Characterization of the rat osteosarcoma UMR-106 using long read methods identifies shared patterns of amplification, mutation, methylation and expression with human disease

DW Mohr[1], AF Scott[1], WA Littrell[1], M Kokosinski[2], R Babu, YS Zou[3], V Stinnett[3], J Madhiwala[4], J Anderson[4], K Gabrielson[4,5]

1 Genetic Resources Core Facility, Johns Hopkins Genomics, Dept. Gen. Med., Johns Hopkins University School of Medicine , 2 Johns Hopkins Genomics, Dept. Gen. Med., Johns Hopkins University School of Medicine, 3 Cytogenetic Laboratory, Department of Pathology, Johns Hopkins Genomics, Johns Hopkins University School of Medicine,4 Department of Molecular and Comparative Pathobiology, Johns Hopkins University School of Medicine,5 Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University

Introduction

The rat osteosarcoma cell line UMR-106 is widely used for the study of bone cancer biology. We have characterized this cell line using a combination of optical genome mapping (OGM), nanopore long-read sequencing and short-read RNAseq. The genome sequence and methylation patterns were compared to a Sprague-Dawley control animal, the strain from which UMR-106 was derived and expression data were compared to a public rat osteoblast dataset. Nanopore sequencing identified UMR-106 specific changes in copy number, structural variants, likely oncogenic mutations in Tp53 and Myc as well as methylation differences in the tumor suppressor CDKN2A each of which have been reported in the COSMIC database for genes most affected in human osteosarcomas. OGM found numerous translocations and other genomic events characteristic of widespread chromothripsis. The combined optical mapping and nanopore sequencing identified a region of approximately 30 Mb with about 5-7 X amplification near Myc on rat chromosome 7. Many genes in this region also showed increased expression vs. that reported in osteoblasts. Several of these have been cited as biomarkers in human osteosarcoma along with Myc, including Mdm2, Angpt1, TRps1, Eif3h, Deptor, Has2, Fam91a1, Tmem65, and Rnf139. The observation that so many loci from the amplified region are seen in human osteosarcoma suggests that this may be more than coincidental. This dataset illustrates the value and complementarity of both long DNA methods, optical mapping and sequencing, for the characterization of cell lines. The study also illustrates how inter-species analyses can inform about the genetic events that underpin specific tumor types. The data should be a valuable resource for investigators studying osteosarcoma and, specifically, the UMR-106 model of this disease.

Methods

HMW DNA was isolated from freshly expanded rat osteosarcoma UMR-106 cells using the NEB Monarch kit and an ultra-long ONT library was prepared using the SQK-ULK114 kit. A Qiagen DNA prep was done on a female Sprague-Dawley rat to correspond to the cell line and a standard ONT library was made (SQK-LSK114). The ULK library was made on the cell line in anticipation of likely structural variants. Both libraries were run on 10.4.1 PromethION flow cells with triple loading. The UMR resulted in 45.3 Gb of aligned bases (to the rat rn7.2 reference) with an N50 of 78 kb.

Data were aligned to reference using minimap2 (1). Methylation calls were generated using modbam2bed(2). Modbamtools was used to plot methylation data (3). Summary Statistics were generated with Nanoplot (4).

OGM methods: DNA from UMR-106 cells and from RBCs of the control animal were isolated. Optical Genome Mapping (OGM) methods: ultra-high-molecular-weight (UHMW) genomic DNA (gDNA) from UMR-106 cells and from RBCs of the control animal were isolated via the blood and cell culture DNA Isolation Kit according to manufacturers' instructions (Bionano Genomics, San Diego, CA, USA). Briefly, cells were treated with lysis-and-binding buffer (LBB) to release gDNA, which was bound to a nanobind disk before it was washed and eluted in the elution buffer. The Direct Label and Stain (DLS) DNA Labeling Kit (Bionano Genomics, San Diego, CA, USA) was used to label UHMW gDNA molecules. 750 ng of gDNA was labeled with Direct Label Enzyme (DLE-1) and DL-green fluorophores. The G3.3 chips were utilized, and samples were processed on the Bionano Saphyr instrument (San Diego, CA, USA). OGM analysis was performed using the Rare Variant Analysis (RVA) pipelines, utilizing the Bionano Access software v1.7.2. CNVs and SVs were manually determined by genetic analysts.

RNASeq methods:

Isolated RNA from UMR-106 was poly-A selected using the NEBNExt Poly(A) isolation module (NEB #E7490) and libraries were made using NEBNext Ultra II RNA Library Prep Kit for Illumina (Cat# E7775). Sequencing was performed on a NovaSeq 6000 SP 2x50bp run. FASTQ data were aligned to reference using the STAR aligner (5) Control osteoblast data were from SRR16368266. A total of 800 million reads from UMR-106 and 45 million reads from the osteoblast control were obtained. Data were down-sampled to match read length and approximate depth using samtools(6).

Results

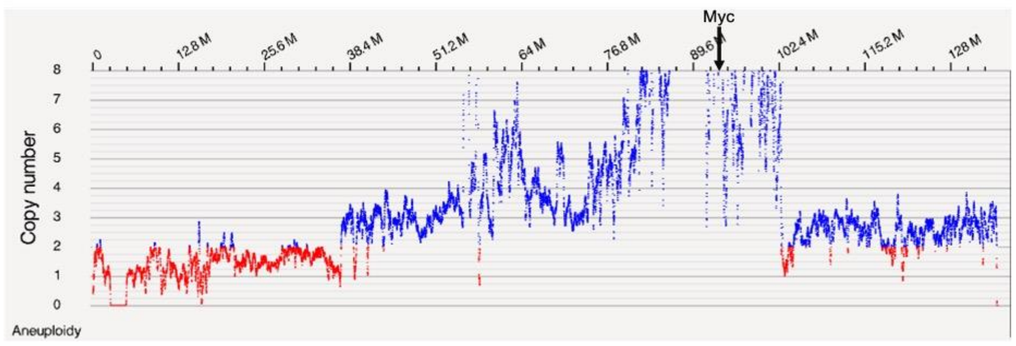


Fig. 1. Optical mapping (OGM) molecule depth shows CN increase along UMR-106 chromosome 7

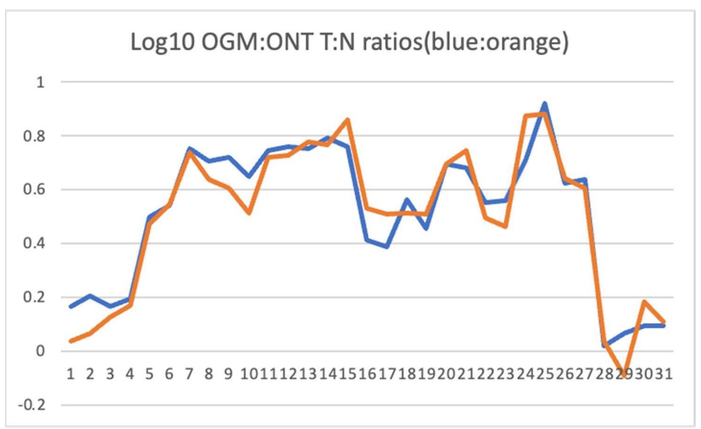


Fig. 2. Concordance of OGM and nanopore (ONT) depth across the amplified block. Data is converted to log(10) values. T=UMR-106, N=sex matched healthy Sprague-Dawley rat.

COSMIC Catalogue Of Somatic Mutations In Cancer

Gene	Change	Comments
H3F3A	DEL	10,995 bp Deletion
TP53	C>T	Nonsense Mutation in COSMIC
BRAF	A>G	Missense Mutation in COSMIC
KMT2D	DEL	7 base deletion followed by 2 base deletion
KMT2D	INS	15 base insertion
MYC	T>C	Missense Mutation in COSMIC, amplified copies

The COSMIC database was searched for the top twenty human genes associated with osteosarcoma. Mutations in Tp53 and Myc were observed in UMR-106. The Tp53 was at an orthologous position as reported in humans.

The Variant Effect Predictor (VEP) was used to produce a list of all variants in both UMR-106 and the control. Variants identified with a high score only in the cell line were manually reviewed. None were found in known cancer genes.

Of the top COSMIC osteosarcoma genes we identified mutations in H3F3A, TP53, BRAF, KMT2D, and MYC.

Chromosome 7 Amplified Region:

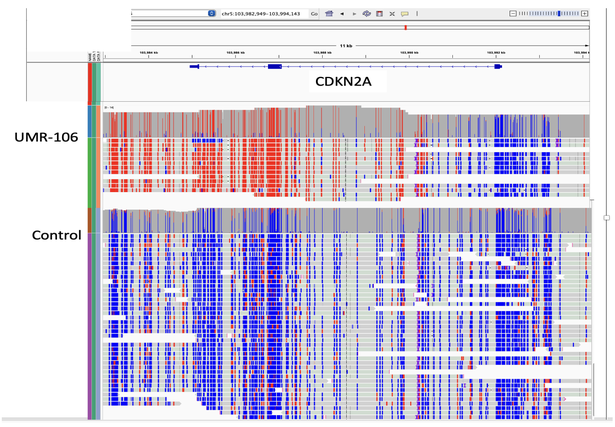
We looked at several genes in the amplified block flanking Myc by calculating the log(10) values for the ratio of T:N OGM reads, ONT depth and RNAseq TPM for UMR-106 compared to a public osteoblast control data set (ref).

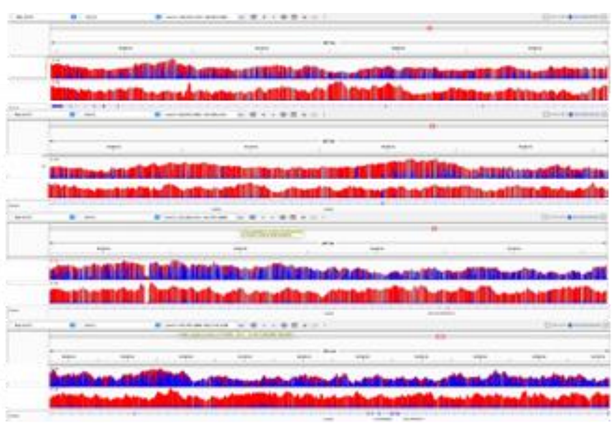
Genes with statistically significant overexpression in UMR vs osteoblasts are underlined.

As shown in the table we noted that 14 of the genes have been cited as osteosarcoma (OS) markers.

				OGM T:N	ONT T:N	RNAseq	T-Test	
				Molecules	Depth	TPM T:N	2 sample	Comments
Gene	Chr	Start	End	Ratio log(10)	Ratio log(10)	Ratio log(10)	UMR:SRR p-value	
Mdm2	7	53290660	53315205	0.167	0.037	-0.209	NS	OS amp
Mdm1	7	53729603	53766034	0.207	0.064	0.419	NS	
Oxr1	7	72528750	72965666	0.167	0.127	-0.109	NS	
Angpt1	7	73528345	73783953	0.196	0.17	-0.703	0.001	OS marker
Csmd3	7	78747322	80066466	0.5	0.473		NS	mut in ESCC
Trps1	7	81916668	82142733	0.541	0.544	0.374	NS	OS marker
Eif3h	7	83091037	83174451	0.753	0.737	0.887	<.001	OS marker
Taf2	7	86422613	86479616	0.706	0.639	0.952	<.001	BRC amp
Deptor	7	86514859	86668817	0.721	0.604	-0.707	0.002	OS marker
Has2	7	88113326	88139337	0.649	0.515	-1.152	<.001	OS marker
Zhx2	7	89226358	89374266	0.744	0.719	-0.222	NS	Cancer related
Fam91a1	7	89969558	90007546	0.758	0.727	0.796	<.001	OS marker
Tmem65	7	90336997	90378930	0.751	0.776	0.838	<.001	OS marker
Rnf139	7	90439726	90450911	0.793	0.767	0.928	<.001	OS marker
Myc	7	93593705	93598633	0.759	0.859	0.544	<.001	OS marker
Gsdmc	7	95594015	95606106	0.412	0.531		NS	Cancer related
Cyrib	7	95633876	95760588	0.389	0.511	0.719	<.001	Cancer related
Asap1	7	95786130	96093111	0.562	0.512	0.58	<.001	Cancer related
Adcy8	7	96417310	96665911	0.457	0.508	0.699	NS	Cancer related
Efr3a	7	97552677	97633369	0.696	0.695	0.755	<.001	Cancer related
Kcnq3	7	97730219	98025652	0.68	0.744		NS	
Phf20l1	7	98330580	98396526	0.552	0.496	0.089	NS	Cancer related
Ccn4	7	98645238	98677253	0.558	0.464	0.748	<.001	OS marker
Ndrg1	7	98684487	98725869	0.708	0.873	0.508	<.001	OS marker
St3gal1	7	98845270	98913409	0.922	0.882	0.236	NS	OS marker
Zfat	7	99886954	100054288	0.624	0.641	0.637	NS	Cancer related
Khdrbs3	7	100837707	100995644	0.636	0.604	1.655	<.001	Cancer related
Col22a1	7	103730939	103968452	0.021	0.037		<.001	OS marker
Trappc9	7	104521593	104998352	0.068	-0.092	-0.073	NS	Cancer related
Chrac1	7	105013047	105016435	0.095	0.185	0.091	NS	Cancer related
Mfng	7	110310810	110328653	0.096	0.111		NS	Cancer related

An 11 kb deletion in H3f3a long ONT reads from UMR-106 relative to the control. Highlighted reads span the deletion and precisely identify the end points.





IGV(7) display showing hypermethylation (red) of the 3' end of Cdkn2A including exons 2 and 3. and Lsamp hypomethylation comparing UMR-106 to SD. Direction of transcription for both genes is right to left. Expression differences in both loci are associated with cancer progression(9).

Conclusion

Only a small portion of the UMR-106 genome was characterized in this study, but there are clearly many similarities between human osteosarcomas and this rat cell line that validate it as a useful model for this cancer. We found relatively few Cosmic OS-associated genes were mutated in UMR-106 with the exception of Tp53 and Myc. The observation that identical somatic mutations were shared in these two orthologous genes argues that these mutations are particularly important, and that characterization of related tumor types between species may offer new further insight into the significance of specific gene modifications and somatic mutations. Perhaps more significantly this study highlights the role that increased copy number plays in cancer. We noted that amplified region including Mcyc also contained 24 other genes of which 13 have been reported as OS markers and 14 implicated in other cancers, two of which are amplified in breast or esophageal cancer.

The combination of long-DNA methods that we used illustrates the power of newer technologies to rapidly characterize genomes. The ability to observe methylation differences with ONT sequencing may be particularly valuable. Because so many cell lines are commonly used in biological studies, we argue that these tools should see widespread use in genomic analysis beyond cancer. Cells should be recharacterized over time and, especially, whenever their phenotypes in *in vitro* or *in vivo* change.

References

1. Li, H. (2017). Minimap2: fast pairwise alignment for long nucleotide sequences. arXiv:1708.01492
2. Modbam2bed: https://github.com/epi2me-labs/modbam2bed
3. Roham Razaghi, Paul W. Hook, Shujun Ou, Michael C. Schatz, Kasper D. Hansen, Miten Jain, Winston Timp. Modbamtools: Analysis of single-molecule epigenetic data for long-range profiling, heterogeneity, and clustering. 2022. doi:10.1101/2022.07.07.499188
4. Wouter De Coster, Rosa Rademakers, NanoPack2: population-scale evaluation of long-read sequencing data. 2023. doi:10.1093/bioinformatics/btad311
5. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013. doi: 10.1093/bioinformatics/bts635
6. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. 2009. doi: 10.1093/bioinformatics/btp352.
7. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative Genomics Viewer. Nature biotechnology. 2011;29(1):24-26. doi:10.1038/nbt.1754.
8. Hua Y et al (2011). Revealing NDRG1 as a marker for osteosarcoma. Tumour Biol PMID: 21706236.
9. 9. Righi A, et al (2016). P16 expression as a prognostic and predictive marker for high-grade localized osteosarcoma. Human Pathol PMID 27544803.
10. Riffo-Campos AL, et al (2018) Role of epigenetic factors in alternative splicing isoforms of human KRAS in colorectal cancer cell lines. Oncotarget 9:20578. PMID 29755673.

Contact information: David Mohr: dwmohr@jhmi.edu, Alan Scott: afscott@jhmi.edu , Kathy Gabrielson: kgabriel@jhmi.edu