions of the MM genome.
- We identified significant mutation in non-coding regions of key hematological genes, including *BCL6* (hypermutated in DLBCL) and *CXCR4* (mutated in Waldenström's macroglobulinemia), indicating that the mutational driver spectrum extends beyond the coding regions of MM.
- We identified complex SVs, involving multiple chromosomes, resulting from chromoplexy indicating that genome instability is key in high risk MM.

Myeloma Institute
University of Arkansas for Medical Sciences

Contact: tcashby@uams.edu