**Analysis Concept Proposals**

1. **Study title:** Whole-genome sequencing of the youngest osteosarcoma cases.

2. **Working group and investigators:**


   - CCSS Working Groups:
     1. Genetics: primary
     2. Epidemiology/Biostatistics: secondary
   - Investigators*:
     1. Lisa Mirabello: mirabellol@mail.nih.gov; fax 240-276-7836
     2. Matthew Gianferante: matthew.gianferante@nih.gov; fax 240-276-7836
     3. Yutaka Yasui: yutaka.yasui@stjude.org
     4. Wendy Leisenring: wleisenr@fredhutch.org
     5. Smita Bhatia: smitabhatia@uabmc.edu
     6. Greg Armstrong: greg.armstrong@stjude.org
     7. Kirsten Ness: kiri.ness@stjude.org
     8. Melissa Hudson: melissa.hudson@stjude.org


   *all investigators will be included as coauthors on any resultant manuscripts.

3. **Background and rationale:**

   Osteosarcoma is the most common primary malignant bone tumor in adolescents and young adults, its etiology is poorly characterized, and metastatic osteosarcoma continues to have poor outcomes. Most cases are sporadic with no known predisposing factor. Osteosarcoma risk factors include prior therapeutic radiation,[1] tall stature,[2, 3] high birth-weight,[2, 4] and at least eight established cancer predisposition syndromes,[5, 6] including Li-Fraumeni Syndrome[7] and hereditary retinoblastoma.[8] Our goal is to determine the underlying germline genetic architecture of pediatric osteosarcoma. Genetic studies to date suggest a role for both common single nucleotide polymorphisms (SNPs) and rare germline variants in osteosarcoma,[9-14] supporting a complex underlying architecture for its genetic etiology that appears to be weighted disproportionately towards rare variants.

   Large sequencing studies suggest that germline mutations in cancer-susceptibility genes may play a major role in osteosarcoma pathogenesis especially in the youngest cases. Epidemiologic patterns of osteosarcoma,[15-17] and our *TP53* and exome sequencing data,[12, 13] suggest that early-onset osteosarcoma has a different etiology than osteosarcoma of adult-onset. We showed that cases carrying a germline pathogenic or likely pathogenic (P/LP) variant were significantly younger than cases without a P/LP variant, and the youngest cases, aged 0-10 years, had the highest frequency of autosomal dominant P/LP variants (25% of cases).[12] There is a paucity of data on the youngest osteosarcoma cases, as they are typically grouped with adolescent and young adult cases due to the rarity of disease. Therefore, since the youngest cases have the strongest genetic predisposition, we propose to perform whole-genome sequencing (and utilize existing WGS data) to try to identify additional pathogenic variants, with a special focus on structural variants, in these young cases. The findings from this study will improve our understanding of the genetic etiology of osteosarcoma and may provide new insight into the biology of osteosarcoma.

4. **Specific aims/objectives/research hypotheses:**

The objective of this study is to determine if young children (aged <10 years old) have an undetected germline SNV, INDEL, or structural variant that predisposed them to osteosarcoma using whole-genome sequencing.

There are a total of 429 cases aged 0-10 years in our current dataset. We hypothesize that these young cases may have other pathogenic variants, including SNVs, INDELs, and structural variants (i.e., copy number changes, rearrangements), that predisposed them to osteosarcoma. Whole-genome sequencing yields more pathogenic variants than exome sequencing, primarily due to improved sequencing coverage and detection of structural variants.[18, 19] We propose to conduct whole-genome sequencing to detect genomic events that are often not detectable with exome sequencing or genotype data on as many of these young cases as possible. There are 33 CCSS first primary osteosarcoma cases from the CCSS original cohort diagnoses 1970-1986, aged 0-10 years at the Cancer Genomics Research Laboratory of DCEG with DNA available for whole-genome sequencing. Here, we are requesting permission to include these 33 young-onset osteosarcoma cases for whole-genome sequencing.

In addition, we also plan to utilize the WGS and WES data available for 39 osteosarcoma cases aged 0-10 years from the St. Jude Cloud (16 CCSS Expansion Cohort Cases + 23 SJLIFE cases). We will be requesting access to these data in parallel.

5. **Analysis framework:**

The study population will include a total of approximately 429 young cases (including 33 CCSS cases with DNA available that have not been whole-genome amplified) with no identified pathogenic predisposing variant from our analyses of the exome sequencing data, plus cases aged 0-10 years from the St. Jude Cloud. Whole-genome sequencing will be performed using indexed Illumina NGS libraries prepared from the germline DNA using the KAPA HyperPlus (KAPA Biosystems, Wilmington, MA) library preparation kit, and then sequenced on an Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA) at an average depth of 30-40× using paired-end read lengths of 2x150bp. A series of quality control steps will be implemented, including evaluating sequencing reads (FastQC), duplications, contamination, individual subject coverage, cohort coverage, sex concordance, and pre-variant-calling. Single nucleotide variants (SNVs) and small insertions and deletions (INDELs) will be detected using the HaplotypeCaller module from Genome Analysis Toolkit (GATK v4.0), Google's deep learning based variant caller DeepVariant (v0.5.2), and Illumina's Strelka2 (v2.9). An in-house developed Ensemble workflow will be applied to filter and integrate variant calling results from these three callers. A post-variant-calling check will also be conducted to assess the variant calling quality. Variant annotation, including pathogenicity classification, will be performed using an in-house custom pipeline.[20] Structural variants will be detected with GATK-SV, a comprehensive multi-algorithm pipeline.[21] Using this pipeline, structural variants will be called with Manta, MELT, Whamg, GATK-gCNV, and cn.MOPS then jointly clustered, filtered, and genotyped across the entire cohort. Structural variants will be annotated with a GATK-SV built-in workflow and with AnnotSV.[22] Pathogenicity will be interpreted according to American College of Medical Genetics (ACMG) guidelines.[23]

SNVs, INDELs, and structural variants of interest will be subjected to manual inspection of the raw bam files for further QC checks to rule out artifacts. After QC analyses, we will estimate the frequency of rare (MAF <0.1% in gnomAD) pathogenic or likely pathogenic (P/LP) SNVs, INDELs, and structural variants in cancer-associated genes of interest. Frequencies will be compared to internal cancer-free controls using Fisher's exact test. Significance will be adjusted for multiple comparisons using a Benjamini-Hochberg false discovery rate adjustment. Further analyses will be planned with the DCEG statistician after we determine the number of events detected in these cases.

- Outcome(s) of interest: Frequencies of rare P/LP variants (SNVs, INDELs, and structural variants).
- Subject population: all young-onset osteosarcoma cases aged 0-10 years with sufficient DNA (total N=429 cases, including 33 CCSS cases). In-house DCEG cancer-free adult controls profiled with the same whole-genome sequencing method and bioinformatic pipeline (N=~500).
- Exploratory variables: we will evaluate the frequencies of SNVs, INDELs, and structural variants by age, ancestry, and sex. Outcome and clinical variables will not be available for all cases, but we plan to conduct exploratory analyses to determine if there are any suggestions of a relationship between pathogenic variants and osteosarcoma tumor location, subtype, metastases, relapse and vital status (as previously performed[12]). Please see example table below from our published exome study.[12]

**Supplemental Table 8**. Characteristics of the cases with *TP53* pathogenic (P) or likely pathogenic (LP) variants and all P/LP cancer-susceptibility gene variants in 1,004 osteosarcoma cases (discovery set).

| Variable[†] | | N individuals evaluable | No *TP53* P/LP variant | | *TP53* P/LP variant | | P | No P/LP variant | | Any P/LP variant[€] | | P | 2 or more P/LP variants | | P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | N | N % | N | N % | | N | N % | N | N % | | N | N % | |
| Age at dx, mean (SD) | | 974 | 16.5 (10) | | 14.7 (6) | | 0.155 | 16.9 (10) | | 15.3 (7) | | **0.015** | 15.6 (7) | | 0.327 |
| Age group (years) | 0-10 | 151 | 142 | 94.0% | 9 | 6.0% | | 107 | 70.9% | 44 | 29.1% | | 6 | 4.0% | |
| | 11-20 | 698 | 672 | 96.3% | 26 | 3.7% | | 521 | 74.6% | 177 | 25.4% | | 25 | 3.6% | |
| | 21-30 | 67 | 62 | 92.5% | 5 | 7.5% | | 48 | 71.6% | 19 | 28.4% | | 5 | 7.5% | |
| | 31-40 | 26 | 25 | 96.2% | 1 | 3.8% | | 19 | 73.1% | 7 | 26.9% | | 1 | 3.8% | |
| | 41+ | 32 | 32 | 100.0% | 0 | 0.0% | 0.316 | 30 | 93.8% | 2 | 6.3% | 0.107 | 0 | 0.0% | 0.396 |
| Gender | Male | 540 | 519 | 96.1% | 21 | 3.9% | | 397 | 73.5% | 143 | 26.5% | | 25 | 4.6% | |
| | Female | 462 | 440 | 95.2% | 22 | 4.8% | 0.490 | 346 | 74.9% | 116 | 25.1% | 0.620 | 13 | 2.8% | 0.165 |
| Ancestry [¥] | EUR | 732 | 702 | 95.9% | 30 | 4.1% | | 544 | 74.3% | 188 | 25.7% | | 24 | 3.3% | |
| | AFR | 54 | 49 | 90.7% | 5 | **9.3%** | | 40 | 74.1% | 14 | 25.9% | | 3 | 5.6% | |
| | ADM | 73 | 71 | 97.3% | 2 | 2.7% | | 60 | 82.2% | 13 | 17.8% | | 2 | 2.7% | |
| | ASN | 3 | 3 | 100.0% | 0 | 0.0% | | 3 | 100.0% | 0 | 0.0% | | 0 | 0.0% | |
| | His | 142 | 136 | 95.8% | 6 | 4.2% | 0.422 | 97 | 68.3% | 45 | 31.7% | 0.196 | 9 | 6.3% | 0.480 |
| AFR vs. non-AFR ancestry | non-AFR | 950 | 912 | 96.0% | 38 | 4.0% | | 704 | 74.1% | 246 | 25.9% | | 35 | 3.7% | |
| | AFR | 54 | 49 | 90.7% | 5 | **9.3%** | **0.060** | 40 | 74.1% | 14 | 25.9% | 0.990 | 3 | 5.6% | 0.387 |
| Osteosarcoma location | Lower long bones | 746 | 719 | 96.4% | 27 | 3.6% | | 562 | 75.3% | 184 | 24.7% | | 28 | 3.8% | |
| | Lower short bones | 7 | 7 | 100.0% | 0 | 0.0% | | 4 | 57.1% | 3 | 42.9% | | 0 | 0.0% | |
| | Upper long bones | 91 | 86 | 94.5% | 5 | 5.5% | | 66 | 72.5% | 25 | 27.5% | | 2 | 2.2% | |
| | Upper short bones | 2 | 2 | 100.0% | 0 | 0.0% | | 2 | 100.0% | 0 | 0.0% | | 0 | 0.0% | |
| | Face or skull | 2 | 2 | 100.0% | 0 | 0.0% | | 2 | 100.0% | 0 | 0.0% | | 0 | 0.0% | |
| | Mandible | 2 | 2 | 100.0% | 0 | 0.0% | | 1 | 50.0% | 1 | 50.0% | | 0 | 0.0% | |
| | Chest region | 10 | 8 | 80.0% | 2 | 20.0% | | 6 | 60.0% | 4 | 40.0% | | 0 | 0.0% | |
| | Pelvic region | 31 | 27 | 87.1% | 4 | 12.9% | | 21 | 67.7% | 10 | 32.3% | | 4 | 12.9% | |
| | Soft tissue | 5 | 4 | 80.0% | 1 | 20.0% | | 3 | 60.0% | 2 | 40.0% | | 0 | 0.0% | |
| | Vertebral column | 1 | 1 | 100.0% | 0 | 0.0% | **0.086** | 1 | 100.0% | 0 | 0.0% | 0.740 | 0 | 0.0% | 0.346 |
| Axial vs. Extremity location | Extremity | 846 | 814 | 96.2% | 32 | 3.8% | | 634 | 74.9% | 212 | 25.1% | | 30 | 3.5% | |
| | Axial | 51 | 44 | 86.3% | 7 | 13.7% | **0.001** | 34 | 66.7% | 17 | 33.3% | 0.188 | 4 | 7.8% | 0.309 |
| Metastases at dx | No | 382 | 365 | 95.5% | 17 | 4.5% | | 277 | 72.5% | 105 | 27.5% | | 11 | 2.9% | |
| | Yes | 138 | 126 | 91.3% | 12 | **8.7%** | **0.060** | 95 | 68.8% | 43 | 31.2% | 0.413 | 9 | 6.5% | **0.089** |
| Relapse | No | 231 | 218 | 94.4% | 13 | 5.6% | | 170 | 73.6% | 61 | 26.4% | | 10 | 4.3% | |
| | Yes | 136 | 126 | 92.6% | 10 | 7.4% | | 97 | 71.3% | 39 | 28.7% | | 5 | 3.7% | |
| | Progression | 4 | 4 | 100.0% | 0 | 0.0% | 0.703 | 3 | 75.0% | 1 | 25.0% | 0.890 | 0 | 0.0% | 0.817 |
| Percent necrosis at surgery | Poor, <90% | 140 | 133 | 95.0% | 7 | 5.0% | | 99 | 70.7% | 41 | 29.3% | | 5 | 3.6% | |
| | Good, >90% | 134 | 130 | 97.0% | 4 | 3.0% | 0.396 | 92 | 68.7% | 42 | 31.3% | 0.711 | 6 | 4.5% | 0.814 |
| Conventional vs. Surface subtype | Conv. | 342 | 327 | 95.6% | 15 | 4.4% | | 238 | 69.6% | 104 | 30.4% | | 14 | 4.1% | |
| | Surface | 22 | 21 | 95.5% | 1 | 4.5% | 0.970 | 19 | 86.4% | 3 | 13.6% | **0.094** | 0 | 0.0% | 0.491 |

dx, diagnosis; EUR, European ancestry; AFR, African ancestry; ADM, admixed; HIS, Hispanic; ASN, Asian; P, pathogenic; LP, likely pathogenic;

P values for the difference between cases with the specified P/LP variants and cases without these variants using a Chi-Square test;

[†] Not all cases had all variable data, counts (% of total) are given for the cases with these data;

[€] includes cases with all pathogenic/likely pathogenic variants, and both AD and AR inheritance gene variants.

[¥] Ancestry based on GWAS data.

6. **Special consideration:** None

**References**


1. Le Vu B, de Vathaire F, Shamsaldin A, et al. Radiation dose, chemotherapy and risk of osteosarcoma after solid tumours during childhood. *Int J Cancer*. Jul 29 1998;77(3):370-7. doi:10.1002/(sici)1097-0215(19980729)77:3<370::aid-ijc11>3.0.co;2-c
2. Mirabello L, Pfeiffer R, Murphy G, et al. Height at diagnosis and birth-weight as risk factors for osteosarcoma. *Cancer Causes Control*. 2011;22(6):899-908.
3. Arora RS, Kontopantelis E, Alston RD, Eden TO, Geraci M, Birch JM. Relationship between height at diagnosis and bone tumours in young people: a meta-analysis. *Cancer Causes & Control*. 2011/05/01 2011;22(5):681-688. doi:10.1007/s10552-011-9740-9
4. Chen S, Yang L, Pu F, et al. High Birth Weight Increases the Risk for Bone Tumor: A Systematic Review and Meta-Analysis. 2015;12(9):11178-11195.
5. Calvert G, Randall R, Jones K, Cannon-Albright L, Lessnick S, Schiffman J. At-risk populations for osteosarcoma: the syndromes and beyond. *Sarcoma*. 2012;2012:152382.
6. Gianferante DM, Mirabello L, Savage SA. Germline and somatic genetics of osteosarcoma — connecting aetiology, biology and therapy. Review Article. *Nature Reviews Endocrinology*. 03/24/online 2017;13:480. doi:10.1038/nrendo.2017.16
7. Mai PL, Best AF, Peters JA, et al. Risks of first and subsequent cancers among TP53 mutation carriers in the National Cancer Institute Li-Fraumeni syndrome cohort. *Cancer*. Aug 6 2016;doi:10.1002/cncr.30248
8. Wong F, Boice JD, Jr, Abramson DH, et al. Cancer incidence after retinoblastoma: Radiation dose and sarcoma risk. *JAMA*. 1997;278(15):1262-1267. doi:10.1001/jama.1997.03550150066037
9. Savage SA, Mirabello L, Wang Z, et al. Genome-wide association study identifies two susceptibility loci for osteosarcoma. Letter. *Nat Genet*. 2013;45(7):799-803.
10. Savage S, Mirabello L. Using epidemiology and genomics to understand osteosarcoma etiology. *Sarcoma*. 2011;2011:548151.
11. Musselman J, Bergemann T, Ross J, et al. Case-parent analysis of variation in pubertal hormone genes and pediatric osteosarcoma: a Children's Oncology Group (COG) study. *Int J Mol Epidemiol Genet*. 2012;3(4):286-93.
12. Mirabello L, Zhu B, Koster R, et al. Frequency of Pathogenic Germline Variants in Cancer-Susceptibility Genes in Patients With Osteosarcoma. *JAMA oncology*. May 1 2020;6(5):724-734. doi:10.1001/jamaoncol.2020.0197
13. Mirabello L, Yeager M, Mai PL, et al. Germline TP53 variants and susceptibility to osteosarcoma. *J Natl Cancer Inst*. Jul 2015;107(7)doi:10.1093/jnci/djv101
14. Akhavanfard S, Padmanabhan R, Yehia L, Cheng F, Eng C. Comprehensive germline genomic profiles of children, adolescents and young adults with solid tumors. *Nat Commun*. May 5 2020;11(1):2206. doi:10.1038/s41467-020-16067-1
15. Wu J, Sun H, Li J, et al. Increased survival of patients aged 0-29 years with osteosarcoma: A period analysis, 1984-2013. *Cancer Med*. Aug 2018;7(8):3652-3661. doi:10.1002/cam4.1659
16. Mirabello L, Troisi RJ, Savage SA. Osteosarcoma incidence and survival rates from 1973 to 2004: data from the Surveillance, Epidemiology, and End Results Program. *Cancer*. Apr 1 2009;115(7):1531-43. doi:10.1002/cncr.24121
17. Cole S, Gianferante DM, Zhu B, Mirabello L. Osteosarcoma: A Surveillance, Epidemiology, and End Results program-based analysis from 1975 to 2017. *Cancer*. Jun 1 2022;128(11):2107-2118. doi:10.1002/cncr.34163
18. Belkadi A, Bolze A, Itan Y, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proceedings of the National Academy of Sciences*. 2015;112(17):5473-5478. doi:doi:10.1073/pnas.1418631112

19. Ewans LJ, Minoche AE, Schofield D, et al. Whole exome and genome sequencing in mendelian disorders: a diagnostic and health economic analysis. *European Journal of Human Genetics*. 2022/08/15 2022;doi:10.1038/s41431-022-01162-2
20. Mirabello L, Zhu B, Koster R, et al. Frequency of Pathogenic Germline Variants in Cancer-Susceptibility Genes in Patients With Osteosarcoma. *JAMA oncology*. Mar 19 2020;doi:10.1001/jamaoncol.2020.0197
21. Collins RL, Brand H, Karczewski KJ, et al. A structural variation reference for medical and population genetics. *Nature*. 2020/05/01 2020;581(7809):444-451. doi:10.1038/s41586-020-2287-8
22. Geoffroy V, Herenger Y, Kress A, et al. AnnotSV: an integrated tool for structural variations annotation. *Bioinformatics*. 2018;34(20):3572-3574. doi:10.1093/bioinformatics/bty304
23. Riggs ER, Andersen EF, Cherry AM, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med*. Feb 2020;22(2):245-257. doi:10.1038/s41436-019-0686-8