UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
DEPARTAMENTO DE GENÉTICA

ALEXANDRA GALVÃO GOMES

Caracterização de alterações genômicas caóticas em osteossarcoma

Characterization of chaotic genomic rearrangements in osteosarcoma

RIBEIRÃO PRETO – SP
2018

ALEXANDRA GALVÃO GOMES

Caracterização de alterações genômicas caóticas em osteossarcoma

Characterization of chaotic genomic rearrangements in osteosarcoma

Tese apresentada à Universidade de São Paulo, como requisito para obtenção do título de Doutorado em Ciências, pelo curso de Pós-graduação em Genética da Faculdade de Medicina de Ribeirão Preto.
Área de concentração: Genética
Orientador: Prof. Dr. Jeremy A. Squire

Versão corrigida. A versão original encontra-se disponível tanto na Biblioteca da Unidade que aloja o Programa, quanto na Biblioteca Digital de Teses e Dissertações da USP (BDTD).

RIBEIRÃO PRETO – SP
2018

FICHA CATALOGRÁFICA

Gomes, Alexandra Galvão

Caracterização de alterações genômicas caóticas em osteossarcoma. Ribeirão Preto, São Paulo, 2018.

122p.: il.;30cm.

Tese de Doutorado, apresentada à Faculdade de Medicina de Ribeirão Preto/USP – Área de concentração: Genética.

Orientador: Squire, Jeremy A.

1.*Chromothripsis*; 2.Osteossarcoma; 3.Citogenômica; 4. Rearranjos caóticos; 5.Instabilidade cromossômica.

Gomes, Alexandra Galvão

Characterization of chaotic genomic rearrangements in osteosarcoma. Ribeirão Preto, São Paulo, 2018.

122p.: il.;30cm.

PhD thesis presented to the Medical School of Ribeirão Preto, University of São Paulo. Area of concentration: Genetics.

Supervisor: Squire, Jeremy A.

1.*Chromothripsis*; 2.Osteosarcoma; 3.Cytogenomics; 4. Chaotic rearrangements; 5. Chromosomic instability.

FOLHA DE APROVAÇÃO

Nome: Alexandra Galvão Gomes

Título: Caracterização de alterações genômicas caóticas em osteossarcoma

Tese apresentada à Universidade de São Paulo, como requisito para obtenção do título de Doutorado em Ciências, pelo curso de Pós-graduação em Genética da Faculdade de Medicina de Ribeirão Preto.
Área de concentração: Genética
Orientador: Prof. Dr. Jeremy A. Squire

Aprovada em: _____|_____|_____

BANCA EXAMINADORA

Prof. Dr.:_____
Julgamento:_____ Assinatura:_____

Prof. Dr.:_____
Julgamento:_____ Assinatura:_____

Prof. Dr.:_____
Julgamento:_____ Assinatura:_____

Prof. Dr.:_____
Julgamento:_____ Assinatura:_____

Prof. Dr.:_____
Julgamento:_____ Assinatura:_____

DEDICATÓRIA

À minha família.
Aos meus amigos.
À família Bloco C.
À todos que contribuíram para tornar este trabalho possível.
Aos pesquisadores de todo o mundo que disponibilizam seus dados e descobertas “open access” em prol da evolução científica.
Aos Pacientes.

AGRADECIMENTOS

In first place I want to thanks my supervisor, professor Jeremy Squire, for everything! You are very supportive and patient. I just wanna thank you, I am very greatfull for your support! I have no words to express my feelings but thank you! I am very proud to say you are my “orientador”. You were born to be a scientist and you are blessed! So many great ideas, too much knowledge, and also so simple. You are a mirror as human being. Sorry if I was one reason to lose more hair..hahaha

À professora Lucia Martelli, pela grande oportunidade que me deu ao me aceitar para um estágio e, posteriormente, o mestrado na USP e encaminhamento à supervisão do professor Jeremy para o doutorado. Pelo constante apoio, amizade, orientações e sugestões e convívio no grupo de pesquisa.

Ao técnico aposentado Sílvio Santos, também tenho muito a agradecer por todo o conhecimento compartilhado, acadêmico e sobre a vida e a amizade... Como fui abençoada!

Ao Reginaldo Vila, de técnico do laboratório ao lado a um bom amigo! Obrigada pelas caronas no RejUber, pelo revezamento na preparação do café e pelas risadas...porque chegou a minha vez de passar, e vc continuar...rsrsrs mas continuará nas minhas boas lembranças também! Pois como diria você: “Os alunos passam, e a gente (funcionário) fica!”

Aos amigos da Família Bloco C e Genética, que desde agosto de 2011 vem fazendo parte da minha rotina... Carrego todos os que passaram pela minha trajetória com carinho nas minhas lembranças: Ciro, Flavinha, Ju Dourado, Clarissa, Carlos, Ju Josahkian, Vivi, Marco, Anderson, Rafaella, Nathalia, Drous, Cris, Simone, Hélida, Murilinho, Simone, Sarah, Paulinha, Larissa, Karina, Dona Mara, Marli, Grace, Helô, Manu... Aos colegas que atualmente mantém o convívio, que breve se tornará saudades e lembranças: Bianca, Artur, Lívia, Mateus, Ju da Bahia, Jorge, Seu Paulo, Seu Luís, Tati Malta, Felipe (da pitchula)... E aqueles colegas que também são amiguez do coração: Thiago Vidotto, Vick, Perla, Carolzinha e Hugo!

Aos colegas que também são aquele suporte técnico de bioinformática: Muito obrigada Thiago Chedraoui e Marcos Abraão!

Ao Professor Dr. Tiezzi, pela disponibilidade em sempre tentar ajudar, mesmo em missões quase impossíveis, como o BRASS...Obrigada!

I want to thanks Sarah McClelland and her research group at Barts Cancer Institute for their support during my PhD. sandwich period in London. I miss you guys (Nadeem, Naoka, Alice, Lizie and Sarah)! I will never forget the great group you are! Também gostaria

de agradecer aos alunos brasileiros em Londres pelo suporte e ajuda, e pelos amigos que fiz lá: Cadu, Manu, Breno e Talita, Ariane, Clark, Renan, Luciana, Gabi..

Aos citogeneticista do HC-FMRP: Lucimar, Rinaldo e Sarah.

Queria agradecer a todas colegas de república que conviveram comigo durante estes anos, foram muitas vivências e pessoas maravilhosas para conviver...

A todos os meus amigos, sem distinção, os de fora e os de dentro do ambiente acadêmico!

À Alexandra Elbakyan, pelo Sci-hub! Obrigada por ser o "Robbin Hood" da ciência!

Aos grupos de apoio, nas redes sociais, de bolsistas de pós-graduação. Pelo suporte e por tornar os desafios dessa vida mais leves...

Aos membros da banca pelo seu tempo e dedicação. Aprecio muito suas opiniões!

Aos membros do Departamento de Genética, na pessoa da Profa. Ester Ramos, e das secretárias Susie e Silvia: OBRIGADA!

Âs agências de fomento que foram o suporte financeiro para a realização do nosso projeto: Capes, CNPq, FAPESP e FAEPA.

Gostaria de agradecer à minha família 'pustiça' em Ribeirão Preto: a família Oliveira. Obrigada pelo apoio Dona Neide, seu Daniel e Leandro (além do resto da família).

Por fim, gostaria de fazer um agradecimento especial às pessoas mais importantes da minha vida: minha família! Mamãe, Lício e Rafaela, Alex, Alexia e Lucas, além de Renatinha... Vocês são o meu porto seguro...a minha paz, os meus amores! Amo vocês, o título é nosso!

“*Andando devagar, porque já tive pressa..*

Levo esse sorriso, porque já chorei demais..

Hoje me sinto mais forte, mais feliz, quem sabe?

Só levo a certeza de que muito pouco eu sei,

***Ou nada sei*.”**

(Renato Teixeira e Almir Sater)

RESUMO

Gomes, A.G.**. Caracterização de alterações genômicas caóticas em osteossarcoma.** 2018, 122p. Tese (Doutorado), Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo - Brasil.

Metodologias de sequenciamento do genoma total para investigação de diferentes tipos de câncer detectaram recentemente uma nova classe de alterações caóticas de DNA, denominada *Chromothripsis*. Este fenômeno de instabilidade genômica é relativamente comum em tumores de Osteossarcoma (OS), mas existem poucos estudos que expliquem esta conexão ou abordem suas causas e consequências. A presente tese iniciou-se com a re-análise de microarrays de dez amostras de OS pediátrico, previamente processadas pelo nosso laboratório, para avaliar a variação de número de cópias de DNA (CNVs). Usando ferramentas de detecção de padrões característicos de *Chromothripsis* (CTLPs), encontramos 3 amostras de OS com *Chromothripsis* , que afetaram quatro cromossomos (2, 10, 14 e 20). As amostras com presença de *Chromothripsis* tiveram uma media de 468 CNVs/amostra, enquanto o grupo sem o fenômeno teve uma média de 255 CNVs/amostra. Após essa avaliação de CNVs, comparamos os níveis de expressão de RNA entre duas amostras com a presença e quatro tumores com ausência de *Chromothripsis*. Cerca de 171 genes estão presentes em regiões de CNVs diferentes entre os grupos avaliados. Destes, a maioria (77 genes) são relacionados com funções de comunicação celular e ao ciclo celular. Um grupo de 43 genes foi relacionado às vias de processo metabólico (principalmente associado ao metabolismo do RNA) e 27 genes associados à organização do componente celular ou biogênese. Tumores com *Chromothripsis* possuiam 4 genes do sistema imune menos expressos (*CADM1; CLEC4A; CCR1; CD164*) e 12 estavam superexpressos (*IL32, LAT, BCL3, FCAR, RFX1, ILIB, CXCL1, SPON2, CCR6, IL6, SEMA3C, GEM*). Os genes pouco expressos também têm um papel na via de adesão celular. A adesão celular está associada à progressão do câncer e metástase. Em seguida, re-analisamos as CNVs de 82 amostras de OS e 35 linhagens celulares de OS, usando microarrays disponíveis em bancos de dados públicos (GEO e arrayexpress), para identificar potenciais regiões cromossômicas comumente envolvidas em alterações caóticas no número de cópias de DNA, especialmente CTLPs. Identificamos *Chromothripsis* em 27 amostras (11 tumores e 16 linhagens), afetando 17 cromossomos diferentes. Os cromossomos 2, 8 e 12 foram alvos frequentes de *Chromothripsis* em OS. Em seguida, foram analisados dados de sequenciamento WGS de 12 tumores de OS disponíveis no banco de dados online dbGaP. Fizemos a avaliação da variação de número de cópias para caracterizar detalhadamente as alterações caóticas e identificar as

regiões cromossômicas alvo envolvidas nas regiões de alterações caóticas no número de cópias do DNA. Encontramos CTPLs em 7 (58%) das 12 amostras de OS analisadas, usando dados de sequenciamento total. Foram encontrados 12 cromossomos diferentes afetados pelo fenômeno de alteração caótica. CTPLs foram detectadas em 62,5% das amostras de pacientes que faleceram em decorrência deste tumor. Os cromossomos 1, 3 e 7 foram um pouco mais afetados por *Chromothripsis* nas amostras disponibilizadas pelo dbGaP. Além disso, os cromossomos 2 e 12 também foram afetados por *Chromothripsis* nessas amostras. Cerca de 700 genes/tumor foram encontrados nas regiões de CTLPs. Um total de 101 genes foram localizados em regiões de alteração de número de cópias que distinguem os grupos com e sem *Chromothripsis*. Estes genes estão relacionados com vias de processo celular (45 genes - os quais 17 estão associados à comunicação celular) e processo metabólico (22 genes - os quais 19 estão associados ao processo metabólico primário). Nós também comparamos os níveis de expressão gênica das amostras disponíbilizadas pelo dbGap, em que foram avaliados dados de expressão de 6 amostras de RNA de OS com *Chromothripsis* e de 3 amostras de RNA de OS sem *Chromothripsis* . Diferentes algoritmos e ferramentas foram utilizadas para avaliação de RNA. Nós analisamos os dados de expressão por dois diferentes mecanismos: EdgeR e Nexus Expression. Ambos mostraram menor expressão de RNA nas vias de comunicação celular e processo metabólico primário em amostras com *Chromothripsis*. Os genes com regulação negativa da resposta do sistema imunológico foram encontrados em ambas ferramentas (*COL8A1, CCL25*). Para estudar os cromossomos envolvidos na formação de micronúcleos na linhagem celular U2OS, foram investigados erros na divisão celular induzidos por drogas (durante a anáfase). Esta etapa foi realizada durante o período de doutorado sanduíche, no Barts Cancer Institute, em Londres-UK. Os cromossomos com erros durante a anáfase foram contados por meio da técnica de FISH centromérica. Os cromossomos mais comumente encontrados com erros foram Chr2, Chr6, Chr11 e Chr12. Estes dados corroboram a ideia de que alguns cromossomos são mais suscetíveis a erros de divisão celular e colaboram para maiores índices de CTPLs em certos tumores. O fenômeno *Chromothripsis* parece estar presente em pelo menos 30% dos tumores de osteossarcoma e pode estar contribuindo para o fenótipo mais agressivo deste tumor ósseo.

Palavras-chave: *Chromothripsis*; Osteossarcoma; Citogenômica; Rearranjos caóticos; Instabilidade cromossômica.

ABSTRACT

Gomes, A.G.**. Characterization of chaotic genomic rearrangements in osteosarcoma**. 2018, 122p. PhD thesis - Medical School of Ribeirão Preto, University of São Paulo. Ribeirão Preto, São Paulo - Brazil.

Whole genome sequencing methods applied to a number of human cancers have detected a new class of chaotic DNA alterations in tumors called *Chromothripsis*. This mechanism of genomic instability is relatively common in the human bone tumor osteosarcoma (OS), but there are few studies in this tumor addressing either its causes or consequences. In this thesis we initially re-analyzed the DNA copy number data using newer software designed to detect signatures of *Chromothripsis*-like Patterns (CTLPs) using ten OS samples previously studied by our laboratory. We found three of the osteosarcomas had *Chromothripsis* signatures that affected four chromosomes (2, 10, 14 and 20). The osteosarcomas with *Chromothripsis* had a median of 468 copy number abnormalities per tumor compared to 255 for OS tumors without *Chromothripsis*. Next, we compared global RNA expression levels from two OS samples with *Chromothripsis* to four tumors without *Chromothripsis* to determine the types of gene expression differences associated with this process. We found that 171 genes mapped to regions of *Chromothripsis* with the majority (77 genes) mainly having functions related to cellular communication and cell cycle. There were 43 genes that were related to metabolic process (mainly associated with RNA metabolism) and 27 genes with cellular component organization or biogenesis. Also, there were four genes associated with the immune system that were underexpressed (*CADM1; CLEC4A; CCR1; CD164*) and 12 were overexpressed (*IL32, LAT, BCL3, FCAR, RFX1, ILIB, CXCL1, SPON2, CCR6, IL6, SEMA3C, GEM*) in the *Chromothripsis* tumors. Interestingly, all the genes underexpressed also have a role in cell adhesion pathway. Cell adhesion is associated with cancer progression and metastasis. We then reanalyzed DNA copy number data from 82 OS tumors and 35 OS cell lines using microarrays datasets available in public databanks (GEO and arrayexpress), to identify potential chromosomal regions commonly involved in chaotic DNA copy number alterations, especially CTLPs. We found *Chromothripsis* in 27 OS samples (11 tumors and 16 cell lines), affecting 17 different chromosomes. Chromosomes 2, 8 and 12 were frequent targets of *Chromothripsis* in OS. Sequentially, the DNA copy number alterations were analyzed using whole genome sequence data of 12 OS tumors available from dbGaP databank to characterize chaotic alterations in detail and identify the target chromosomal regions involved in *Chromothripsis*. We found *Chromothripsis* patterns in 7 (58%) of the 12 OS samples analyzed using whole genome sequence data. In total there were

12 different chromosomes involved affecting 62.5% of samples from patients that died from OS. Chromosomes 1, 2, 3, 7 and 12 were slightly more often *Chromothripsis* target locations. Nearly 700 genes per tumor were found in the CTLPs regions. A total of 101 genes were located in regions of copy number change that distinguished the group of OS with *Chromothripsis* in comparison to OS without *Chromothripsis*. These genes are related with cellular process (45 genes – which 17 are associated with cell communication) and metabolic process (22 genes – which 19 are associated with primary metabolic process). We were also able to compare the RNA levels from the dbGap samples when expression data was available: comparing 6 OS RNA samples with *Chromothripsis* to 3 OS RNA samples without *Chromothripsis*. Both the EdgeR and Nexus Expression pipelines showed downregulation in cell communication pathway and primary metabolic process in samples with *Chromothripsis*. Genes downregulated of immune system response pathway were found in both pipeline (*COL8A1, CCL25*). To study the chromosomes involved in micronucleus formation in the OS cell line U2OS, errors in cell division induced by drugs during the anaphase were evaluated during the sandwich period at Barts Cancer Institute in London-UK. The lagging chromosomes were counted and the most common chromosomes with errors were Chr2, Chr6, Chr11, and Chr12. These data provide further support to the idea that some chromosomes are more susceptible to cell division errors and corroborate with the chromosomes affected by CTPLs in some tumors.

Keywords: *Chromothripsis*; Osteosarcoma; Cytogenomics; Chaotic rearrangements; Chromosomic instability.

LIST OF ILLUSTRATIONS

Figure 1 - Osteosarcoma affects meanly the long bones, such as the femur or the tibia. The knee is the region more affected by OS. Usually the site of metastases is the lung, and the image shows the tumor cells pathways by the vascularity of the blood system (arrow). Adaptation from Clinic Universidad de Navarra, 2014. Available in: <https://www.cun.es/dam/cun/infograficos/COT/2014_infco_osteosarcomas_copyright.pdf> ……..……... 19

Figure 2 - Genetic alterations in osteosarcoma. Genes with mutations (DNA and/or RNA level) are indicated with red (Toguchida, 2016)…….……... 21

Figure 3 - Osteogenesis and Osteosarcomagenesis. (A) Initiation of osteogenic differentiation from mesenchymal stem cells (MSCs); (B) Defects in osteogenesis lead to osteosarcomagenesis (Lin *et al.*, 2017)……..... 23

Figure 4 - Mesenchymal osteoblastic differentiation process can have implications in the heterogeneous tumor mass resistant to chemotherapeutic treatment and may contribute to metastasis formation (Botter et al., 2014)…….……... 25

Figure 5 - (A) Example of a sequence of progressive rearrangements disrupting a model chromosome (B) Example of how a chromosomal catastrophe might break the chromosome into many pieces that are then stitched back together haphazardly (Stephens *et al.*, 2011) 26

Figure 6 - Chromosomic rearrangements patterns representative of *Chromoplexy*, *Chromothripsis* and Chromoanasynthesis. Modified from (Baca *et al.*, 2013; Weckselblatt and Rudd, 2015)..... 28

Figure 7 - *Chromothripsis* from Ruptured Micronuclei (Hatch and Hetzer, 2015).... 29

Figure 8 - Mechanistic models of *Chromothripsis* initiation (Rode *et al.*, 2015)….. 31

Figure 9 - Parameters for accurate detection of *Chromothripsis* by CTPLScanner…. 43

Figure 10- RNA samples compared……... 44

Figure 11 dbGap analysis: pipeline…….…… 49

Figure 12- U2-OS SKY karyotype (Janssen and Medema, 2012)……. 51

Figure 13- *Chromothripsis* Explorer showing *Chromothripsis* rates in OS human tumors……... 55
Figure 14- Overview of the *Chromothripsis*DB for Osteosarcoma……... 56
Figure 15- Plots of OS samples (by chromosome) characterized with *Chromothripsis* in *Chromothripsis*DB……. 57
Figure 16- *Chromothripsis*DB shows some affected genes by *Chromothripsis* in OS but not in Germline…….. 58
Figure 17- The top 20 mutated COSMIC genes in Bone Osteosarcoma samples by whole genome and target screen…….. 58
Figure 18- The top non mutated COSMIC genes in Bone Osteosarcoma samples by whole genome and target screen. …… 59
Figure 19- Overview of the 10 OS samples showing the high rate of copy number changes between the 10 OS samples (GEO #12830) genome, by Nexus 9.0. ……... 59
Figure 20- Overview of the 10 OS samples showing the high rate of copy number changes between the 10 OS samples (GEO #12830) per chromosome, by Nexus 9.0. ……... 60
Figure 21 Chromosomes affected by *Chromothripsis* in GEO study (#12830): Chr2 (sample OS180); Chr10 (sample OS183); Chr14 and Chr20 (sample OS87B) 61
Figure 22- Chromosome region affected by *Chromothripsis* on zoom in Chromosome 10 (sample OS183), where are located importante genes as *PTEN*. …….. 61
Figure 23 Comparation between groups by the diferent rates of CN distribution… 63
Figure 24 Heatmap of different immune response genes expression between CTLP+ x CTLP- groups. …… 67
Figure 25- Heatmap of different bone related genes expression between CTLP+ x CTLP- groups. …… 69
Figure 26- Overview of the 12 OS samples showing the high rate of copy number changes between the WGS samples (dbGap phs000699), by Nexus 9.0…. 71
Figure 27 - Overview of the 12 OS samples showing the high rate of copy number changes between the WGS samples (dbGap phs000699) by chromosomes (Nexus 9.0)……... 72

Figure 28 – Frames with COSMIC genes present in the *Chromothripsis* regions by sample and chromosome. …… 74
Figure 29 - Comparison between samples CTLP+ and CTLP-, by Nexus 9.0……... 75
Figure 30 Overview of the 12 OS samples showing the high rate of copy number changes between the WGS samples CTLP+ and CTLP-, by Nexus 9.0….. 75
Figure 31- Chromosomes affected on sample SRR1701133 (Chr 1 and Chr 10)……. 79
Figure 32- Chromosomes affected on sample SRR1701235 (Chr 1 and Chr 6)…….... 80
Figure 33 - Chomosome 3 affected on sample SRR1701366……. 82
Figure 34 Chomosomes affected on sample SRR1701388. …… 83
Figure 35- Chomosome 19 affected on sample SRR11701470……. 84
Figure 37 Chomosomes 2 and 9 affected on sample SRR11701617…….. 85
Figure 38- Chomosomes 3,4,5,7, 12 and 16 affected on sample SRR11701727 (indicative of *Chromoplexy*). ……... 87
Figure 39– Volcano Plot of expression data from dbGap samples (RNA-seq) - EdgeR…….. 89
Figure 40- Immune system pathway affected by underexpressed genes in CTLP+ samples: 8 genes from innate immune pathway; 5 genes of adaptive immune system; and 4 genes in cytokine signaling pathway (by Reactome). …….. 91
Figure 41 – Immune system pathway affected by underexpressed genes in CTLP+ samples (Nexus Expression): 8 genes from immune response (by panther). …….. 92
Figure 42 – Heatmap of expression genes related with Immune system, by Nexus…... 92
Figure 43 – Heatmap and Enrichment by GSEA: Positive correlation between the CTLP+ expression and CTLP- in the pathways of RNA pol. III and G2 Checkpoints. …… 93
Figure 44- Number of anaphases found: with and without errors during the mitosis according with the treatment. …….. 94
Figure 45 – Example of FISH technique in slide showing oneU2-OS lagging anaphase. The image shows two lagging chromosomes in red (chr12) and a bridge. Threatment of Nocodozol 8h. …….. 95
Figure 46- Number of anaphases found: with and without errors during the mitosis 95

by Immnunofluorescense. ……...
Figure 47- U2-0S cell untreated by Imunofluorescence. …….. 96
Figure 48 – Number of chromosomes found in error anaphases per slides counted (mean of 50 anaphases counted by probe) – Nocodozol treatment……. 96
Figure 49– Number of chromosomes found in error anaphases with the correction of the genome size function. …… 97
Figure 50 – Picture from a live cell movie. We could see micronuclei and cells with strange format. This video is available in supplementary data folder <https://tinyurl.com/yby5xjsj>……. 97

LIST OF TABLE

Table 1 - Summary results of FISH, centrosome, and TP53 data for seven OS patient samples and control fibroblast (Bayani *et al.*, 2003). …….. 20

Table 2- Genetic alterations in osteosarcoma and hereditary syndromes related (Taran et al., 2017). ……. 22

Table 3 - Overview of different classes of complex chromosomal rearrangements (Pellestor, 2018)…… 27

Table 4 - Overview of study GEO #12830…….. 40

Table 5- Example of the header of one output file from nexus…….. 42

Table 6- Input segmented data to CTPLScanner. …….. 42

Table 7- Summary of arrays samples analyzed. ……. 45

Table 8- Summary of dbGap samples from study phs000699…… 49

Table 9- Summary of publications about complex rearrangements in OS classified as chaotic events…… 54

Table 10- CTLPScanner results showing CTLP+ samples, chromosome regions and the CN status. ……... 60

Table 11- Total CN aberrations by sample of GEO study (#12830), quality of arrays samples and CTLPScanner classification……. 62

Table 12- Comparation of CNVs regions between CTLP+ samples and CTLP- samples, and the genes present in each region……. 64

Table 13- Genes with different expression between CTLP+ x CTLP- comparation (imune response pathway). p-value< 0.01…… 65

Table 14- Genes with different expression between CTLP+ x CTLP- comparation (bone related pathway). p-value< 0.01……. 68

Table 15- The 27 Samples CTLP+, according with the platform, sample type, and chromosome affected…… 69

Table 16 - The 7 Samples CTLP+, according with the CTLPs region size, Copy number status, and chromosome affected…… 72

Table 17 - Summary of the genes from the dbGAP samples – CTLP+……. 73

Table 18- Report of genes found in CTLP regions of WGS samples from dbGap ……. 76

Table 19 - Comparation between CTLP(+) x CTLP(-) by different regions of copy 77

number alterationss showing the event and the genes related……..
Table 20 - CTLPScanner results showing SRR1701133 sample, by chromosome regions and the CN status……. 78
Table 21- List of genes in CTLPs regions of sample SRR1701133 with important Biological process. ……... 79
Table 22 - CTLPScanner results showing SRR1701235 sample, by chromosome regions and the CN status. ……... 80
Table 23 - List of genes in CTLPs of regions with important Biological process from SRR1701235. ……... 80
Table 24 - CTLPScanner results showing SRR1701366 sample, by chromosome regions and the CN status. ……... 82
Table 25 List of genes in CTLPs regions with important Biological process in sample SRR1701366. ……... 82
Table 26 CTLPScanner results showing SRR17011388 sample, by chromosome regions and the CN status. ……... 83
Table 27 List of genes in CTLPs regions with important biological process……. 83
Table 28- CTLPScanner results showing sample SRR1701470, by chromosome regions and the CN status. Table 31- – List of genes in CTLPs regions with important Biological process – sample SRR1701617…….. 83
Table 29- List of genes in CTLPs regions with important Biological process in sample SRR1101470…….…… 84
Table 30 CTLPScanner results showing SRR1701617 sample, by chromosome regions and the CN status. ……... 85
Table 31- List of genes in CTLPs regions with important Biological process – sample SRR1701617. ……... 85
Table 32- CTLPScanner results showing SRR1701727 sample, by chromosome regions and the CN status. ……... 85
Table 33 - List of genes in CTLPs regions with important Biological process in SRR1701727……. 86
Table 34– List of RNA samples analyzed from dbGap……. 87
Table 35- List of samples CTPL+ to compare with samples CTPL-…… 88
Table 36 - Genes downregulated in CTLP+ Samples…… 89
Table 37- Genes upregulated inn CTLP+ Samples……... 90
Table 38 - Genes evaluated in (Davoli *et al.*, 2017) publication that have different 91

expression between the groups

SUMMARY

1 - Introduction 18

1.1- Osteosarcoma ……19

1.2- Chromosomal Instability (CIN)……..……22

1.3- Chaotic genomic rearrangements25

2 - Rationale 32

3 - Hypothesis 35

4- Aims 36

5 - Methodology.....39

6- Results and Discussion 53

7 - Conclusion 99

8 - References.....101

Attachments 109

Introduction

1.0 - INTRODUCTION

1.1 – Osteosarcoma

Osteosarcoma (OS) is the most common type of malignancy in bone tissue, with an incidence of 1-4 cases/million, affecting mainly children and adolescents (75%), with majority being males (ratio 1.5:1) (Kansara *et al.*, 2014; Durfee *et al.*, 2016). Considered a rare tumor, OS is an aggressive malignancy originating from mesenchymal stem cells that produce osteoid or immature bone. OS mainly occurs in the long bones (80-90%), usually affecting the femur (42%) or the tibia (19%). The most common sites of metastases are the lung (>85%) and bone (Wang *et al.*, 2016; Taran *et al.*, 2017). In figure 1 the regions more affected by this type of tumor are shown (Lin *et al.*, 2017).

Figure 1- Osteosarcoma affects meanly the long bones, such as the femur or the tibia. The knee is the region more affected by OS. Usually the site of metastases is the lung, and the image shows the tumor cells pathways by the vascularity of the blood system (arrow). Adaptation from Clinic Universidad de Navarra, 2014. Available in: <https://www.cun.es/dam/cun/infograficos/COT/2014_infco_osteosarcomas_copyright. pdf>.

The treatment for newly diagnosed OS patients includes preoperative chemotherapy, surgical resection, radiotherapy (mainly when the tumor cannot be removed completely) and postoperative chemotherapy. Chemotherapy is successful for many patients, but 30-40% of

patients with localized disease fail to respond to this treatment. These patients have a worse prognosis and usually progress to advanced/metastatic tumors, with the 5-year survival rate less than 20% (Anninga *et al*., 2011; Martin *et al*., 2012; Selvarajah *et al.*, 2014; Mathias *et al*., 2016; Wang *et al*., 2016).

Bayani and colaborators, in 2003, investigated OS chromosomal complexity, using spectral karyotyping (SKY) to analyse 14 OS tumors and 4 OS cell lines. The study identified a multiple rearrangements and a high number of chromosomal breakpoints, with between 30-40 stuctural alterations per primary tumor. Chromosome 20 had the highest number of structural alterations, and chromosomal rearrangements of chromosome 8 were also frequent. Regions 8q23-24 and 17p11-13 had increased copy number by chromosomal comparative genomic hybridization (CGH) or had partial genomic gains. Chromosomes 1 and 6 presented with recurrent structural rearrangements. Chromosomes 1, 6, 13, 14, 17, and 20 had high rate of centromeric rearrangements. The very high frequency of structural and numerical alterations found in OS tumor, including changes even in ploidy, are presented in the table 1. This high level of chromosome complexity in OS likely has a role in the development and progression of this cancer.

Table 1 - Summary results of FISH, centrosome, and TP53 data for seven OS patient samples and control fibroblast (Bayani *et al.*, 2003)

					Centromere signals per cell (%)									
Case	Age/ sex	*TP53*	Centrosome aberrations (%)	Aneu-ploidy range	Signals/ chromo-some	1	2	3	4	5	6	7	CGH[a]	Ploidy established by SKY[b] or cytogenetics
Control		wt	3	2	1	5	94	1	0	0	0	0	46,XX	46⟨2n⟩
					6	2	96	0	2	0	0	0		
					7	1	91.5	6	1.5	0	0	0		
					8	1	95	1.5	2.5	0	0	0		
					17	2	92	4	2	0	0	0		
OS1[b]	17/F	mut	NM	1–4	1	6	79	10	5	0	0	0	+X, +1p31→q24, −2q,	59–75⟨3n+⟩
					6	5	71	9	15	0	0	0	amp8q, +9p, +10q, −11p,	
					8	4	55	14	21	0	0	0	−16p, +18q21→qter,	
OS2[b]	20/F	NM	NM	1–7	6	0	5	17	68	1	5	4	+1p13→p31, +5p13→p14	85–92⟨4n−⟩
					8	1	13	23	55	1	4	3		
					17	0	28	24	48	0	0	0		
OS3b	11/M	mut	NM	1–4	6	51	44	5	0	0	0	0	+1p32→pter, −3p,	30⟨2n−⟩
													+5p14→pter, −6, −10,	
													−12q21→q15, −13,−18,	
					8	1	97	2	0	0	0	0	+19, +20, +22	
					18	40	59	1	0	0	0	0		
OS4[b]	13/M	wt	NM	1–6	6	17	75	7	1	0	0	0	+1p35→pter, −4p,	66–70⟨3n⟩
					8	3	83	10	3	1	0	0	−5q32→qter, +9p	
					17	0	71	22	7	0	0	0		
OS13[b]	20/F	NM	24	1–4	7	2	71.5	17.5	9	0	0	0	NM	46⟨2n⟩
					8	1	87.5	8.5	3	0	0	0		
					17	0	95	3.4	1.6	0	0	0		
OS14[b]	22/F	NM	26	1–7	7	0	69.8	11.9	8.9	6.4	1.5	1.5	NM	46⟨2n⟩
					8	3.9	74.7	6.9	12	2	0	0		
					17	4	94	1	1	0	0	0		
OS19	17/F	NM	4	1–6	1	0	98	0	1	0	0	0	NM	No analyzable metaphases
					8	0	96	2	2	3	16	0		
					17	5	90	4	1	0	0	0		

Abbreviations: CGH, comparative genomic hybridization; mut, mutation; NM, no material available; SKY, spectral karyotyping; wt, wild-type.
[a] Ploidy as established by previous cytogenetic work by CGH [20].
[b] Ploidy as established by previous cytogenetic work by SKY [4].

OS tumors are more complex than other sarcomas, however publications about the genetic cause of OS are still restricted given the rare incidence of the tumor (Durfee *et al.*, 2016; Yang *et al.*, 2018).

In addition to being a highly aggressive tumour, OS is characterized by having an unusually high level of genomic alteration and chromosomal instability. OS also presents with frequent cytogenetic rearrangements but without apparent recurrent translocations or fusion genes. Tetraploidy is often also present with non-specific chromosome gains and losses (see section 1.2) leading to higher levels of cytogenetic heterogeneity. Tumors usually have complex chromosome aberrations with high incidence of numerical DNA copy number gains (regions 1p, 6p, 8q, 12q and 17p are commonly reported) and losses (regions 2q, 3q, 6q, 10, 13q and 17p are commonly reported) (Martin *et al.*, 2012; Rosenberg *et al*., 2013). More than 20 genetic alterations were related with the molecular mechanisms of growth and progression in OS as showed in the figure 2 (Rosenberg *et al.*, 2013; Toguchida, 2016).

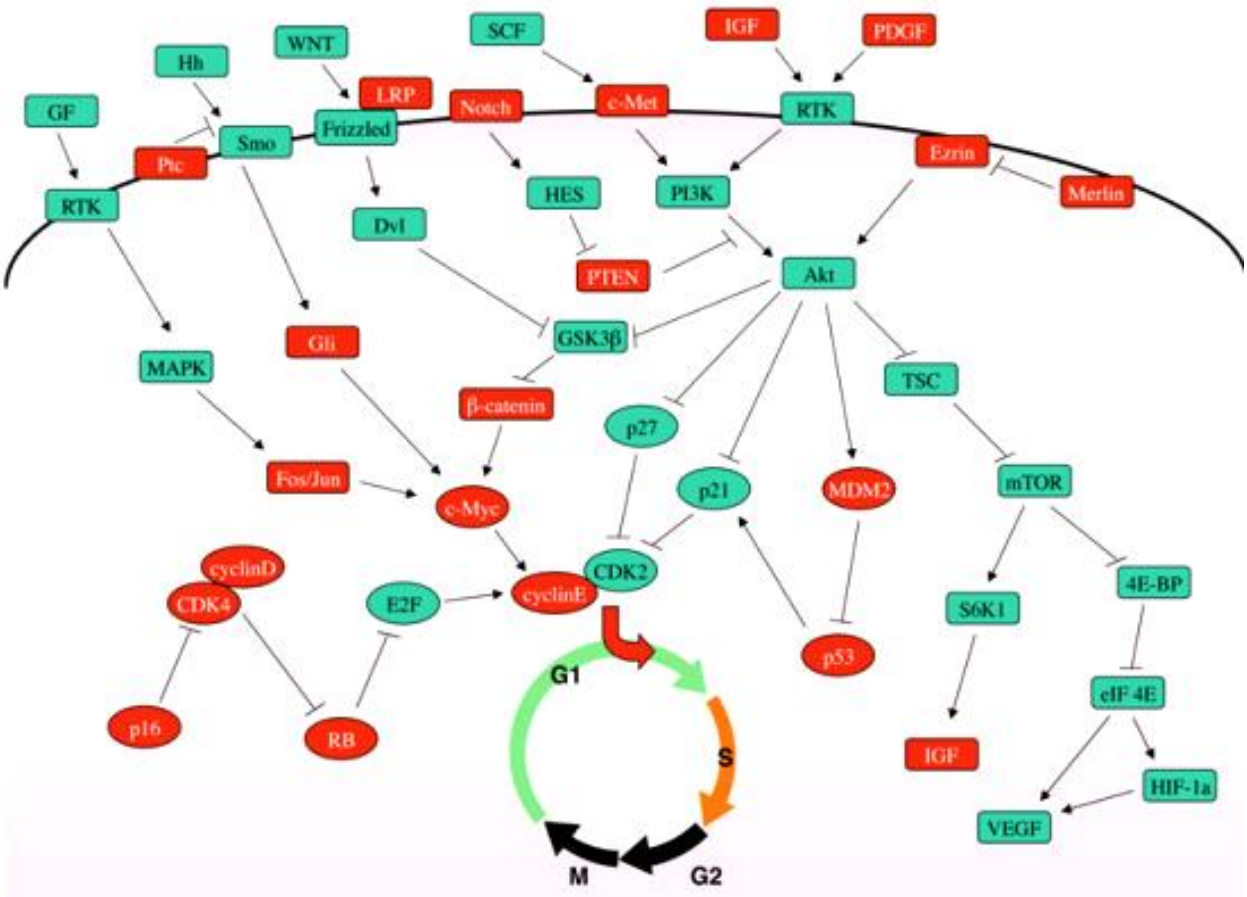

Figure 2 - Genetic alterations in osteosarcoma. Genes with mutations (DNA and/or RNA level) are indicated with red (Toguchida, 2016).

1.2 – Chromosomal instability (CIN)

Chromosomal instability (CIN) describes the excessive rate of numerical and structural genomic change in tumors (Bayani *et al.*, 2007). At the cellular level this genetic diversity provides the selective advantage that leads to the acquisition of genomic changes such as gene amplification and chromosomal gains that can be associated with the aggressive and drug-resistant behavior in tumors like OS (Birkbak *et al.*, 2011; Kovtun *et al.*, 2015).

CIN is thought to result from a combination of stress (i.e., replicative or oxidative stress) and mutations in cancer pathways associated with genome stability. At the cytogenetic level OS is characterized by having many complex structurally abnormal chromosomes as well as gene amplification, dicentric chromosomes, multiple marker chromosomes, double minutes (dmin), homogeneously staining regions (hsr), and/or ploidy changes and anaphase bridges that can lead to micronuclei, as seen in other human cancers with a high rate of CIN(Al-Romaih *et al.*, 2003; Donley and Thayer, 2013).

Some hereditary genetic syndromes increases the risk of developing OS, such as hereditary retinoblastoma, Rothmund–Thomson syndrome, Li-Fraumeni syndrome, and Werner syndrome. Genes associated with these syndromes (*RB1*, *RECQL4*, *TP53*, and *WRN*) are reported in the table 2, and possibly might influence in the pathogenesis of OS (Moriarity *et al.*, 2015; Taran *et al.*, 2017). Moreover, other genes were reported related with OS, as *RUNX2* (6p), *MYC* (8q), and *PTEN* (10q) (Rosenberg *et al.*, 2013).

Table 2 - Genetic alterations in osteosarcoma and hereditary syndromes related (Taran et al., 2017).

Gene	Percentage affected	Tumor suppressors	References
Tumor suppressors			
p53	20-50 (or more)	Li-Fraumeni	McIntyre *et al.*, Lonardo *et al.*, Gokgoz *et al.*, Hauben *et al.*
Rb	Up to 70	Retinoblastoma	Eng *et al.*
p16INK4A/p14ARF	~10%	Dysplastic nevus syndrome	Lopez-Guerrero *et al.*, Shimizu *et al.*
Oncogenes			
MDM2	6-14	SNP309 of MDM2 have accelerated tumor formation	Bond *et al.*
AP-1 (c-jun/c-fos)	40-60 for both c-fos and c-jun	None known	David *et al.*
Notch	Unknown	No	Engin *et al.*

Previous array CGH (comparative genomic hybridization microarray) and spectral karyotyping studies have demonstrated that OS has one of the highest rates of CIN with copy number gains and structural changes affecting more than 50% of the genome (Al-Romaih *et al.*, 2003; Bayani *et al.*, 2003; Selvarajah *et al.*, 2008; Sadikovic *et al.*, 2009).

The high rate of CIN in OS has been attributed in part to the role of *MYC*, *RB1* and *TP53* in the maintenance of genomic stability (Martin *et al.*, 2012). The presence of

abnormally complex chromosomes in OS was considered to be predominantly associated with dicentric chromosomes and the bridge-breakage-fusion cycle (Selvarajah *et al.*, 2006), until large-scale sequencing was applied to this tumour (Stephens *et al.*, 2011) as described in section 1.3.

The occurrence of OS arises in the second decade of life, which is thought to be related to the period of fast bone growth experienced by adolescents (Al-Romaih *et al.*, 2003). The development of the bones occurs by osteogenic differentiation, which is a process closely regulated by different genetic pathways, transcriptional regulators and cell-cycle controllers. Gene expression differs constantly through the various stages of differentiation. Some genes can be analysed as markers: *COL1A* and *ALP* for osteoblastic progenitors and pre-osteoblasts; *PTH1R* and *BGLAP* for mature osteoblasts, and *FGF23* and *MEPE* for osteocytes. These cell types exist in regions of active bone cell progenitor proliferation called bone growth plates (Cheng, 2018). Some genomic alterations may affect the normal developmental process in these regions, causing incomplete differentiation in bone progenitors. It is thought that these genetic alterations may lead to inbalance between proliferation and differentiation of bone progenitors, and can cause uncontrolled proliferation within the developing bone growth plates. Osteosarcoma precursors cells possibly will arise from these cells and multiply to form osteosarcoma (Lin *et al.*, 2017). The figure 3 shows one scheme of the osteogenesis and osteosarcomagenesis processes.

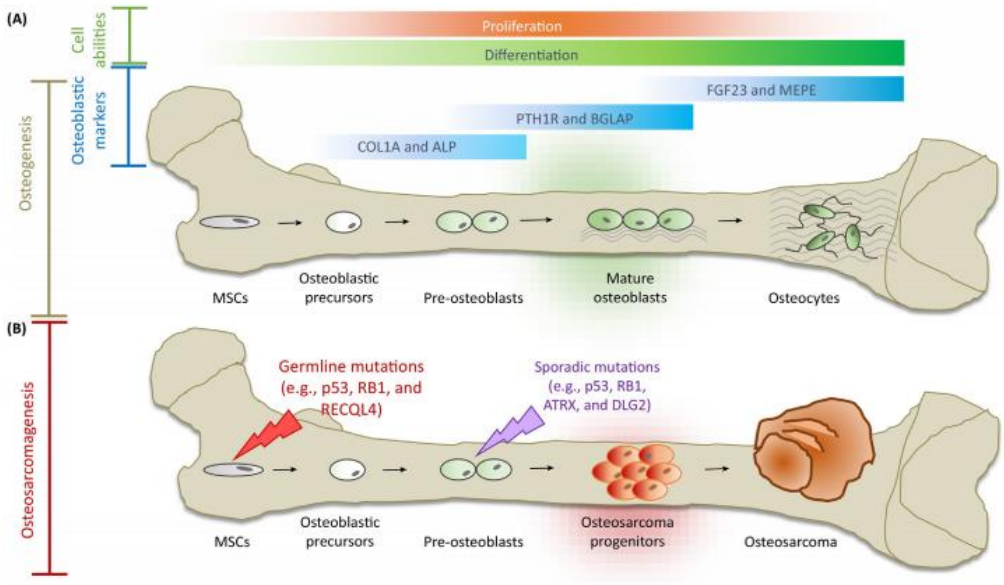

Figure 3 - Osteogenesis and Osteosarcomagenesis. (A) Initiation of osteogenic differentiation from mesenchymal stem cells (MSCs); (B) Defects in osteogenesis lead to osteosarcomagenesis (Lin *et al.*, 2017).

More than 60% of OS tumors have copy number loss of *PTEN*. Some authors strongly suggest that loss of both genes *PTEN* and *TP53* is a cooperative event driving osteosarcomagenesis (Moriarity *et al.*, 2015).

There is also an association between CIN and molecular defects in mitotic checkpoints. High-grade OS, for example, has a highly abnormal mitotic rate that can be attributed to dysregulation of the controlling mechanisms on chromosomal segregation. This same characteristic has been seen in other tumors (Al-Romaih *et al*., 2003)

The resistance of the tumors to chemotherapy can be associated to some of the pathways controlling cellular responses to drugs: innate resistance, acquired resistance and adaptive resistance. Innate resistance permits the progression of the tumor cells even during chemotherapy which is connected to the continued unchecked errors and failure stop proliferation (non-responsive tumors). Acquired resistance is associated with the resistant cell clones present in the primary tumor that survive after chemotherapy (tumors usually regress in the beginning and relapse in the future) and subsequently repopulate locally or spread to another location. Adaptive resistance occurs when a different mechanisms (e.g. acquired genetic alterations) permits the development of novel traits associated with chemotherapy resistance. Genome alterations (as mutations and rearrangements) can be promoted by the drugs used for treatment and may facilitate the generation of cells with different phenotypes such as drug-resistance, and the ability for tumor regrowth. These different pathways to tumor resistance can show altered stages of heterogeneity or CIN according with the predominant pathway in the course of the tumour regrowth. Figure 4 shows a scheme with the mutations that may take place during each phase of the mesenchymal osteoblastic differentiation process and can have implications in the heterogeneous tumor mass resistant to chemotherapeutic treatment and may contribute to metastasis formation (Botter *et al*., 2014; Mcclelland, 2017).

The types of unusual genetic change that characterizes OS provides new therapeutic opportunities for cancer control. Acquisition of aneuploidy by tumors may induce cell death instead of promoting tumorigenesis. The involvement of immune system and its surveillance can recognize and eliminate aneuploid cells in tumors. This natural mechanism of tumor resistance must be evaluated. Another pertinent example, is the tetraploidy that is typical of OS. These cells may express specific cell surface antigens that could be recognized and be used as targets for cancer treatment. The link between CIN and immune function has been suggestive and the role in carcinogenesis requires further study. Major nonspecific pathways that may be considered CIN sign ºatures are involved in oxidative stress response and immune functions. Until recently tumors with CIN were interpreted as a source of recurrence

and tumor progression, and the involvement of the immune system in recurrence was not a consideration. The transcriptomic reprogramming in some tumors, such as lung cancer, can affect multiple pathways and may reduce the immune surveillance, with adaptive immunity and NK-cell-mediated cytotoxicity decreased (Yamada *et al.*, 2016). These pathways have not yet been investigated in OS therapeutics.

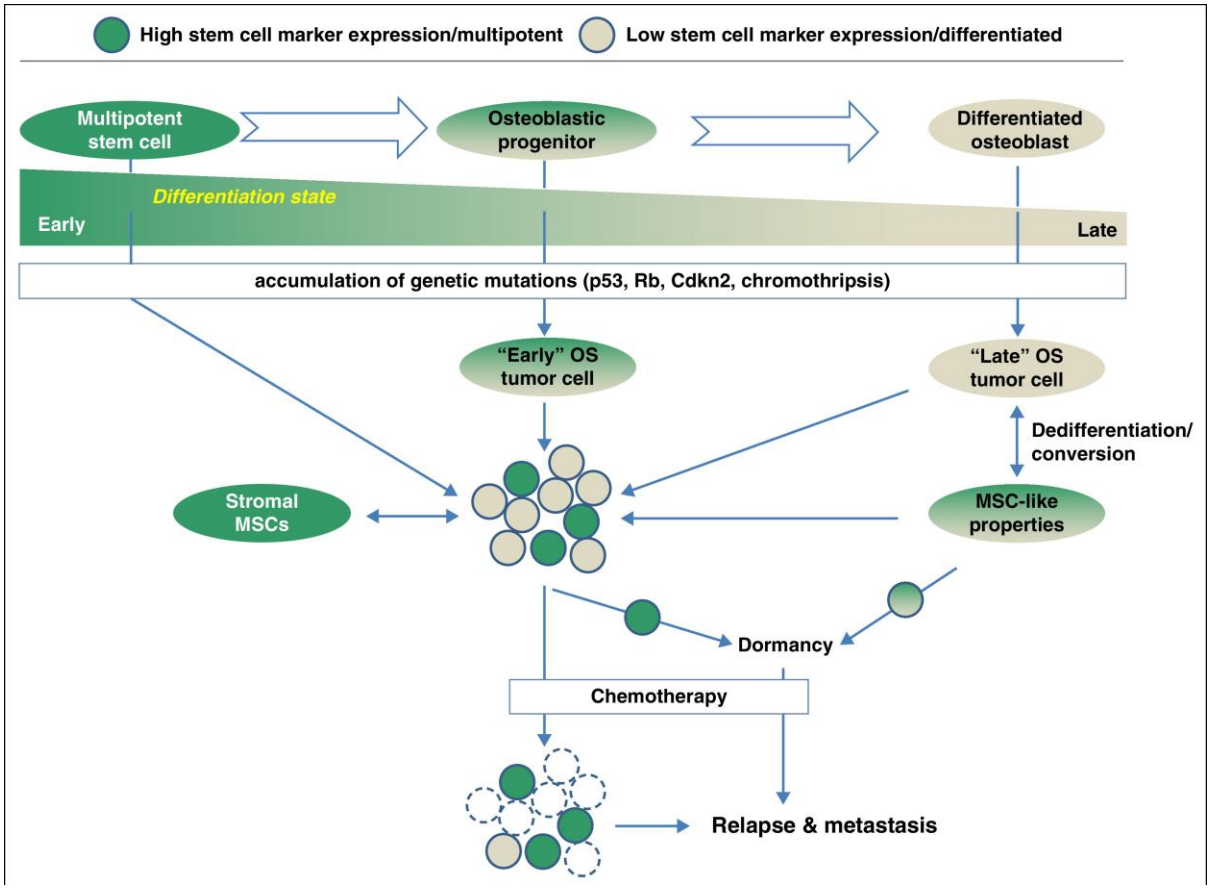

Figure 4 -Mesenchymal osteoblastic differentiation process can have implications in the heterogeneous tumor mass resistant to chemotherapeutic treatment and may contribute to metastasis formation (Botter et al., 2014).

1.3 – Chaotic genomic rearrangements

Whole genome sequencing has provided the first comprehensive picture of all somatic mutations in cancer genomes, identifying patterns of mutations and genomic alterations that provide insights concerning the mechanism of mutational diversity in human cancers (Willis *et al.*, 2015). These methods detected a previously unrecognized class of catastrophic genomic rearrangement called *Chromothripsis*. The genomic breakpoints associated with *Chromothripsis* occur in tens to hundreds and are usually restricted to discrete regions on one or two chromosomes. *Chromothripsis* seems to occur as a single event to one cell in contrast to the bridge-breakage-fusion cycle in which is a progressive mechanism inducing multiple

genomic changes to a chromosome over many cell generations, we can observe in figure 5 an scheme of how occurs both processes (Stephens *et al*., 2011; Forment *et al*., 2012).

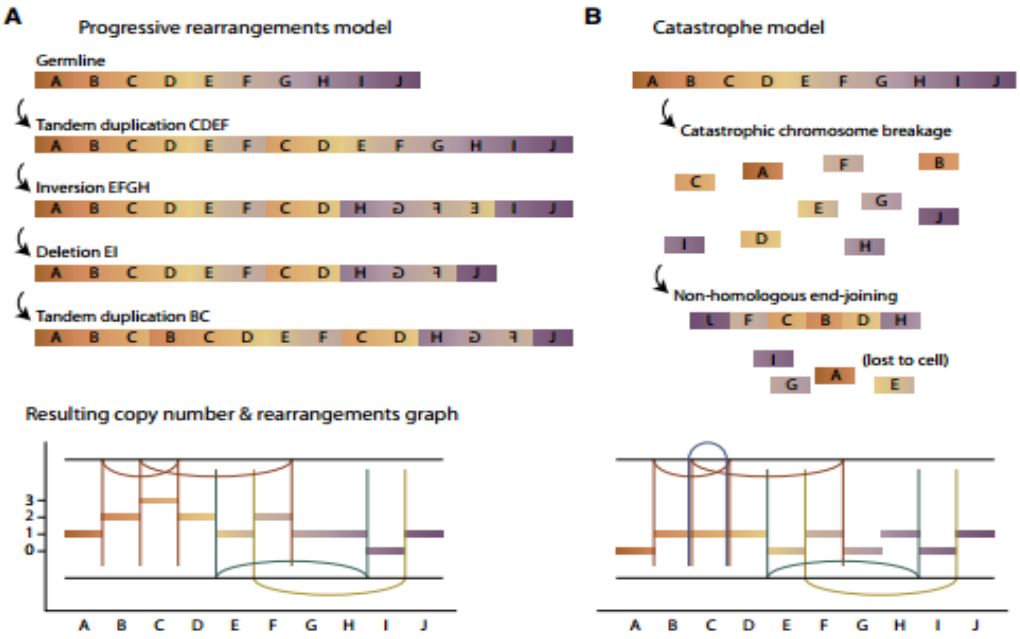

Figure 5- (A) Example of a sequence of progressive rearrangements disrupting a model chromosome (B) Example of how a chromosomal catastrophe might break the chromosome into many pieces that are then stitched back together haphazardly (Stephens *et al.*, 2011).

These type of rearrangements are better studied by genomic high resolution techniques such as whole genome sequencing, mate-pair sequencing, and DNA microarray analysis, to evaluate the complexity of the genome reorganizations in detail (Pellestor, 2018).

Analysis of microarray data of single copy nucleotide polymorphisms (SNP array CGH) and next-generation paired-end sequencing across a range of tumor cell types, has suggested that *Chromothripsis* occurs in 2-3% of primary tumors. However, the frequency of this phenomenon may be greater than 33% in OS (Stephens *et al.*, 2011).

Korbel & Campbel (Korbel and Campbell, 2013), published the criteria to define a genome region with complex rearrangement as *Chromothripsis*. The region must have this characteristics:

(1) multiple and complex rearrangements primarily alter a single chromosome, chromosomal arm, or region and, in some instances, concurrent rearrangements between chromosomes;

(2) many regions show copy number changes alternating between two states, one copy (heterozygous deletion) or two copy (no loss or gain);

(3) regions of single copy are not necessarily from simple deletions but are the byproduct of complex rearrangements spanning the region;

(4) pronounced clustering of breakpoints;

(5) the fragments residing in the clustered breakpoint regions do not reside in close proximity in the germline;

(6) breakpoints involving multiple chromosomes also show clustering.

More recently other classes of chaotic complex rearrangements called *Chromoplexy* and Chromoanasynthesis have been detected by whole genome sequencing methods (Zhang *et al.*, 2013). The term *Chromoplexy* was first used in 2013, to describe a new type of intrinsic genomic rearrangement that occurs in an interleaved fashion and in conjunction of several chromosomal regions (Baca *et al.*, 2013). This phenomenon may be responsible for many of genomic alterations known to be present in tumors, leading to the generation of fusion genes and disruption or deletion of genes next breakpoints regions. These types of rearrangements have not been considered to have an independent occurrence but to take place in a coordinated and simultaneous way (Shen, 2013).

Liu *et. al.* (2011) argues that the phenomenon termed *Chromothripsis* also might be better referred to as ‘‘Chromoanasynthesis’’ (chromosome reconstitution or chromosome reassortment). However, Chromoanasynthesis differs from *Chromothripsis* and *Chromoplexy*, because it may be part of a continuum of segmental amplification mechanisms, the tandem segmental duplication serving as the simplest element (Willis *et al*., 2015). The table 3 shows a table published by Pellestor *et al.* (2018), with the shared and distinguishing features of *Chromothripsis*, chromoanasynthesis, and *Chromoplexy*. Figure 5 shows the chromosomic rearrangements patterns of each of these three fenomena.

Table 3 – Overview of different classes of complex chromosomal rearrangements (Pellestor, 2018).

	Chromothripsis	Chromoanasynthesis	Chromoplexy
Number of events	Single	Single or multiple	Multiple
Structural variation	Balanced or deletions	Balanced, deletions, duplications/triplications	Balanced with occasional deletions
Involved chromosomes	Few (1–4)	Few (often 1)	Multiple (>5)
Number of breakpoints	Many (more in cancer than germline rearrangements)	Fewer (usually 5–25)	Fewer (usually 5–25)
Breakpoint signature	Blunt ends (small insertions)	Microhomology	Blunt ends
Proposed mechanism	NHEJ	MMBIR/FoSTes	NHEJ

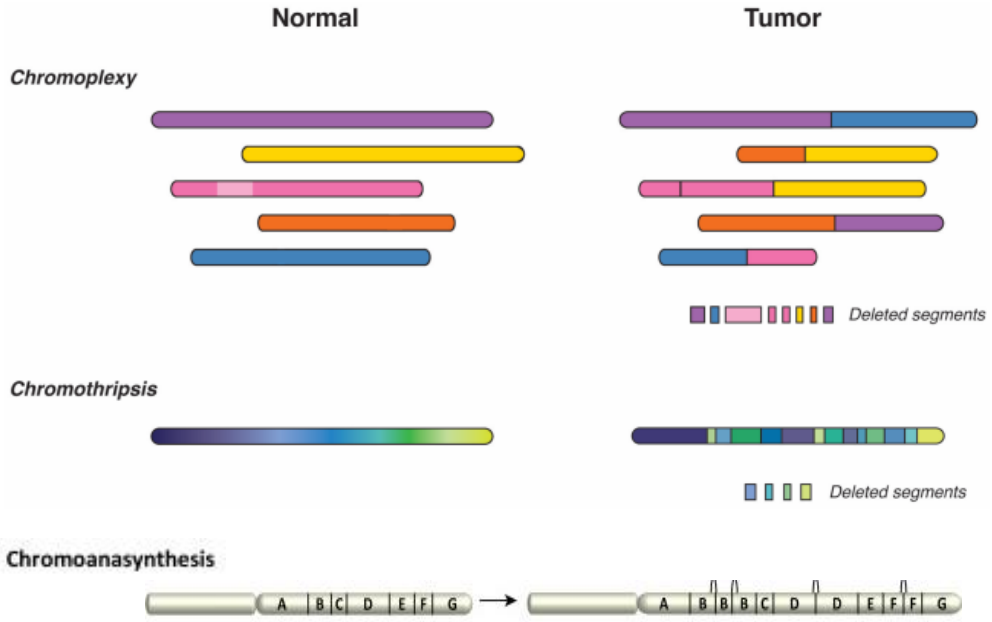

Figure 6 - Chromosomic rearrangements patterns representative of *Chromoplexy*, *Chromothripsis* and Chromoanasynthesis. Modified from (Baca *et al.*, 2013; Weckselblatt and Rudd, 2015)

However, the distinction between *Chromothripsis* and *Chromoplexy* is not well defined, and it is likely that some coordinated structural rearrangements may have intermediate properties. Interestingly OS is also characterized by high levels of localized hypermutations called "kataegis" which typically occur in up to 50% of OS tumors as clusters close to regions of genomic rearrangement (Chen *et al.*, 2014). The mechanism responsible for kataegis remains unknown, but it has been proposed that localized replication-timing changes may occur near chromosomal breakpoints and these could be responsible for the focal mutagenesis observed in kataegis (Donley and Thayer 2015). *MYC* is often overexpressed or amplified in OS and it is well established that increased expression of this oncogene can lead to DNA replication stress (Dominguez-Sola and Gautier 2014).

At the molecular level *Chromothripsis* events might be the result of chromosome shattering followed by end-joining of the double strand breaks via non-homologous end-joining or alternative end-joining. However the gains and duplications observed in *Chromothripsis* suggest replication may also be involved in the mechanism (Gelot *et al.*, 2015).

Microhomology-mediated break-induced replication creates a stress that could lead to replication stalling, and could produce complex rearrangements by template switching when close to regions of microhomology (Forment *et al.*, 2012; Donley and Thayer, 2013; Dominguez-Sola and Gautier, 2014). Such replication forks could undergo several rounds of

template switching, generating the type of complex clustered rearrangements observed in *Chromothripsis* (Gelot *et al.*, 2015). These events may be more likely to occur in OS with amplified or increased copies of *MYC*. *PTEN* inactivation may be responsible for DNA damage-induced multinucleation and chemo-resistance features (Mukherjee *et al.*, 2013).

At the cellular level *Chromothripsis* may initiate as a result of segregation errors in mitosis and stress during replication (Holland and Cleveland, 2012). There is increasing evidence in support of the micronucleus mechanism for the origin of *Chromothripsis*. When a chromosome mis-segregates during mitosis, it may product a daughter cell with two nuclei, the primary nucleus and the micronucleus (having a mis-segregated chromosome or part of a chromosome). After that, the cell enters in S phase and the DNA replication can happen on the micronucleated chromatin. The disturbance of the nuclear envelope during replication causes DNA damage, as well as double-stranded DNA breaks. The damaged chromatin is re-enclosed in a nuclear envelope after mitosis, and the DNA damage repair pathways can identify the shattered chromatin and arbitrarily reassemble the fragments to form a new chromosome. Unassembled pieces can be lost from the chromosome or may become circularized and persist in the genome. Since only one copy of the micronucleated chromatin is present at mitosis areas subject to *Chromothripsis* will only affect one homologue (Hatch and Hetzer, 2015). This process is showed in figure 7.

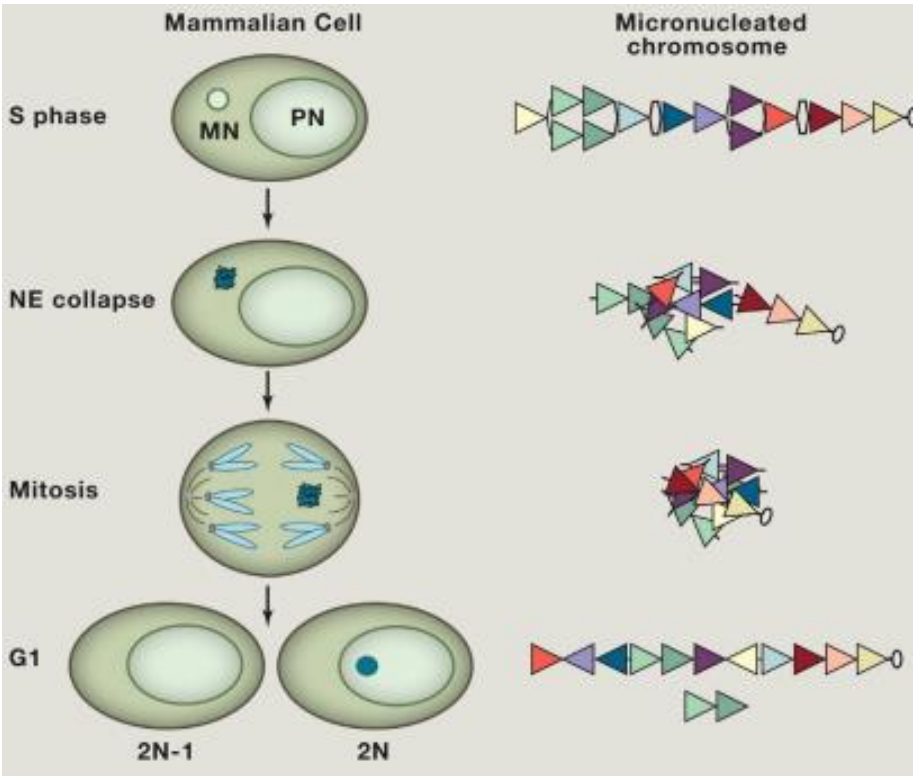

Figure 7 - *Chromothripsis* from Ruptured Micronuclei (Hatch and Hetzer, 2015).

Micronucleus formation is a peculiar feature of genomic instability, and loss of p53 appears as a result of increased *Chromothripsis* frequency (Hatch *et al.*, 2013). Interestingly some chromosomes appear to be more prone to mis-segregation (Worrall *et al.*, 2018), and thus are more likely to become micronuclei. Thus if re-incorporation of chromosomal DNA from micronuclei back into the tumour genome is an important *Chromothripsis* mechanism, then those chromosomes that are more prone to mis-segregation errors will be more commonly subject to this type of chaotic rearrangement (Zhang *et al.*, 2015).

The mechanism leading to chaotic rearrangements is presently unknown, however numerous models have been suggested that have some conceptual overlap. The figure 8 shows five models in an squematic view. In the first model the generated micronucleus (figure 8.a) provides a source of locally delimited damaged DNA. The chromosomal DNA contained within micronuclei suffers aberrant DNA replication and can undergo extensive DNA fragmentation. Then, the returning of the DNA segments into the genome leads to derivative chromosomes that contain DNA derived from the micronucleus reincorporated into the tumor genome in one distinct location. There is no consensus regarding the timing of the reincorporation into the primary nucleus. The premature chromosome condensation (PCC) hypothesis (figure 8.b) proposes the idea that the asynchronous cell-cycle progression between primary nucleus and micronucleus might induce an early condensation of replicating DNA and lead to instability of the condensed DNA fragments. The breakage-fusion-bridge (BFB) cycles and telomere dysfunction hypothesis (figure 8.c) suggests that the vulnerable chromosome ends could join to form an unstable derivative chromosome, which gets shattered in the successive cell cycles and promotes multiple rounds of local rearrangements until the derivative chromosome stabilizes. The ionizing radiation (IR) model (figure 8.d) suggest that external causes such as IR can to produce multiple DNA DSBs, which could be repaired erroneously and initiate *Chromothripsis*. Finally, the aborted apoptosis idea (figure 8.e) proposes that the beginning of apoptosis leads to DNA shattering and, in exceptional situations, apoptosis can be initiated and then aborted, and the cell escapes complete DNA fragmentation and just undergoes partial fragmentation and repair (Rode *et al.*, 2015).

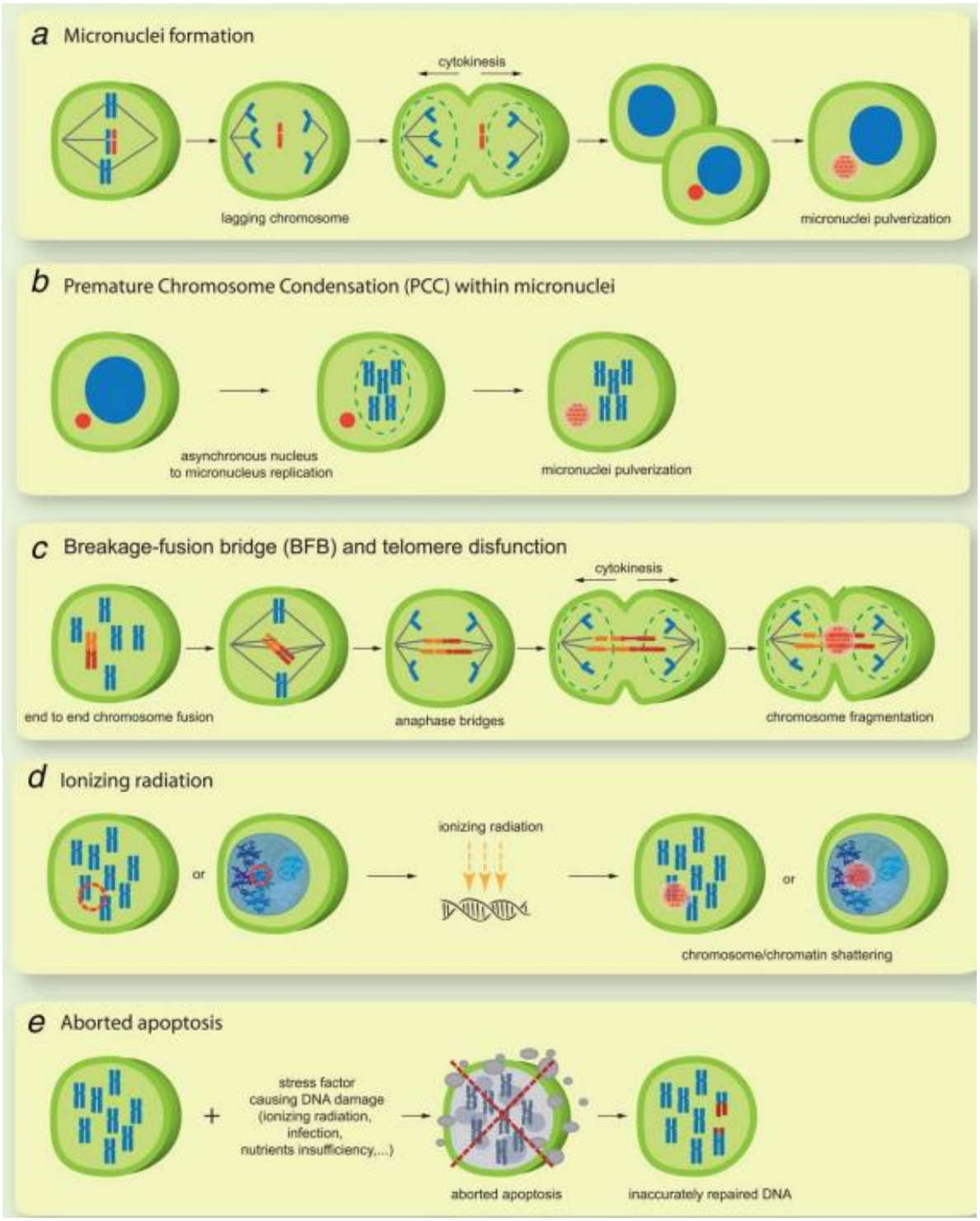

Figure 8. Mechanistic models of *Chromothripsis* initiation (Rode *et al.*, 2015).

Rationale

2.0 - RATIONALE

There are a limited number of studies in the scientific literature addressing the role of *Chromothripsis* in osteosarcoma. There is little information on the possible mechanisms that allow its occurrence or explaining why OS tumors have very high rate of these type of rearrangements. It seems possible that OS with *Chromothripsis* have an atypical repair process that contributes to the occurrence of these phenomena. This thesis addresses the correlation between the mechanisms involved in these complex chaotic chromosomal rearrangements and the regions of genomic instability in OS. Thus, OS is a unique model tumour to study somatic chaotic alterations in human cancer to advance our understanding of how these chromosomal lesions are initiated and can contribute to tumorigenesis.

Hypothesis

3.0 - HYPOTHESIS

Osteosarcoma tumors have multiple defects affecting chromosomal segregation, the DNA repair system (pairing process of homologous and non-homologous), and genome stability, which allows a chaotic chromosomal reorganization forming complex rearrangements in specific regions of the genome. Clues concerning the mechanism of *Chromothripsis* can be obtained by studying genomic locations subject to chaotic alterations in OS. The selective advantage of these types of genomic alterations in OS can be better understood by studying changes in gene expression caused by *Chromothripsis*.

Aims

4.0 – AIMS

4.1 – General Aim

We intend to identify and characterize chaotic chromosomal rearrangements in tumour samples and cell lines of osteosarcoma, using genomic public databases and micronucleus studies. We will characterize the class of chaotic chromosomal rearrangements present in OS by its signature based on copy number changes and we will to evaluate the difference of the RNA expression between the genes of the samples with/without the chaotic rearrangements.

4.2 – Specific aims

Summarize scientific publications that recognized chaotic chromosomal rearrangements in samples of Osteosarcoma by array CGH copy number analysis;

Reanalyze DNA copy number data by array CGH techniques of OS samples previously processed by our laboratory research group, focusing in to identify potential chromosomal regions commonly involved in chaotic DNA copy number alterations;

Reanalyze DNA copy number data by microarrays datasets available in public databanks (GEO and arrayexpress), focusing in to identify potential chromosomal regions commonly involved in chaotic DNA copy number alterations;

Reanalyze DNA copy number variation of whole genome sequence data of OS samples available from dbGaP databank to characterize chaotic alterations in OS and identify chromosomal regions involved in chaotic genomic alterations and compare them with published data;

To study the chromosomes involved in micronucleus formation in the OS cell line U2OS;

To determine whether specific chromosomal regions of the OS subject to chaotic genomic rearrangements contain tumour suppressor genes, oncogenes or other genes causally associated with OS oncogenesis;

Reanalyze Expression Array data of OS samples available in public databank, from samples previously analyzed by copy number variation, focusing in to identify genes with different RNA expression in samples characterized with chaotic DNA copy number alterations by comparation with samples without this characteristic, and check the pathways involved.

Methodology

5.0 – METHODOLOGY

5.1- Ethics Comittee

The local USP Ethics Committee was consulted and as this project involves the use of public databases ethical approval is not required. This dispensation was approval by the HCRP Ethics Comittee in October 1, 2016. The document is on Attachments section as Attachment A.

5.2 - Publications Summary

Our literature searches up to May 30, 2018, yielded 12 publications, which contained the keywords '*Chromothripsis*', 'chaotic genomic rearrangements' and 'osteosarcoma' in the title, abstract or full text, using digital library search engines as Pubmed, Scielo, Google, and Google Scholar. These studies presented analyses characterizing samples with complex rearrangements as chaotic events.

5.3 – Tools online

Some tools online were used to analyse the background of *Chromothripsis* in Osteosarcoma. We used the *Chromothripsis* Explorer (available at <http://compbio.med.harvard.edu/*Chromothripsis*/>), *ChromothripsisDB* (available at <http://cgma.scu.edu.cn/*Chromothripsis*DB/>), and COSMICv85 (available at <https://cancer.sanger.ac.uk/cosmic>).

Chromothripsis Explorer is the result of the partnership between PCAWG project, The Cancer Genome Atlas, and The International Cancer Genome Consortium. When someone search for one tumor type, the software online shows the *Chromothripsis* rates for that specific cancer (according the publication of (Cortes-Ciriano *et al.*, 2018)).

*Chromothripsis*DB is the first repository providing convenient public access to *Chromothripsis* data. It curated and integrated hundreds of *Chromothripsis* samples from the published literature into the database per type of tumor.

COSMIC, the Catalogue of Somatic Mutations In Cancer, is the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.

We search in these tools data choosing array or WGS of human OS samples (all types of Bone osteosacoma).

5.4 – GEO #12830 array reanalysis

We reanalyzed the DNA copy number data by array CGH technique of 10 OS human pediatric tumors previously processed by our laboratory research group (Sadikovic *et al.*, 2009), already available in GEO public functional genomics data repository (available at <https://www.ncbi.nlm.nih.gov/geo/>). These raw data is in the study GEO #12830. The reevaluating of these data was performed focusing on the identification of potential chromosomal regions commonly involved in chaotic DNA copy number alterations.

Table 4 – Overview of study GEO #12830.

#GEO Study	Sample	Sample ID	Platform Build 35	Sample Type
GSE12830	GSM322064	OS87B	Agilent FE	OS pediatric tumor
	GSM322072	OS138	Agilent FE	OS pediatric tumor
	GSM322074	OS177	Agilent FE	OS pediatric tumor
	GSM322076	OS178	Agilent FE	OS pediatric tumor
	GSM322078	OS179	Agilent FE	OS pediatric tumor
	GSM322086	OS180	Agilent FE	OS pediatric tumor
	GSM322088	OS182	Agilent FE	OS pediatric tumor
	GSM322090	OS183	Agilent FE	OS pediatric tumor
	GSM322092	OS2336	Agilent FE	OS pediatric tumor
	GSM322094	OS2960	Agilent FE	OS pediatric tumor

Nexus copy number software version 9.0 (obtained from BioDiscovery, Inc.) was used to process the aCGH platform Agilent 244k txt files (build 35), stringent, and as a mosaic sample.

Nexus analisys followed this status:

#Step = Systematic Correction

Type = Linear Correction

#File = /Applications/BioDiscovery/Nexus 9.0/Organisms/Human/NCBI Build 35/SystematicCorrection/ Agilent/Catalog_Agilent_244k_20101116.txt

#Step = Recenter Probes

Type = Median

#Step = Combine Replicates Between Arrays
Type = None
#Step = Analysis
Type = FASST2 Segmentation
Significance Threshold = 1.0E-7
Max Contiguous Probe Spacing (Kbp) = 1000
Min number of probes per segment = 3
High Gain = 1.14
Gain = 0.2
Loss = -0.23
Big Loss = -1.1
#Organism = Human
#Build = NCBI Build 35

According to the Nexus 9.0 manual (Biodiscovery, 2017), the copy number alterations were called using Fast Adaptive States Segmentation Technique (FASST2) algorithm together with quadratic correction implemented in Nexus. Nexus software uses in aCGH evaluation the FASST2 algorithm which were developed to address the needs of increased density of array technology in the adaptation of new high-throughput sequencing technology. Although a number of algorithms have been proposed based on the well know Hidden Markov Model (HMM) approach which have linear time requirements, these methods often rely on rather restrictive assumptions that are not satisfied in common types of real world samples (e.g. cancer data which often contains significant mosaicism and normal cell contamination). On the other hand recursive segmentation methods, such as Circular Binary Segmentation (CBS) and Rank Segmentation, do not require such restrictive assumption and have performed well in comparison studies, but have at least quadratic time performance. The FASST2 approach achieves a balance between these previous methods by using an HMM model not to estimate the copy number or allelic event states but rather a large number of possible segment levels that might fall between the expected states. Subsequent processing is performed to combine these basic segments into copy number and allelic event calls.

The output files showed the CNVs results and the specific regions, events and probe mean (log 2), as the example below in table 5.

Table 5 – Example of the header of one output file from nexus.

Chromosome Region	**Event**	**Length**	**Cytoband**	**% of CNV Overlap**	**Probe Median**	**Probes**
chr1:1,533,651-19,010,391	CN Gain	17476741	p36.33 - p36.13	68.8483779	0.555805176	1660

5.4.1 - GEO #12830 array by CTLPScanner

CTLPScanner is a web server we used for the detection of *Chromothripsis*-like patterns (CTLP) in genomic data (available at <http://cgma.scu.edu.cn/CTLPScanner/>). The output interface presents intuitive graphical representations of detected chromosome pulverization region, as well as detailed results in table format. CTLPScanner also provides additional information for associated genes in *Chromothripsis* region to help identify the potential candidates involved in tumorigenesis (Yang *et al.*, 2015). There is a script in R language to download from website.

The uploaded data file should be plain text format with tab separator. The system supports file types include: .txt, .csv, .tab, .zip. The minimum required data fields for CTLPScanner:

1. sample: The name of the data file;
2. chro: The chromosome identifier;
3. start: The starting position of the segment;
4. stop: The ending position of the segment;
5. mean: The normalized fluorescence intensity of the segment (log2 transformed).

The output files from Nexus 9.0 were modified using Excel software as the input model below (table 6) to be used in CTLPScanner. Just copy number alterations were considering, all results about allelic imbalance was ignored.

Table 6- Input segmented data to CTPLScanner.

sample	chro	start	stop	mean	probes
GSM681798	1	51599	346406	0.2925	30
GSM681798	1	394035	593454	-1.2171	6
GSM681798	1	615334	12784829	-0.5975	6738

CTLPScanner offers a set of parameters for accurate detection of *Chromothripsis*. The web server provides optimized default values for all parameters, which may also be adjusted for customized screening (figure 9).

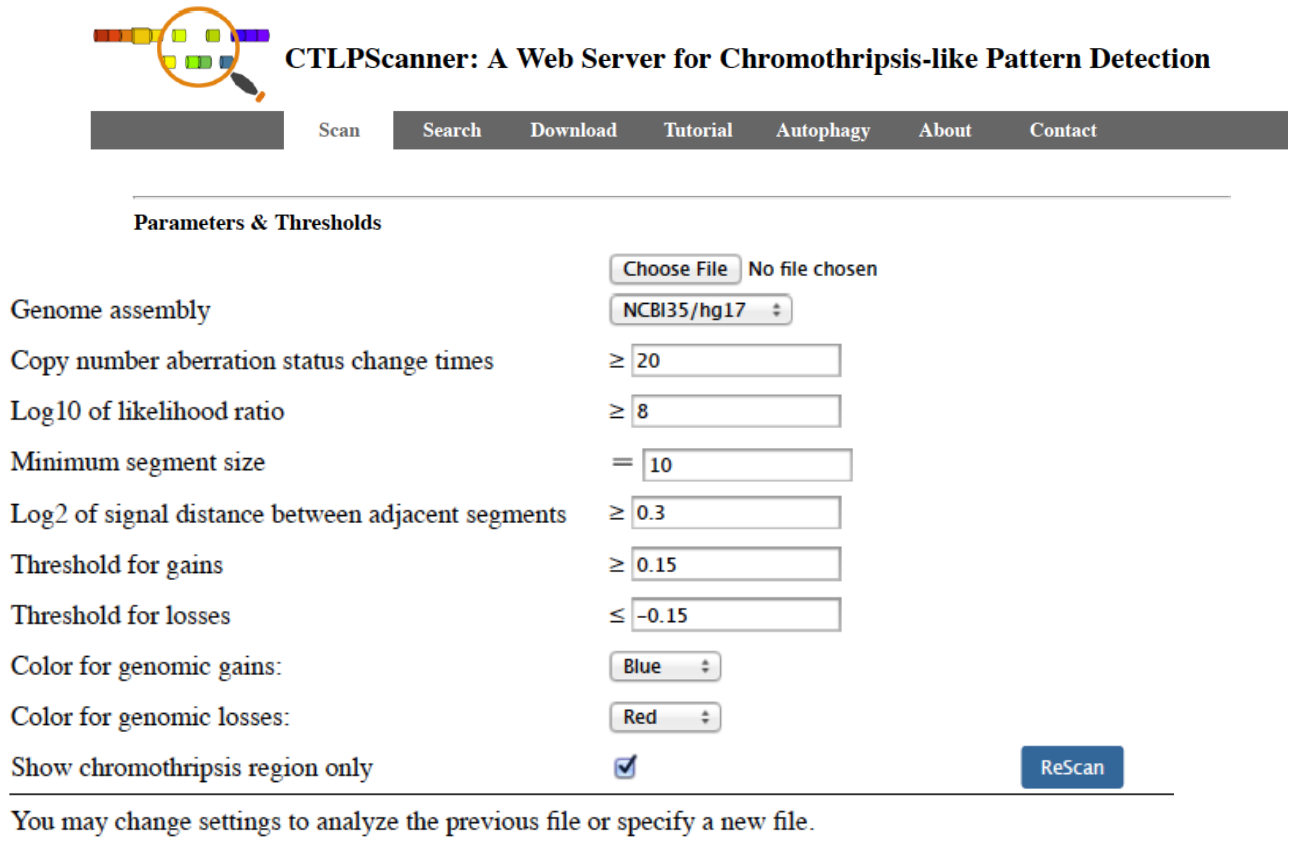

Figure 9 - Parameters for accurate detection of *Chromothripsis* by CTPLScanner.

To detect *Chromothripsis*-like patterns (CTLPs) the algorithm described by (Korbel and Campbell, 2013), was applied to identify clustering of copy number changes in the genome.

The samples were evaluated by CTLPScanner and then separated in 2 groups: CTLP+ (yes – with *Chromothripsis*) and CLTP- (no – without *Chromothripsis*). After that, the groups were compared using Nexus, and the genes in differents chromosome regions with CNVs were evaluated between the groups focusing in their biological process and if they are listed in COSMIC (Catalogue of Somatic Mutations in Cancer)(Institute, 2018). We also evaluate each CTLP+ sample alterations.

5.4.2 - GEO #12830 Expression Data

The GEO #12865 study presents the Expression HuGene Array of 6 OS samples from the study GEO#12830. And we used the software Nexus Expression 3.0 to compare 2 RNA samples CTLP+ with 4 RNA samples CTLP-, According with figure 10.

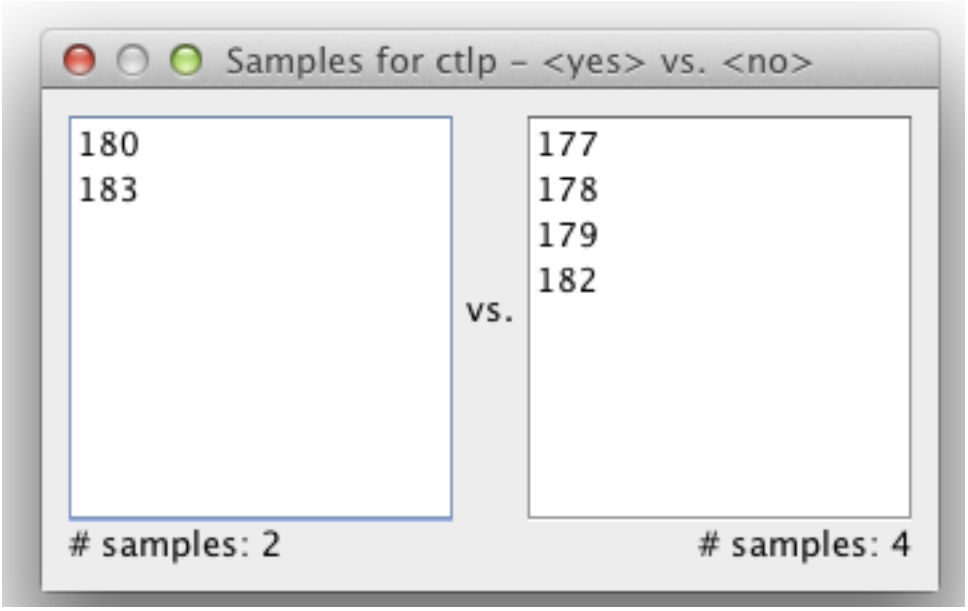

Figure 10: RNA samples compared.

We made comparative analysis using some pathways of importance, as immune response pathway and bone formation.

5.5 – Other arrays

Other 4 set of arrays also was evaluated, according with the same conditions of the GEO #GSE12830, adjusted by the platform type and build. They are: GSE67125 (Affymetrix CytoScan HD Array); GSE3153 and GSE36003 (Affymetrix Genome-Wide Human SNP 6.0 Array); and GSE12789 (Agilent-014698 Human Genome CGH Microarray 105A). The samples are in table 7.

Table 7 – Summary of arrays samples analyzed (Continue).

Sample	Build	Platform	#Study databank	Sample Type	Gender	Age (years)	Source
GSM1639703	37	Affymetrix	GSE67125	OS cell line	female	11	SAOS (HTB-85) early passage
GSM1639704	37	Affymetrix	GSE67125	OS cell line	female	11	SAOS (HTB-85) late passage
GSM1639705	37	Affymetrix	GSE67125	OS cell line	female	11	LM5 (SAOS derived) - early passage
GSM1639706	37	Affymetrix	GSE67125	OS cell line	female	11	LM5 (SAOS derived) - late passage
GSM820994	37	Affymetrix	GSE33153	OS tumor	male	18	Tibia (left proximal)
GSM820995	37	Affymetrix	GSE33153	OS tumor	male	14	Femur (left distal)
GSM820996	37	Affymetrix	GSE33153	OS tumor	male	18	Femur (left distal)
GSM820997	37	Affymetrix	GSE33153	OS tumor	male	16	Femur (right distal)
GSM820998	37	Affymetrix	GSE33153	OS tumor	female	13	Tibia (right proximal)
GSM820999	37	Affymetrix	GSE33153	OS tumor	male	14	Tibia (right distal)
GSM821000	37	Affymetrix	GSE33153	OS tumor	male	8	Femur (left diaphyseal)
GSM821001	37	Affymetrix	GSE33153	OS tumor	male	11	Tibia (left proximal)
GSM821002	37	Affymetrix	GSE33153	OS tumor	male	16	Tibia (left proximal)
GSM821003	37	Affymetrix	GSE33153	OS tumor	male	25	Tibia (left proximal)
GSM821004	37	Affymetrix	GSE33153	OS tumor	female	20	Femur (right distal)
GSM821005	37	Affymetrix	GSE33153	OS tumor	male	12	Tibia (right proximal)
GSM821006	37	Affymetrix	GSE33153	OS tumor	female	15	Tibia (left proximal)
GSM821007	37	Affymetrix	GSE33153	OS tumor	male	15	Tibia (right proximal)
GSM821008	37	Affymetrix	GSE33153	OS tumor	male	16	Femur (right proximal)
GSM821009	37	Affymetrix	GSE33153	OS tumor	male	17	Femur (left distal)
GSM821010	37	Affymetrix	GSE33153	OS tumor	male	15	Femur (distal)
GSM821011	37	Affymetrix	GSE33153	OS tumor	male	18	Tibia (right proximal)
GSM821012	37	Affymetrix	GSE33153	OS tumor	male	32	Humerus (right)
GSM821013	37	Affymetrix	GSE33153	OS tumor	male	14	Femur (left distal)
GSM821014	37	Affymetrix	GSE33153	OS tumor	male	14	Femur (left distal)
GSM821015	37	Affymetrix	GSE33153	OS tumor	male	14	Femur (left distal)
GSM821016	37	Affymetrix	GSE33153	OS tumor	male	41	Ilium
GSM821017	37	Affymetrix	GSE33153	OS tumor	male	18	Femur (distal)
GSM821018	37	Affymetrix	GSE33153	OS tumor	male	15	Femur (left distal)
GSM821019	37	Affymetrix	GSE33153	OS tumor	female	10	Femur (left distal)
GSM821020	37	Affymetrix	GSE33153	OS tumor	male	14	Femur (left)
GSM821021	37	Affymetrix	GSE33153	OS tumor	male	23	Femur (distal)
GSM821022	37	Affymetrix	GSE33153	OS tumor	male	13	Fibula (right proximal)
GSM821023	37	Affymetrix	GSE33153	OS tumor	NA	17	Humerus (left)
GSM821024	37	Affymetrix	GSE33153	OS tumor	NA	11	Femur (left distal)
GSM821025	37	Affymetrix	GSE33153	OS tumor	NA	NA	NA
GSM879206	37	Affymetrix	GSE36003	OS cell line	female	13	143B OS cell line
GSM879207	37	Affymetrix	GSE36003	OS cell line	male	15	HAL cell line

Table 7 – Summary of arrays samples analyzed (Continue).

GSM879208	37	Affymetrix	GSE36003	OS cell line	female	13	HOS cell line
GSM879209	37	Affymetrix	GSE36003	OS cell line	male	15	IOR/OS9 cell line
GSM879210	37	Affymetrix	GSE36003	OS cell line	female	10	IOR/OS10 cell line
GSM879211	37	Affymetrix	GSE36003	OS cell line	male	13	IOR/OS14 cell line
GSM879212	37	Affymetrix	GSE36003	OS cell line	female	12	IOR/OS15 cell line
GSM879213	37	Affymetrix	GSE36003	OS cell line	male	33	IOR/OS18 cell line
GSM879214	37	Affymetrix	GSE36003	OS cell line	female	13	IOR/MOS cell line
GSM879215	37	Affymetrix	GSE36003	OS cell line	male	25	IOR/SARG cell line
GSM879216	37	Affymetrix	GSE36003	OS cell line	male	7	KPD cell line
GSM879217	37	Affymetrix	GSE36003	OS cell line	male	14	MG-63 cell line
GSM879218	37	Affymetrix	GSE36003	OS cell line	female	41	MHM cell line
GSM879219	37	Affymetrix	GSE36003	OS cell line	female	13	MNNG/HOS cell line
GSM879220	37	Affymetrix	GSE36003	OS cell line	male	14	OHS cell line
GSM879221	37	Affymetrix	GSE36003	OS cell line	male	19	OSA cell line
GSM879222	37	Affymetrix	GSE36003	OS cell line	female	11	SAOS-2 cell line
GSM879223	37	Affymetrix	GSE36003	OS cell line	female	15	U2OS cell line
GSM879224	37	Affymetrix	GSE36003	OS cell line	male	21	ZK-58 cell line
GSM320781	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320782	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320783	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320784	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320785	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320786	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320787	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320788	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320789	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320790	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320791	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320792	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320793	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320794	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320795	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320796	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320797	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320798	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320799	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320800	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320801	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320802	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320803	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320804	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA

Table 7 – Summary of arrays samples analyzed (End).

GSM320805	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320806	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320807	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320808	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320809	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320810	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320811	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320812	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320813	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320814	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320815	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320816	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320817	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320818	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320819	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320820	35	Agilent FE	GSE12789	OS tumor	NA	NA	NA
GSM320821	35	Agilent FE	GSE12789	OS cell line	NA	NA	NA
GSM320822	35	Agilent FE	GSE12789	OS cell line	NA	NA	NA
GSM320823	35	Agilent FE	GSE12789	OS cell line	NA	NA	NA
GSM320824	35	Agilent FE	GSE12789	OS cell line	NA	NA	NA
GSM320825	35	Agilent FE	GSE12789	OS cell line	NA	NA	NA
GSM320826	35	Agilent FE	GSE12789	OS cell line	NA	NA	NA
GSM320827	35	Agilent FE	GSE12789	OS cell line	NA	NA	NA
GSM320828	35	Agilent FE	GSE12789	OS cell line	NA	NA	NA
GSM320829	35	Agilent FE	GSE12789	OS cell line	NA	NA	NA
GSM320830	35	Agilent FE	GSE12789	OS cell line	NA	NA	NA
GSM320831	35	Agilent FE	GSE12789	OS cell line	NA	NA	NA
GSM320832	35	Agilent FE	GSE12789	OS cell line	NA	NA	NA
GSM322064	35	Agilent FE	GSE12830	OS tumor	NA	NA	NA
GSM322072	35	Agilent FE	GSE12830	OS tumor	NA	NA	NA
GSM322074	35	Agilent FE	GSE12830	OS tumor	NA	NA	NA
GSM322076	35	Agilent FE	GSE12830	OS tumor	NA	NA	NA
GSM322078	35	Agilent FE	GSE12830	OS tumor	NA	NA	NA
GSM322086	35	Agilent FE	GSE12830	OS tumor	NA	NA	NA
GSM322088	35	Agilent FE	GSE12830	OS tumor	NA	NA	NA
GSM322090	35	Agilent FE	GSE12830	OS tumor	NA	NA	NA
GSM322092	35	Agilent FE	GSE12830	OS tumor	NA	NA	NA
GSM322094	35	Agilent FE	GSE12830	OS tumor	NA	NA	NA

*NA= Non-available data

We collected 153 samples from 11 studies, but the arrays set without the presence of CTLPs in at least one sample were excluded. Probably the array platforms of these datasets may have not well resolution to detect chaotic rearrangements.

Nexus analisys followed this status:

-Agilent platforms:

Max Contiguous Probe Spacing (Kbp) = 1000
Min number of probes per segment = 3
High Gain = 1.14
Gain = 0.2
Loss = -0.23
Big Loss = -1.1

-Affymetrix platforms:

Max Contiguous Probe Spacing (Kbp) = 1000
Min number of probes per segment = 3
High Gain = 0.7
Gain = 0.1
Loss = -0.15
Big Loss = -1.1

Together, the 5 arrays set analysed in this thesis (including #GSE12830) have 117 OS samples: 82 tumors and 35 cell lines. The analysis on Nexus 9.0 was realized according with each microarray platform.

5.6 – DbGap WGS samples

We first submited a project to have access to the project phs000699 at dbGap databank, available at <https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id= phs000699.v1.p1>.

We used the SRA toolkit 2.8.2, Samtools 1.7 and Aspera Connect 3.7.4 softwares as indicated in NCBI manual (Ncbi, 2011). The pipeline of dbGap data analysis is in figure 11. The platform used was the Illumina Hiseq 2000, paired-end, matched, hg 19.

We had access to 13 WGS samples, and 35 RNA-seq samples of OS. However one WGS sample had to be excluded because it showed an error file. WGS samples analyzed are in table 8.

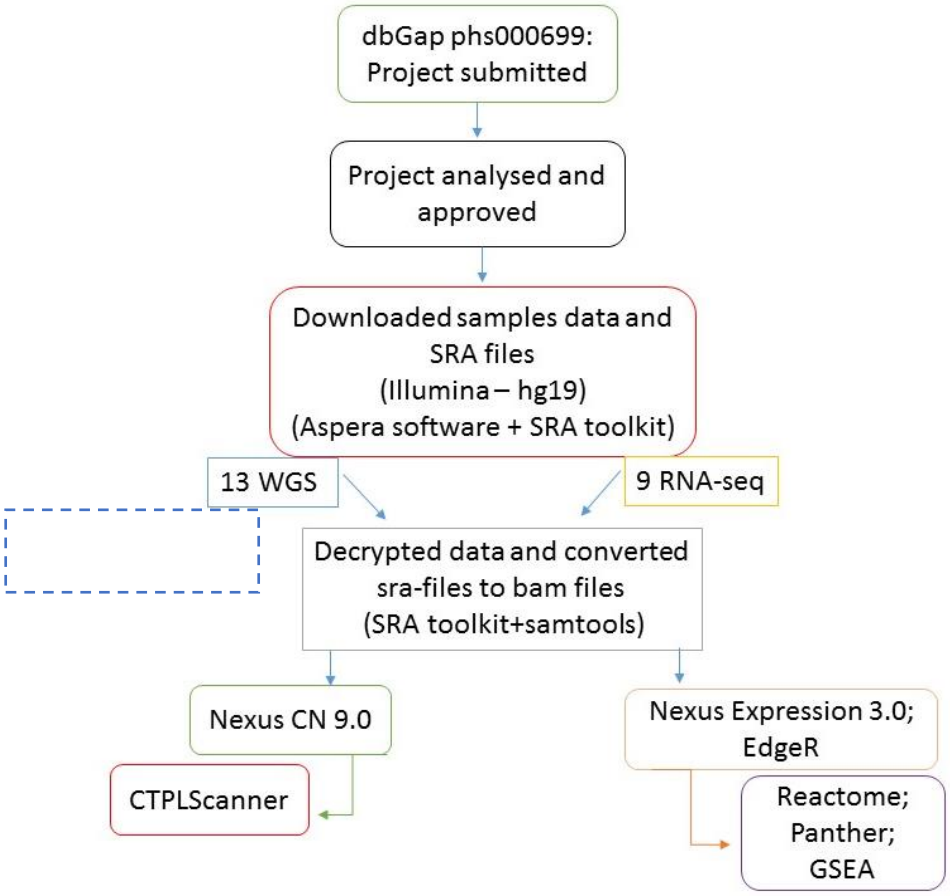

Figure 11- dbGap analysis: pipeline.

Table 8 – Summary of dbGap samples from study phs000699

Sample	Experiment	Gender	Last Known Outcome	Age at diagnosis	Months of survival
SRR1701133	BZ06-Tumor	Male	Deceased	19	5
SRR1701169	BZ10-Tumor	Male	Living	10	3
SRR1701188	BZ32-Tumor	Female	Deceased	12	7
SRR1701235	BZ17-Tumor	Female	Living	12	43
SRR1701245	BZ04-Tumor	Female	Living	8	44
SRR1701366	BZ15-Tumor	Female	Deceased	14	13
SRR1701388	BZ18-Tumor	Male	Deceased	15	33
SRR1701470	BZ30-Tumor	Male	Living	19	64
SRR1701546	BZ23-Tumor	Female	Deceased	13	12
SRR1701617	BZ11-Tumor	Male	Deceased	22	11
SRR1701703	MX02-Tumor	Male	Deceased	17	9
SRR1701727	BZ36-Tumor	Male	Deceased	10	18

The WGS samples were analysed by Nexus 9.0 to the evaluation of copy number patterns, using the pipeline of Biodiscovery (sample matched with control). The experiment was the DNA from tumor, and the control was the DNA sample from blood of the same

patient. BAM files were processed directly via comparisons to identify statistically significant differences between subgroups, plotting of BAF (B-allele frequency) for the BAM ngCGH (matched) algorithm.

We used just the Copy Number output file from Nexus 9.0 with adaptions to imput the data in CTLPScanner web server, and analyse the patterns of *Chromothripsis* searching for the copy number alterations. This process was similar with the arrays data. We did not have success using other tools to check *Chromothripsis* as Pathwork, Shatterproff, and CNomplexity to analyse also the structural variants. During the conclusion of this thesis were found another tool, called Shatterseek (Harvard bioinformatics group), which is new and will be used to try the analysis with copy number and structural variants together.

We just analysed 9 samples of RNA-seq, of the same samples previously analysed by CNVs and classified as with or without CTLPs. The RNA data was analysed by two pipelines: by EdgeR and Nexus expression 3.0. Samples with and without CTLPs were compared to check the diferential expression between the groups.

Tables with all altered genes tables are in supplementary data to better visualization. It is in attachment D.

Some tools to check the pathways affected by this alterations were used as: Reactome < https://reactome.org/> and Panther <www.pantherdb.org/pathway/>, using the lisk of the genes in the digital platforms online. GSEA software was used to check pathways too, using the output from EdgeR, the gene sets database used was: gseaftp.broadinstitute.org://pub/gsea/gene_sets_final/c2.cp.reactome.v6.1.symbols.gmt, and the p-value <0.06.

5.7 – PhD.Sandwich at Barts Cancer Institute (Supervisor: Dr. Sarah McClelland)

The Sandwich PhD. period was realized at Barts Cancer Institute of Queen Mary University of London (QMUL) in London - UK, under the supervision of Dr. Sarah McClelland, during 4 months (May-Aug 2017).

5.7.1- Cell culture and treatment:

The U2-OS cell line is from the tibia of a Caucasian girl (15 years old), deceased because the OS tumor. The line originally 2T was derived in 1964, and has a chromosomally highly altered: chromosome counts in the hypertriploid range, high number of stable marker

chromosomes and different chromosomal rearrangements involving the same chromosomes (N1, N7, N9, and N11 particularly) and 22 markers are found including: t(9qter--->9q21::1p36--->1p::?), 7p+, iso(17q), t(15q;?), 4q+, del(3)(q21), 5q(aberrant) and others (Atcc, 2018). The U2-OS SKY karyotype is in figure 12.

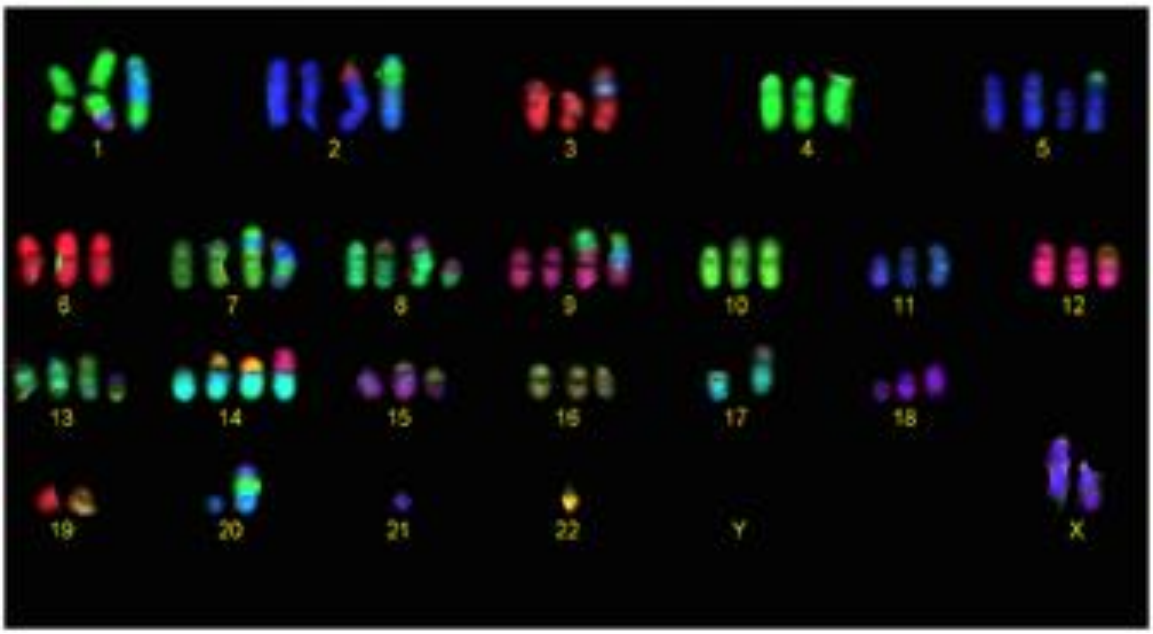

Figure 12: U2-OS SKY karyotype (Janssen and Medema, 2012).

The U2-OS cell line did grown under different conditions that promoted chromosome missegregation for example to induce whole chromosome lagging and missegregation we used a nocodazole washout strategy. Cells in log phase growth were treated with 100 ng/mL nocodazole for 8 hr and released following mitotic shake-off into fresh medium for 12 hr.

Such abnormal chromosomes are subject to unequal segregation at anaphase, thereby contributing to changes in chromosome number between daughter cells. To induce this type of error we treated cells with aphidicolin 0,1ng/ml for 24hours to cause replication stress-induced aneuploidy (Burrell et al., 2013).

5.7.2 - Fluorescence In-Situ hybridisation (FISH):

To analyse aneuploidy rates cells were harvested and treated with Barts laboratory's standard protocol to perform conventional FISH on microscopy slides, performed following the protocols used at Barts Institute.

We also tried the FISH to Imagestream cytometer to analyse the cells treated with aphidicolin. However, the technique was not working well at the lab because of the

background in the results. They are also performing single-cell sequencing (SCS) and aneuploidy detection using AneuFinder to validate their analysis what we did not have time to do.

5.7.3 - Microscopy techniques:

To observe behaviour of individual chromosomes during mitosis and determine likely mechanisms driving missegregation we analyzed the frequency at which aneuploidy-prone chromosomes lag at anaphase with almost all-centromere FISH probes in combination. We also made live cell imaging to try to find the tracking of specific chromosomes through missegregation at anaphase, incorporation in MN, reincorporation in major nucleus and other events.

All detailed protocols used at Barts Cancer Institute and live cell imaging are available in Supplementary Data (attachment D).

Results and Discussion

6.0 – RESULTS AND DISCUSSION

6.1 - Publications Summary

We surveyed all the publications about chaotic rearrangements in OS involving 351 OS samples (comprising 11 cell lines and 340 tumours) based on sequencing technology or CGH/SNP arrays to determine the incidence of the various classes of chaotic genomic rearrangement (table 9). Microarrays were used just in 4 publications, and NGS (Next Generation Sequencing) techniques in 10 publication. We found 156 samples (4 cell lines and 152 tumors) with *Chromothripsis, Chromothripsis*-like or *Chromoplexy*. The rate of chaotic rearrangements in all OS samples found was 44%.

Table 9– Summary of publications about complex rearrangements in OS classified as chaotic events

Reference	Technology	OS Samples	Altered Samples (%)	Phenomenon
Cell (Stephens *et al.*, 2011)	SNP array ; Sequencing	9	3 (33%)	*Chromothripsis*
Genome Res (Kim *et al.*, 2013)	CGH array	7	0 (0%)	-
Human Genetics (Reimann *et al.*, 2014)	Exome sequencing	1	1 (100%)	*Chromothripsis*
Cell (Chen *et al.*, 2014)	WGS	34	4 (~11.8%)	*Chromothripsis*
PNAS (Perry *et al.*, 2014)	WGS, WES, RNA Sequencing	13	11 (84%)	Indicative of *Chromoplexy*
Nat Commun (Kovac *et al.*, 2015)	WES; SNP array	31	2 (~6.45%)	*Chromoplexy*/ *Chromothripsis*-like
Oncotarget (Lorenz *et al.*, 2016)	WGS	11 (cell lines)	4 (~36%)	*Chromothripsis*-like
Nat Commun (Behjati *et al.*, 2017)	WGS	37	33 (89%)	*Chromothripsis* (11) + *Chromothripsis* and amplification (22)
IJC (Smida *et al.*, 2017)	CGH + SNP arrays	157	52 (33%)	*Chromothripsis*-like
Leukemia (Ratnaparkhe *et al.*, 2017)	WGS	3	3 (100%)	*Chromothripsis*
Nature (Gröbner *et al.*, 2018)	WGS	14	14 (100%)	*Chromothripsis*
Biorxiv (Cortes-Ciriano *et al.*, 2018)	WGS	34	29 (85%)	*Chromothripsis*

There was a high variation between the rate of chaotic events througout the studies (varying from 0 to 100%). This large range can be explained by the different criteria used to

assign *Chromothripsis* in the different publications (Kinsella et al., 2014), and the variation in resolution of the platforms used in the last several years.

6.2– Tools online

Using the tool *Chromothripsis* Explorer, and searching for Bone Osteosarcoma, we found the rates shown in the figure 13. These data are the same that were found in Cortes-Ciriano et al. (2018) publication, with 85% OS samples (29 of 34 OS tumors) presenting *Chromothripsis*.

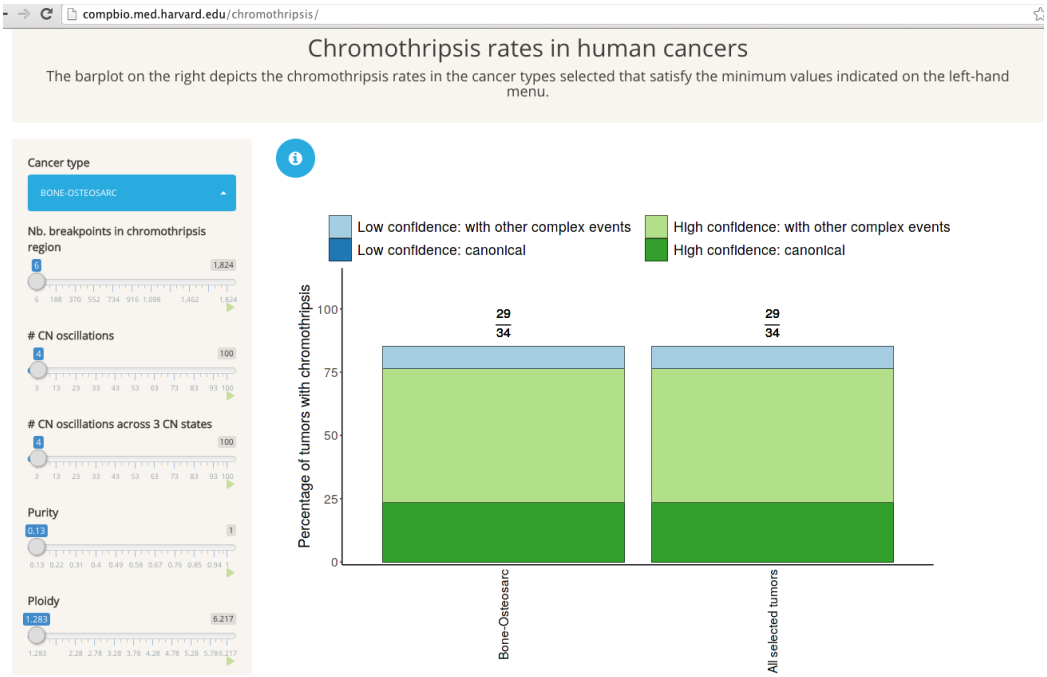

Figure 13: *Chromothripsis* Explorer showing *Chromothripsis* rates in OS human tumors.

The chaotic rearrangements rate found in Cortes-Ciriano *et al*. (2018) publication is very high, more than the double of the first and bench-mark publication by (Stephens *et al.*, 2011), which analyzed just 9 OS samples.

Chromothripsis Explorer seems a potencial tool as a repository which can be improved with new data from future publications with the same pipeline, using the Shatterseek.

Using the tool *Chromothripsis*DB, we found 4 OS studies registered. Some results of NGS techniques of 9 OS samples were involved (Figure 14).

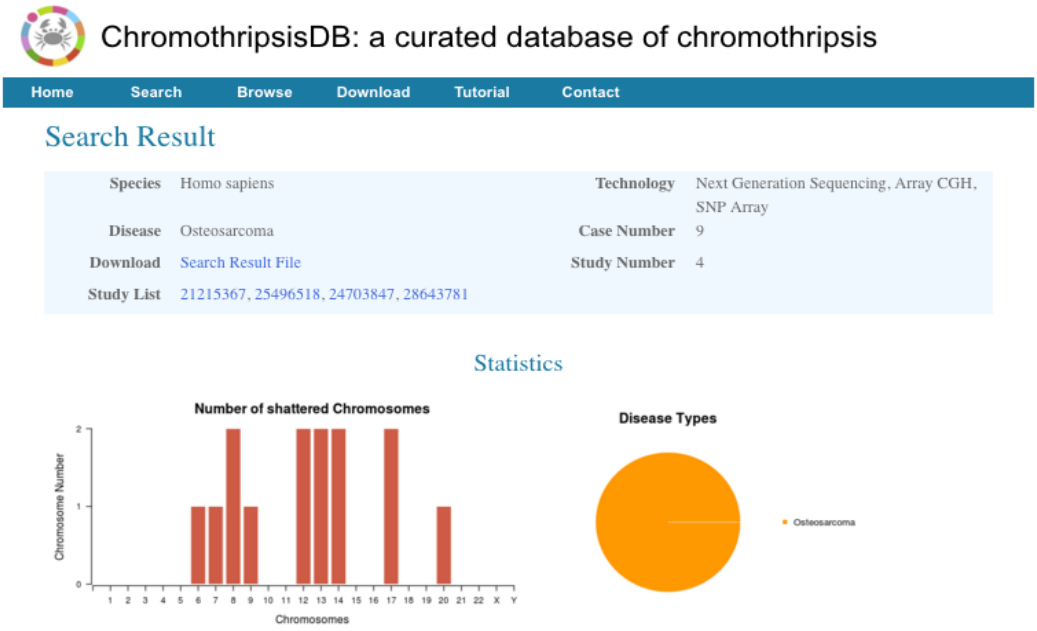

Figure 14: Overview of the *Chromothripsis*DB for Osteosarcoma.

The results presented are from published data: Pubmed 21215367 (Stephens *et al.*, 2011)- with 3 samples); Pubmed 25496518 (Reimann *et al.*, 2014)- 1 sample); Pubmed 28643781 (Behjati *et al.*, 2017)- 1 sample); and Pubmed 24703847 (Chen *et al.*, 2014) - with 4 samples).

The figure 15 shows the 9 chromosomes with *Chromothripsis* : chromosomes 6, 7, 8, 9, 12, 13, 14, 17 and 20. Chromosomes 8, 12, 13, 14, and 17 were related in two different studies.

*Chromothripsis*DB has potencial to be a good repository, however this tool does not seems to be updated regularly.

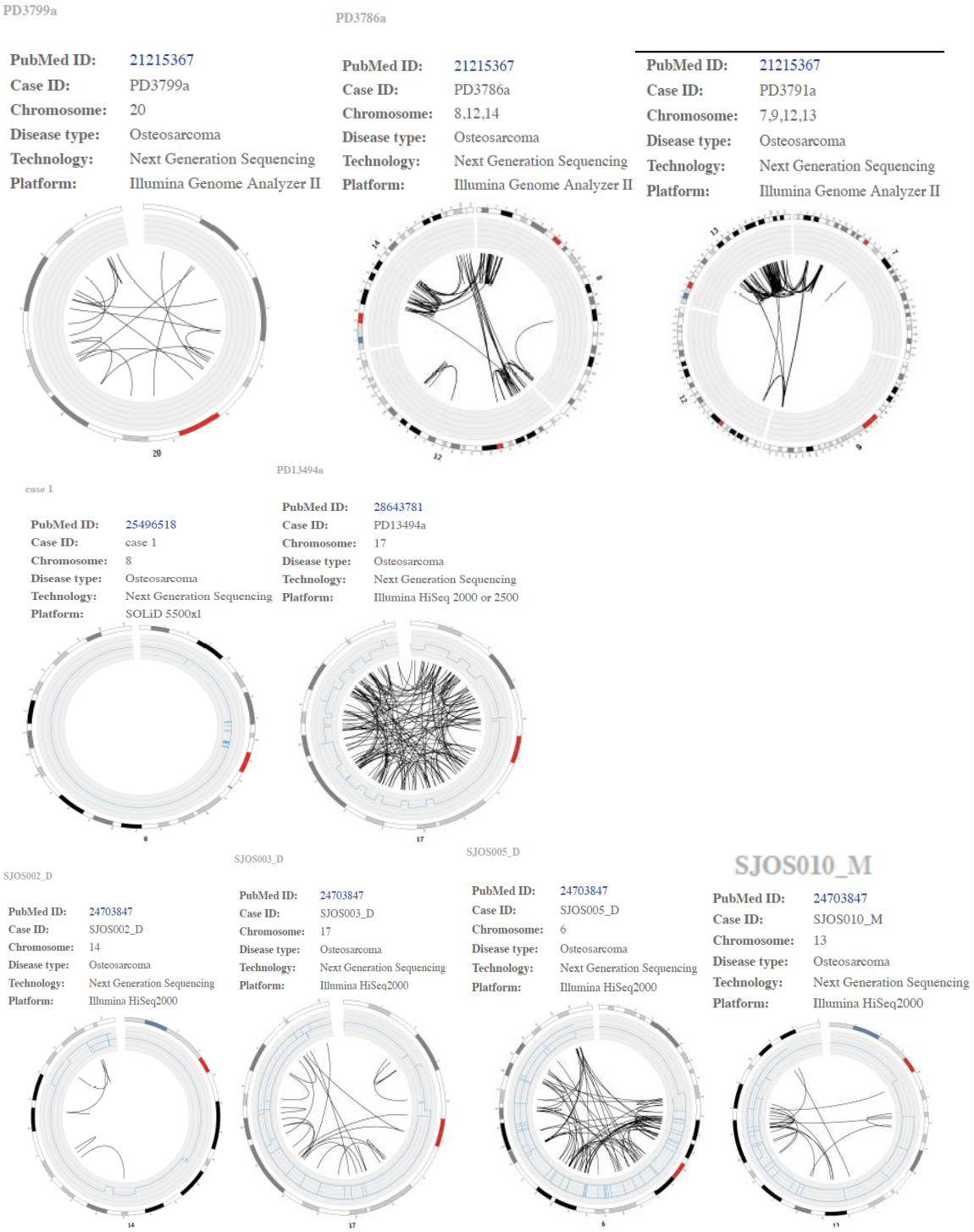

Figure 15: Plots of OS samples (by chromosome) characterized with *Chromothripsis* in *Chromothripsis*DB.

*Chromothripsis*DB also shows the genes altered in one tumor in comparation with another. We made a comparison between OS and the germline findings (figure 16). Genes present just in OS were: *GZF1*, *STK4*, *GFRA2*, *SPTAN1*, *ETV6*, *TP53*, *MAP2K4* and *NF*. These genes are related with biological process pathways of cell communication, biological regulation (e.g. regulation of cell cycle), development (e.g. cell death), and response to stimulus (e.g. response to stress) (Mi *et al.*, 2013).

Gene Results

Affected genes in **Osteosarcoma**, but **NOT** in **Germline**.

Gene Symbol	Entrez Gene	Chromosome	Start	End	Description
GZF1	64412	20	23342787	23353700	GDNF-inducible zinc finger protein 1
STK4	6789	20	43595115	43708600	serine/threonine kinase 4
GFRA2	2675	8	21547915	21669869	GDNF family receptor alpha 2
SPTAN1	6709	9	131314866	131395941	spectrin, alpha, non-erythrocytic 1
ETV6	2120	12	11802788	12048336	ets variant 6
TP53	7157	17	7565097	7590856	tumor protein p53
MAP2K4	6416	17	11924141	12047147	mitogen-activated protein kinase kinase 4
NF1	4763	17	29421945	29709134	neurofibromin 1

Figure 16: *Chromothripsis*DB shows some affected genes by *Chromothripsis* in OS but not in Germline.

Using the tool Cosmic Cancer Browser, we identified the most frequently altered genes in Osteosarcoma. Cosmic database analysed the 269 samples of Bone Osteosarcoma with all screens (whole genome and target). The genes mutated more frequently in OS are *TP53, CDKN2A, RB1, ATRX, KMT2C, LRP1B, GNAS, MLLT3, ARID1A, PTEN* etc. (figure 17).

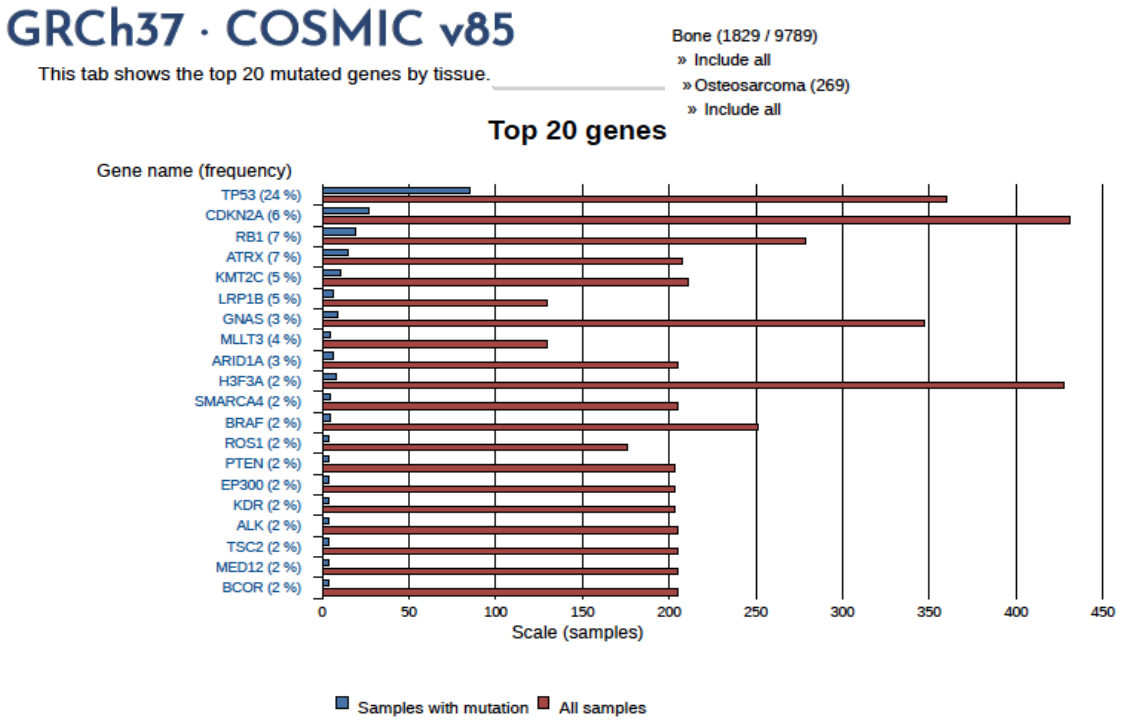

Figure 17 – The top 20 mutated COSMIC genes in Bone Osteosarcoma samples by whole genome and target screen.

The genes more frequent in OS samples without mutations are *IDH1*, *NRAS*, *KRAS*, *HRAS*, *KIT*, *CDKN2C*, *PDGFRA*, *FGFR2*, *CTNNB1*, *ERBB2*, *JAK2*, *CEBPA*, *NFE2L2*, *EZH2*, *RAC1*, etc. (figure 18).

Genes

Top 20 genes | Genes with mutations | Genes without mutations

This tab shows genes that have no mutations for the current tissue/histology selections. Read more on our help pages.

Show 25 entries

Gene	Samples tested
IDH1	503
NRAS	303
KRAS	289
HRAS	284
KIT	267
CDKN2C	266
PDGFRA	243
FGFR2	242
CTNNB1	239
ERBB2	233
JAK2	225
CEBPA	223
NFE2L2	212
EZH2	211
RAC1	205
CDC73	205
PRKAR1A	205

Figure 18 – The top non mutated COSMIC genes in Bone Osteosarcoma samples by whole genome and target screen.

The gene *TP53*, located in chromosome region 17p13.1, is related not just with OS but also with *Chromothripsis* in OS. Copy number alterations and mutations in this gene can interfere in the maintenance of genomic stability (Martin et al., 2012). *Chromothripsis* was regularly observed in hyperdiploid cancers, specially when *TP53* mutations are present (Gröbner et al., 2018).

6.3 – GEO #GSE12830 array reanalysis

All ten samples evaluated from the GEO #GSE12830 study showed 3020 CN aberrations in total, mean of 275 CN aberration per sample (figures 19 and 20).

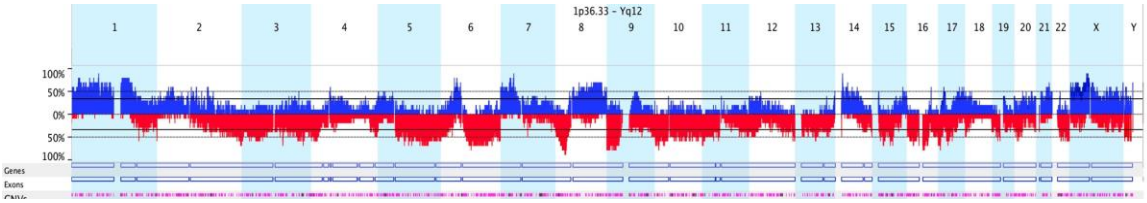

Figure 19 – Overview of the 10 OS samples showing the high rate of copy number changes between the 10 OS samples (GEO #12830) genome, by Nexus 9.0.

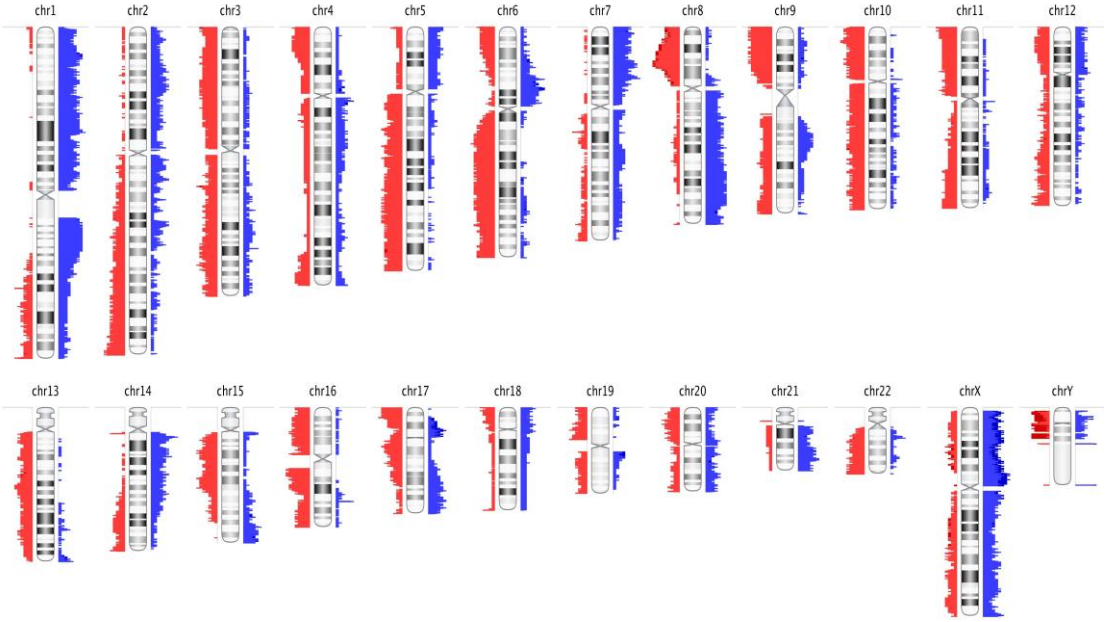

Figure 20 – Overview of the 10 OS samples showing the high rate of copy number changes between the 10 OS samples (GEO #12830) per chromosome, by Nexus 9.0.

Copy number alterations affected all chromosomes, with 1547 gains and 1473 losses in total. We found more gains than losses in this GEO study. The high rate of copy number alterations in these OS samples is compatible with the literature. OS is characterized by having an unusually high level of genomic alteration and chromosomal instability (Martin *et al.*, 2012; Rosenberg *et al.*, 2013).

6.3.1 - GEO #12830 array by CTLPScanner

The results obtained from the screening of the 10 OS samples in CTLPScanner are showed in table 10.

Table 10 – CTLPScanner results showing CTLP+ samples, chromosome regions and the CN status.

Array ID	CTLP Region (Mb)	Chromosome	Start	Stop	CNA status change times	Likelihood ratio (log10)
GSM322086CTLP	88.83	2	140000001	228827254	38	29
GSM322064CTLP	62.44	14	20000001	82435964	21	8
	62.44	20	1	62435964	24	10
GSM322090CTLP	78.77	10	56638886	135413628	27	21

We found 3 OS samples (30%) with CTLP+. This data is similar with the first publication about *Chromothripsis* (Stephens et al., 2011). The chromosomes affected by the chaotic events were Chr2, Chr10, Chr14 and Chr20 (figure 21), which are frequent related in OS. Chromosomes 14 and 20 have the high rate of centromeric rearrangements, and

chromosome 20 is classified as the chromosome with highest number of copy number alterations in OS tumors (Bayani et al., 2003). Furthermore, DNA copy number alterations in chromosomes Chr2 and Chr10 also are commonly reported in OS (Rosenberg *et al.*, 2013; Martin *et al.*, May 2012). Some regions involved in *Chromothripsis* present important genes, as *PTEN* (figure 22).

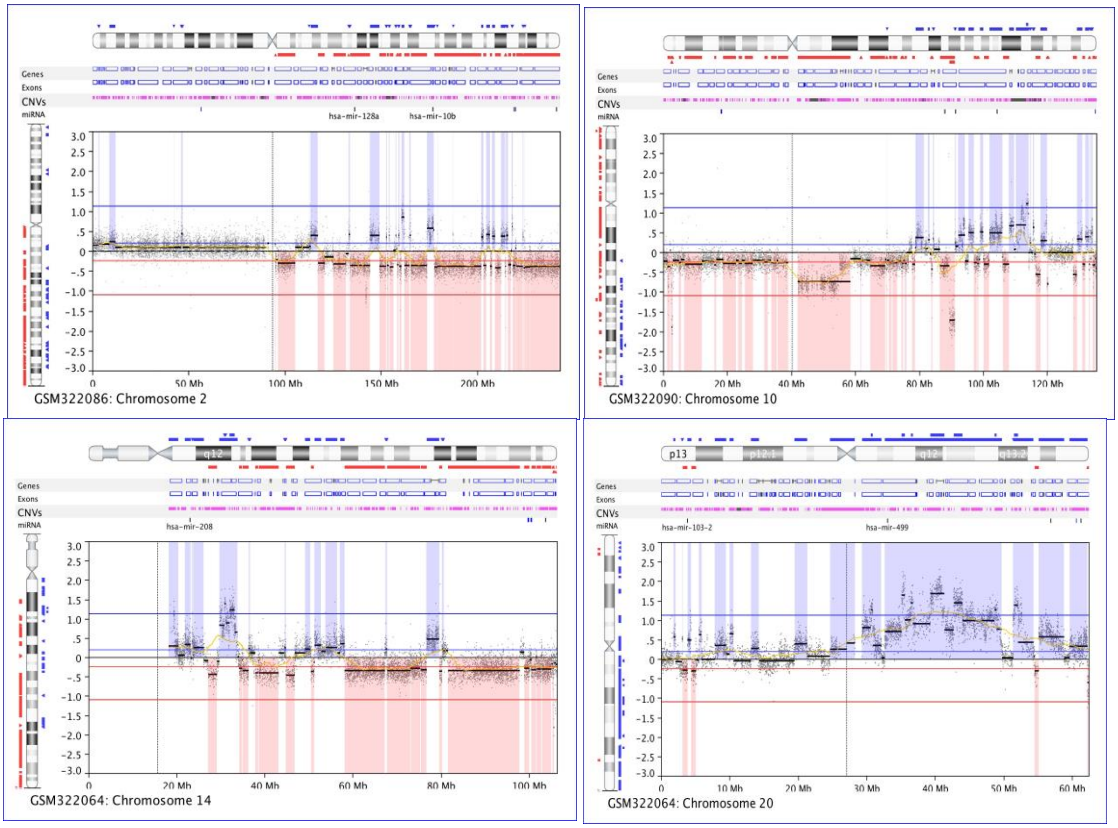

Figure 21 – Chromosomes affected by *Chromothripsis* in GEO study (#12830): Chr2 (sample OS180); Chr10 (sample OS183); Chr14 and Chr20 (sample OS87B).

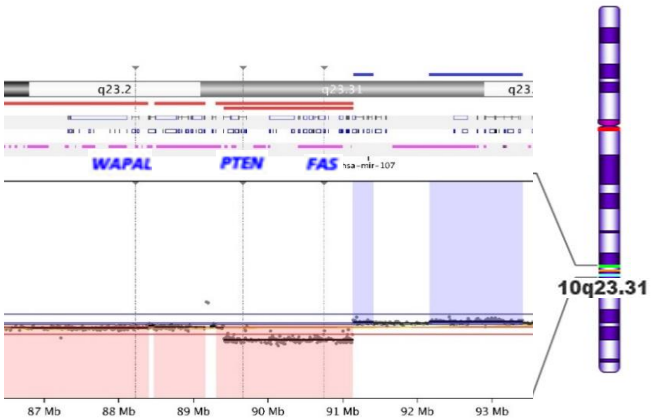

Figure 22 – Chromosome region affected by *Chromothripsis* on zoom in Chromosome 10 (sample OS183), where are located importante genes as *PTEN*.

Complex rearrangements may promote deletions of tumor suppressor genes (such as *PTEN* and *FAS*) and oncogenes amplification (such as *MYC*). These genes are related to OS pathogenesis, and contribute to accelerate the tumor development (Moriarity et al., 2015).

Among the ten OS samples, seven had alterations in chromosome 8q, with MYC duplication. All three samples related to CTLPs had this gain. 75% of OS cases have gains in 8q region, generally with MYC amplification (Kovac et al., 2015). This gene can act as a substrate to accelerate the evolution and progression of OS tumors (Stephens et al., 2011).

CTLP+ samples presented 1236 CN, mean of 468 CN per sample. CTLP- samples presented 1784 CN, mean of 255 CN per sample (table 11). The comparation between groups by the diferent rates of CN distribution is showed in figure 23.

Table 11 – Total CN aberrations by sample of GEO study (#12830), quality of arrays samples and CTLPScanner classification

#GEO Study	Sample	Sample ID	Total CN aberrations	Quality	Probes Discarded %	**CTLP**
GSE12830	GSM322064	OS87B	468	0.0554	0.0312	Yes
	GSM322072	OS138	388	0.0672	0.0459	No
	GSM322074	OS177	389	0.0495	0.0703	No
	GSM322076	OS178	272	0.1709	0.0417	No
	GSM322078	OS179	192	0.0643	0.0198	No
	GSM322086	OS180	278	0.045	0.0265	Yes
	GSM322088	OS182	255	0.2095	0.016	No
	GSM322090	OS183	490	0.0369	0.0206	Yes
	GSM322092	OS2336	52	0.044	0.0131	No
	GSM322094	OS2960	236	0.0625	0.1254	No

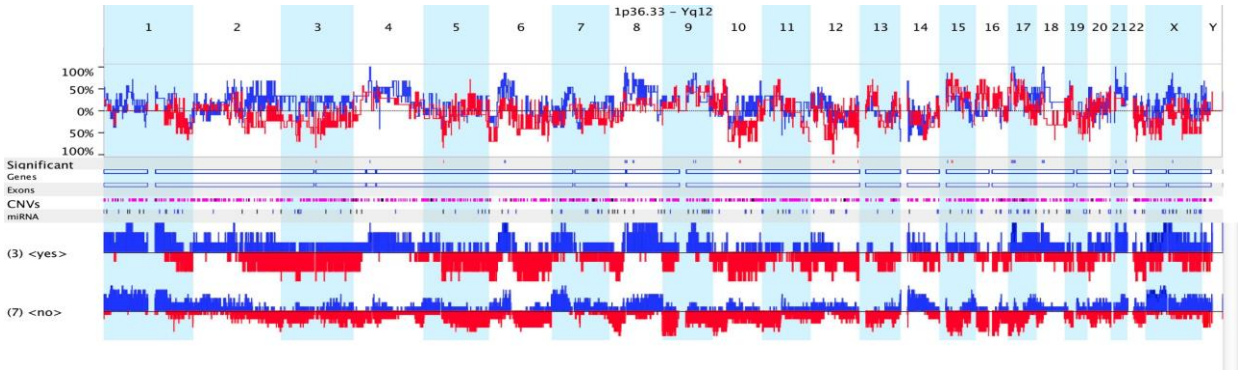

Figure 23 - Comparation between groups by the diferent rates of CN distribution.

When we compared the CNVs events between the groups CTLP+ and CTLP-, and the chromosome regions, we found the genes evolved in this CN alterations (table 12). CTLP+ samples present more losses than gains events (mean of 183 gains/sample and 228 losses/sample). CTLP- samples present more gains than losses (mean of 142gains/sample and 112 losses/sample).

Table 12 – Comparation of CNVs regions between CTLP+ samples and CTLP- samples, and the genes present in each region (continue).

Region	Event	Freq. in <yes> (%)	Freq. in <no> (%)	≠	p-value	CNA	Gene Symbols
chr3:95,021,157-95,152,842	CN Loss	100	14.3	85.7	0.03	1	PROS1
chr4:43,856,333-45,282,560	CN Gain	100	0	100	0.008	1	KCTD8, YIPF7, GUF1, GNPDA2
chr6:41,117,007--44,380,641	CN Gain	100	14.3	85.7	0.03	3	TSPO2, APOBEC2, OARD1, NFYA, C6orf132, GUCA1A, GUCA1B, MRPS10, TRERF1 LOC100132354, C6orf223, MRPL14, TMEM63B, CAPN11, SLC29A1, HSP90AB1, MIR4647, SLC35B2, NFKBIE, TMEM151B, TCTE1, AARS2 SLC20A2, C8orf40, CHRNB3, CHRNA6, THAP1, RNF170, MIR4469, HOOK3, FNTA, SGK196, HGSNAT, POTEA
chr8:42,498,582-43,647,122	CN Gain	100	14.3	85.7	0.03	2	SLC20A2, C8orf40, CHRNB3, CHRNA6, THAP1, RNF170, MIR4469, HOOK3, FNTA, SGK196, HGSNAT, POTEA
chr8:48,003,671-48,745,096	CN Gain	100	14.3	85.7	0.03	1	LOC100287846, KIAA0146
chr8:63,583,529-65,697,223	CN Gain	100	14.3	85.7	0.03	1	NKAIN3, LOC100130155, MIR124-2, LOC401463, BHLHE22, CYP7B1
chr9:84,925,229-91,593,701	CN Gain	100	14.3	85.7	0.03	2	GADD45G, UNQ6494, MIR4290, LOC286370, LOC340515, DIRAS2, SYK, LOC100129316, AUH, NFIL3, MIR3910-2, MIR3910-1, ROR2
chr10:72,980,218-75,637,374	CN Loss	100	14.3	85.7	0.03	2	CDH23, VCL, AP3M1, ADK
chr12:127,874,421-128,178,249	CN Loss	100	14.3	85.7	0.03	1	GLT1D1, TMEM132D
chr12:57,944,175-61,809,692	CN Loss	100	0	100	0.008	3	FAM19A2, USP15, MON2, C12orf61, MIRLET7I, PPM1H
chr15:20,070,027-20,235,180	CN Gain	100	14.3	85.7	0.03	1	REREP3, MIR4509-1, MIR4509-2, MIR4509-3

Table 12 – Comparation of CNVs regions between CTLP+ samples and CTLP- samples, and the genes present in each region (end).

chr15:31,518,485-34,698,093	CN Loss	0	85.7	85.7	0.03	3	RYR3, AVEN, CHRM5, AQR, C15orf41
chr17:10,047,750-15,059,837	CN Gain	100	0	100	0.008	1	MYH13, MYH8, MYH4, MYH1, MYH2, MYH3, SCO1, ADPRM, MAGOH2, TMEM220, TMEM220-AS1, LINC00675, PIRT, SHISA6, DNAH9, ZNF18, MIR744, MAP2K4, LINC00670, MYOCD, ARHGAP44, ELAC2, HS3ST3A1, CDRT15P1, COX10-AS1, COX10, CDRT15, MGC12916, HS3ST3B1, CDRT7
chr17:15,059,837-19,364,790	CN Gain	100	14.3	85.7	0.03	3	PMP22, MIR4731, TEKT3, CDRT4, TVP23C-CDRT4, TVP23C, CDRT1, TRIM16, ZNF286A, TBC1D26, CDRT15P2, MEIS3P1, ADORA2B, ZSWIM7, TTC19, NCOR, CCDC144A, FAM106CP, USP32P1, KRT16P2, TNFRSF13B LLGL1, FLII, SMCR7, TOP3A, SMCR8, SHMT1, EVPLL, LOC339240, KRT16P1, LGALS9C, USP32P2, FAM106A, CCDC144B, TBC1D28, FOXO3B, ZNF286B, TRIM16L, FBXW10, TVP23B, PRPSAP2, SLC5A10, FAM83G, GRAP, GRAPL, EPN2-IT1, EPN2, EPN2-AS1, MIR1180, B9D1, MAPK7, MFAP4, RNF112
chr17:7,602,229-8,694,313	CN Loss	0	85.7	85.7	0.03	2	DNAH2, KDM6B, TMEM88, LSMD1, CYB5D1, CHD3, SCARNA21, LOC284023, KCNAB3, TRAPPC1, CNTROB, GUCY2D, PFAS, RANGRF, SLC25A35, ARHGEF15, ODF4, LOC100128288, KRBA2, RPL26, RNF222, NDEL1, MYH10, CCDC42, SPDYE4, MFSD6L, PIK3R6
chr18:14,562,532-16,100,000	CN Gain	100	0	100	0.008	1	ANKRD30B, MIR3156-2, LOC644669
chr18:14,303,026-14,562,532	CN Gain	100	14.3	85.7	0.03	1	CYP4F35P, CXADRP3, POTEC
chr18:17,654,794-18,245,461	CN Gain	100	0	100	0.008	1	MIB1, MIR133A1, MIR1-2, GATA6
chr18:16,100,000--18,758,031	CN Gain	100	14.3	85.7	0.03	3	ROCK1, GREB1L, ESCO1, SNRPD1, ABHD3, MIR320C1, MIB1,CTAGE1,
chr21:41,320,910-41,676,644	CN Gain	100	14.3	85.7	0.03	1	LINC00323, MIR3197, PLAC4, BACE2, FAM3B, MX2
chrX:71,999,653-72,455,305	CN Gain	100	14.3	85.7	0.03	1	PABPC1L2B, PABPC1L2A, NAP1L6, NAP1L2, CDX4

We found that 171 genes mapped to regions of *Chromothripsis* with the majority (77 genes) mainly having functions related to cellular communication and cell cycle (Mi et al., 2013). There were 43 genes that were related to metabolic process (mainly associated with RNA metabolism) and 27 genes with cellular component organization or biogenesis. Numerous pathways regulate cell proliferation, motility, and survival, and the alterations that happen in cancer cells are the consequence of multiple alterations in cellular signaling machinery (Martin, 2003).

6.3.2 - GEO #12830 Expression data

Six RNA samples of the two groups were compared (CTLP+ x CTLP-) using Nexus Expression 3.0. Two samples CTLP+ (OS180, OS183) were compared with four samples CTLP- (OS182, OS179, OS178, OS177). The differential expression of some genes of immune system pathway is showed in table 13 and the heatmap in figure 24.

Table 13 –Genes with different expression between CTLP+ x CTLP- comparation and biological process involved in immune response (p-value< 0.01) (continue).

Gene Symbol	Comparison	Biological Process
CADM1	**DOWN**	T cell mediated cytotoxicity, activated T cell proliferation, apoptosis, cell adhesion, cell differentiation, cell recognition, detection of stimulus, heterophilic cell adhesion, homophilic cell adhesion, immune response, multicellular organismal development, negative regulation of cell cycle, positive regulation of cytokine secretion, spermatogenesis, susceptibility to natural killer cell mediated cytotoxicity
CLEC4A	**DOWN**	cell adhesion, cell surface receptor linked signal transduction, immune response
CCR1	**DOWN**	G-protein signaling; coupled to cyclic nucleotide second messenger, cell adhesion, cell-cell signaling, chemotaxis, cytokine and chemokine mediated signaling pathway, elevation of cytosolic calcium ion concentration, immune response, inflammatory response
CD164	**DOWN**	cell adhesion, hemopoiesis, heterophilic cell adhesion, immune response, multicellular organismal development, negative regulation of cell adhesion, negative regulation of cell proliferation, signal transduction
IL32	**UP**	cell adhesion, defense response, immune response
LAT	**UP**	Ras protein signal transduction, calcium-mediated signaling, immune response, integrin-mediated signaling pathway, intracellular signaling cascade, mast cell degranulation, regulation of T cell activation, transport

Table 13 –Genes with different expression between CTLP+ x CTLP- comparation and biological process involved in immune response (p-value< 0.01) (end).

BCL3	**UP**	DNA damage response; signal transduction by p53 class mediator resulting in induction of apoptosis, I-kappaB kinase/NF-kappaB cascade, T-helper 1 type immune response, T-helper 2 cell differentiation, antimicrobial humoral response, defense response to bacterium, defense response to protozoan, extracellular matrix organization and biogenesis, follicular dendritic cell differentiation, germinal center formation, humoral immune response mediated by circulating immunoglobulin, maintenance of protein location in nucleus, marginal zone B cell differentiation, negative regulation of apoptosis, negative regulation of interleukin-8 biosynthetic process, negative regulation of transcription, negative regulation of tumor necrosis factor biosynthetic process, positive regulation of interferon-gamma production, positive regulation of interleukin-10 biosynthetic process, positive regulation of transcription, positive regulation of transcription from RNA polymerase II promoter, positive regulation of translation, protein import into nucleus; translocation, regulation of DNA binding, regulation of NF-kappaB import into nucleus, regulation of transcription; DNA-dependent, response to DNA damage stimulus, response to UV-C, response to virus, spleen development
FCAR	**UP**	immune response
RFX1	**UP**	immune response, regulation of transcription; DNA-dependent
IL1B	**UP**	activation of MAPK activity, angiogenesis, anti-apoptosis, apoptosis, cell-cell signaling, cytokine and chemokine mediated signaling pathway, elevation of cytosolic calcium ion concentration, fever, immune response, inflammatory response, leukocyte migration, negative regulation of cell proliferation, neutrophil chemotaxis, positive regulation of I-kappaB kinase/NF-kappaB cascade, positive regulation of JNK cascade, positive regulation of chemokine biosynthetic process, positive regulation of interleukin-6 biosynthetic process, positive regulation of interleukin-6 production, positive regulation of protein amino acid phosphorylation, positive regulation of transcription factor activity, signal transduction
CXCL1	**UP**	G-protein coupled receptor protein signaling pathway, actin cytoskeleton organization and biogenesis, chemotaxis, immune response, inflammatory response, intracellular signaling cascade, negative regulation of cell proliferation, nervous system development
SPON2	**UP**	axon guidance, cell adhesion, immune response
CCR6	**UP**	G-protein coupled receptor protein signaling pathway, cell motion, cellular defense response, chemotaxis, elevation of cytosolic calcium ion concentration, humoral immune response, signal transduction
IL6	**UP**	acute-phase response, cell surface receptor linked signal transduction, cell-cell signaling, defense response to protozoan, humoral immune response, inflammatory response, negative regulation of apoptosis, negative regulation of cell proliferation, negative regulation of chemokine biosynthetic process, negative regulation of hormone secretion, neutrophil apoptosis, positive regulation of MAPKKK cascade, positive regulation of T-helper 2 cell differentiation, positive regulation of cell proliferation, positive regulation of peptidyl-serine phosphorylation, positive regulation of peptidyl-tyrosine phosphorylation, positive regulation of transcription from RNA polymerase II promoter, positive regulation of translation, response to glucocorticoid stimulus
SEMA3C	**UP**	immune response, multicellular organismal development, response to drug, transmembrane receptor protein tyrosine kinase signaling pathway
GEM	**UP**	cell surface receptor linked signal transduction, immune response, small GTPase mediated signal transduction

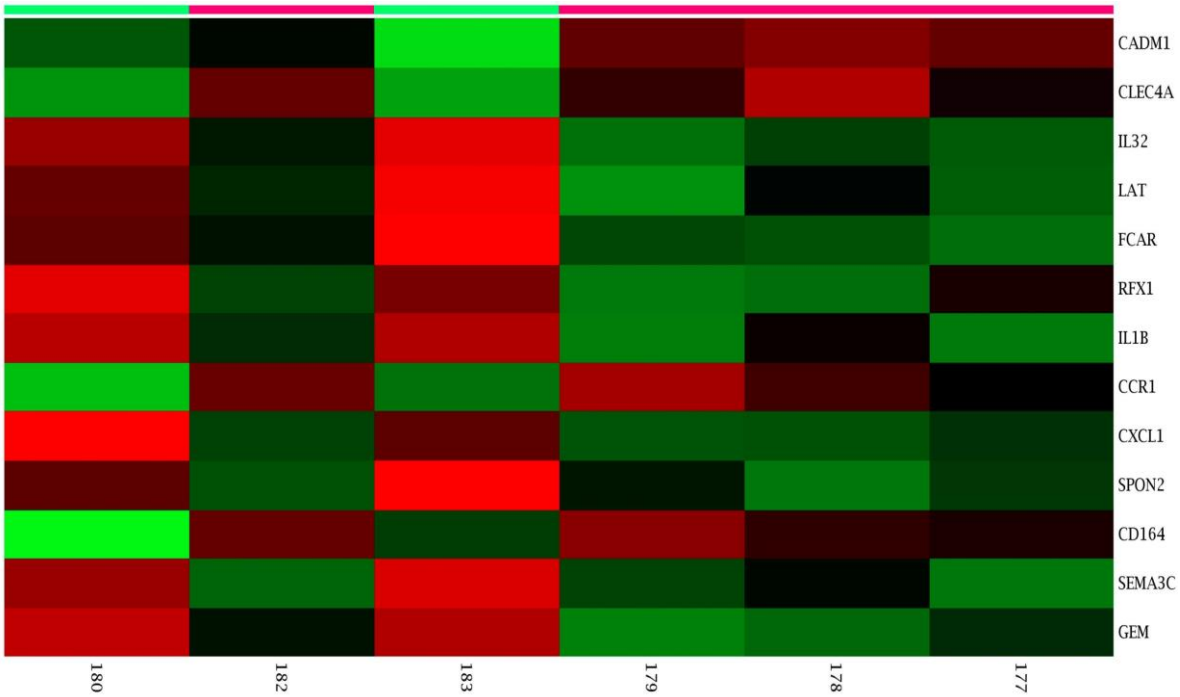

Figure 24- Heatmap of different immune response genes expression between CTLP+ x CTLP- groups.

There were four genes associated with the immune system that were underexpressed (*CADM1; CLEC4A; CCR1; CD164*) and 12 were overexpressed (*IL32, LAT, BCL3, FCAR, RFX1, ILIB, CXCL1, SPON2, CCR6, IL6, SEMA3C, GEM*) in the *Chromothripsis* tumors. Interestingly, all the genes underexpressed also have a role in cell adhesion pathway. Cell adhesion is associated with cancer progression and metastasis (Bendas and Borsig, 2012). Loss of intercellular adhesion can permits malignant cells to escape from their location of origin, damage the extracellular matrix, obtain a more motile and invasion phenotype, and metastasize (Okegawa et al., 2004; Bendas and Borsig, 2012).

The differential expression of genes related to bone pathways were compared between the two groups, and we can see the results in table 14 and the heatmap in figure 25.

Table 14 – Genes with different expression between CTLP+ x CTLP- comparation (bone related pathway). p-value< 0.01

Gene Symbol	Comparison	Biological Process
CALCA	**UP**	cell-cell signaling, cellular calcium ion homeostasis, cytosolic calcium ion homeostasis, , elevation of cytosolic calcium ion concentration during G-protein signaling;,endothelial cell migration, endothelial cell proliferation, inflammatory response, leukocyte adhesion, **negative regulation of bone resorption**, negative regulation of ossification, negative regulation of osteoclast differentiation, negative regulation of smooth muscle contraction, negative regulation of transcription, positive regulation of macrophage differentiation, positive regulation of ossification.
CARTPT	**UP**	negative regulation of bone resorption, negative regulation of osteoclast differentiation, , **regulation of bone remodeling,** signal transduction
ADRB2	**UP**	**bone resorption**, brown fat cell differentiation, negative regulation of calcium ion transport via voltage-gated calcium channel, negative regulation of inflammatory response, negative regulation of multicellular organism growth, negative regulation of ossification, negative regulation of smooth muscle contraction, positive regulation of apoptosis, positive regulation of bone mineralization, positive regulation of cell proliferation, positive regulation of heart contraction, positive regulation of skeletal muscle growth, positive regulation of transcription from RNA polymerase II promoter, positive regulation of vasodilation, receptor-mediated endocytosisregulation of sodium ion transport

Bone deposition and resorption are processes of bone remodeling. OS tumors have focal bone deposition. Changes in expression levels in the complex network of genes that controls bone remodeling (as *CALCA, CARTPT* and *ADRB2*) can influence the genesis and progression of bone diseases, and may be connected with OS biology and clinical features (Toledo et al., 2010).

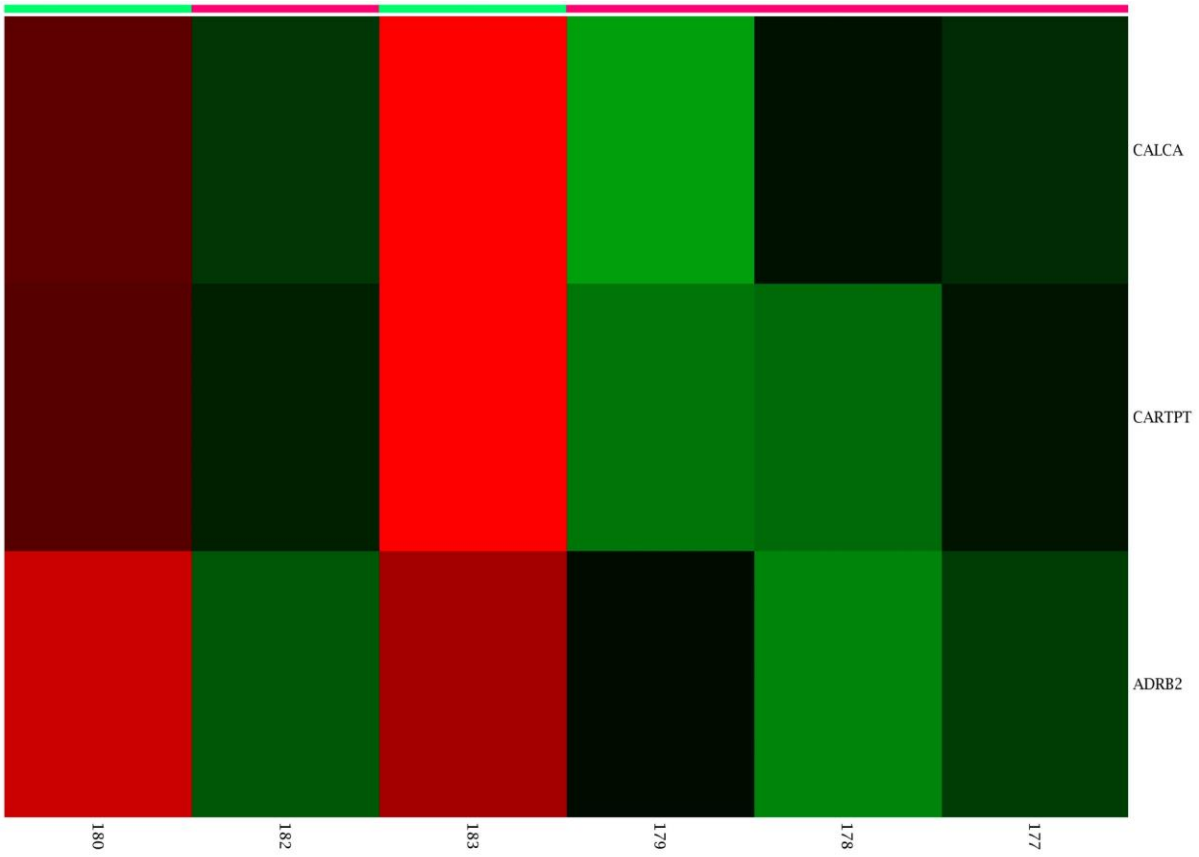

Figure 25- Heatmap of different bone related genes expression between CTLP+ x CTLP- groups.

6.4 – Other OS arrays studies at public Databank

Other 4 set of arrays (GSE67125; GSE3153; GSE36003; and GSE12789) also were evaluated, according with the same conditions of the GEO #GSE12830.

Together, the 5 arrays set analysed in this thesis (including #GSE12830) have 117 OS samples: 82 tumors and 35 cell lines. We reanalyzed these DNA copy number data to identify potential chromosomal regions commonly involved in chaotic DNA copy number alterations, especially CTLPs. We found *Chromothripsis* in 27 OS samples (23%). There were 11 tumors CTLP+ (13%) and 16 cell lines (45%) (table 15).

Table 15 – The 27 Samples CTLP+, according with the platform, sample type, and chromosome affected.

Sample	Build	Platform	#Study databank	Sample Type	chr affected
GSM1639703	37	Affymetrix	GSE67125	OS cell line	chr13
GSM821012	37	Affymetrix	GSE33153	OS tumor	chr5
GSM821017	37	Affymetrix	GSE33153	OS tumor	chr2
GSM821019	37	Affymetrix	GSE33153	OS tumor	chr2
GSM879206	37	Affymetrix	GSE36003	OS cell line	chr11
GSM879209	37	Affymetrix	GSE36003	OS cell line	chr8; chr9
GSM879210	37	Affymetrix	GSE36003	OS cell line	chr17
GSM879212	37	Affymetrix	GSE36003	OS cell line	chr4
GSM879214	37	Affymetrix	GSE36003	OS cell line	chr1, chr3, chr15
GSM879215	37	Affymetrix	GSE36003	OS cell line	chr11
GSM879216	37	Affymetrix	GSE36003	OS cell line	chr6
GSM879217	37	Affymetrix	GSE36003	OS cell line	chr4; chr8
GSM879218	37	Affymetrix	GSE36003	OS cell line	chr12
GSM879220	37	Affymetrix	GSE36003	OS cell line	chr8
GSM879221	37	Affymetrix	GSE36003	OS cell line	chr13
GSM879222	37	Affymetrix	GSE36003	OS cell line	chr1
GSM879224	37	Affymetrix	GSE36003	OS cell line	chr8
GSM320781	35	Agilent FE	GSE12789	OS tumor	chr3; chr5, chr10
GSM320790	35	Agilent FE	GSE12789	OS tumor	chr12
GSM320792	35	Agilent FE	GSE12789	OS tumor	chr6
GSM320795	35	Agilent FE	GSE12789	OS tumor	chr16
GSM320807	35	Agilent FE	GSE12789	OS tumor	chr2
GSM320824	35	Agilent FE	GSE12789	OS cell line	chr2
GSM320831	35	Agilent FE	GSE12789	OS cell line	chr12
GSM322064	35	Agilent FE	GSE12830	OS tumor	chr14,chr20
GSM322086	35	Agilent FE	GSE12830	OS tumor	chr2
GSM322090	35	Agilent FE	GSE12830	OS tumor	ch10

We found 17 different chromosomes reported with *Chromothripsis*. Except chromosomes 7, 18, 19, 21 and the sexual chromosomes were not affected by CTLPs. Chromosomes 2, 8 and 12 were frequent targets of *Chromothripsis* in OS arrays. Chromosome 2 were reported in five CTLP+ samples, chromosome 8 in four CTLP+ samples, and chromosome 12 in three CTLP+ samples. Chromosomes 1, 3, 4, 5, 6, 10, 11 and 13 were present in two different CTLP+ samples.

OS tumors usually have complex chromosome aberrations with high incidence of numerical DNA copy number alterations in chromosomes 1, 2, 3, 6, 8, 12, 10, 13, 14, 17, and 20 (Bayani et al., 2003; Martin et al., 2012; Rosenberg et al., 2013). We can consider that these alterations are according with literature, and these chromosomes may also be more susceptible to *Chromothripsis*.

We have to consider the variation in resolution of the platforms used in the last several years. The most CTLP+ were build 37, however had more samples build 35. We found 55 arrays build 37 (17 were CTLP+) and 62 arrays build 35(10 were CTLP+). The recent technologies used to have better resolution and sensibility enough to detect these type of chaotic genomic rearrangements.

6.5 – dbGap data

The DNA copy number alterations were analyzed using whole genome sequence data of 12 OS tumors available from dbGaP databank. The average age at diagnosis of patients were 14.25 yers old. Five samples are female (41.67%), and seven male (58.33%). Until the last outcome were reported four patients living (33.33%) and eight deceased (66.67%) of OS tumor. The deceased patients lived for 13.5 months (mean) after the diagnosis. The figures 26 and 27 show the copy number alterations between the OS samples from dbGap databank (Using Nexus). We can observe the high number of gains and losses, even between the chromosomes.

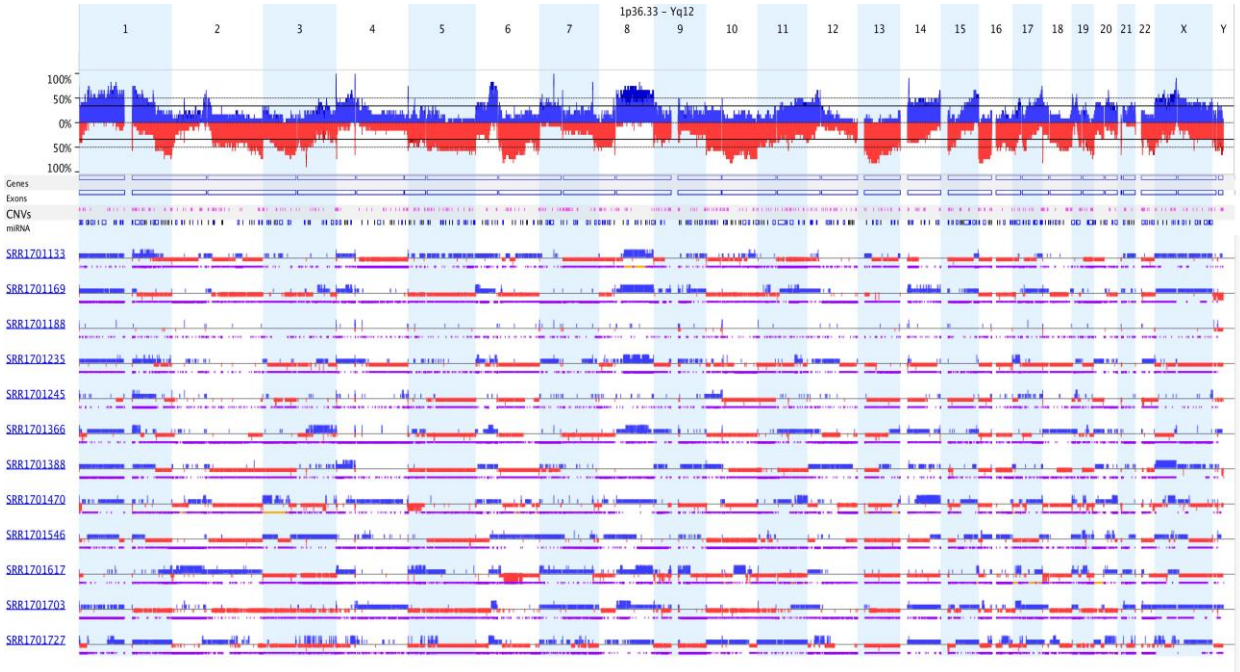

Figure 26 - Overview of the 12 OS samples showing the high rate of copy number changes between the WGS samples (dbGap phs000699), by Nexus 9.0.

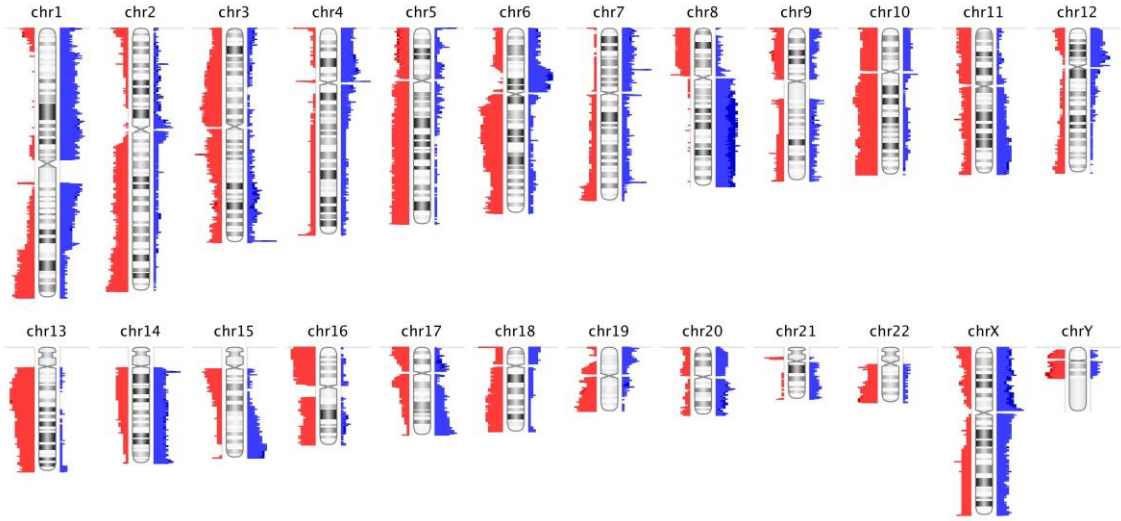

Figure 27 - Overview of the 12 OS samples showing the high rate of copy number changes between the WGS samples (dbGap phs000699) by chromosomes (Nexus 9.0).

We found CTLPs in 7 (58%) of the 12 OS samples analyzed using whole genome sequence data (table 16). In total there were 12 different chromosomes involved, affecting 62.5% of samples from patients who died of OS. Chromosomes 1, 2, 3, 7 and 12 were slightly more often *Chromothripsis* target locations.

Table 16 – The 7 Samples CTLP+, according with the CTLPs region size, Copy number status, and chromosome affected.

Sample	Size (Mb)	Chr	Start	End	CN	Log10
SRR1701727CTPL	81.19	3	110000001	191195210	65	30.8
	30	4	20000001	50000000	29	17.6
	51.30	5	30000001	81304566	42	20.63
	30	7	105000001	135000000	21	10.11
	48.12	12	15000001	63129895	39	18.02
	63.02	16	25000001	88025520	38	13.22
SRR1701388CTLP	59.12	7	1	59128983	21	20.35
SRR1701366CTPL	81.19	3	115000001	196195210	61	40.77
SRR1701617CTLP	30	2	1	30000000	21	10.13
	102.53	9	1	102531392	81	40.49
SRR1701470CTLP	59.12	19	1	59128983	32	14.66
SRR1701133CTLP	81.19	1	145000001	226195210	57	41.03
	90.35	10	30000001	120354753	31	11.99
SRR1701235CTLP	48.12	1	200000001	248129895	38	27.93
	59.12	6	1	59128983	23	9.576

In table 17 we can show the number of genes in CTLP regions: common genes and Cosmic genes, by chromosome. Apparently, the sample SRR1701727 is affected by *Chromoplexy*, with six chromosomes affected.

Table 17 - Summary of the genes from the dbGAP samples – CTLP+

Sample	**Genes in *CTLP* regions**	Number of Cancer Genes (COSMIC)*	**Chromosome**	**Genes** (cosmic)
SRR1701133	846	20	Chr1	455 (10)
			Chr10	391 (10)
SRR1701235	875	15	Chr1	296 (5)
			Chr6	579 (10)
SRR1701366	349	11	Chr3	349 (11)
SRR1701388	267	7	Chr7	267 (7)
SRR1701470	1064	19	Chr19	1064 (19)
SRR1701617	423	12	Chr2	77 (2)
			Chr9	346 (10)
SRR1701727	1170	30	chr3	370 (12)
			chr4	78 (2)
			chr5	160 (4)
			chr7	63 (0)
			chr12	267 (7)
			chr16	232 (5)

The COSMIC genes in CTLP regions, by chromosome and samples are available in figure 28.

chr01		chr02	Chr03		Chr04
SRR1701133	SRR1701235	SRR1701617	SRR1701366	SRR1701727	SRR1701727
ABL2 ARNT BCL9 ELK4 FCGR2B MDM4 PBX1 SDHC SLC45A3 TPR	ELK4 FH H3F3A MDM4 SLC45A3	C2orf44 MYCN CD274	BCL6 ETV5 FOXL2 GATA2 GMPS LPP MLF1 PIK3CA RPN1 SOX2 WWTR1	BCL6 EIF4A2 ETV5 FOXL2 GATA2 GMPS LPP MLF1 PIK3CA RPN1 SOX2 WWTR1	PHOX2B SLC34A2

Ch05	Ch06	Chr7		Chr9	Chr10
SRR1701727	SRR1701235	SRR1701388	SRR1701727	SRR1701617	SRR1701133
IL6ST IL7R LIFR PIK3R1	CCND3 DAXX DEK FANCE HIST1H4I HMGA1 PIM1 POU5F1 TFEB TRIM27	CARD11 EGFR ETV1 HNRNPA2B1 IKZF1 JAZF1 PMS2	No cosmic genes	FANCC FANCG GNAQ JAK2 MLLT3 NFIB OMD PAX5 XPA	BMPR1A CCDC6 NCOA4 NFKB2 PRF1 PTEN SUFU TCF7L2 TLX1 VTI1A

chr12	chr16	chr19
SRR1701727	SRR1701727	SRR1701470
CDK4 DDIT3 HOXC11 HOXC13 KRAS LRIG3 NACA	CBFB CDH1 CDH11 CYLD MAF	AKT2 BCL3 BRD4 CBLC CCNE1 CD79A CEBPA CIC ELL ERCC2 FSTL3 JAK3 KLK2 PPP2R1A STK11 TCF3 TFPT TPM4 ZNF331

Figure 28 – Frames with COSMIC genes present in the *Chromothripsis* regions by sample and chromosome.

We made the comparison between the samples with and without CTLPs (figure 29).

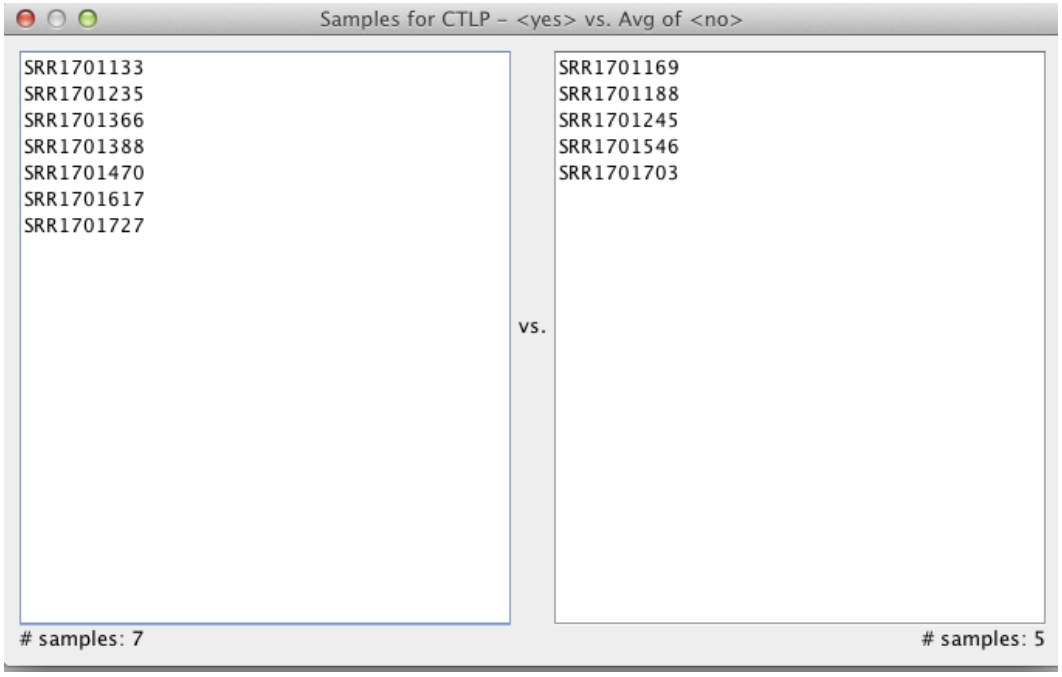

Figure 29 - Comparison between samples CTLP+ and CTLP-, by Nexus 9.0

A high level of copy number alterations between the groups is showed in figure 30, and by chromosomes in figure 31 and 32.

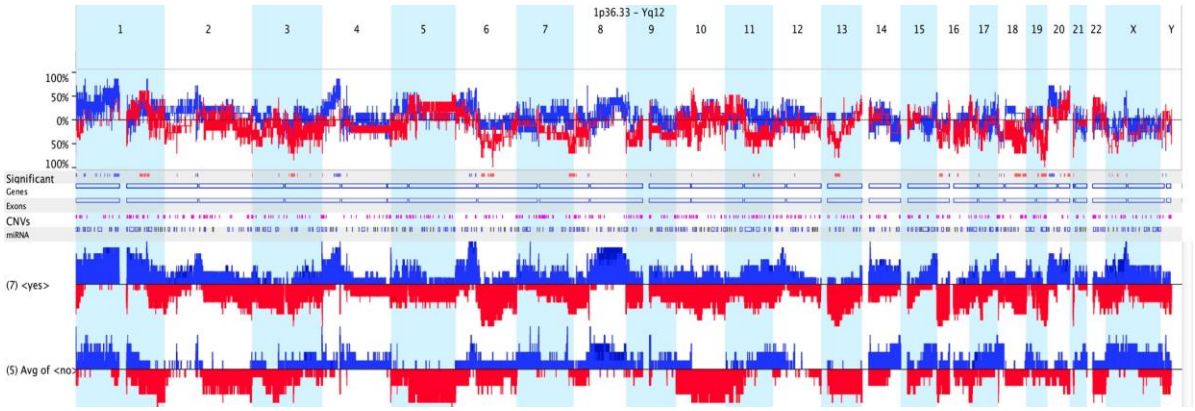

Figure 30 - Overview of the 12 OS samples showing the high rate of copy number changes between the WGS samples CTLP+ and CTLP-, by Nexus 9.0.

The table 18 reports the number of alterations in each sample by each chromosome in the WGS samples.

Table 18 – Report of genes found in CTLP regions of WGS samples from dbGap

Sample	SRR 1701133	SRR 1701235	SRR 1701366	SRR 1701388	SRR 1701470	SRR 1701617	SRR 1701727
Number genes	846	875	349	267	1064	423	1170
Number Cosmic genes	20	15	11	7	19	12	30
Other relevant genes	32	63	1	7	21	19	20
Number of chromosomes affected	2	2	1	1	1	2	6
Chromosome	**Number of genes in *Chromothripsis* regions**						
chr1	455	296					
chr2						77	
chr3			349				370
chr4							78
chr5							160
chr6		579					
chr7				267			63
chr9						346	
chr10	391						
chr12							267
chr16							232
chr19					1064		

Nearly 700 genes per tumor were found in the CTLPs regions. A total of 101 genes were located in regions of copy number change that distinguished the group of OS with *Chromothripsis* in comparison to OS without *Chromothripsis* (table 19). These genes are related with cellular process (45 genes – which 17 are associated with cell communication) and metabolic process (22 genes – which 19 are associated with primary metabolic process).

The samples had their rearrangements and chromosomes affected by chaotic events individualy listed and shown in the next tables (20 until 33) and figures (figure 31 until 38).

Table 19 - Comparation between CTLP(+) x CTLP(-) by different regions of copy number alterationss showing the event and the genes related

Region	Cytoband Location	Event	Number of CN alterations	Freq. in <yes> (%)	Freq. in Avg of <no> (%)	≠	Genes
chr1: 108,691,416-116,851,364	p13.1-p13.3	CN Gain	23	85.7	0	85.7	SLC25A24, AKNAD1, CLCC1, WDR47, GPR61, GNAI3, RBM15, SLC16A4, LAMTOR5, TMIGD3, RAP1A, ST7L, CAPZA1, MOV10, MAGI3, MAGI3, MAGI3, PHTF1, RSBN1, PTPN22, AP4B1-AS1, BCL2L15, HIPK1, TRIM33, BCAS2, DENND2C, AMPD1, SYCP1, CASQ2, LINC01649, LOC101928977, SLC22A15
chr1: 119,038,964-119,183,907	p12	CN Gain	2	85.7	0	85.7	No genes
chr1: 198,820,748-199,602,876	q32.1	CN Gain	3	85.7	0	85.7	MIR181B1, MIR181A1, MIR181A1HG
chr1: 25,190,083-25,215,783	p36.11	CN Gain	2	85.7	0	85.71	No genes
chr1: 87,265,175-87,271,412	p22.3	CN Gain	1	85.7	0	85.7	No genes
chr3: 116,951,916-116,974,404	q13.31	CN Loss	1	85.7	0	85.7	No genes
chr4: 39,119,379-47,288,341	p12-p14	CN Gain	7	85.7	0	85.7	KLHL5, ATP8A1, GABRG1, GABRB1
chr6: 103,090,396-103,737,894	q16.3	CN Loss	3	85.7	0	85.7	No genes
chr6: 73,247,796-74,459,196	q13	CN Loss	7	100	20	80	RIMS1, KCNQ5-IT1, KCNQ5, MIR4282, KCNQ5-AS1, KHDC1L, KHDC1, DPPA5, KHDC3L, OOEP, DDX43, CGAS, MTO1, SNORD141A, SNORD141B, EEF1A1, SLC17A5, LOC101928489, CD109
chr6: 77,310,122-77,327,347	q14.1	CN Loss	1	100	20	80	No genes
chr13: 38,868,582-38,874,421	q13.3	CN Loss	1	100	20	80	No genes
chr13: 48,107,616-49,986,699	q14.2	CN Loss	3	100	20	80	CAB39L
chr19: 42,077,895-50,517,090	q13.2 - q13.31	CN Loss	17	85.7	0	85.7	CEACAM21, CEACAM5, LIPE-AS1, PSG1, PSG6, PSG7, PSG7, CD177, STRN4, ARHGAP35, NPAS1, TMEM160, ZC3H4, CABP5, CARD8, CARD8-AS1, ZNF114, CCDC114, CGB3, SNAR-G2, CGB2, CGB1, SNAR-G1, CGB5, CGB8, CGB7, SLC6A16, MIR4324, FLT3LG, RPL13AP5, RPL13A, SNORD32A, SNORD33, SNORD34, SNORD35A, SNORD35B, RPS11, MIR150,

							PRR12, RRAS, SCAF1, IRF3, BCL2L12, MIR5088, PRMT1, ADM5, CPT1C, TBC1D17, MIR4750, IL4I1, NUP62, SIGLEC11, SIGLEC16, VRK3
chr19: 50,598,251-50,790,295	q13.33	CN Loss	3	100	20	80	SNAR-A3, SNAR-A4, SNAR-A5, SNAR-A6, SNAR-A7, SNAR-A8, SNAR-A9, SNAR-A10, SNAR-A11, SNAR-A14, SNAR-A3, SNAR-A4, SNAR-A5, SNAR-A6, SNAR-A7, SNAR-A8, SNAR-A9, SNAR-A10, SNAR-A11, SNAR-A14, SNAR-A3, SNAR-A4, SNAR-A5, SNAR-A6, SNAR-A7, SNAR-A8, SNAR-A9, SNAR-A10, SNAR-A11, SNAR-A14, MYH14
chr19: 51,567,361-53,894,646	q13.41-q13.42	CN Loss	7	85.7	0	85.7	KLK13, KLK14, CTU1, LOC101928517, SIGLEC17P, LOC101928517, MIR8074, CD33, SIGLECL1, LINC01872, CEACAM18, SIGLEC12, SIGLEC6, FPR2, FPR3, ZNF350-AS1, ZNF350, ZNF615, ZNF614, ZNF432, ZNF841, ZNF616, ZNF525
chr19: 51,679,687-51,702,002	q13.41	CN Loss	1	100	0	100	LOC101928517
chr21: 9,874,300-9,876,037	p11.2	CN Loss	1737	100	20	80	No genes

To acces all tables with the all genes in *Chromothripsis* region, access:

https://tinyurl.com/yby5xjsj

Table 20 - CTLPScanner results showing SRR1701133 sample, by chromosome regions and the CN status.

Array ID	CTLP Region (Mb)	Chromosome	Start	Stop	CNA status change times	Likelihood ratio (log10)
SRR1701133CTLP	81.2	1	145000001	226195210	57	41
SRR1701133CTLP	90.35	10	30000001	120354753	31	11

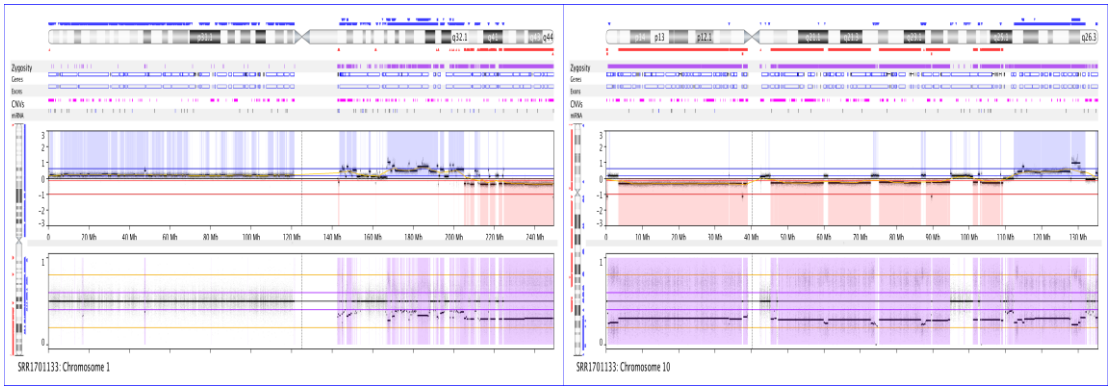

Figure 31- Chromosomes affected on sample SRR1701133 (Chr 1 and Chr 10).

Table 21 – List of genes in CTLPs regions of sample SRR1701133 with important Biological process.

Symbol	Name	Event	Chromosome	Other Biological process
BRINP2	bone morphogenetic protein/retinoic acid inducible neural-specific 2	Gain	chr1	Bone formation
BRINP3	bone morphogenetic protein/retinoic acid inducible neural-specific 3	Gain		
CD160	CD160 molecule	Gain		Immune response
CD1A	CD1a molecule	Gain		
CD1B	CD1b molecule	Gain		
CD1C	CD1c molecule	Gain		
CD1D	CD1d molecule	Gain		
CD1E	CD1e molecule	Gain		
CD247	CD247 molecule	Gain		
CD34	CD34 molecule	Loss		
CD55	CD55 molecule, decay accelerating factor for complement (Cromer blood group)	Loss		
CDC42SE1	CDC42 small effector 1	Gain		
CDC73	cell division cycle 73	Gain		cell cicle
CDK18	cyclin-dependent kinase 18	Loss		
CENPF	centromere protein F, 350/400kDa	Loss		CIN
CENPL	centromere protein L	Gain		
CEP350	centrosomal protein 350kDa	Gain		
BMPR1A	**bone morphogenetic protein receptor, type IA**	**Loss**	chr10	Bone formation
CCAR1	cell division cycle and apoptosis regulator 1	Loss		cell cicle
CCDC172	coiled-coil domain containing 172	Gain		
CCDC6	**coiled-coil domain containing 6**	**Loss**		
CCNY	cyclin Y	Loss		
CDK1	cyclin-dependent kinase 1	Loss		
CNNM1	cyclin and CBS domain divalent metal cation transport mediator 1	Loss		
CNNM2	cyclin and CBS domain divalent metal cation transport mediator 2	Loss		
DCLRE1A	DNA cross-link repair 1A	Gain		DNA repair
DDIT4	DNA-damage-inducible transcript 4	Gain		
DDX21	DEAD (Asp-Glu-Ala-Asp) box helicase 21	Loss		
DDX50	DEAD (Asp-Glu-Ala-Asp) box polypeptide 50	Loss		
DNA2	DNA replication helicase/nuclease 2	Loss		
ENTPD7	ectonucleoside triphosphate diphosphohydrolase 7	Loss		
ERCC6	excision repair cross-complementation group 6	Loss		

Table 22 - CTLPScanner results showing SRR1701235 sample, by chromosome regions and the CN status.

Array ID	CTLP Region (Mb)	Chromosome	Start	Stop	CNA status change times	Likelihood ratio (log10)
SRR1701235CTLP	48.13	1	200000001	248129895	38	27
SRR1701235CTLP	59.13	6	1	59128983	23	9

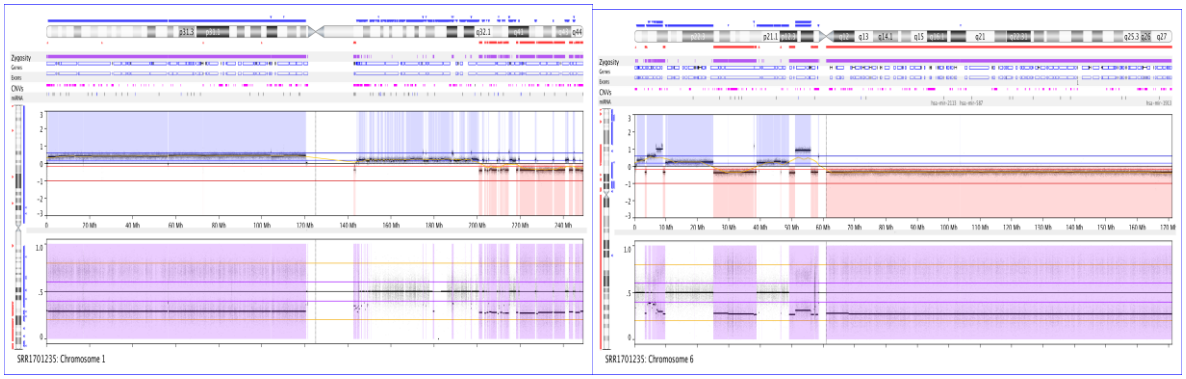

Figure 32- Chromosomes affected on sample SRR1701235 (Chr 1 and Chr 6).

Table 23 - List of genes in CTLPs of regions with important Biological process from SRR1701235.

Symbol	Name	Event	Chromosome	Biological Process
CCSAP	centriole, cilia and spindle-associated protein	Loss	chr1	CIN
CD34	CD34 molecule	Loss		Immune Response
CD46	CD46 molecule, complement regulatory protein	Loss		
CD55	CD55 molecule, decay accelerating factor for complement (Cromer blood group)	Loss		
CDK18	cyclin-dependent kinase 18	Gain		Cell cycle
CENPF	centromere protein F, 350/400kDa	Gain		CIN
CEP170	centrosomal protein 170kDa	Loss		
EXO1	exonuclease 1	Gain		DNA repair
FAIM3	Fas apoptotic inhibitory molecule 3	Loss		Immune Response
G0S2	G0/G1 switch 2	Loss		Cell cycle
IGFN1	immunoglobulin-like and fibronectin type III domain containing 1	Loss		Immune Response
IKBKE	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase epsilon	Loss		
IL10	interleukin 10	Loss		
IL19	interleukin 19	Loss		
IL20	interleukin 20	Loss		
IL24	interleukin 24	Loss		
INTS7	integrator complex subunit 7	Loss		

IPO9	importin 9	Loss		
IRF2BP2	interferon regulatory factor 2 binding protein 2	Loss		
IRF6	interferon regulatory factor 6	Gain		
MAP10	microtubule-associated protein 10	Loss		CIN
				Tumor supressor
MAP1LC3C	microtubule-associated protein 1 light chain 3 gamma	Gain		
TP53BP2	tumor protein p53 binding protein 2	Loss		
BMP5	bone morphogenetic protein 5	Gain	chr6	Bone formation
BMP6	bone morphogenetic protein 6	Gain		
CCND3	cyclin D3	Gain		Cell cycle
CD2AP	CD2-associated protein	Gain		
CD83	CD83 molecule	Gain		
CDC5L	cell division cycle 5-like	Gain		OS related
CDKAL1	CDK5 regulatory subunit associated protein 1-like 1	Gain		CIN
CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	Loss		
CENPQ	centromere protein Q	Loss		
ETV7	ets variant 7	Loss		Oncogenesis
HIVEP1	human immunodeficiency virus type I enhancer binding protein 1	Gain		Immune Response
HLA-A	major histocompatibility complex, class I, A	Loss		
HLA-B	major histocompatibility complex, class I, B	Loss		
HLA-C	major histocompatibility complex, class I, C	Loss		
HLA-DMA	major histocompatibility complex, class II, DM alpha	Loss		
HLA-DMB	major histocompatibility complex, class II, DM beta	Loss		
HLA-DOA	major histocompatibility complex, class II, DO alpha	Loss		
HLA-DOB	major histocompatibility complex, class II, DO beta	Loss		
HLA-DPA1	major histocompatibility complex, class II, DP alpha 1	Loss		
HLA-DPB1	major histocompatibility complex, class II, DP beta 1	Loss		
HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	Loss		
HLA-DQA2	major histocompatibility complex, class II, DQ alpha 2	Loss		
HLA-DQB1	major histocompatibility complex, class II, DQ beta 1	Loss		
HLA-DQB2	major histocompatibility complex, class II, DQ beta 2	Loss		
HLA-DRA	major histocompatibility complex, class II, DR alpha	Loss		
HLA-DRB1	major histocompatibility complex, class II, DR beta 1	Loss		
HLA-DRB5	major histocompatibility complex, class II, DR beta 5	Loss		
HLA-E	major histocompatibility complex, class I, E	Loss		
HLA-F	major histocompatibility complex, class I, F	Loss		

HLA-G	major histocompatibility complex, class I, G	Loss		
LTA	lymphotoxin alpha	Loss		
LTB	lymphotoxin beta (TNF superfamily, member 3)	Loss		
LY6G5B	lymphocyte antigen 6 complex, locus G5B	Loss		
LY6G5C	lymphocyte antigen 6 complex, locus G5C	Loss		
LY6G6C	lymphocyte antigen 6 complex, locus G6C	Loss		
LY6G6D	lymphocyte antigen 6 complex, locus G6D	Loss		
LY6G6F	lymphocyte antigen 6 complex, locus G6F	Loss		
LY86	lymphocyte antigen 86	Gain		
RUNX2	runt-related transcription factor 2	Gain		OS related
WRNIP1	Werner helicase interacting protein 1	Gain		DNA repair

Table 24 - CTLPScanner results showing SRR1701366 sample, by chromosome regions and the CN status.

Array ID	CTLP Region (Mb)	Chromosome	Start	Stop	CNA status change times	Likelihood ratio (log10)
SRR1701366CTPL	81.2	3	115000001	196195210	61	40

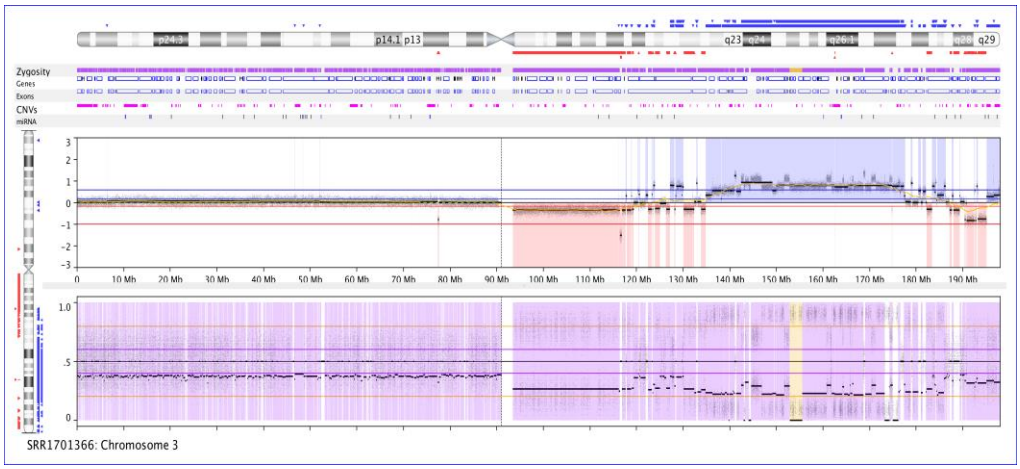

Figure 33- Chomosome 3 affected on sample SRR1701366

Table 25- List of gene in CTLPs regions with important Biological process in sample SRR1701366.

Symbol	Name	Event	Chromosome	Biological Process
HLTF	helicase-like transcription factor	Gain	Chr 3	DNA Repair

Table 26 - CTLPScanner results showing SRR17011388 sample, by chromosome regions and the CN status.

Array ID	CTLP Region (Mb)	Chromosome	Start	Stop	CNA status change times	Likelihood ratio (log10)
SRR1701388CTLP	59.13	7	1	59128983	21	20

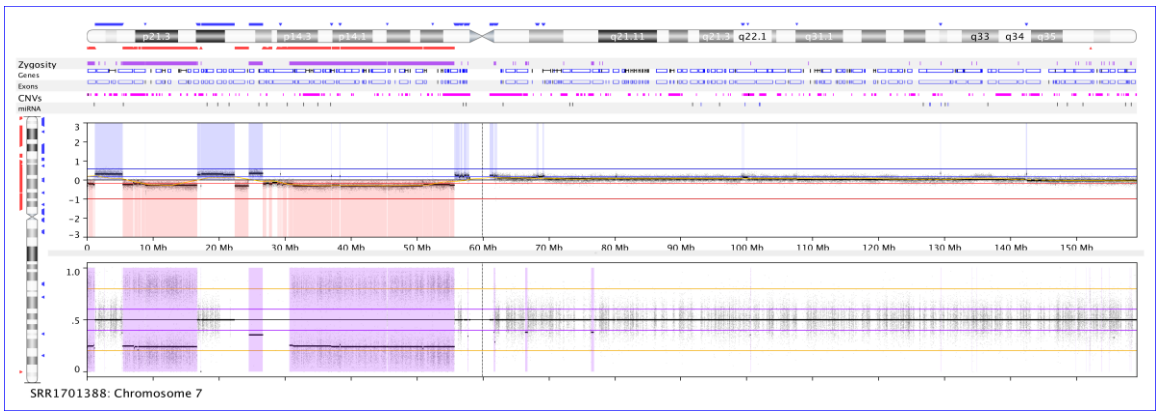

Fig 34- Chomosomes affected on sample SRR1701388.

Table 27 – List of genes in CTLPs regions with important biological process.

Symbol	Name	Event	chromosome	Biological Process
BMPER	BMP binding endothelial regulator	Loss	chr7	Bone formation
BRAT1	BRCA1-associated ATM activator 1	Gain		DNA repair
CDCA7L	cell division cycle associated 7-like	Gain		Cell cycle
CDK13	cyclin-dependent kinase 13	Loss		
DDX56	DEAD (Asp-Glu-Ala-Asp) box helicase 56	Loss		DNA repair
GBAS	glioblastoma amplified sequence	Gain		Cancer related
MEOX2	mesenchyme homeobox 2	Loss		Bone formation

Table 28- CTLPScanner results showing sample SRR1701470, by chromosome regions and the CN status.

Array ID	CTLP Region (Mb)	Chromosome	Start	Stop	CNA status change times	Likelihood ratio (log10)
SRR1701470CTLP	59.13	19	1	59128983	32	14

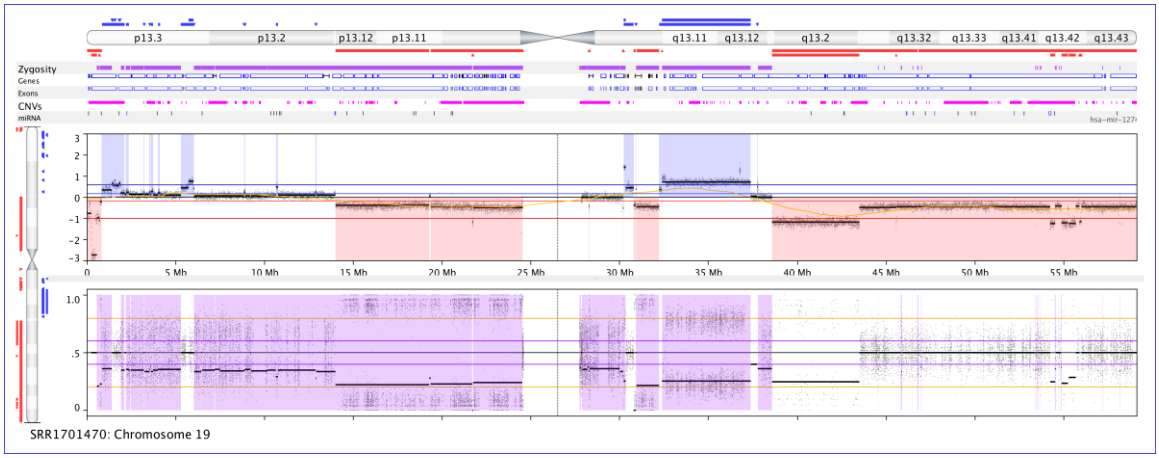

Figure 35 – Chomosome 19 affected on sample SRR11701470.

Table 29-– List of genes in CTLPs regions with important Biological process in sample SRR1101470.

Symbol	**Name**	**Event**	**chromosome**	**Biological Process**
CD97	CD97 molecule	Loss	chr19	Cell cycle
CDC34	cell division cycle 34	Loss		
CDC42EP5	CDC42 effector protein (Rho GTPase binding) 5	Loss		
CEP89	centrosomal protein 89kDa	Gain		CIN
CILP2	cartilage intermediate layer protein 2	Loss		Bone
COMP	cartilage oligomeric matrix protein	Loss		
DDX39A	DEAD (Asp-Glu-Ala-Asp) box polypeptide 39A	Loss		Repair
DDX49	DEAD (Asp-Glu-Ala-Asp) box polypeptide 49	Loss		
ERCC1	excision repair cross-complementation group 1	Loss		
ERCC2	excision repair cross-complementation group 2	Loss		
FOSB	FBJ murine osteosarcoma viral oncogene homolog B	Loss		OS related
GLTSCR1	glioma tumor suppressor candidate region gene 1	Loss		Tumor supressor
GLTSCR2	glioma tumor suppressor candidate region gene 2	Loss		
JUNB	jun B proto-oncogene	Gain		Oncogene
JUND	jun D proto-oncogene	Loss		
MIER2	mesoderm induction early response 1, family member 2	Loss		Mesoderm
OSCAR	osteoclast associated, immunoglobulin-like receptor	Loss		Bone
POLD1	polymerase (DNA directed), delta 1, catalytic	Loss		Repair

	subunit			
POLR2E	polymerase (RNA) II (DNA directed) polypeptide E, 25kDa	Gain		
POLR2I	polymerase (RNA) II (DNA directed) polypeptide I, 14.5kDa	Gain		
POLRMT	polymerase (RNA) mitochondrial (DNA directed)	Loss		

Table 30 - CTLPScanner results showing SRR1701 sample, by chromosome regions and the CN status.

Array ID	CTLP Region (Mb)	Chromosome	Start	Stop	CNA status change times	Likelihood ratio (log10)
SRR1701617CTLP	30	2	1	30000000	21	10
SRR1701617CTLP	102.53	9	1	102531392	81	40

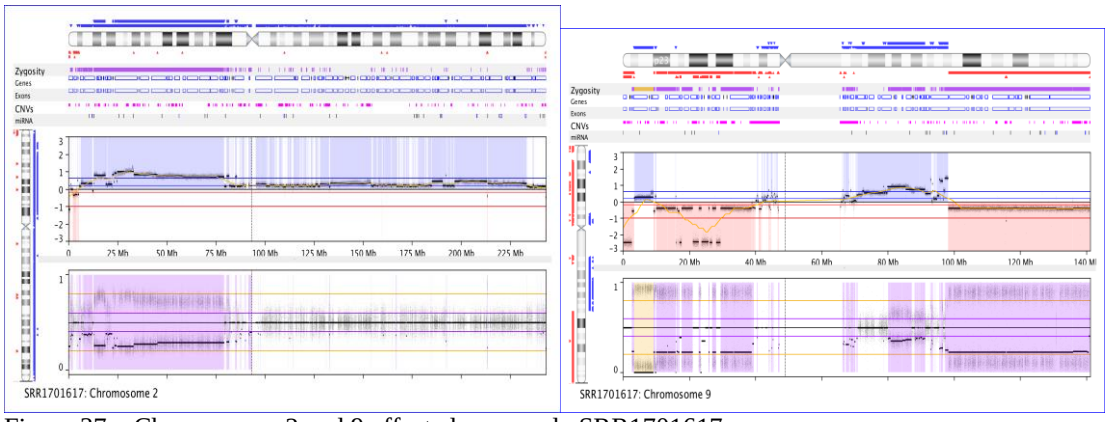

Figure 37 – Chomosomes 2 and 9 affected on sample SRR1701617.

Table 31- – List of genes in CTLPs regions with important Biological process – sample SRR1701617.

Symbol	Name	Event	Chromosome	Biological Process
DDX1	DEAD (Asp-Glu-Ala-Asp) box helicase 1	Gain	chr2	DNA repair
CCL19	chemokine (C-C motif) ligand 19	Loss	chr9	Immune response
CCL21	chemokine (C-C motif) ligand 21	Loss	chr9	
CCL27	chemokine (C-C motif) ligand 27	Loss	chr9	
CD274	CD274 molecule	Gain	chr9	
CD72	CD72 molecule	Loss	chr9	
CDC14B	cell division cycle 14B	Loss	chr9	Cell Cycle
CDC37L1	cell division cycle 37-like 1	Gain	chr9	

CDK20	cyclin-dependent kinase 20	Gain	chr9	
CDKN2A	cyclin-dependent kinase inhibitor 2A	Loss	chr9	
CDKN2B	cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)	Loss	chr9	
CENPP	centromere protein P	Gain	chr9	CIN
CEP78	centrosomal protein 78kDa	Gain	chr9	
CER1	cerberus 1, DAN family BMP antagonist	Loss	chr9	Bone formation
DDX58	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58	Loss	chr9	DNA repair
OGN	osteoglycin	Gain	chr9	Bone formation
OMD	**osteomodulin**	**Gain**	chr9	
OSTF1	osteoclast stimulating factor 1	Gain	chr9	
RMI1	RecQ mediated genome instability 1	Gain	chr9	CIN

Table 32 - CTLPScanner results showing SRR1701727 sample, by chromosome regions and the CN status.

Array ID	CTLP Region (Mb)	Chromosome	Start	Stop	CNA status change times	Likelihood ratio (log10)
SRR1701727CTPL	81.2	3	110000001	191195210	65	30
SRR1701727CTPL	30	4	20000001	50000000	29	17
SRR1701727CTPL	51.3	5	30000001	81304566	42	20
SRR1701727CTPL	30	7	105000001	135000000	21	10
SRR1701727CTPL	48.13	12	15000001	63129895	39	18
SRR1701727CTPL	63.03	16	25000001	88025520	38	13

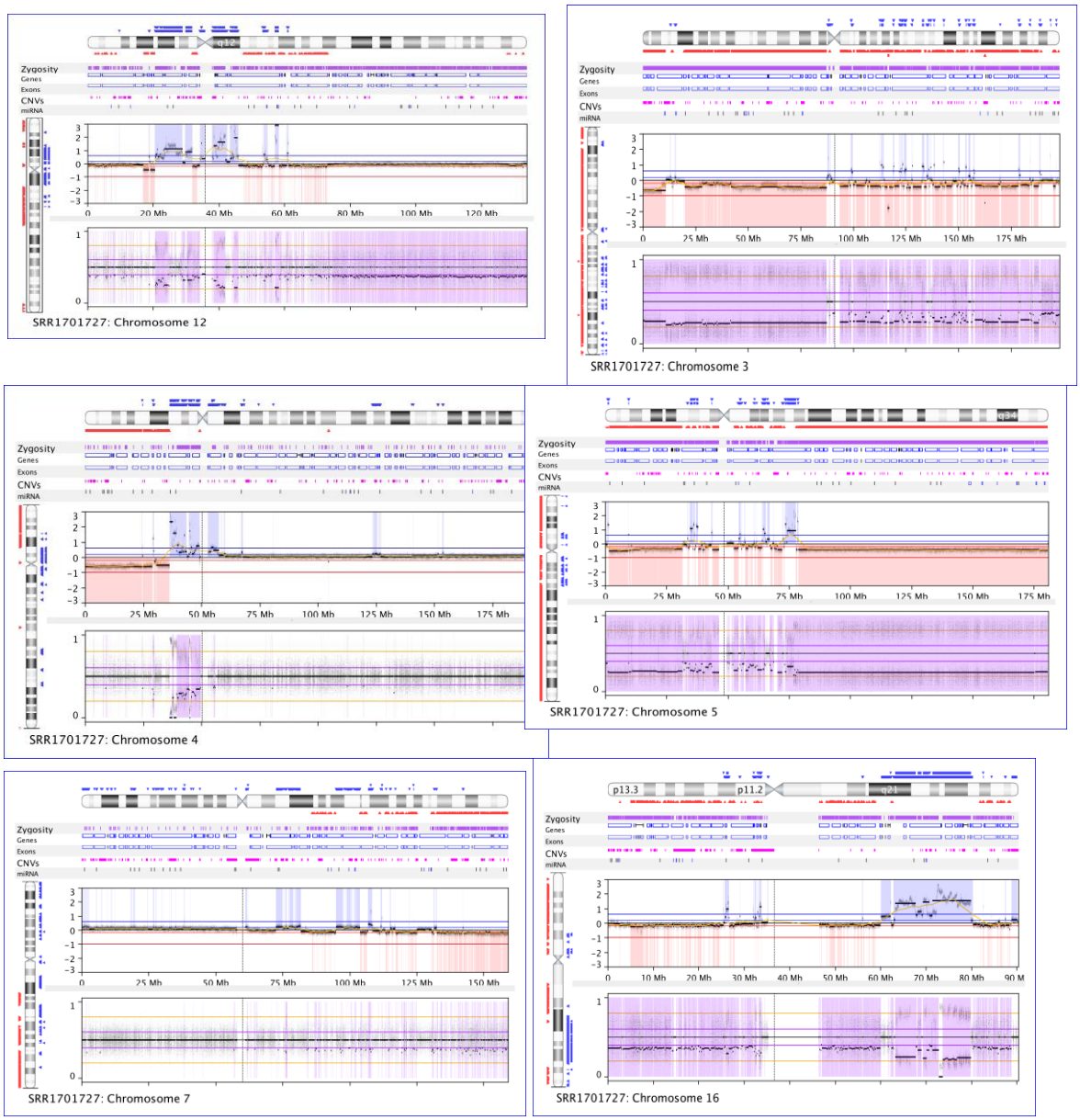

Figure 38 - Chomosomes 3,4,5,7, 12 and 16 affected on sample SRR1701727 (indicative of *Chromoplexy*).

Table 33 - List of genes in CTLPs regions with important Biological process in SRR1701727.

Symbol	Name	Event	chromosome	Biological Process
ANAPC13	anaphase promoting complex subunit 13	Loss	chr3	CIN
CD200R1	CD200 receptor 1	Loss	chr3	Immune response
CD200R1L	CD200 receptor 1-like	Loss	chr3	
CD80	CD80 molecule	Gain	chr3	
CD86	CD86 molecule	Loss	chr3	
CEP63	centrosomal protein 63kDa	Loss	chr3	CIN
CEP70	centrosomal protein 70kDa	Loss	chr3	
FAIM	Fas apoptotic inhibitory molecule	Gain	chr3	Immune response

POT1	protection of telomeres 1	Loss	chr7	CIN
GLI1	GLI family zinc finger 1	Gain	chr12	p53 related
OS9	osteosarcoma amplified 9, endoplasmic reticulum lectin	Gain	chr12	OS related
RECQL	RecQ helicase-like	Gain	chr12	DNA repair
CENPT	centromere protein T	Gain	chr16	CIN
DDX19A	DEAD (Asp-Glu-Ala-Asp) box polypeptide 19A	Gain	chr16	DNA repair
DDX19B	DEAD (Asp-Glu-Ala-Asp) box polypeptide 19B	Gain	chr16	
DDX28	DEAD (Asp-Glu-Ala-Asp) box polypeptide 28	Gain	chr16	
TP53TG3	TP53 target 3	Loss	chr16	p53 related
TP53TG3B	TP53 target 3B	Loss	chr16	
TP53TG3C	TP53 target 3C	Loss	chr16	
TP53TG3D	TP53 target 3D	Loss	chr16	

RNA-seq

We were also able to compare the RNA levels from the dbGap samples when expression data was available: comparing 6 OS RNA samples with *Chromothripsis* (CTLP+) to 3 OS RNA samples without *Chromothripsis* (CTLP-) (tables 34 and 35).

Table 34– List of RNA samples analyzed from dbGap.

Sample	Experiment	Gender	Outcome	Sample	CTLP
SRR1701102	BZ06-Tumor	Male	Deceased	RNA	yes
SRR1701137	BZ36-Tumor	Male	Deceased	RNA	yes
SRR1701291	BZ17-Tumor	Female	Living	RNA	yes
SRR1701307	BZ15-Tumor	Female	Deceased	RNA	yes
SRR1701369	BZ32-Tumor	Female	Deceased	RNA	no
SRR1701556	BZ23-Tumor	Female	Deceased	RNA	no
SRR1701662	BZ10-Tumor	Male	Living	RNA	no
SRR1701777	BZ30-Tumor	Male	Living	RNA	yes
SRR1701796	BZ11-Tumor	Male	Deceased	RNA	yes

Table 35 – List of samples CTPL+ to compare with samples CTPL-.

To Compare	
Group positive	Group negative
SRR1701796	SRR1701662
SRR1701307	SRR1701556
SRR1701102	SRR1701369
SRR1701291	
SRR1701777	
SRR1701137	

Both the EdgeR and Nexus Expression pipelines showed downregulation in cell communication pathway and primary metabolic process in samples with *Chromothripsis*.

Pipeline: nfcore/rnaseq (edgeR script)

Comparation CTLP(yes) X CTLP(no)

edgeR: FDR 0.05, log2FC > 2, pvalue<0.01

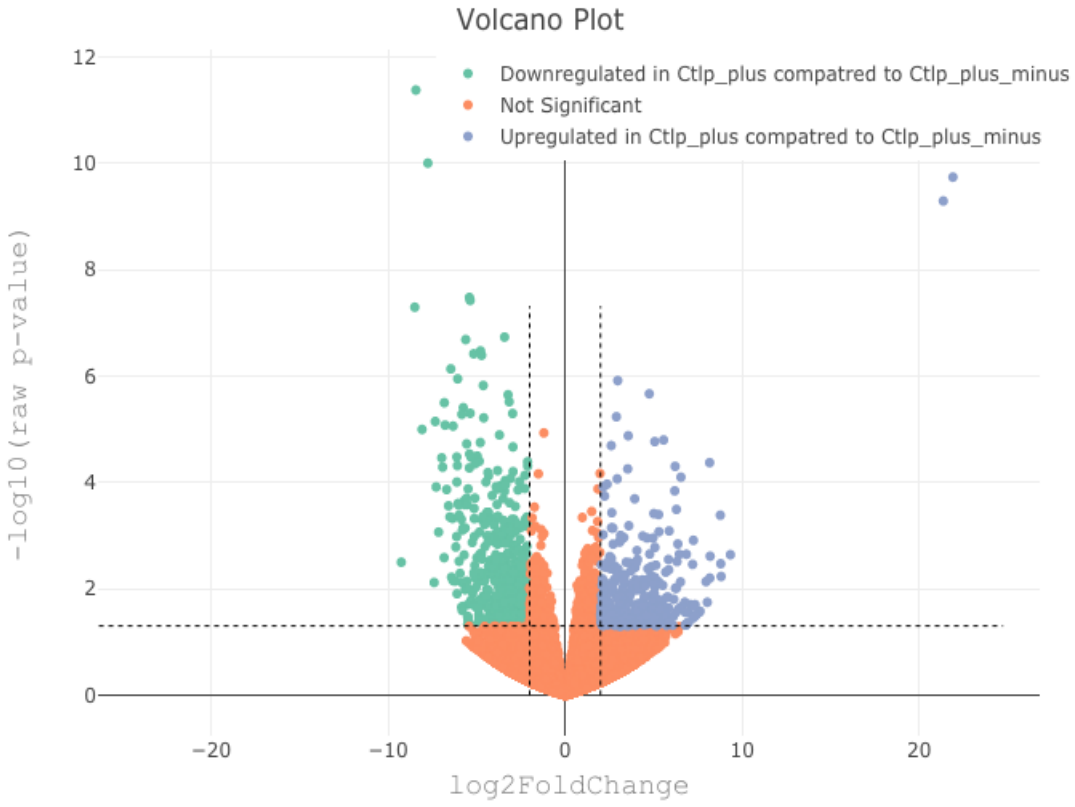

Figure 39 – Volcano Plot of expression data from dbGap samples (RNA-seq) - EdgeR

Table 36 - Genes downregulated in CTLP+ samples.

67 genes downregulated		
gene_symbol		
COMP	MOXD1	LINC00925

IGHA1	C2orf40	NRCAM
CRTAC1	CCL25	POU2AF1
SCUBE1	HEPH	LMTK3
TBX5	MRGPRF	ITGB4
MPO	GRIK5	MEIS3
PAK3	OR2A1-AS1	IRF4
COL8A1	IGJ	IGKV1-5
ACAN	DUSP26	VEPH1
RP11-773H22.2	HCN1	TAGLN
DGKK	AC012360.6	SNED1
IGHA2	MIA	NKX2-5
SERPINE1	C1QTNF3	CHADL
ARHGEF34P	COL26A1	PLA2G5
GOLGA8M	CD109	PEAR1
RP4-800G7.2	RP11-343B18.2	IGLV3-19
RSPO2	ALPK3	LIPG
OLFM1	TMEM132C	AMIGO2
SORCS2	GALNT9	KNDC1
BGLAP	IGHM	APOB
TTC9	CAMSAP3	AC104809.4
ZFP42	MYO5C	PHKA1
		FCRL5

Table 37 – Genes upregulated inn CTLP+ Samples.

20 genes Upregulated
gene_symbol
PSPHP1
RP11-81H3.2
POU3F4
UBE2C
PHACTR3
TUBB2B
TFAP2A
PDE6A
TDO2
SPC24
RP4-792G4.2
AP000525.10
RBM20
PADI3
C5orf38
IRX2
AURKB
BMS1P17
TTLL7
HIST1H1B

Table 38 - Genes evaluated in (Davoli et al., 2017) publication that have different expression between the groups

Gene	Expression profile	Biological process
AMIGO2	Downregulated	CD4.mature
FLT3	Downregulated	DIFERENTIATION AND DEVELOPMENT
MFGE8	Downregulated	METABOLISM
CD79A	Downregulated	**B.cells**
POU6F2	Upregulated	Immune system in melanoma samples

The table 38 shows the genes with different expression between the groups compared (CTLP+ x CTLP-) also found in Davoli et al (2017) publication. This publication shown that highly aneuploid tumors have reduced expression of markers of cytotoxic infiltrating immune cells, and increased expression of cell proliferation markers.

Genes downregulated of immune system response pathway were found in both pipeline (*COL8A1*, *CCL25*) (figure 40, 41 and 42).

Genes downregulated – Immune System

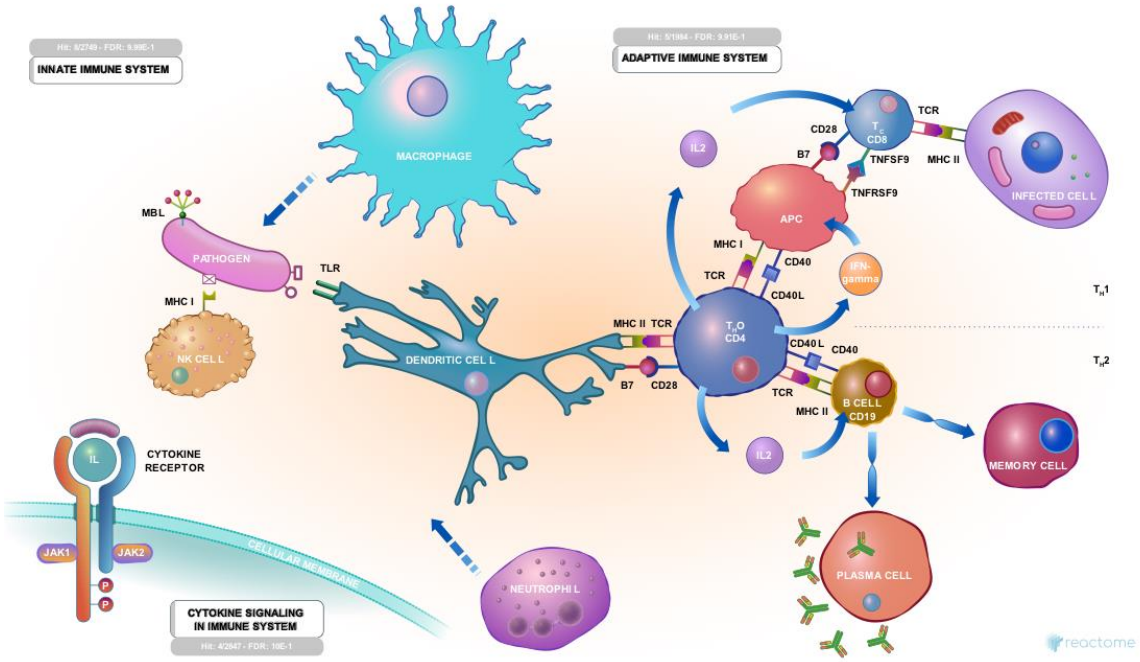

Figure 40 – Immune system pathway affected by underexpressed genes in CTLP+ samples: 8 genes from innate immune pathway; 5 genes of adaptive immune system; and 4 genes in cytokine signaling pathway (by Reactome).

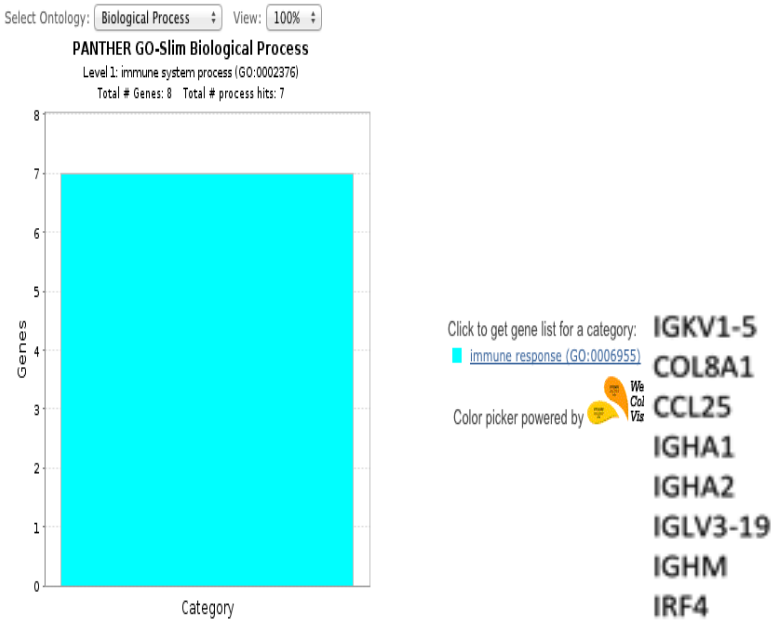

Figure 41 – Immune system pathway affected by underexpressed genes in CTLP+ samples (Nexus Expression): 8 genes from immune response (by panther).

Immune System

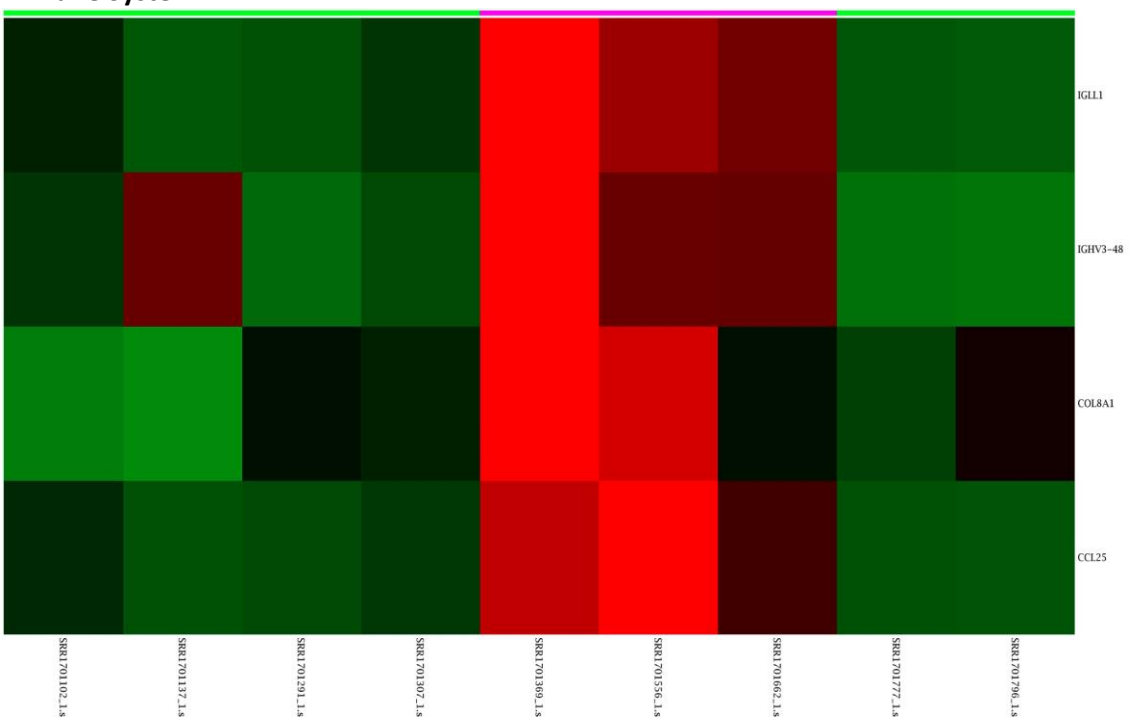

Figure 42 – Heatmap of expression genes related with Immune system, by Nexus.

The GSEA analysis used a p-value <0.06, and the gene sets database: gseaftp.broadinstitute.org://pub/gsea/gene_sets_final/c2.cp.reactome.v6.1.symbols.gmt. shown positive correlation between the different expression CTLP+ x CTLP- in the pathways of RNA pol. III and G2 Checkpoints (figure 43).

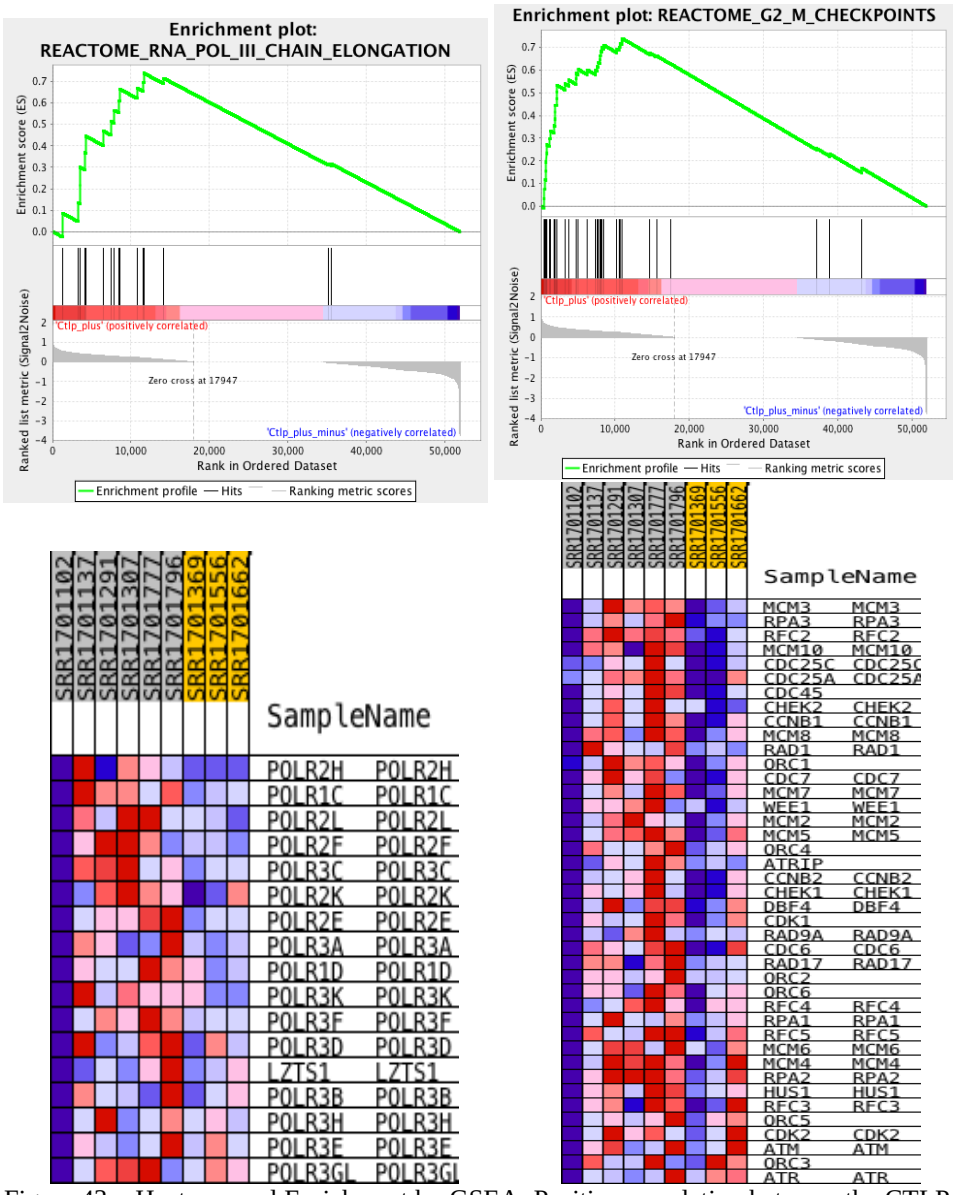

Figure 43 – Heatmap and Enrichment by GSEA: Positive correlation between the CTLP+ expression and CTLP- in the pathways of RNA pol. III and G2 Checkpoints.

6.6 - Sandwich Results:

To study the chromosomes involved in micronucleus formation in the OS cell line U2OS, errors in cell division induced by drugs during the anaphase were evaluated during the sandwich period at Barts Cancer Institute in London-UK.

We used the U2-OS cell line to check the missegregation errors using drugs like Nocodazole and Aphidicolin to stress the cells, and made conventional FISH to analyze of chromosome-specific levels of structural aberrations. We used microscope analysis to determine which FISH probes in combination, and the live cell imaging movies to evaluate missegregation and the micronucleus formation. Figure 57 shows the number of anaphases found (with and withour errors), according to the treatment.

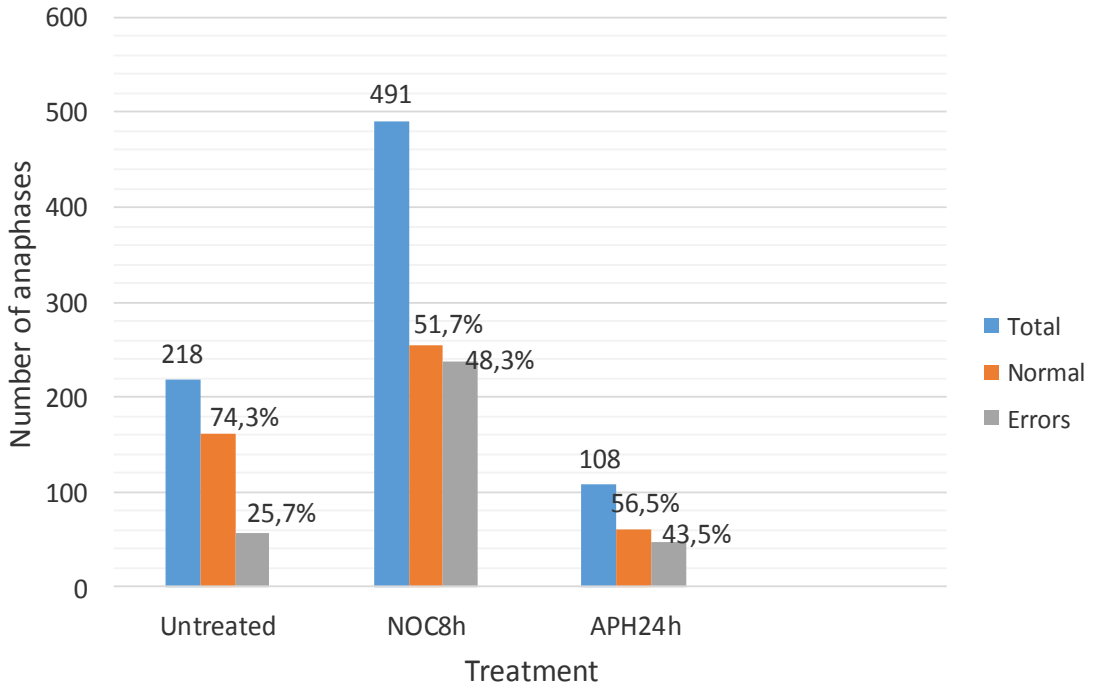

Figure 44 – Number of anaphases found: with and without errors during the mitosis according with the treatment.

Nocodozole is a microtubule depolymerising agent in cycling cells nocodazole causes mitotic arrest after drug washout cells resume mitosis but mis-segregate chromosome. Nocodozle depolymerises microtubules, leading to disassembly of the mitotic spindle. Upon drug washout, spindles reassemble in an error-prone manner, leading to improper chromosome-spindle attachments and chromosome non-disjunction. In addition to mitosis defects, aneuploidy can arise following the generation of chromosomes that are structurally abnormal due to defective DNA replication or repair (Siegel and Amon, 2012). The figure 58

shows one error anaphase using FISH, with lagging chromosome 12 and a bridge. Cells were treated with nocodozole.

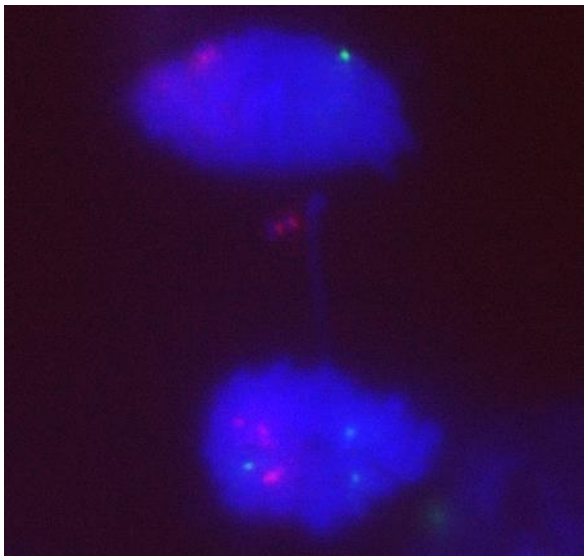

Figure 45 – Example of FISH technique in slide showing one U2-OS lagging anaphase. The image shows two lagging chromosomes in red (chr12) and a bridge. Threatment of Nocodozol 8h.

Aphidicolin is a potent and specific inhibitor of B-family DNA polymerases, halting replication. Using low doses induced replication stress but allows continued progress into the next mitosis with chromosome segregation errors (Baranovskiy et al., 2014).

We found, using Immunofluorescence (IF), different rate of error anaphases. This can be justified because there are more telophases cells discarded, and is very difficult to check this using FISH. The figure 60 shows the tubulines and how it is easier to differenciate anaphase of early telophase.

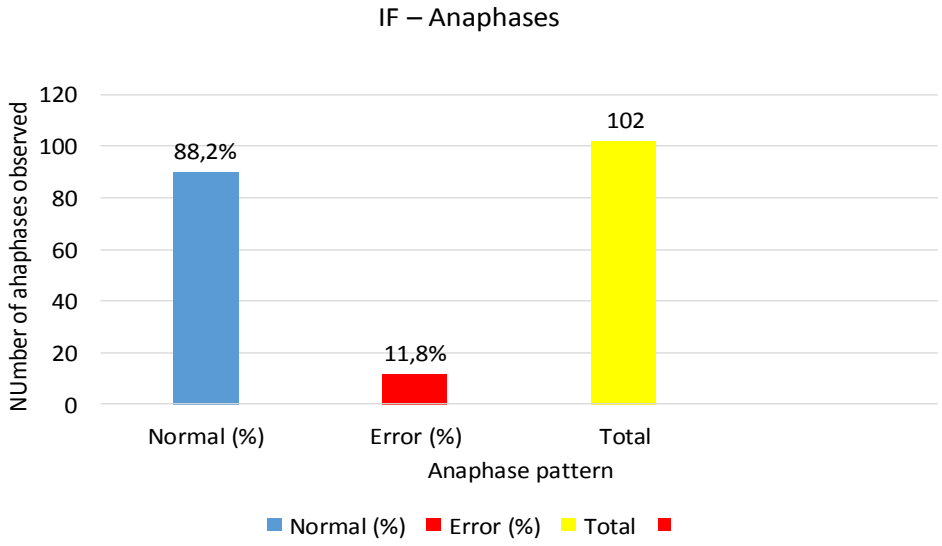

Figure 46 – Number of anaphases found: with and without errors during the mitosis by Immnunofluorescense.

CREST- far red
CENPA – green
BETATUBUL – red
DAPI- Blue

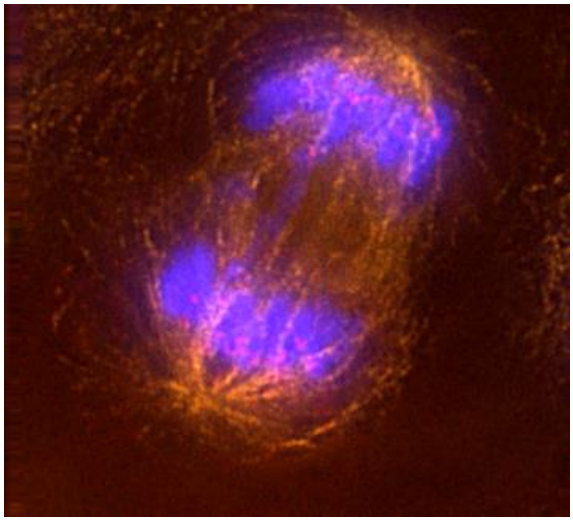

Figure 47: U2-0S cell untreated by Imunofluorescence

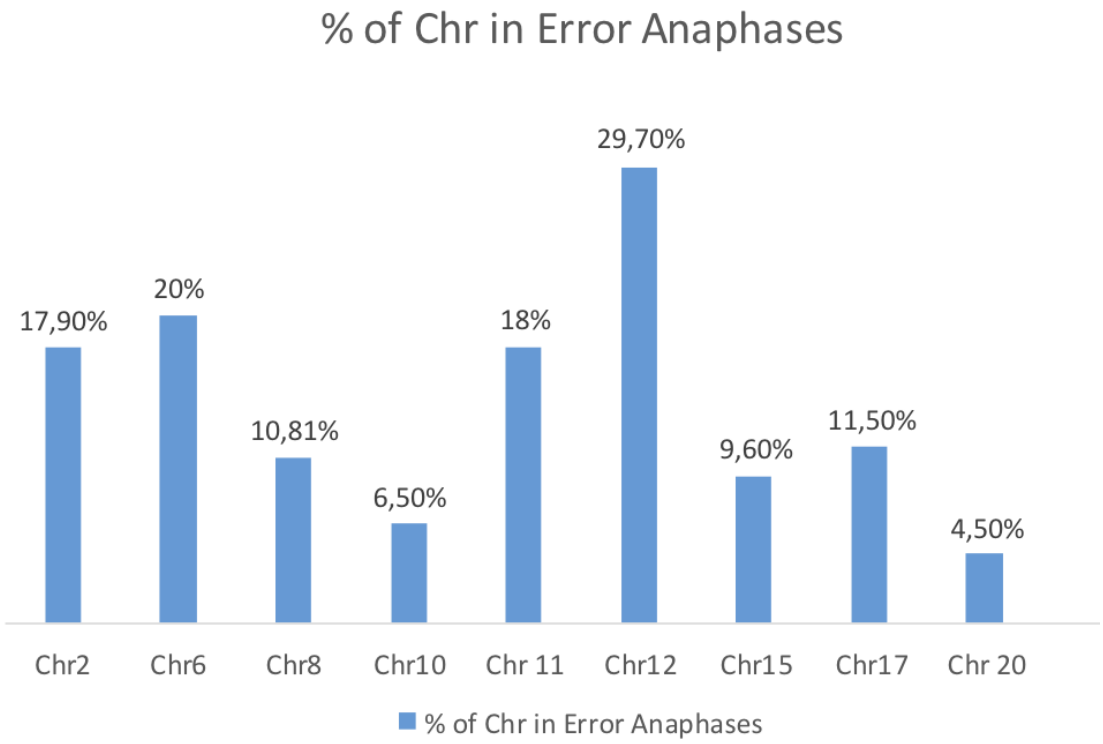

Figure 48 – Number of chromosomes found in error anaphases per slides counted (mean of 50 anaphases counted by probe) – Nocodozol treatment.

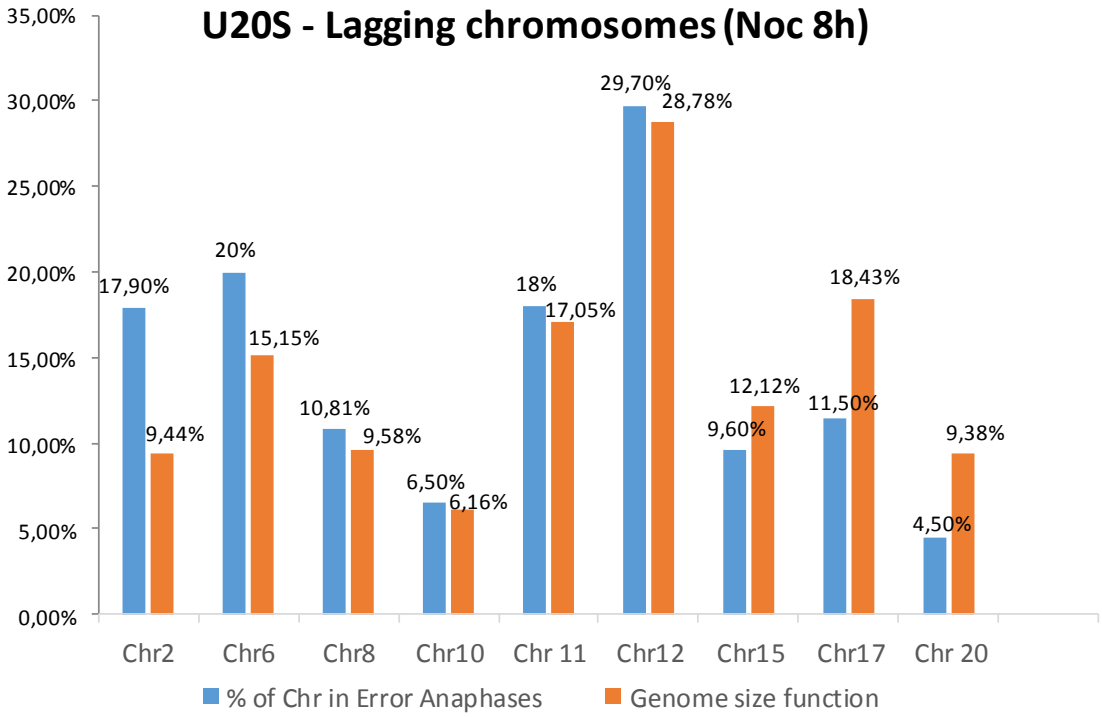

Figure 49– Number of chromosomes found in error anaphases with the correction of the genome size function.

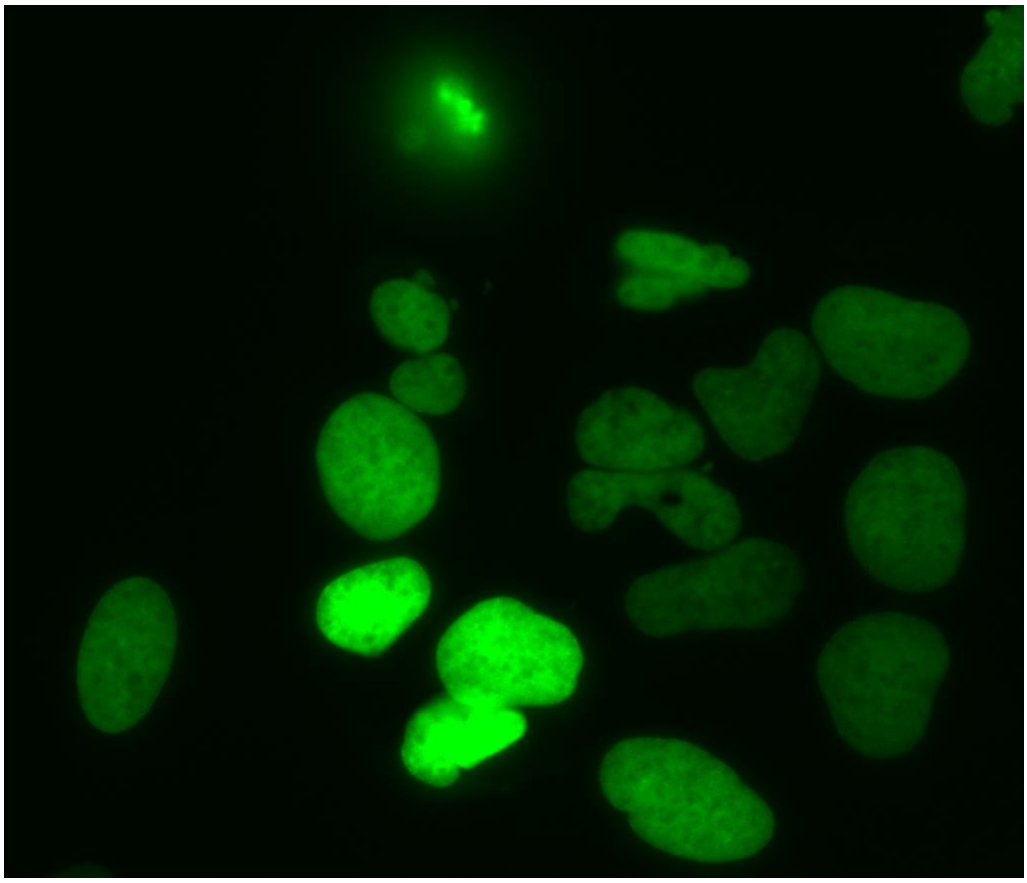

Figure 50 – Picture from a live cell movie. We could see micronuclei and cells with strange format. This video is available in supplementary data folder <https://tinyurl.com/yby5xjsj>.

The lagging chromosomes were counted and the most common chromosomes with errors were Chr2, Chr6, Chr11, and Chr12. These data provide further support to the idea that some chromosomes are more susceptible to cell division errors (Worrall *et al.*, 2018) and corroborate with the chromosomes affected by CTPLs in some tumors.

Conclusion

7.0 - CONCLUSION

This study demonstrates that the summary of publications shows the rate of chaotic rearrangements in OS samples of 44% (with a high variation).

The OS microarrays samples from our group revealed 3 OS samples (30%) with CTLP+. Chromosomes 2, 10, 14 and 20 were related with CTLP.

Moreover, we found 17 different chromosomes with CTLPs by microarrays from public databanks. Only chromosomes 7, 18, 19, 21 and the sex chromosomes were not affect by CTLPs. Chromosomes 2, 8 and 12 were frequent target with *Chromothripsis*. Chromosomes 1, 3, 4, 5, 6, 10, 11 and 13 were also related with CTLPs.

We found CTLPs in 58% of the 12 OS samples analyzed using whole genome sequence data. Chromosomes 1, 2, 3, 7 and 12 were slightly more often. Genes located in regions of copy number change that distinguished the groups of OS (with and without CTLPs) are connected with cellular process and metabolic process.

U2-OS cell line treated to produce anaphases with errors present lagging chromosomes, commonly: chromosomes 2, 6, 11, and 12.

Gene *TP53* is related not just with OS but also with *Chromothripsis* in OS, especially when TP53 mutations are present (Gröbner *et al.*, 2018). We found complex rearrangements in tumor suppressor genes (*PTEN* and *FAS*) and in oncogenes amplification (*MYC*). All three samples related to CTLPs (GSE12830) had *MYC* amplification. These genes may accelerate the evolution and progression of OS tumors according to the literature.

We found genes associated with the immune system that were underexpressed *(CADM1; CLEC4A; CCR1; CD164)* in *Chromothripsis* positive tumors in arrays samples, and (*COL8A1, CCL25*) in WGS samples. Genes *CALCA, CARTPT* and *ADRB2* were upregulated and can controls bone remodeling such as *Chromothripsis* can influence the genesis and progression of bone diseases.

In conclusion, our study corroborated with the idea that the complex genomic rearrangements are an integral part of mutation mechanisms contributing to cancer development, affecting different and important pathways and should be considered as a critical process in bone oncogenesis, like in OS.

.

References

REFERENCES

AL-ROMAIH, K. et al. Chromosomal instability in osteosarcoma and its association with centrosome abnormalities. **Cancer Genet Cytogenet,** v. 144, n. 2, p. 91-9, Jul 15 2003. ISSN 0165-4608 (Print)0165-4608. Available at: < http://dx.doi.org/ >.

ANNINGA, J. K. et al. Chemotherapeutic adjuvant treatment for osteosarcoma: where do we stand? **Eur J Cancer,** v. 47, n. 16, p. 2431-45, Nov 2011. ISSN 0959-8049. Available at: < http://dx.doi.org/10.1016/j.ejca.2011.05.030 >.

ATCC. U-2 OS ATCC HTB-96TM: Homo sapiens bone osteosarcoma. 2018. Avaialble at: < https://www.atcc.org/Products/All/HTB-96.aspx#generalinformation >.

BACA, S. C. et al. Punctuated evolution of prostate cancer genomes. **Cell,** v. 153, n. 3, p. 666-77, Apr 25 2013. ISSN 0092-8674. Available at: < http://dx.doi.org/10.1016/j.cell.2013.03.021 >.

BARANOVSKIY, A. G. et al. Structural basis for inhibition of DNA replication by aphidicolin. **Nucleic Acids Res,** v. 42, n. 22, p. 14013-21, Dec 16 2014. ISSN 0305-1048. Available at: < http://dx.doi.org/10.1093/nar/gku1209 >.

BAYANI, J. et al. Genomic mechanisms and measurement of structural and numerical instability in cancer cells. **Semin Cancer Biol,** v. 17, n. 1, p. 5-18, Feb 2007. ISSN 1044-579X (Print)1044-579x. Available at: < http://dx.doi.org/10.1016/j.semcancer.2006.10.006 >.

_____. Spectral karyotyping identifies recurrent complex rearrangements of chromosomes 8, 17, and 20 in osteosarcomas. **Genes Chromosomes Cancer,** v. 36, n. 1, p. 7-16, Jan 2003. ISSN 1045-2257 (Print)1045-2257. Available at: < http://dx.doi.org/10.1002/gcc.10132 >.

BEHJATI, S. et al. Recurrent mutation of IGF signalling genes and distinct patterns of genomic rearrangement in osteosarcoma. **Nat Commun,** v. 8, p. 15936, Jun 23 2017. ISSN 2041-1723. Available at: < http://dx.doi.org/10.1038/ncomms15936 >.

BENDAS, G.; BORSIG, L. Cancer cell adhesion and metastasis: selectins, integrins, and the inhibitory potential of heparins. **Int J Cell Biol,** v. 2012, p. 676731, 2012. ISSN 1687-8876. Available at: < http://dx.doi.org/10.1155/2012/676731 >.

BIODISCOVERY. **Nexus Copy Number 9.0 Users Manual**. Nexus software - USA. 1**:** 544 p. 2017.

BIRKBAK, N. J. et al. Paradoxical relationship between chromosomal instability and survival outcome in cancer. **Cancer Res,** v. 71, n. 10, p. 3447-52, May 15 2011. ISSN 0008-5472. Available at: < http://dx.doi.org/10.1158/0008-5472.can-10-3667 >.

BOTTER, S. M.; NERI, D.; FUCHS, B. Recent advances in osteosarcoma. **Curr Opin Pharmacol,** v. 16, p. 15-23, Jun 2014. ISSN 1471-4892. Available at: < http://dx.doi.org/10.1016/j.coph.2014.02.002 >.

BURRELL, R. A. et al. Replication stress links structural and numerical cancer chromosomal instability. **Nature,** v. 494, n. 7438, p. 492-6, Feb 28 2013. ISSN 0028-0836. Available at: < http://dx.doi.org/10.1038/nature11935 >.

CHEN, X. et al. Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. **Cell Rep,** v. 7, n. 1, p. 104-12, Apr 10 2014. Available at: < http://dx.doi.org/10.1016/j.celrep.2014.03.003 >.

CHENG, E. Y. Osteosarcoma – An Introduction. **Rein in Sarcoma (RIS)**, 2018. Avaialble at: < https://www.reininsarcoma.org/what-is-sarcoma-2/bone-sarcomas/osteosarcoma-an-introduction/ >.

CORTES-CIRIANO, I. et al. Comprehensive analysis of *Chromothripsis* in 2,658 human cancers using whole-genome sequencing. 2018-05-30 2018. Available at: < https://www.biorxiv.org/content/early/2018/05/30/333617 >.

DAVOLI, T. et al. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. 2017-01-20 2017. Available at: < http://science.sciencemag.org/content/355/6322/eaaf8399/tab-pdf >.

DOMINGUEZ-SOLA, D.; GAUTIER, J. MYC and the Control of DNA Replication. **Cold Spring Harb Perspect Med,** v. 4, n. a014423, 2014-06-01 2014. Available at: < http://perspectivesinmedicine.cshlp.org/content/4/6/a014423.abstract >.

DONLEY, N.; THAYER, M. J. DNA replication timing, genome stability and cancer: late and/or delayed DNA replication timing is associated with increased genomic instability. **Semin Cancer Biol,** v. 23, n. 2, p. 80-9, Apr 2013. ISSN 1044-579x. Available at: < http://dx.doi.org/10.1016/j.semcancer.2013.01.001 >.

DURFEE, R. A.; MOHAMMED, M.; LUU, H. H. Review of Osteosarcoma and Current Management. In: (Ed.). **Rheumatol Ther**, v.3, 2016. p.221-43. ISBN 2198-6576 (Print)2198-6584 (Electronic).

FORMENT, J. V.; KAIDI, A.; JACKSON, S. P. *Chromothripsis* and cancer: causes and consequences of chromosome shattering. **Nature Reviews Cancer,** v. 12, n. 10, p. 663-670, 2012-09-13 2012. ISSN 1474-175X. Available at: < http://www.nature.com/nrc/journal/v12/n10/abs/nrc3352.html >.

GELOT, C.; MAGDALOU, I.; LOPEZ, B. S. Replication stress in Mammalian cells and its consequences for mitosis. **Genes (Basel),** v. 6, n. 2, p. 267-98, 2015. ISSN 2073-4425. Available at: < http://dx.doi.org/10.3390/genes6020267 >.

GRÖBNER, S. N. et al. The landscape of genomic alterations across childhood cancers. **Nature,** v. 555, n. 7696, p. 321, 2018-02-28 2018. ISSN 1476-46871476-4687. Available at: < https://www.nature.com/articles/nature25480 >.

HATCH, E. M. et al. Catastrophic nuclear envelope collapse in cancer cell micronuclei. **Cell,** v. 154, n. 1, p. 47-60, Jul 3 2013. ISSN 0092-8674. Available at: < http://dx.doi.org/10.1016/j.cell.2013.06.007 >.

HATCH, E. M.; HETZER, M. W. *Chromothripsis*. **Current Biology,** v. 25, n. 10, p. R397–R399, 2015. Available at: < http://dx.doi.org/10.1016/j.cub.2015.02.033 >.

HOLLAND, A. J.; CLEVELAND, D. W. Chromoanagenesis and cancer: mechanisms and consequences of localized, complex chromosomal rearrangements. **Nat Med,** v. 18, n. 11, p. 1630-8, Nov 2012. ISSN 1078-8956. Available at: < http://dx.doi.org/10.1038/nm.2988 >.

INSTITUTE, S. COSMIC: Catalogue of Somatic Mutations in Cancer. 2018. Avaialble at: < https://cancer.sanger.ac.uk/cosmic >.

JANSSEN, A.; MEDEMA, R. H. Genetic instability: tipping the balance. **Oncogene,** v. 32, n. 38, p. 4459, 2012-12-17 2012. ISSN 1476-55941476-5594. Available at: < https://www.nature.com/articles/onc2012576 >.

KANSARA, M. et al. Translational biology of osteosarcoma. **Nature Reviews Cancer,** v. 14, n. 11, p. 722-735, Nov 2014. ISSN 1474-175X. Available at: < <Go to ISI>://WOS:000344430300009 >.

KIM, T. M. et al. Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes. **Genome Res,** v. 23, n. 2, p. 217-27, Feb 2013. ISSN 1088-9051. Available at: < http://dx.doi.org/10.1101/gr.140301.112 >.

KINSELLA, M.; PATEL, A.; BAFNA, V. The elusive evidence for *Chromothripsis*. **Nucleic Acids Res,** v. 42, n. 13, p. 8231-42, Sep 01 2014. ISSN 0305-1048 (Print)1362-4962 (Electronic). Available at: < http://dx.doi.org/10.1093/nar/gku525 >.

KORBEL, J. O.; CAMPBELL, P. J. Criteria for inference of *Chromothripsis* in cancer genomes. **Cell,** v. 152, n. 6, p. 1226-36, Mar 14 2013. ISSN 0092-8674. Available at: < http://dx.doi.org/10.1016/j.cell.2013.02.023 >.

KOVAC, M. et al. Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of BRCA deficiency. **Nature Communications,** v. 6, 2015-12-03 2015. ISSN 2041-1733. Available at: < http://www.nature.com/ncomms/2015/151203/ncomms9940/full/ncomms9940.html >.

KOVTUN, I. V. et al. Chromosomal catastrophe is a frequent event in clinically insignificant prostate cancer. **Oncotarget,** v. 6, n. 30, p. 29087-96, Oct 6 2015. ISSN 1949-2553. Available at: < http://dx.doi.org/10.18632/oncotarget.4900 >.

LIN, Y. H. et al. Osteosarcoma: Molecular Pathogenesis and iPSC Modeling. **Trends Mol Med,** v. 23, n. 8, p. 737-755, Aug 2017. ISSN 1471-4914. Available at: < http://dx.doi.org/10.1016/j.molmed.2017.06.004 >.

LIU, P. et al. Chromosome catastrophes involve replication mechanisms generating complex genomic rearrangements. **Cell,** v. 146, n. 6, p. 889-903, Sep 16 2011. ISSN 0092-8674. Available at: < http://dx.doi.org/10.1016/j.cell.2011.07.042 >.

LORENZ, S. et al. Unscrambling the genomic chaos of osteosarcoma reveals extensive transcript fusion, recurrent rearrangements and frequent novel TP53 aberrations.

Oncotarget, v. 7, n. 5, p. 5273-88, Feb 2 2016. ISSN 1949-2553. Available at: < http://dx.doi.org/10.18632/oncotarget.6567 >.

MARTIN, G. S. Cell signaling and cancer. **Cancer Cell,** v. 4, n. 3, p. 167-174, 2003/09/01/ 2003. ISSN 1535-6108. Available at: < http://www.sciencedirect.com/science/article/pii/S1535610803002162 >.

MARTIN, J. W.; SQUIRE, J. A.; ZIELENSKA, M. The genetics of osteosarcoma. **Sarcoma,** v. 2012, n. 627254, 2012. ISSN 1357-714x. Available at: < http://dx.doi.org/10.1155/2012/627254 >.

_____. The Genetics of Osteosarcoma. **Sarcoma,** v. 2012, n. 627254, May 2012.

MATHIAS, M. et al. Osteosarcoma with apparent Ewing sarcoma gene rearrangement. 2016-02-18 2016. Available at: < http://biorxiv.org/content/early/2016/02/18/039834 >.

MCCLELLAND, S. E. Role of chromosomal instability in cancer progression. **Endocr Relat Cancer,** v. 24, n. 9, p. T23-t31, Sep 2017. ISSN 1351-0088. Available at: < http://dx.doi.org/10.1530/erc-17-0187 >.

MI, H. et al. Large-scale gene function analysis with the PANTHER classification system. **Nature Protocols,** v. 8, n. 8, p. 1551, 2013-07-18 2013. ISSN 1750-27991750-2799. Available at: < https://www.nature.com/articles/nprot.2013.092 >.

MORIARITY, B. S. et al. A Sleeping Beauty forward genetic screen identifies new genes and pathways driving osteosarcoma development and metastasis. **Nat Genet,** v. 47, n. 6, p. 615-24, Jun 2015. ISSN 1061-4036. Available at: < http://dx.doi.org/10.1038/ng.3293 >.

MUKHERJEE, A. et al. Multinucleation regulated by the Akt/PTEN signaling pathway is a survival strategy for HepG2 cells. **Mutat Res,** v. 755, n. 2, p. 135-40, Aug 15 2013. ISSN 0027-5107 (Print)0027-5107. Available at: < http://dx.doi.org/10.1016/j.mrgentox.2013.06.009 >.

NCBI, H. M. SRA Knowledge Base. 2011 2011. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/ >.

OKEGAWA, T. et al. The role of cell adhesion molecule in cancer progression and its application in cancer therapy. **Acta Biochim Pol,** v. 51, n. 2, p. 445-57, 2004. ISSN 0001-527X (Print)0001-527x. Available at: < http://dx.doi.org/035001445 >.

PELLESTOR, F. ***Chromothripsis* - Methods and Protocols**. New York: Humana Press, 2018. 367 Available at: < //www.springer.com/us/book/9781493977796 >.

PERRY, J. A. et al. Complementary genomic approaches highlight the PI3K/mTOR pathway as a common vulnerability in osteosarcoma. **Proc Natl Acad Sci U S A,** v. 111, n. 51, p. E5564-73, Dec 23 2014. ISSN 0027-8424. Available at: < http://dx.doi.org/10.1073/pnas.1419260111 >.

RATNAPARKHE, M. et al. Genomic profiling of Acute lymphoblastic leukemia in ataxia telangiectasia patients reveals tight link between ATM mutations and *Chromothripsis*. **Leukemia,** v. 31, n. 10, p. 2048-2056, Oct 2017. ISSN 0887-6924. Available at: < http://dx.doi.org/10.1038/leu.2017.55 >.

REIMANN, E. et al. Whole exome sequencing of a single osteosarcoma case—integrative analysis with whole transcriptome RNA-seq data. **Human Genomics,** v. 8, n. 1, p. 1, 2014-12-11 2014. ISSN 1479-7364. Available at: < https://humgenomics.biomedcentral.com/articles/10.1186/s40246-014-0020-0 >.

RODE, A. et al. *Chromothripsis* in cancer cells, an update. **International Journal of Cancer**, 2015. ISSN 1097-0215. Available at: < http://onlinelibrary.wiley.com/doi/10.1002/ijc.29888/abstract >.Available at: < http://onlinelibrary.wiley.com/doi/10.1002/ijc.29888/pdf >.

ROSENBERG, A. E. et al. Conventional Osteosarcoma. In: FLETCHER, C. D. M.;BRIDGE, J. A., *et al* (Ed.). **WHO Classification of Tumours of Soft Tissue and Bone**. 4°: World Health Organization, 2013. chap. 16 - Osteogenic Tumors, p.427. ISBN 9283224345.

SADIKOVIC, B. et al. Identification of interactive networks of gene expression associated with osteosarcoma oncogenesis by integrated molecular profiling. **Hum Mol Genet,** v. 18, n. 11, p. 1962-75, Jun 1 2009. ISSN 0964-6906. Available at: < http://dx.doi.org/10.1093/hmg/ddp117 >.

SELVARAJAH, S. et al. Genetic aspects of bone tumors. In: HEYMANN, D. (Ed.). **Bone Cancer: Primary Bone Cancers and Bone Metastases**. 2nd. Nantes, FR: Elsevier, v.1, 2014. chap. 27, p.736. ISBN 0124167284.

_____. Genomic signatures of chromosomal instability and osteosarcoma progression detected by high resolution array CGH and interphase FISH. **Cytogenet Genome Res,** v. 122, n. 1, p. 5-15, 2008. ISSN 1424-8581. Available at: < http://dx.doi.org/10.1159/000151310 >.

_____. The breakage-fusion-bridge (BFB) cycle as a mechanism for generating genetic heterogeneity in osteosarcoma. **Chromosoma,** v. 115, n. 6, p. 459-67, Dec 2006. ISSN 0009-5915 (Print)0009-5915. Available at: < http://dx.doi.org/10.1007/s00412-006-0074-4 >.

SHEN, M. M. *Chromoplexy*: a new category of complex rearrangements in the cancer genome. **Cancer Cell,** v. 23, n. 5, p. 567-9, May 13 2013. ISSN 1535-6108. Available at: < http://dx.doi.org/10.1016/j.ccr.2013.04.025 >.

SIEGEL, J. J.; AMON, A. New Insights into the Troubles of Aneuploidy. **Annu Rev Cell Dev Biol,** v. 28, p. 189-214, 2012. ISSN 1081-0706 (Print)1530-8995 (Electronic). Available at: < http://dx.doi.org/10.1146/annurev-cellbio-101011-155807 >.

SMIDA, J. et al. Genome-wide analysis of somatic copy number alterations and chromosomal breakages in osteosarcoma. **Int J Cancer,** v. 141, n. 4, p. 816-828, Aug 15 2017. ISSN 0020-7136. Available at: < http://dx.doi.org/10.1002/ijc.30778 >.

STEPHENS, P. J. et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. **Cell,** v. 144, n. 1, p. 27-40, Jan 7 2011. ISSN 0092-8674. Available at: < http://dx.doi.org/10.1016/j.cell.2010.11.055 >.

TARAN, S. J.; TARAN, R.; MALIPATIL, N. B. Pediatric Osteosarcoma: An Updated Review. In: (Ed.). **Indian J Med Paediatr Oncol**, v.38, 2017. p.33-43. ISBN 0971-5851 (Print)0975-2129 (Electronic).

TOGUCHIDA, J. Genetics of Osteosarcoma. In: (Ed.). **Osteosarcoma**: Springer Japan, 2016. p.3-17. ISBN 978-4-431-55696-1.

TOLEDO, S. R. et al. Bone deposition, bone resorption, and osteosarcoma. **J Orthop Res,** v. 28, n. 9, p. 1142-8, Sep 2010. ISSN 0736-0266. Available at: < http://dx.doi.org/10.1002/jor.21120 >.

WANG, Z. et al. T-Cell-Based Immunotherapy for Osteosarcoma: Challenges and Opportunities. **Front Immunol,** v. 7, p. 353, 2016. ISSN 1664-3224 (Print)1664-3224. Available at: < http://dx.doi.org/10.3389/fimmu.2016.00353 >.

WECKSELBLATT, B.; RUDD, M. K. Human Structural Variation: Mechanisms of Chromosome Rearrangements. **Trends Genet,** v. 31, n. 10, p. 587-599, Oct 2015. ISSN 0168-9525 (Print)0168-9525. Available at: < http://dx.doi.org/10.1016/j.tig.2015.05.010 >.

WILLIS, N. A.; RASS, E.; SCULLY, R. Deciphering the Code of the Cancer Genome: Mechanisms of Chromosome Rearrangement. **Trends Cancer,** v. 1, n. 4, p. 217-230, Dec 1 2015. ISSN 2405-8033 (Print). Available at: < http://dx.doi.org/10.1016/j.trecan.2015.10.007 >.

WORRALL, J. T. et al. Non-random Mis-segregation of Human Chromosomes. **Cell Rep,** v. 23, n. 11, p. 3366-3380, Jun 12 2018. Available at: < http://dx.doi.org/10.1016/j.celrep.2018.05.047 >.

YAMADA, H. Y. et al. Systemic chromosome instability in Shugoshin-1 mice resulted in compromised glutathione pathway, activation of Wnt signaling and defects in immune system in the lung. In: (Ed.). **Oncogenesis**, v.5, 2016. p.e256-. ISBN 2157-9024 (Electronic).

YANG, J.; DENG, G.; CAI, H. *Chromothripsis*DB: a curated database of *Chromothripsis*. **Bioinformatics**, Dec 31 2015. ISSN 1367-4803. Available at: < http://dx.doi.org/10.1093/bioinformatics/btv757 >.

YANG, Y. et al. A Bayesian Gene-Based Genome-Wide Association Study Analysis of Osteosarcoma Trio Data Using a Hierarchically Structured Prior:. **https://doi.org/10.1177/1176935118775103**, 2018-05-21 2018. Available at: < http://journals.sagepub.com/doi/full/10.1177/1176935118775103 >.

ZHANG, C.-Z. et al. *Chromothripsis* from DNA damage in micronuclei. **Nature,** v. 522, p. 179-184, 2015-05-27 2015. ISSN 0028-0836. Available at: < http://www.nature.com/nature/journal/v522/n7555/full/nature14493.html >.

ZHANG, C. Z.; LEIBOWITZ, M. L.; PELLMAN, D. *Chromothripsis* and beyond: rapid genome evolution from complex chromosomal rearrangements. **Genes Dev,** v. 27, n. 23, p. 2513-30, Dec 1 2013. ISSN 0890-9369. Available at: < http://dx.doi.org/10.1101/gad.229559.113 >.

Attachments

Attachment A – Ethics committee referee

FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
UNIVERSIDADE DE SÃO PAULO

COMITÊ DE ÉTICA EM PESQUISA - HCRP
16:50 h
Recebi
29/07/16

Ilma. Senhora Profa. Dra. Marcia Guimarães Villanova,
(Coordenadora do Comitê de Ética em Pesquisa do HCFMRP-USP)

Venho solicitar dispensa de apreciação ética do projeto de pesquisa de Doutorado intitulado "Caracterização de Alterações Genômicas Caóticas em Osteosarcoma" ao Comitê de Ética em Pesquisa do Hospital das Clinicas da Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo (HCFMRP-USP). A solicitação de dispensa justifica-se uma vez que esta pesquisa apresenta caráter retrospectivo, utilizando dados genômicos de amostras humanas de tumores de Osteosarcoma, que estão publicamente disponiveis à comunidade científica em bancos de dados internacionais. Além disso, também serão utilizados dados da literatura publicados em revistas científicas indexadas nos principais bancos de dados públicos. Os dados disponíveis online preservam total sigilo de identidade dos pacientes envolvidos. Além disso, os dados genômicos públicos foram gerados após aprovação de cada pesquisa pela avaliação do Comitê de Ética em Pesquisa locais. Este projeto de Doutorado será desenvolvido no Departamento de Genética da FMRP-USP, pela aluna de Pós-Graduação Alexandra Galvão Gomes, número USP 7376110, bolsista CNPq (processo 142192/2014-7), sob orientação do Professor da Faculdade de Medicina de Ribeirão Preto Dr. Jeremy Andrew Squire. O projeto se constituirá de análise retrospectiva de amostras de tumores de Osteosarcoma avaliados por microarranjos genômicos, e os resultados serão validados por análise retrospectiva comparativa com amostras de tumores de Osteosarcoma avaliados por sequenciamento de nova geração. O objetivo do trabalho é analisar e caracterizar alterações genômicas caóticas em Osteosarcoma, com foco em rearranjos cromossômicos complexos que evidenciem regiões de instabilidade genômica.

Ribeirão Preto, 28 de julho de 2016.

Aluna: Ms. Alexandra Galvão Gomes
Doutoranda do PPG Genética FMRP

Orientador: Prof. Dr. Jeremy A. Squire
Depto. Genética

Não há necessidade de submissão ao Comitê de Ética em Pesquisa.

Dra. Marcia Guimarães Villanova
Coordenadora do Comitê de Ética em Pesquisa do HC e FMRP-USP
01/08/2016

Attachment B – Publication during the PhD period.

Mol Syndromol. 2017 Jan;8(1):45-49. doi: 10.1159/000452681. Epub 2016 Nov 17.

Complex Mosaic Ring Chromosome 11 Associated with Hemizygous Loss of 8.6 Mb of 11q24.2qter in Atypical Jacobsen Syndrome.

Galvão Gomes A[1], Paiva Grangeiro CH[2], Silva LR[3], Oliveira-Gennaro FG[1], Pereira CS[4], Joaquim TM[1], Panepucci RA[5], Squire JA[6], Martelli L[2].

Author information

Abstract

Jacobsen syndrome (JBS) is a contiguous gene deletion syndrome involving terminal chromosome 11q. The haploinsufficiency of multiple genes contributes to the overall clinical phenotype, which can include the variant Paris-Trousseau syndrome, a transient thrombocytopenia related to *FLI1* hemizygous deletion. We investigated a boy with features of JBS using classic cytogenetic methods, FISH and high-resolution array CGH. The proband was found to have a mosaic ring chromosome 11 resulting in a hemizygous 11q terminal deletion of 8.6 Mb, leading to a copy number loss of 52 genes. The patient had a hemizygous deletion in the *FLI1* gene region without apparent thrombocytopenia, and he developed diabetes mellitus type I, which has not previously been described in the spectrum of disorders associated with JBS. The relationship of some of the genes within the context of the phenotype caused by a partial deletion of 11q has provided insights concerning the developmental anomalies presented in this patient with atypical features of JBS.

KEYWORDS: Comparative genomic hybridization; Deletion 11q; FLI1; Jacobsen syndrome; Ring chromosome; Thrombocytopenia Paris-Trousseau type

PMID: 28232783 PMCID: PMC5260599 DOI: 10.1159/000452681

Free PMC Article

The entire publication is open access and available online at <https://www.karger.com/Article/FullText/45268>.

Attachment C – Manuscript (to submit to Molecular and Clinical Oncology)

Attachment C – Manuscript (to submit to Molecular and Clinical Oncology)

MOLECULAR AND CLINICAL ONCOLOGY: 2018

Evaluation of genes and RNA expression changes in Osteosarcomas samples with and without Chromothripsis

ALEXANDRA GALVÃO GOMES[1], MAISA YOSHIMOTO[2], and JEREMY A. SQUIRE[1]

1- Genetics Department, Medical School of Ribeirão Preto, University of Sao Paulo- Ribeirão Preto, Brazil.

2- Department of Medical Genetics, University of Alberta – Alberta, Canada.

Received:

DOI:

Correspondence to: Dr. Jeremy Squire. Av. Bandeirantes, 3900. USP: Departamento de Genética – Bloco C. Monte Alegre - Ribeirão Preto – SP (Brazil). CEP: 14049-900

E-mail: alexandragalvao@usp.br

Key-words: Osteosarcoma, Chromothipsis, array CGH, expression, chaotic rearrangements

Abstract

Methods of whole genome sequencing have recently detected unusual types of catastrophic genomic rearrangements in human tumors called chromothripsis. Osteosarcoma has one of the highest rates of chromothripsis of all human cancers, but at the present time there has been no systematic analysis of the chromosomal regions and genes that are commonly affected by these chaotic rearrangements. In this study we performed an analysis of array CGH copy number data from ten OS tumour DNA datasets to determine the incidence of chromothripsis. We found three of the osteosarcomas had chromothripsis that affected four chromosomes (2, 10, 14 and 20) in total. The osteosarcomas with chromothripsis had a median of 468 copy number abnormalities per tumor compared to 255 for tumors without chromothripsis. We then compared global RNA expression levels from two OS samples with chromothripsis to four tumors without chromothripsis to determine the types of gene expression differences associated with this process. We found that 171 genes mapped to regions of chromothripsis with the majority (77 genes) being mainly related to cellular communication and cell cycle. There were 43 genes being related to metabolic process (mainly associated with RNA metabolism) and 27 genes with cellular component organization or biogenesis. Also, there were four genes associated with the immune

system that were underexpressed (*CADM1; CLEC4A; CCR1; CD164*) and 12 were overexpressed (*IL32, LAT, BCL3, FCAR, RFX1, ILIB, CXCL1, SPON2, CCR6, IL6, SEMA3C, GEM*) in the chromothripsis tumors. Interestingly, all the genes underexpressed also have a role in cell adhesion pathway. Cell adhesion is associated with cancer progression and metastasis. Chromothripsis seems to affect at least 30% of osteosarcomas and may be contributing to the more aggressive phenotype of this bone tumor.

Introduction

Osteosarcoma (OS) is the most common type of malignancy in bone tissue, with an incidence of 1-4 cases/million, mainly affecting children and adolescents (75%), with males being more frequently affected (ratio 1.5:1) [1,2]. Considered a rare tumor, OS is an aggressive malignancy originating from mesenchymal stem cells that produce osteoid or immature bone [3,4]. OS tumors are more complex than other sarcomas, however publications about the genetic cause of OS are still restricted given the rare incidence of the tumor [2,5].

In addition to being an aggressive tumor, OS is characterized by having an unusually high level of genomic alteration and chromosomal instability (CIN). At the cytogenetic level OS is characterized by having many complex structurally abnormal chromosomes as well as gene amplification, dicentric chromosomes, multiple marker chromosomes, double minutes (dmin), homogeneously staining regions (hsr), and/or ploidy changes and anaphase bridges that can lead to micronuclei, as seen in other human cancers with a high rate of CIN[6,7]. Tumors usually have complex chromosome aberrations with high incidence of numerical DNA copy number gains (regions 1p, 6p, 8q, 12q and 17p are commonly reported) and losses (regions 2q, 3q, 6q, 10, 13q and 17p are commonly reported) [8,9].

Some hereditary genetic syndromes increase the risk of developing OS, such as hereditary retinoblastoma, Rothmund–Thomson syndrome, Li-Fraumeni syndrome, and Werner syndrome. Genes associated with these syndromes (*RB1, RECQL4, TP53,* and *WRN*) possibly might influence in the pathogenesis of OS [3,10]. Moreover, other genes were reported related with OS, as *RUNX2* (6p), *MYC* (8q), and *PTEN* (10q) [9].

Previous array comparative genomic hybridization (aCGH) and spectral karyotyping studies have demonstrated that OS has one of the highest rates of CIN with copy number gains and structural changes affecting more than 50% of the genome [6,11-13]. Analysis of microarray data of single copy nucleotide polymorphisms (SNP array CGH) and next-generation paired-end sequencing across a range of tumor cell types, has proposed a new class of catastrophic genomic rearrangement called Chomothripsis. The genomic breakpoints associated with Chromothripsis occur in tens to hundreds and are

usually restricted to discrete regions on one chromosome, and seems to occur as a single event to one cell[14,15]. Chromothripsis occurs in 2-3% of primary tumors. However, the frequency of this phenomenon may be greater than 33% in OS [15].

There are a limited number of studies in the scientific literature addressing the role of chromothripsis in osteosarcoma. There is little information on the possible mechanisms that allow its occurrence or explaining why OS tumors have very high rate of these types of rearrangements.

Methods

We analyzed the DNA copy number data by array CGH technique of 10 OS human pediatric tumors samples previously processed by our laboratory research group (table 1), already available in GEO public functional genomics data repository (available at <https://www.ncbi.nlm.nih.gov/geo/>). These raw data is in the study #GSE12830 (platform aCGH Agilent 244k - build 35).

Table 1 – Overview of study GSE12830

#GEO Study	Sample	Sample ID	Platform Build 35	Sample Type
GSE12830	GSM322064	OS87B	Agilent FE	OS pediatric tumor
	GSM322072	OS138	Agilent FE	OS pediatric tumor
	GSM322074	OS177	Agilent FE	OS pediatric tumor
	GSM322076	OS178	Agilent FE	OS pediatric tumor
	GSM322078	OS179	Agilent FE	OS pediatric tumor
	GSM322086	OS180	Agilent FE	OS pediatric tumor
	GSM322088	OS182	Agilent FE	OS pediatric tumor
	GSM322090	OS183	Agilent FE	OS pediatric tumor
	GSM322092	OS2336	Agilent FE	OS pediatric tumor
	GSM322094	OS2960	Agilent FE	OS pediatric tumor

The evaluation of these data was performed focusing on the identification of potential chromosomal regions commonly involved in chaotic DNA Copy Number Alterations (CNAs). Nexus copy number software version 9.0 (obtained from BioDiscovery, Inc.) was used to process the txt files, choosing the stringent form as a mosaic sample. The output files from Nexus 9.0 were modified using Excel software as the input form required to be used in CTLPScanner[16]. The CTLPScanner tool offers a group of factors for chromothripsis-like patterns (CTLPs) detection in microarrays data, which parameters can be adjusted and can be used online as a web server or in R script. To detect CTLPs was used the algorithm described by [17], with the following parameters and thresholds: copy number status change times $\geq$20; log10 of likelihood ratio $\geq$ 8; minimum segment size (Kb):10; Signal distance between adjacent segments 0.3; Signal value for genomic gains $\geq$0.15; Signal value for genomic losses $\leq$ -0.15.

Subsequently, we separated the samples in 2 groups: CTLP+ (with chromothripsis) and CLTP- (without chromothripsis). Then, genes in different chromosome regions with Copy Number Variations (CNVs) were evaluated between the groups focusing in their biological process. Then, we analyzed the RNA expression array data (HuGene platform) of 6 OS samples from this study (also available in GEO #GSE12865). We used the software Nexus Expression 3.0 to compare the RNA of CTLP+ samples versus the RNA of CTLP- samples.

Results

All 10 OS samples evaluated showed 3020 CNVs aberrations in total, median of 275 CNVs aberration per sample (figure 1), by Nexus 9.0.

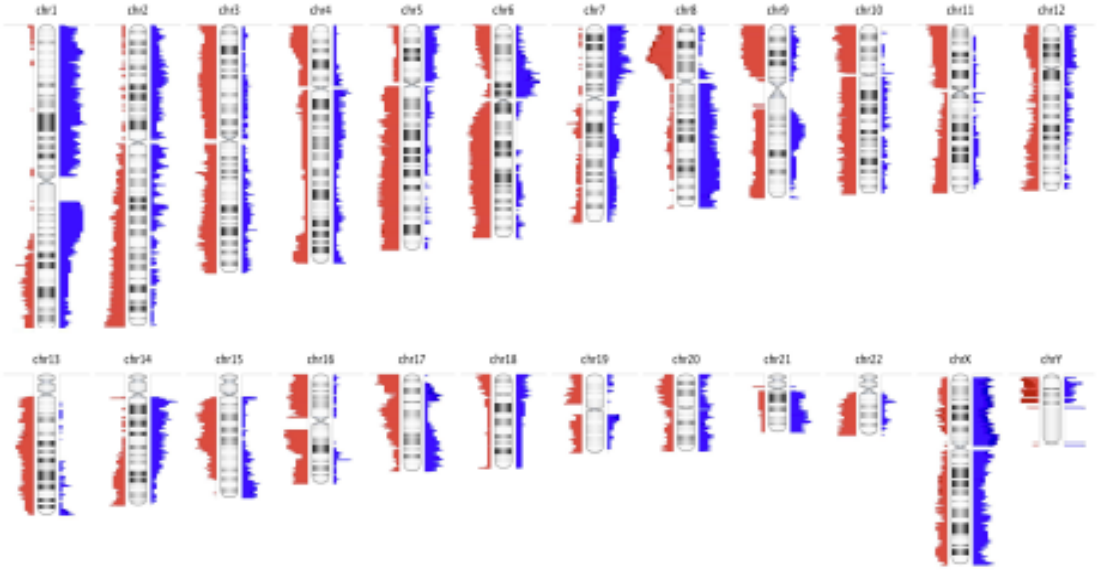

Figure 1 – Overview of the 10 OS samples showing the high rate of copy number changes between the 10 OS samples (GEO #12830) per chromosome, by Nexus 9.0.

The screening of the 10 OS samples in CTLPScanner were realized at web served and confirmed by R script, and the results obtained is showed in table 2, with 3 samples with CTLP+ comprising 4 different chromosomes (figure 2).

Table 2 – CTLPScanner results showing CTLP+ samples, chromosome regions and the CN status.

Array ID	CTLP Region (Mb)	Chromosome	Start	Stop	CNA status change times	Likelihood ratio (log10)
GSM322086CTLP	88.83	2	140000001	228827254	38	29
GSM322064CTLP	62.44	14	20000001	82435964	21	8
	62.44	20	1	62435964	24	10
GSM322090CTLP	78.77	10	56638886	135413628	27	21

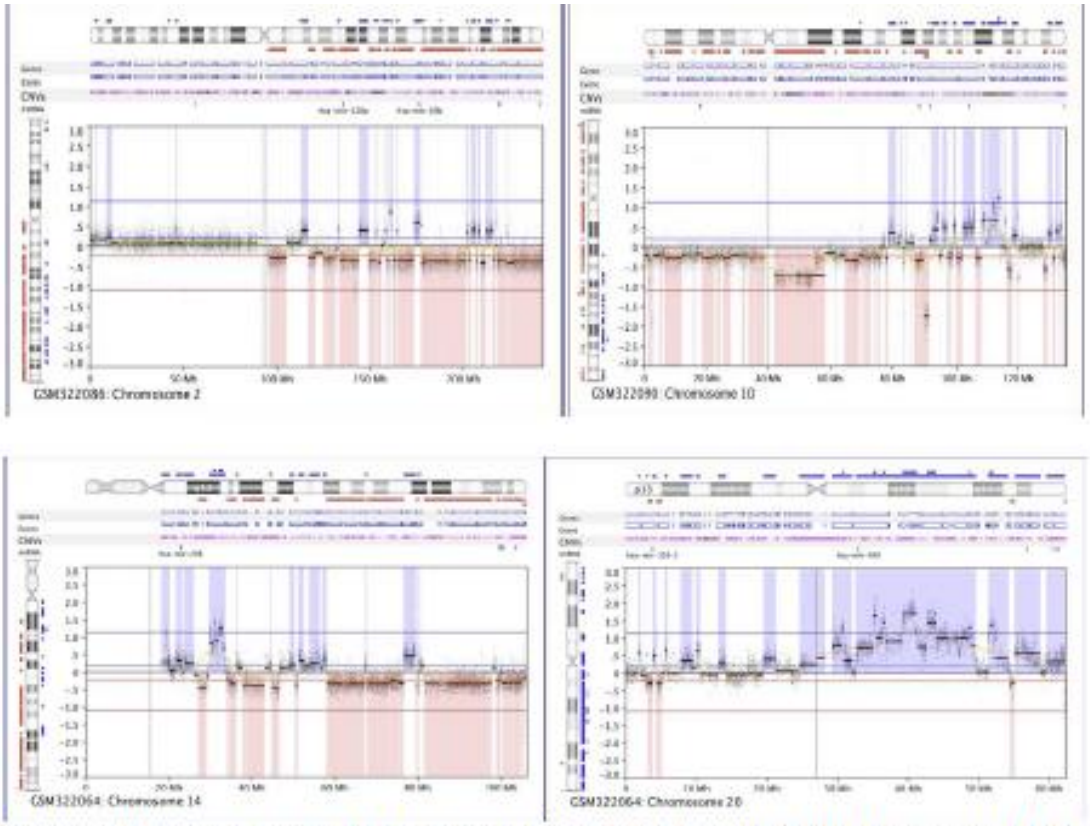

Figure 2 – Chromosomes affected by chromothripsis in GEO study (#12830): Chr2 (sample OS180); Chr10 (sample OS183); Chr14 and Chr20 (sample OS87B).

CTLP+ samples presented 1236 CNAs median of 468 CNAs per sample. CTLP- samples presented 1784 CNAs, median of 255 CNAs per sample (table 3). The results of the CNVs regions comparation between CTLP+ samples and CTLP- samples, and the genes present in each region ($p<0.05$) are in table 3, using Nexus 9.0.

Table 3 – Comparation of CNVs regions between CTLP+ samples and CTLP- samples, and the genes present in each region ($p<0.05$)

Region	CNA	Freq. in <yes> (%)	Freq. in <no> (%)	Gene Symbols
chr3:95,021,157-95,152,842	CN Loss	100	14.3	*PROS1*
chr4:43,856,333-45,282,560	CN Gain	100	0	*KCTD8, YIPF7, GUF1, GNPDA2*
chr6:41,117,007-44,380,641	CN Gain	100	14.3	*TSPO2, APOBEC2, OARD1, NFYA, C6orf132, GUCA1A, GUCA1B, MRPS10, TRERF1,LOC100132354, C6orf223, MRPL14, TMEM63B, CAPN11, SLC29A1, HSP90AB1, MIR4647, SLC35B2, NFKBIE, TMEM151B, TCTE1, AARS2,SLC20A2, C8orf40, CHRNB3, CHRNA6, THAP1, RNF170, MIR4469, HOOK3, FNTA, SGK196, HGSNAT, POTEA*
chr8:42,498,582-43,647,122	CN Gain	100	14.3	*SLC20A2, C8orf40, CHRNB3, CHRNA6, THAP1, RNF170, MIR4469, HOOK3, FNTA, SGK196, HGSNAT, POTEA*
chr8:48,003,671-48,745,096	CN Gain	100	14.3	*LOC100287846, KIAA0146*
chr8:63,583,529-65,697,223	CN Gain	100	14.3	*NKAIN3, LOC100130155, MIR124-2, LOC401463, BHLHE22, CYP7B1*
chr9:84,925,229-91,593,701	CN Gain	100	14.3	*GADD45G, UNQ6494, MIR4290, LOC286370, LOC340515, DIRAS2, SYK, LOC100129316, AUH, NFIL3, MIR3910-2, MIR3910-1, ROR2*
chr10:72,980,218-75,637,374	CN Loss	100	14.3	*CDH23, VCL, AP3M1, ADK*
chr12:127,874,421-128,178,249	CN Loss	100	14.3	*GLT1D1, TMEM132D*
chr12:57,944,175-61,809,692	CN Loss	100	0	*FAM19A2, USP15, MON2, C12orf61, MIRLET7I, PPM1H*
chr15:20,070,027-20,235,180	CN Gain	100	14.3	*REREP3, MIR4509-1, MIR4509-2, MIR4509-3*
chr15:31,518,485-34,698,093	CN Loss	0	85.7	*RYR3, AVEN, CHRM5, AQR, C15orf41*
chr17:10,047,750-15,059,837	CN Gain	100	0	*MYH13, MYH8, MYH4, MYH1, MYH2, MYH3, SCO1, ADPRM, MAGOH2, TMEM220, TMEM220-AS1, LINC00675, PIRT, SHISA6, DNAH9, ZNF18, MIR744, MAP2K4, LINC00670, MYOCD, ARHGAP44, ELAC2, HS3ST3A1, CDRT15P1, COX10-AS1, COX10, CDRT15, MGC12916, HS3ST3B1, CDRT7*
chr17:15,059,837-19,364,790	CN Gain	100	14.3	*PMP22, MIR4731, TEKT3, CDRT4, TVP23C-CDRT4, TVP23C, CDRT1, TRIM16, ZNF286A, TBC1D26, CDRT15P2, MEIS3P1, ADORA2B, ZSWIM7, TTC19, NCOR, CCDC144A, FAM106CP, USP32P1, KRT16P2, TNFRSF13B, LLGL1, FLII, SMCR7, TOP3A, SMCR8, SHMT1, EVPLL, LOC339240, KRT16P1, LGALS9C, USP32P2, FAM106A, CCDC144B, TBC1D28, FOXO3B, ZNF286B, TRIM16L, FBXW10, TVP23B, PRPSAP2, SLC5A10, FAM83G, GRAP, GRAPL, EPN2-IT1, EPN2, EPN2-AS1, MIR1180, B9D1, MAPK7, MFAP4, RNF112*

chr17:7,602,229-8,694,313	CN Loss	0	85.7	*DNAH2, KDM6B, TMEM88, LSMD1, CYB5D1, CHD3, SCARNA21, LOC284023, KCNAB3, TRAPPC1, CNTROB, GUCY2D, PFAS, RANGRF, SLC25A35, ARHGEF15, ODF4, LOC100128288, KRBA2, RPL26, RNF222, NDEL1, MYH10, CCDC42, SPDYE4, MFSD6L, PIK3R6*
chr18:14,562,532-16,100,000	CN Gain	100	0	*ANKRD30B, MIR3156-2, LOC644669*
chr18:14,303,026-14,562,532	CN Gain	100	14.3	*CYP4F35P, CXADRP3, POTEC*
chr18:17,654,794-18,245,461	CN Gain	100	0	*MIB1, MIR133A1, MIR1-2, GATA6*
chr18:16,100,000--18,758,031	CN Gain	100	14.3	*ROCK1, GREB1L, ESCO1, SNRPD1, ABHD3, MIR320C1, MIB1,CTAGE1,*
chr21:41,320,910-41,676,644	CN Gain	100	14.3	*LINC00323, MIR3197, PLAC4, BACE2, FAM3B, MX2*
chrX:71,999,653-72,455,305	CN Gain	100	14.3	*PABPC1L2B, PABPC1L2A, NAP1L6, NAP1L2, CDX4*

Six RNA samples of the two groups were compared (CTLP+ x CTLP-) using Nexus Expression 3.0. Two samples CTLP+ (OS180, OS183) were compared with four samples CTLP- (OS182, OS179, OS178, OS177). The differential expression of some genes of immune system pathway is showed in the heatmap in figure 3.

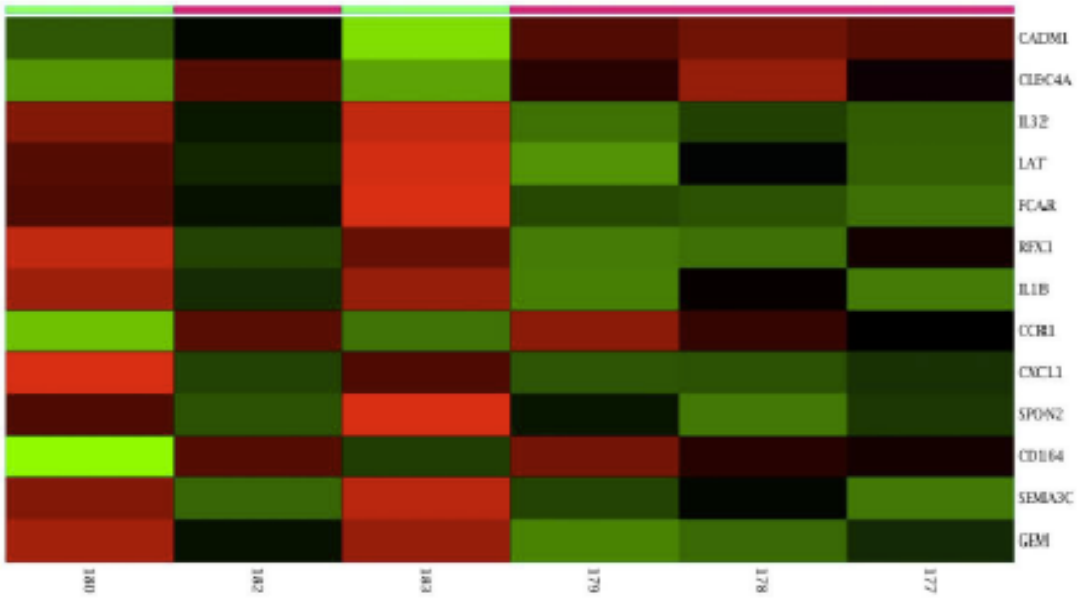

Figure 3- Heatmap of different immune system pathway genes expression between CTLP+ x CTLP- groups.

Discussion

Methods of whole genome sequencing have recently detected unusual types of catastrophic genomic rearrangements in human cancer. These catastrophic

events appear to be sudden, and to be generated by an unexpected use of cellular DNA repair machinery. Chromothripsis is the best described class of catastrophic alteration to date and Osteosarcoma (OS) has one of the highest rates of chromothripsis of all human cancers.

In the present study, we found 3 OS samples (30%) with CTLP+. This data is similar with the first publication about Chromothipsis[15], and we could observe the higher rate of CNAs in CTLP+ samples (median of 468 CNAs/sample versus 255 CNAs/sample). The chromosomes affected by the chaotic events were Chr2, Chr10, Chr14 and Chr20 which are frequent related in OS. Chromosomes 14 and 20 have the high rate of centromeric rearrangements, and chromosome 20 is classified as the chromosome with highest number of CNAs in OS tumors[11]. Furthermore, DNA copy number alterations in chromosomes Chr2 and Chr10 also are commonly reported in OS [8,9].

We found that 171 genes are present in the CNVs comparation between CTLP+ and CTLP- samples focusing in the different chromosome regions (by Nexus 9.0). The majority, 77 genes, are related with cellular process (mainly cellular communication and cell cycle). Also, 43 genes are related with metabolic process (mainly primary metabolic process associated with RNA metabolic process). 27 genes were related with cellular component organization or biogenesis. These biological processes are very important to the development of cancer background.

The expression data revealed that four genes immune system related are underexpressed in CTLP+ samples (*CADM1; CLEC4A; CCR1; CD164*) and 12 are overexpressed (*IL32, LAT, BCL3, FCAR, RFX1, ILIB, CXCL1, SPON2, CCR6, IL6, SEMA3C, GEM*). It is important to notice that all the genes underexpressed also have a role in cell adhesion pathway (just one gene overexpressed has the same role). Cell adhesion is associated with cancer progression and metastasis. Adhesion molecules performance a critical part in the progress of recurrent, invasive, and metastasis during the cancer development. Loss of intercellular adhesion can permits malignant cells to escape from their location of origin, damage the extracellular matrix, obtain a more motile and invasion phenotype, and metastasize[18,19].

In conclusion, complex genomic rearrangements are an integral part of mutation mechanisms contributing to cancer development, affecting different and important pathways and should be considered as a critical process in bone oncogenesis, like in OS.

Acknowlegements

The present study was supported by CAPES, FAEPA and CNPq.

References

1. Kansara M, Teng MW, Smyth MJ, et al: Translational biology of osteosarcoma. Nature Reviews Cancer 14:722-735, 2014
2. Durfee RA, Mohammed M, Luu HH: Review of Osteosarcoma and Current Management, Rheumatol Ther, 2016, pp 221-43
3. Taran SJ, Taran R, Malipatil NB: Pediatric Osteosarcoma: An Updated Review, Indian J Med Paediatr Oncol, 2017, pp 33-43
4. Wang Z, Li B, Ren Y, et al: T-Cell-Based Immunotherapy for Osteosarcoma: Challenges and Opportunities. Front Immunol 7:353, 2016
5. Yang Y, Basu S, Mirabello L, et al: A Bayesian Gene-Based Genome-Wide Association Study Analysis of Osteosarcoma Trio Data Using a Hierarchically Structured Prior:. https://doi.org/10.1177/1176935118775103, 2018
6. Al-Romaih K, Bayani J, Vorobyova J, et al: Chromosomal instability in osteosarcoma and its association with centrosome abnormalities. Cancer Genet Cytogenet 144:91-9, 2003
7. Donley N, Thayer MJ: DNA replication timing, genome stability and cancer: late and/or delayed DNA replication timing is associated with increased genomic instability. Semin Cancer Biol 23:80-9, 2013
8. Martin JW, Squire JA, Zielenska M: The genetics of osteosarcoma. Sarcoma 2012, 2012
9. Rosenberg AE, Cleton-Jansen A-M, Pineux Gd, et al: Conventional Osteosarcoma, in Fletcher CDM, Bridge JA, Hogendoorn PCW, et al (eds): WHO Classification of Tumours of Soft Tissue and Bone (ed 4°), World Health Organization, 2013, pp 427
10. Moriarity BS, Otto GM, Rahrmann EP, et al: A Sleeping Beauty forward genetic screen identifies new genes and pathways driving osteosarcoma development and metastasis. Nat Genet 47:615-24, 2015
11. Bayani J, Zielenska M, Pandita A, et al: Spectral karyotyping identifies recurrent complex rearrangements of chromosomes 8, 17, and 20 in osteosarcomas. Genes Chromosomes Cancer 36:7-16, 2003

12. Selvarajah S, Yoshimoto M, Ludkovski O, et al: Genomic signatures of chromosomal instability and osteosarcoma progression detected by high resolution array CGH and interphase FISH. Cytogenet Genome Res 122:5-15, 2008

13. Sadikovic B, Yoshimoto M, Chilton-MacNeill S, et al: Identification of interactive networks of gene expression associated with osteosarcoma oncogenesis by integrated molecular profiling. Hum Mol Genet 18:1962-75, 2009

14. Forment JV, Kaidi A, Jackson SP: Chromothripsis and cancer: causes and consequences of chromosome shattering. Nature Reviews Cancer 12:663-670, 2012

15. Stephens PJ, Greenman CD, Fu B, et al: Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell 144:27-40, 2011

16. Yang J, Liu J, Ouyang L, et al: CTLPScanner: a web server for chromothripsis-like pattern detection. Nucleic Acids Res 44:W252-8, 2016

17. Korbel JO, Campbell PJ: Criteria for inference of chromothripsis in cancer genomes. Cell 152:1226-36, 2013

18. Okegawa T, Pong RC, Li Y, et al: The role of cell adhesion molecule in cancer progression and its application in cancer therapy. Acta Biochim Pol 51:445-57, 2004

19. Bendas G, Borsig L: Cancer cell adhesion and metastasis: selectins, integrins, and the inhibitory potential of heparins. Int J Cell Biol 2012:676731, 2012

Attachment D – Supplementary Data

Extra data/files are available online at USP cloud. You can access this data using the link: <https://tinyurl.com/yby5xjsj>.

Data List:

- Tables: Summary of results: arrays reevalueted from public databanks (results of arrays OS samples file, genes reported, etc.);
- Files extracted by Nexus 9.0 software (.txt) from dbGaP samples (dbgap by Nexus folder);
- CTLPscanner results files from dbGaP analysis (List of COSMIC genes, and entire list of genes of CTLP regions found, etc.);
- Samples clustered by GEO study in the input form to CTLPscanner;
- Dbgap pipeline and manual;
- Barts Cancer Protocols;
- Live cell movie.

Occasionally, more data can be add or removed from this folder.