

# Signature of Chromosomes Instability in Different Diseases as Accessed on Illumina Miseq Platform using Depth of Coverage Metrics for Variant Evaluation by GATK

Edem Nuglozeh

Molecular and Diagnostics Personalized Therapeutic Unit, University of Hail School of Medicine, Hail, Kingdom of Saudi Arabia.

The author declares that there is no conflict of interest regarding the publication of this manuscript

**Abstract:** *Next-generation sequencing (NGS) has been widely applied to clinical diagnosis. Target-gene capture followed by deep sequencing provides unbiased enrichment of the target sequences, which not only accurately detects single-nucleotide variations (SNVs) and small insertion/deletions (indels) but also provides the opportunity for the identification of exonic copy-number variants (CNVs) and large genomic rearrangements. The use of NGS allow to directly distinguish the underlying causative diseases genes via a systematic filtering, in which the identified gene variants are checked for novelty and diseases association. This maneuver is possible, provided that, we reach proper depth of coverage metrics during sequencing. Here we use the depth of coverage metrics to assess genome stability in different diseases. Our results confirm other results observed in the past about genome instability and rearrangement in multiple cancers while shedding the light about chromosome instability in hypercholesterolemia.*

**Keywords:** Exome sequencing, Depth of Coverage, Chromosomes Instability, Oligonucleotides binding

## 1. Introduction

A proper delineation of diseases associated mutation in functional elements in human genome represents a central ambitious and challenging approach in genomics-based medicine (1). Many diseases like cancers are the results of the inerrant appearance of SNPs and Indels (2) and these later are subjected to chromosomes instability (3-5), chromothripsis events and chromosomes shattering (3-5). The majority of human cancer cells are highly aneuploid harboring chromosome numbers deviating from their normal modal number of 46. In cancer, aneuploidy is a consequence of an increased rate of whole chromosome missegregation during mitosis, a process known as chromosomal instability (CIN) (6). There is now evidence that most cancers may indeed be genetically unstable, but that the instability exists at two distinct levels. In a small subset of tumor cells, the instability is observed at the nucleotides level as results of base substitutions, deletions or insertions. In most other cancers, the instability is observed at the chromosome level, resulting in losses and gains of whole chromosomes or large portions of karyotypes (7). There is evidence that most cancers may indeed be genetically unstable. Recognition and comparison of these instabilities are leading to new insights into tumor pathogenesis (8). The stamp of chromothripsis or genomic instability have been seen in many cancers and occurs in 25% of bone cancers.

The first use of massive parallel pair-ended sequencing in the elucidation of the mechanisms underlying this phenomenon has been conducted by Stephens and colleagues (9). We currently have only a partial repertoire of these alerted sequences in proteins coding sequences and it is practically impossible to identify all these mutations using conventional Sanger Sequencing Methodology. The Next

Generation Sequencing (NGS) technologies parented by few scientists have permitted to put these challenges into reach. Whole-genome sequencing and Exome sequencing and analysis are becoming important part of a translational medicine research toolkit (10) to investigate small-scale changes such as single-nucleotide variants (SNVs) and indels (11-13) in addition to large-scale events such as chromosomal rearrangements and chromosomes shattering (14). For both basic genomics sciences and personalized medicine, the compromise between data quality and quantity will determine what constitutes the accuracy of whole genome analysis, especially for detecting SNPs and Indels. As whole-genome sequencing becomes a commodity, it will be important to determine the descriptive and quantitative metrics that can shed the light on individual's genome sequence. However, not such standards currently exist in NGS sciences.

While population geneticists will be contended with sequencing targeting a low-coverage pooled data sets, in this approach where we are targeting diseases like cancers and hypercholesterolemia we will put the beam on highly accurate SNVs calls from an individual's genome.

Secondly, we seek to understand whether the probe coverage of different chromosomes is peculiar to particular diseases types, inerrant to the chemical identity of the ligands or imputed to both phenomena.

To answer these questions, we undertook human exome sequencing on gDNA from different types of cancers: breast cancer, human cancer cell lines, and ALCL and gDNA originating from blood sample of individual diagnosed blood pressure, hypercholesterolemia and diabetes on MiSeq Illumina platform to a total of 350 GB (representing 39 X average sequenced depth).

Volume 6 Issue 2, February 2017

[www.ijsr.net](http://www.ijsr.net)

Licensed Under Creative Commons Attribution CC BY

We next focused on data plotting from depth coverage using Excel program to generate curve for each chromosome that build a compromise between scatter plot and area under the curve. We meticulously analyzed the magnitude and intensity of the oligos binding sites on each chromosome which ultimately leads us to identify the nature and magnitude of chromosome instability.

## 2. Methods

### Libraries preparation and sequencing

#### 1) Genomic DNA (gDNA) libraries construction for Exomes sequencing on Illumina Platform

##### a) Tagmentation reaction

High quality gDNA was purified from cells and whole blood using Qiagen kits (QIAamp DNA Blood Mini Kit from QIAGEN, Hilden Germany). The construction of the library was done using the NexteraTransposome Kit based on Tn transposase chemistry. Brief, 50 ng of gDNA was used in tagmentation reaction which involves the transposase that simultaneously fragments the gDNA and adds adapters sequences at the ends the fragments allowing amplification the generated DNA fragments by PCR in the subsequent steps. After the fragmentation and tagging reaction, the transposase is removed from the reaction product using magnetic beads before performing limited-cycle PCR (suppression PCR). This limited PCR step adds index 1 (i7) and index 2 (i5) as well as common adapters (P5 and P7) required for libraries DNA population to be recognized by the flowcell leading cluster density generation and sequencing. The PCR product is two times enriched two in successive subtractive-hybridization in presence of a cocktail of Exome oligos or expended Exome oligos followed by capture using streptavidin beads.

##### b) Validation of the library

To achieve a highest data quality on the Illumina platform, it is important to create optimum cluster densities across every lane of the flow cell and this requires accurate quantitation of gDNA library templates using Qubit fluorometer (Invitrogen).

##### c) Quality Assessment

Post-enriched library quality assessment and size distribution are assessed on an Agilent Technologies 2100 Bioanalyzer using an Agilent High Sensitivity DNA Chip. Depending on the level of indexing, an initial dilution of the sample can be necessary. For a 12-plex pool, it is recommends to use 1:10 dilution of sample before loading it onto the Chip. Preferably, loading 300 to 400 ng of the post-enriched products on the Chip put the pick of library molecules sandwiched in the middle high molecular and low molecular weight markers thereby allowing proper estimation of fragments size distribution.

##### d) Library Sequencing

Depending On the concentrations and libraries sizes distribution as assessed by Qubit and Agilent Bioanalyzer, 3-6 pM of each library was sequenced for 151 cycles on Illumina MiSeq using V-3 Chemistry Kit, prior to denaturation of libraries DNA in presence of 0.2N NaOH. To ensure molecular diversity, all samples were spiked with 10% phiX. Samples are loaded into the cartridge and sequenced on MiSeq flow cells using Nextera XT program.

##### 2) Bioinformatics Methodology

Sequence data from Illumina MiSeq were processed using the bioinformatics pipeline outlined in **Figure 1** which included concatenation, reference alignment, and removal of duplicate reads, indel realignment and base recalibration. A quality control analysis was performed using FastQC tool to generate an HTML format report which provides information about Basic Statistics, Per base sequence quality report, Per tile sequence quality, Per sequence quality scores, per base sequence content, Per sequence GC content, Per base N content, Sequence length distribution, Sequence duplication levels, Overrepresented sequences, Adapter content and Kmer content. From the FastQC analysis, the first 15 base pairs were trimmed off as these were the adapter sequences. After performing adapter and quality trimming, FastQC analysis was performed again to check the presence of any bases having low scores or adapter regions in the sequence. Post FastQC analysis, reads were mapped to the human genome reference assembly (GRCh37/hg19) using Bowtie2 aligner algorithm which uses paired-end alignment mode for aligning paired-end data to reference genome. Bowtie 2 outputs the alignments into SAM format. The SAM file thus obtained was converted to BAM file using samtools program. The Depth of Coverage was obtained by inputting the BAM files with proper headers into GenomeAnalysisTK.jar tool for coverage statistics analysis per gene and per sample and with the application of all the following Read Filters like: NotPrimaryAlignmentFilter, FailsVendorQualityCheckFilter, DuplicateReadFilter, UnmappedReadFilter, MalformedReadFilter and BadCigarFilter) are automatically applied to the data by the Engine before processing the DepthOfCoverage analysis. Note that, the NotPrimaryAlignmentFilter recognizes the SAM flag that identifies secondary alignments. It is intended to ensure that only reads that are likely to be mapped in the right place and therefore to be informative will be used in the analysis; FailsVendorQualityCheckFilter recognizes the SAM flag corresponding to the vendor quality check; DuplicateReadFilter recognizes the SAM flag set by MarkDuplicates. Like NotPrimaryAlignmentFilter, it recognizes the SAM flag corresponding to being unmapped. MalformedReadFilter is also applied automatically by all GATK tools in order to protect them from crashing on reads that are malformed. There are a few types of malformation (such as the absence of sequence bases) that are not filtered out by default and can cause errors, but these cases can be preempted by setting flags that cause the problem reads to also be filtered.

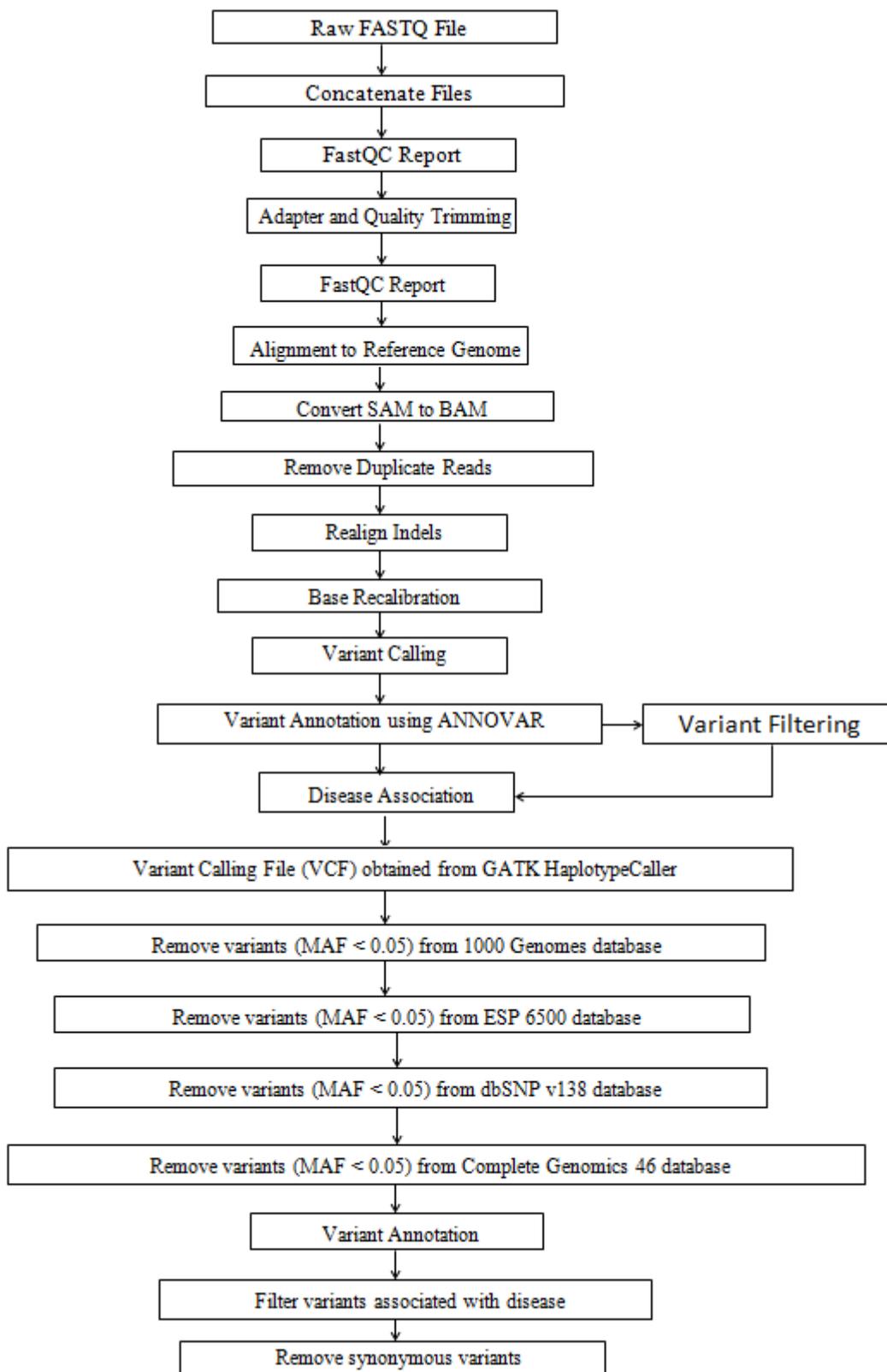


Figure 1: Data Analysis Pipeline

Figure 1: Exome sequencing analysis pipeline listing various steps and tools for processing Miseq data.

### 3) Plotting Chromosomes Binding Curves

Chromosomes binding curves were plotted from the depth of coverage data produced on the spreadsheets after bioinformatics analysis. Each curve was plotted using probes Metrix data on then X axis and the density of the

curve on Y axis. The curves were generated using Excel program that integrates a combination of scattered plot and the area under the curve. To avoid a scattered distribution on the pics of curves, we joined the oligos that depicted the minimum binding sites to their coordinates on the Y-axis and maximum binding sites. This obviates a random distribution of the pics of the curves thereby producing a qualitative curves depicting the chromosomes integrity.

### 3. Results

The **fig1** presents **chromosome 1** binding curves in four different pathologies: anaplasia large lymphoma, hypercholesterolemia, breast cancer and cancer cell line, SUPM2 treated with crizotinib. The gDNA libraries were blended with the same cocktail of oligos and each oligo has its position on X-axis with its corresponding pic on Y-axis. For Anaplasia large lymphoma, the oligo with coordinate **68613** on the X-axis represents the minimum of the binding curve and the oligo **12965** the maximum of the curve. For hypercholesterolemia, the oligo **2264** represents the minimum of the curve and the oligo **5285** the maximum of the curve. In cancer cell line SUPM2, the oligo **631** represents the minimum of the curve and the oligo **148** the maximum of the curve. In breast cancer, the oligo **14748** represents the minimum of the curve and the oligo **14150** the maximum of the curve.

The **fig2** presents **chromosome 3** binding curves with four different pathologies as **fig1**. For anaplasia large lymphoma, the oligo **12884** represents the minimum of the curve against the oligo **12965** representing the maximum of the curve. In hypercholesterolemia, the oligo **4270** represents the minimum of the curve and the oligo **1548** the maximum of the curve. In the cancer cell line SUPM2, the minimum of the curve is represented by the oligo **1968** against the oligo **7176** on the maximum of the curve. In the breast cancer, the minimum of the curve is represented by oligo **7305** against the oligo **11953** on the maximum of the curve.

The **fig3** presents the **chromosome 4** with the same four pathologies. In anaplasia lymphoma, the nature of oligo binding to the chromosome 4 is presented by oligo **15685** at the bottom of the curve against the oligo **10350** on the maximum of the curve. In hypercholesterolemia, the oligo **533** shows the minimum of the curve and the maximum of the curve is represented by the oligo **794**. The nature of chromosome binding in SUPM2 is described by a curve presenting at its minimum the oligo **1071** and its maximum the oligo **1142**. The oligos on Chromosome 4 binding site are represented by the oligo **38** at the minimum of the curve and oligo **41** at the maximum of the curve.

The **fig4** depicts chromosome 6 binding curves in the four different state of pathologies. Likewise on the chromosome 4, breast cancer presents more chromosome shattering as compare to chromosome 3 and chromosome 1.

The **fig5**, 6, 7, 8 and presents respectively: chromosomes 11, 12, 13, 19 and chromosome M, with chromosome M and 13, much more shattered as compared with chromosomes 11, 12 and 19.

### 4. Discussion

The majority of human cancer cells are highly aneuploid harboring chromosome numbers deviating from their normal modal number of 46 (15, 16). In cancer, aneuploidy is a consequence of an increased rate of whole chromosome missegregation during mitosis, a process known as chromosomal instability (CIN) (17). Chromosome instability may be associated with aberrations in mitotic spindle

checkpoints (18) and genes such as *hBUB1* and *MAD2* (19, 20), aberrant sister chromatid exchange (21), DNA repair pathways (22), and abnormal centrosome copy numbers or amplification (23–26). Recently, studies have implicated breakage-fusion-bridge cycles in chromosome instability. Albeit the existence of plethora of studies linking CIN to breast cancer as well as other cancer in general (27–31), there is no direct evidences implicating other diseases like hypercholesterolemia and anaplasia lymphoma to CIN. In breast cancer, aneuploidy (32), absence of *Brca2* gene, the over-expression of *STK15/BTAK* and conditional mutation in *Bracl* exhibits gross genomic instability (33–35). There are two main classes of genomic instability: nucleotide instability and chromosomal instability (CIN) (28). While nucleotide mutations include base substitutions, deletions and insertions, mutations at the cytogenetic level include gains and losses of whole or parts of chromosomes as well as simple or complex chromosomal rearrangements. Here we used the of depth of coverage metrics to demonstrate that the genome instability is much more pronounced in breast tumors as compared with SUPM2 cancer cell line, hypercholesterolemia and anaplasia large lymphoma. A higher population of probe oligos that bind to their sites denote the state of integrity or the stability of the genome before transposase digestion leading to the generation of the libraries. It is important to recall that we used Illumina/Nextera kit to generate these libraries and it is also known that other companies like: Agilent or NimbleGen kits will generate libraries with different magnitudes of depth of coverage resulting from the probes binding efficiency (36–38). In either case, the exome sequencing analysis has proven to demonstrate the stability of the genome. Indeed, other studies have initially reported a comparison performance between four different kits for libraries building from four commercial platforms: Roche/NimbleGen's SeqCap EZ Human Exome Library v3.0, Illumina's Nextera Rapid Capture Exome (v1.2), Agilent's SureSelect XT Human All Exon v5 and Agilent's SureSelect QXT, using the same DNA samples (37). From these studies, it was reported Agilent XT showed the highest target enrichment efficiency and the best SNV and short indel detection sensitivity in coding regions with the least amount of sequencing. Agilent QXT had slightly inferior target enrichment than Agilent XT. Illumina, with additional sequencing, detected SNVs and short indels at the same quality as Agilent XT, and showed the best performance in coverage of medically interesting mutations. NimbleGen detected more SNVs and indels in untranslated regions than the others. Our study was not aimed at comparing the capture performance of each library kit, but the capture performance of Illumina inside different pathologies.

Mitochondrial defects have long been suspected to play an important role in the development and progression of cancer and a key event in carcinogenesis involved the development of an injury to the respiratory machinery, resulting in compensatory increases in glycolytic ATP production (39, 40). Malignant cells produce a large amount of their ATP through glycolytic mechanisms rather than through oxidative phosphorylation. Because glycolytic ATP generation is inefficient by nature, this leads to a high consumption of glucose to fulfill cellular energy requirements, mechanism that contrast with normal cells. In

our study, we reported on fig9 that the mitochondrial genome of patients suffering from hypercholesterolemia and breast cancer, is practically degraded as exemplified by nature of the probes capture compared to large anaplasia lymphoma and the cancer cells line SUPM2.

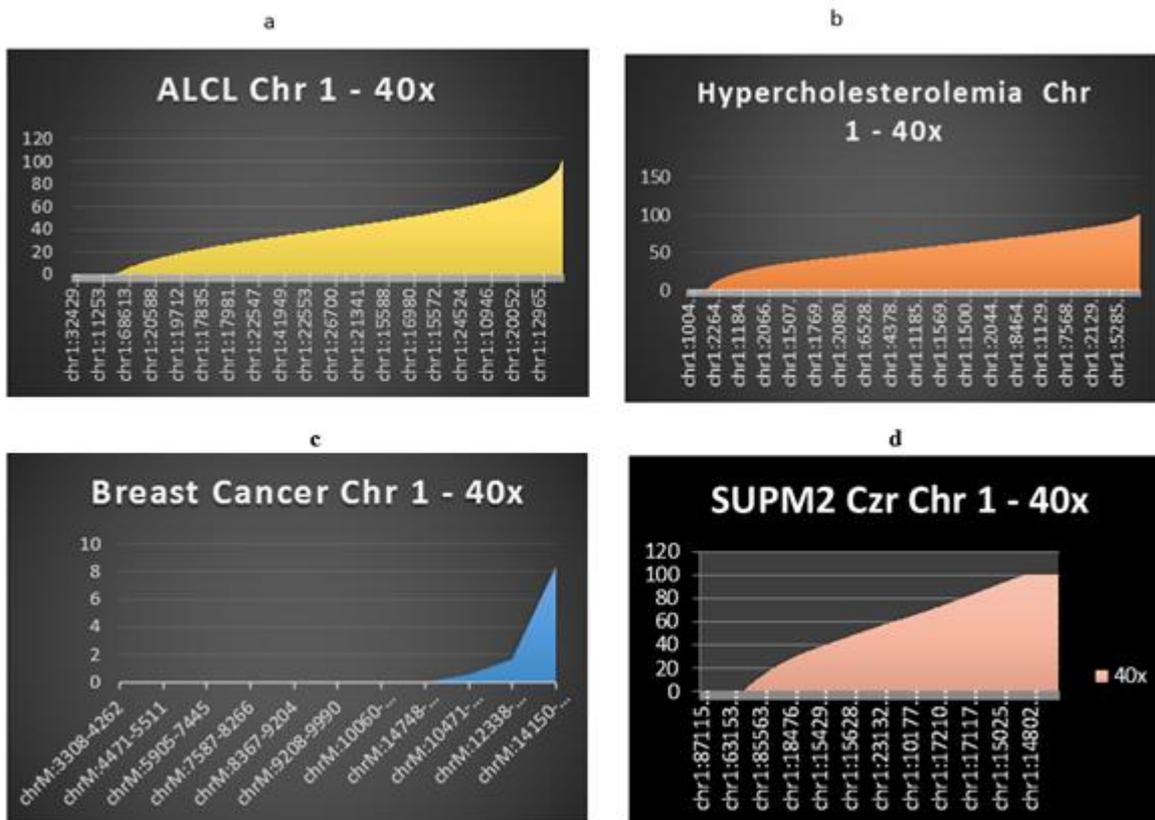
## 5. Conclusion

The depth of coverage is under influence of many factors that are required to properly address biological questions using exome sequencing. The experimental design needs meticulous planning of some intrinsic issues that relate to the integrity of genome structure and its mappability *vis a viso* the relative abundance of reads that inform us about the biological question and the trade-off between controlled replicated designs and sequencing depth. Such an exercise in the quest of depth of coverage in itself constitutes an approximation as it requires a use of presumably true positive control of set of transcripts or binding sites, a setting that the design of the exome sequencing does not have. Here, we used the metrics of depth of coverage derived from our FastQ data analysis on GATK pipeline in combination with excel plotting to access the state of genome stability. More works are required to decipher the nature of interactions between the binding sites on the genome and the probes.

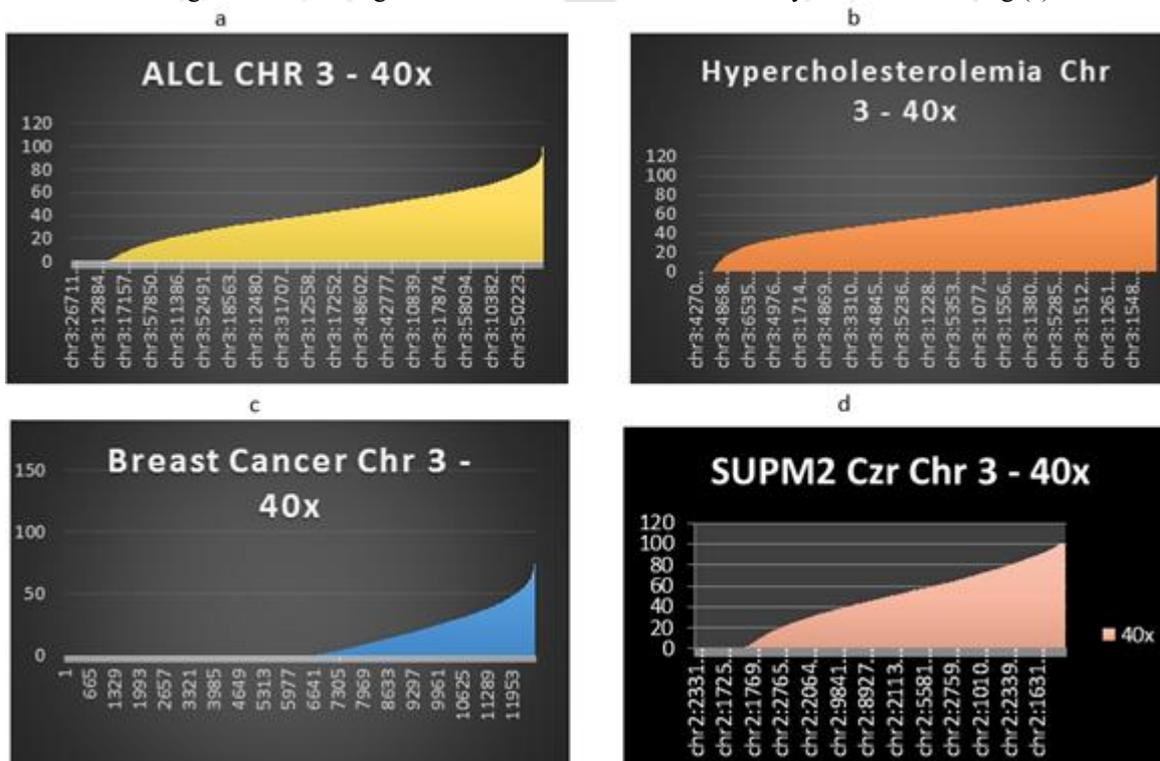
## References

- [1] Francis S. Collins, Eric D. Green, Alan E. Guttmacher and Mark S. **A vision for the future of genomics research.** *Nature.* (2003) Vol 422 835-847.
- [2] Christoph Lengauer, Kenneth W. Kinzler & Bert Vogelstein **Genetic instabilities in human cancers.** *Nature* 396, 643-649 (17 December 1998) (7).
- [3] Hubing Shi, Gatien Moriceau, Xiangju Kong, Mi-Kyung Lee, Hane Lee, Richard C. Koya, Charles Ng, Thinle Chodon, Richard A. Scolyer, Kimberly B. Dahlman, Jeffrey A. Sosman, Richard F. Kefford, Georgina V. Long, Stanley F. Nelson, Antoni Ribas, and Roger S. Lo. **Melanoma whole exome sequencing identifies V600E-RAF amplification-mediated acquired B-RAF inhibitor resistanc.** (2012) *Nat Commun.* ; 3: 724. (3)
- [4] Yaping Yang, Donna M. Muzny, Jeffrey G. Reid, Matthew N. Bainbridge, Alecia Willis, Patricia A. Ward, Alicia Braxton, Joke Beuten, Fan Xia, ZhiyvNiu, Matthew Hardison, Richard Person, Mir Reza Bekheirnia, Magalie S. Leduc, Amelia Kirby, Peter Pham, Jennifer Scull, Min Wang, Yan Ding, Sharon E. Plon, James R. Lupski, Arthur L. Beaudet, Richard A. Gibbs, Christine M. **Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders.** (2013). 17; 369(16): 1502–151.
- [5] Yaping Yang, Donna M. Muzny, Jeffrey G. Reid, Matthew N. Bainbridge, Alecia Willis, Patricia A. Ward, Alicia Braxton, Joke Beuten, Fan Xia, ZhiyvNiu, Matthew Hardison, Richard Person, Mir Reza Bekheirnia, Magalie S. Leduc, Amelia Kirby, Peter Pham, Jennifer Scull, Min Wang, Yan Ding, Sharon E. Plon, James R. Lupski, Arthur L. Beaudet, Richard A. Gibbs, Christine M. **Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders.** (2013). 17; 369(16): 1502–151.
- [6] Bastians H. **Causes of Chromosomal Instability.** *Recent Results Cancer Res.* 2015; 200:95-113.
- [7] Beheshti B<sup>1</sup>, Park PC, Sweet JM, Trachtenberg J, Jewett MA, Squire JA. **Evidence of Chromosomal Instability in Prostate Cancer determined by Spectral Karyotyping (SKY) and Interphase Fish Analysis.** *Neoplasia.* Vol. 3. N.1 2001 pp: 62-69.
- [8] Christoph Lengauer<sup>1</sup>, Kenneth W. Kinzler<sup>1</sup> & Bert Vogelstein. **Genetic instabilities in human cancers.** 1998, *Nature* 396, 643-649.
- [9] Yaping Yang, Donna M. Muzny, Jeffrey G. Reid, Matthew N. Bainbridge, Alecia Willis, Patricia A. Ward, Alicia Braxton, Joke Beuten, Fan Xia, ZhiyvNiu, Matthew Hardison, Richard Person, Mir Reza Bekheirnia, Magalie S. Leduc, Amelia Kirby, Peter Pham, Jennifer Scull, Min Wang, Yan Ding, Sharon E. Plon, James R. Lupski, Arthur L. Beaudet, Richard A. Gibbs, Christine M. **Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders.** (2013). 17; 369(16): 1502–151.
- [10] Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DCY, Nazareth L, Bainbridge M, Dinh H, Jing C, Wheeler DA, et al. 2010. **Whole genome sequencing in a patient with Charcot-Marie-Tooth Neuropathy.** *N Engl J Med* 362: 1181–1191.
- [11] Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, et al. 2008. **Accurate whole human genome sequencing using reversible terminator chemistry.** *Nature* 456: 53–59.
- [12] Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Zhang J, et al. 2008. **The diploid genome sequence of an Asian individual.** *Nature* 456: 60–65.
- [13] Kim JI, Ju YS, Park H, Kim S, Lee S, Yi J-H, Mudge J, Miller NA, Hong D, Bell CJ, et al. 2009. **A highly annotated whole-genome sequence of a Korean individual.** *Nature* 460: 1011–1015.
- [14] Campbell PJ, Stephens PJ, Pleasance ED, O'Meara S, Li H, Santarius T, Stebbings LA, Leroy C, Edkins S, Hardy C, et al. 2008. **Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing.** *Nat Genet* 40: 722–729.
- [15] Duesberg P, Rausch C, Rasnick D, Hehlmann R. **Genetic instability of cancer cells is proportional to their degree of aneuploidy.** *Proc Natl AcadSci USA.* 1998; 95:13692–13697.
- [16] Duesberg P, Rasnick D, Li R, Winters L, Rausch C, Hehlmann R. **How aneuploidy may cause cancer and genetic instability.** *Anticancer Res.* 1999; 19:4887–4906.
- [17] Geigl JB, Obenauf AC, Schwarzbraun T, Speicher MR. **Defining chromosomal instability.** *Trends Genet.* 2008; 24(2):64-69.
- [18] Skibbens RV, Hieter P. **Kinetochores and the checkpoint mechanism that monitors for defects in the chromosome segregation machinery.** *Annu Rev Genet.* 1998; 32:307–337.

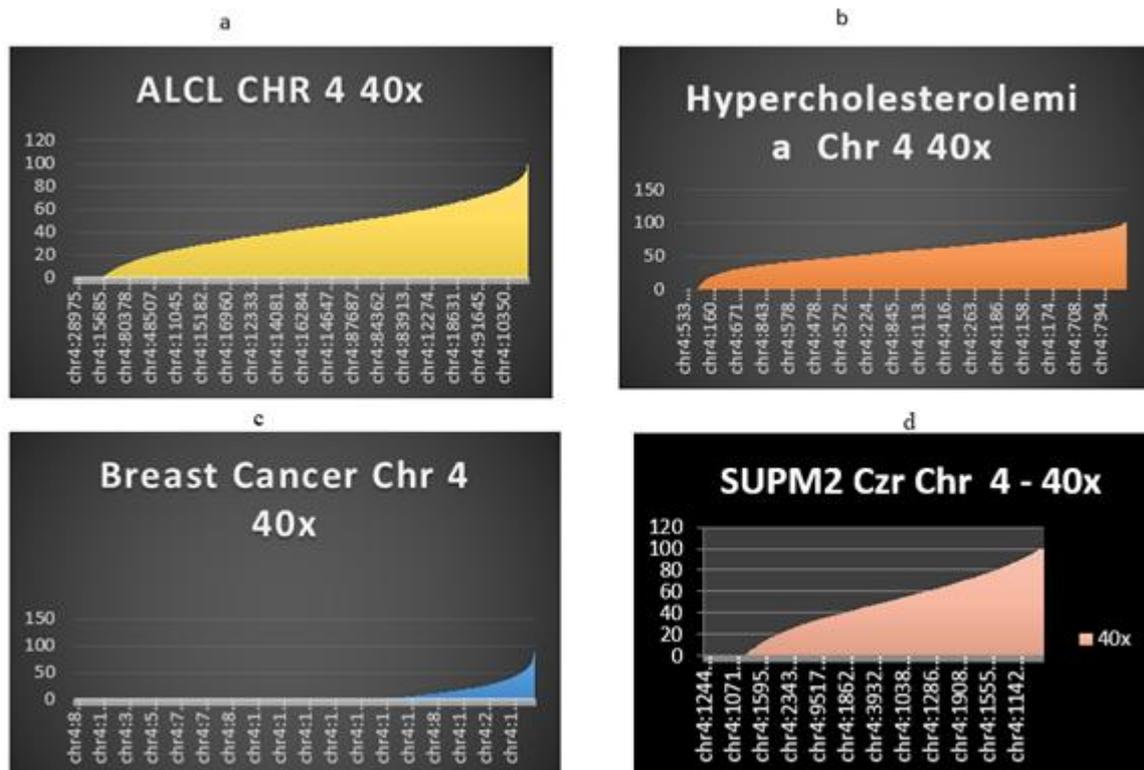
- [19] Pihan GA, Doxsey SJ. **The mitotic machinery as a source of genetic instability in cancer.** *Semin Cancer Biol.* 1999; 9:289–302.
- [20] Cahill DP, da Costa LT, Carson-Walter EB, Kinzler KW, Vogelstein B, Lengauer C. **Characterization of MAD2B and other mitotic spindle checkpoint genes.** *Genomics.* 1999; 58:181–187.
- [21] Dhillon VS, Dhillon IK. **Chromosome aberrations and sister chromatid exchange studies in patients with prostate cancer: possible evidence of chromosome instability.** *Cancer Genet Cytogenet.* 1998; 100:143–147.
- [22] Difilippantonio MJ, Zhu J, Chen HT, Meffre E, Nussenzweig MC, Max EE, Ried T, Nussenzweig A. **DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation.** *Nature.* 2000; 404:510–514.
- [23] Salisbury JL, Whitehead CM, Lingle WL, Barrett SL. **Centrosomes and cancer.** *Biol Cell.* 1999; 91:451–460.
- [24] Ghadimi BM, Sackett DL, Difilippantonio MJ, Schrock E, Neumann T, Jauho A, Auer G, Ried T. **Centrosome amplification and instability occurs exclusively in aneuploid, but not in diploid colorectal cancer cell lines, and correlates with numerical chromosomal aberrations.** *Genes, Chromosomes Cancer.* 2000; 27:183–190.
- [25] Whitehead CM, Salisbury JL. **Regulation and regulatory activities of centrosomes.** *J Cell Biochem Suppl.* 1999; 32–33:192–199.
- [26] Zhou H, Kuang J, Zhong L, Kuo WL, Gray JW, Sahin A, Brinkley BR, Sen S. **Tumour-amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation.** *Nat Genet.* 1998; 20:189–193.
- [27] Mertens F, Johansson B, Hoglund M, Mitelman F. **Chromosomal imbalance maps of malignant solid tumors: a cytogenetic survey of 3185 neoplasms.** *Cancer Res.* 1997; 57(13):2765–80.
- [28] Lengauer C, Kinzler KW, Vogelstein B. **Genetic instabilities in human cancers.** *Nature.* 1998;396(6712):643–9
- [29] Stratton MR, Campbell PJ, Futreal PA. **The cancer genome.** *Nature.* 2009; 458(7239):719–24.
- [30] Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz Jr LA, Kinzler KW. **Cancer genome landscapes.** *Science.* 2013; 339(6127):1546–58.
- [31] Hanahan D, Weinberg RA. **Hallmarks of cancer: the next generation.** *Cell.* 2011; 144(5):646–74.
- [32] Andrew Tutt, Anastasia Gabriel, David Bertwistle, Connor, Hugh Paterson, John Peacock, Gillian Ross and Alan Ashworth. **Absence of Brac2 causes genome instability by chromosome breakage and loss associated with centrosome amplification.**
- [33] Sen S, Zhou H, White RA. **A putative serine/threonine kinase encoding gene BTAK on chromosome 20q13 is amplified and over-expressed in human breast cancer cell lines.** *Oncogene* 1997; 14:2195–200.
- [34] Zhou H, Kuang J, Zhong L, Kuo WL, Gray JW, Sahin A, et al. **Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation.** *Nat Genet* 1998; 20:189–93.
- [35] Tanaka T, Kimura M, Matsunaga K, Fukada D, Mori H, Okano Y. **Centrosomal kinase AIK1 is over-expressed in invasive ductal carcinoma of the breast.** *Cancer Res* 1999; 59:2041– 4.
- [36] Chandra Sekhar Reddy Chilamakuri, Susanne Lorenz, Mohammed-Amin Madoui, Daniel Vodák, Jinchang Sun, EivindHovig, Ola Myklebost and Leonardo A Meza-Zepeda. **Performance comparison of four exome capture systems for deep sequencing.** *BMC Genomics* 2014, 15:449.
- [37] DaichiShigemizu, YukihideMomozawa, Testuo Abe, Takashi Morizono, Keith A. Boroevich, SadaakiTakata, KyotoAshikawa, Michiaki Kubo2, and TatsuhikoTsunoda. **Performance comparison of four commercial human whole-exome capture platforms.** *Scientific Reports | 5:12742 | 2015.*
- [38] Silvia Bonfiglio, Irene Vanni2, Valeria Rossella, Anna Truini, Dejan Lazarevic, Maria Giovanna Dal Bello, Angela Alama, Marco Mora, Erika Rijavec2, Carlo Genova, DavideCittaro, Francesco Grossi and Simona Coco. **Performance comparison of two commercial human whole-exome capture systems on formalin-fixed paraffin embedded lung adenocarcinoma samples.**
- [39] Warburg O: **The metabolism of tumors.** *London Constable Co. Ltd* 1930.
- [40] Warburg O: **On the origin of cancer cells.** *Science* 1956, 123:309-314



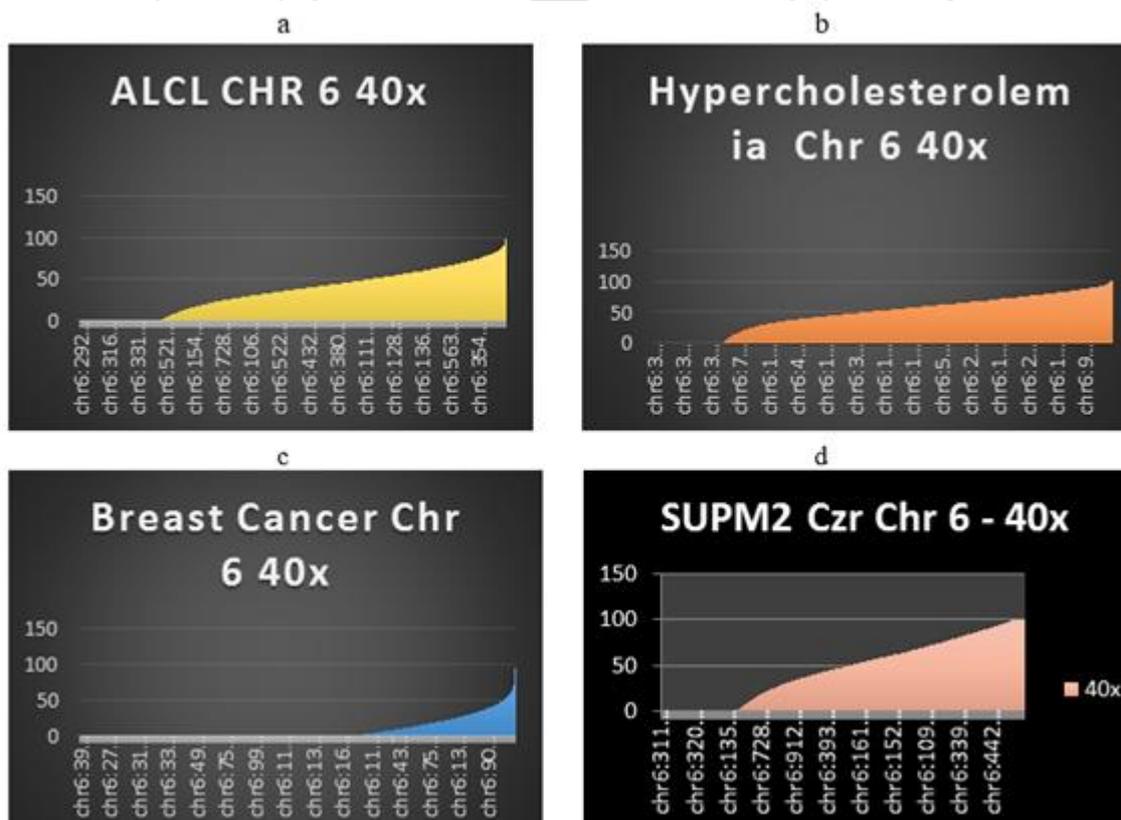
**Figure 1:**Chromosome 1 binding curve for ALCL, Hypercholesterolemia Breast cancer and SUPM2 C2r. The X axis represent the probes coordinates and the Y axis represent the % of binding to a specific locus. The magnitude of CIN is predicted by the ability of the probes on X-axis to bind to its site in the fragments of the library. The relative integrity of the sites will show binding, whereas the degradation of the sites is characterized by absence of binding (c).



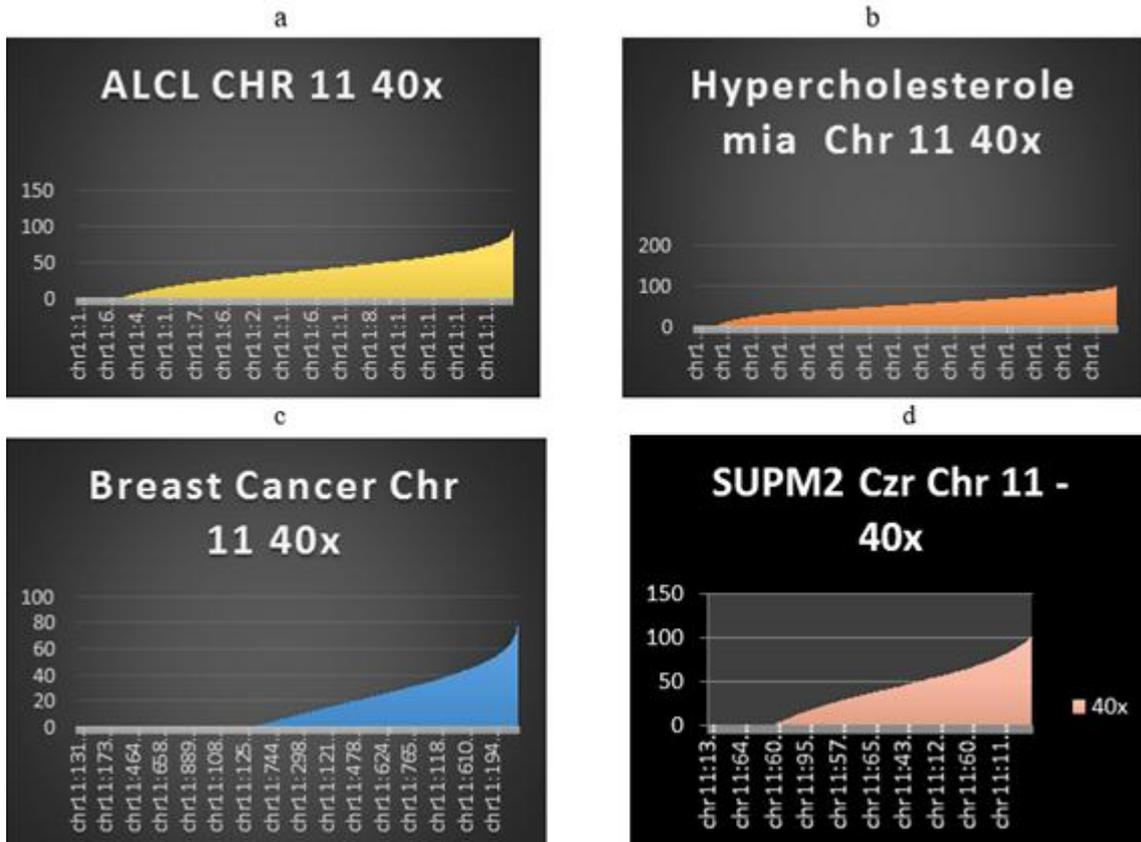
**Figure 2:** Chromosome 3 binding curve for ALCL, Hypercholesterolemia Breast cancer and SUPM2 C2r. The X axis represent the probes coordinates and the Y axis represent the % of binding to a specific locus. The magnitude of CIN is predicted by the ability of the probes on X-axis to bind to its site in the fragments of the library. The relative integrity of the sites will show binding, whereas the degradation of the sites is characterized by absence of binding (c).



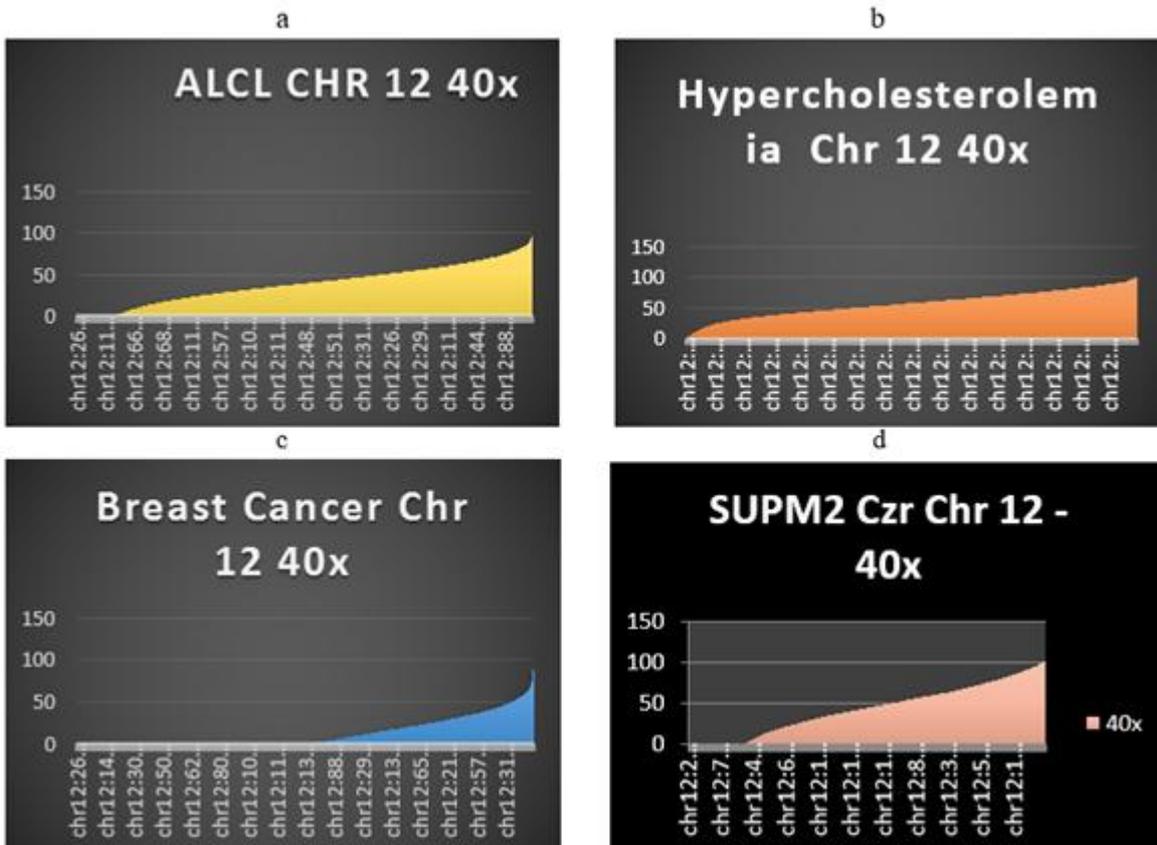
**Figure 3:** Chromosome 4 binding curve for ALCL, Hypercholesterolemia Breast cancer and SUPM2 C2r. The X axis represent the probes coordinates and the Y axis represent the % of binding to a specific locus. The magnitude of CIN is predicted by the ability of the probes on X-axis to bind to its site in the fragments of the library. The relative integrity of the sites will show binding, whereas the degradation of the sites is characterized by absence of binding (c).



**Figure 4:** Chromosome 6 binding curve for ALCL, Hypercholesterolemia Breast cancer and SUPM2 C2r. The X axis represent the probes coordinates and the Y axis represent the % of binding to a specific locus. The magnitude of CIN is predicted by the ability of the probes on X-axis to bind to its site in the fragments of the library. The relative integrity of the sites will show binding, whereas the degradation of the sites is characterized by absence of binding (b and c).

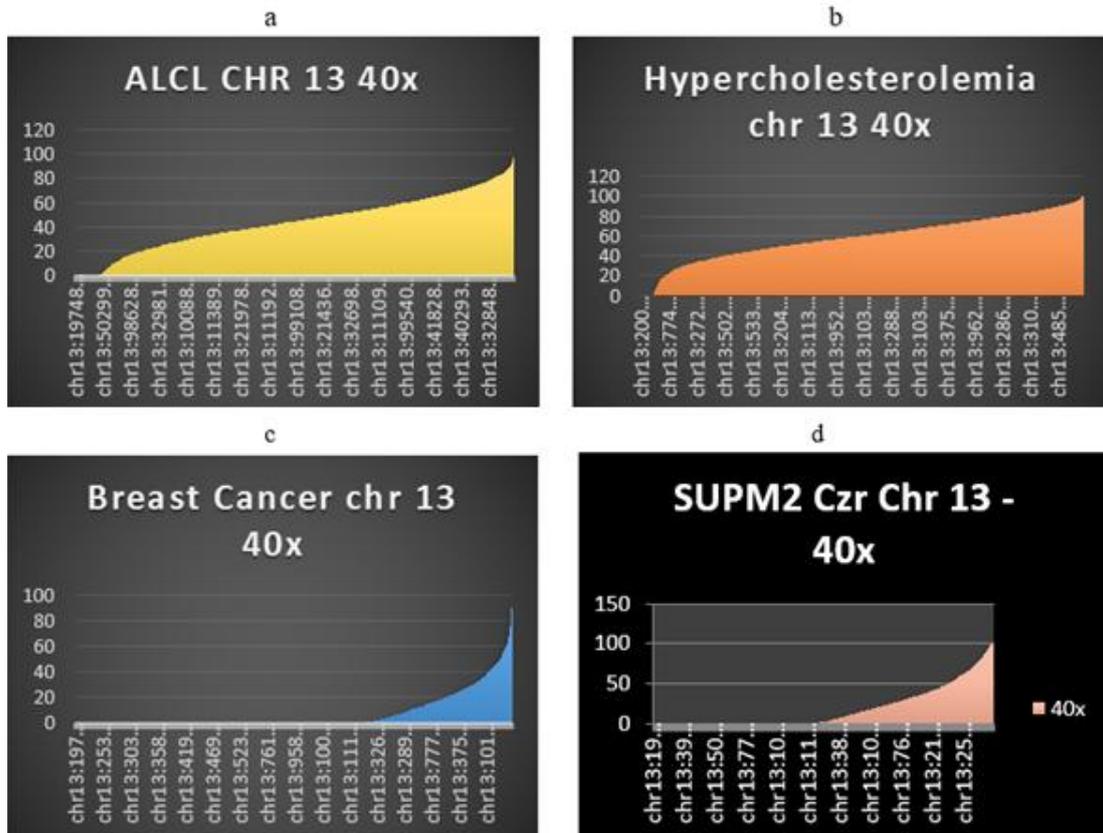


**Figure 5:** Chromosome 11 binding curve for ALCL, Hypercholesterolemia Breast cancer and SUPM2 Czir. The X axis represent the probes coordinates and the Y axis represent the % of binding to a specific locus. The magnitude of CIN is predicted by the ability of the probes on X-axis to bind to its site in the fragments of the library. The relative integrity of the sites will show binding, whereas the degradation of the sites is characterized by absence of binding (b and c).

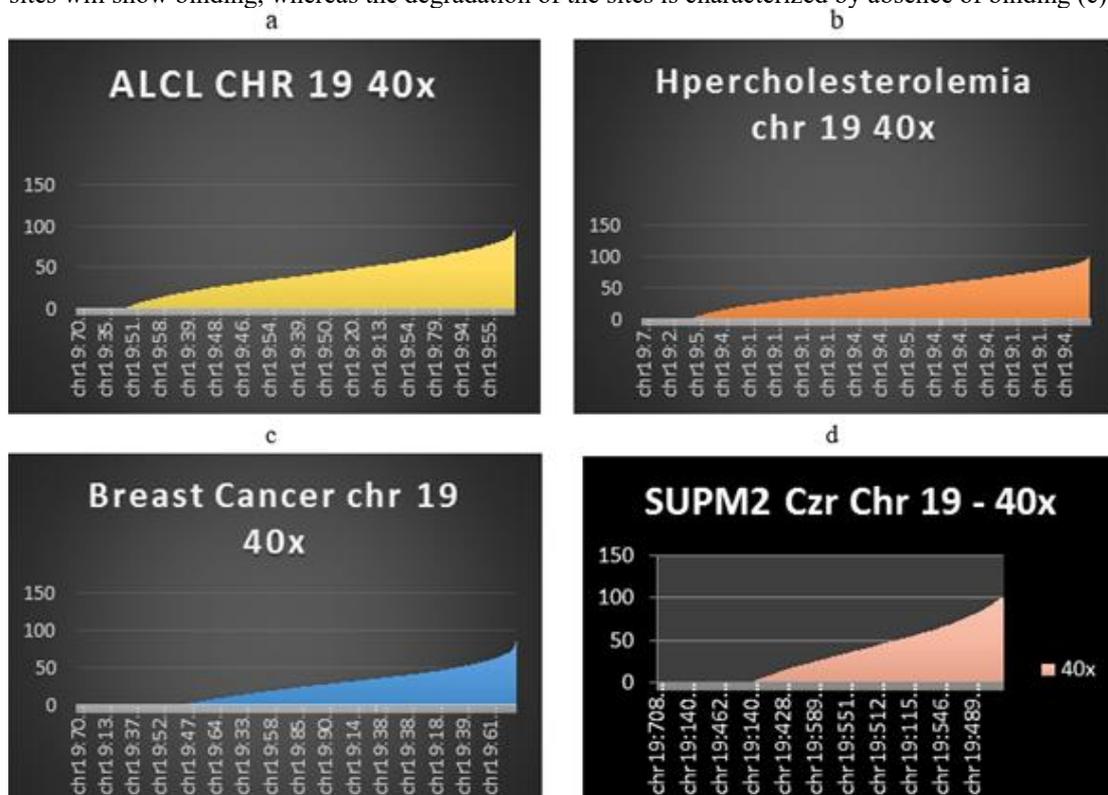


**Figure 6:** Chromosome 12 binding curve for ALCL, Hypercholesterolemia Breast cancer and SUPM2 Czir. The X axis represent the probes coordinates and the Y axis represent the % of binding to a specific locus. The magnitude of CIN is

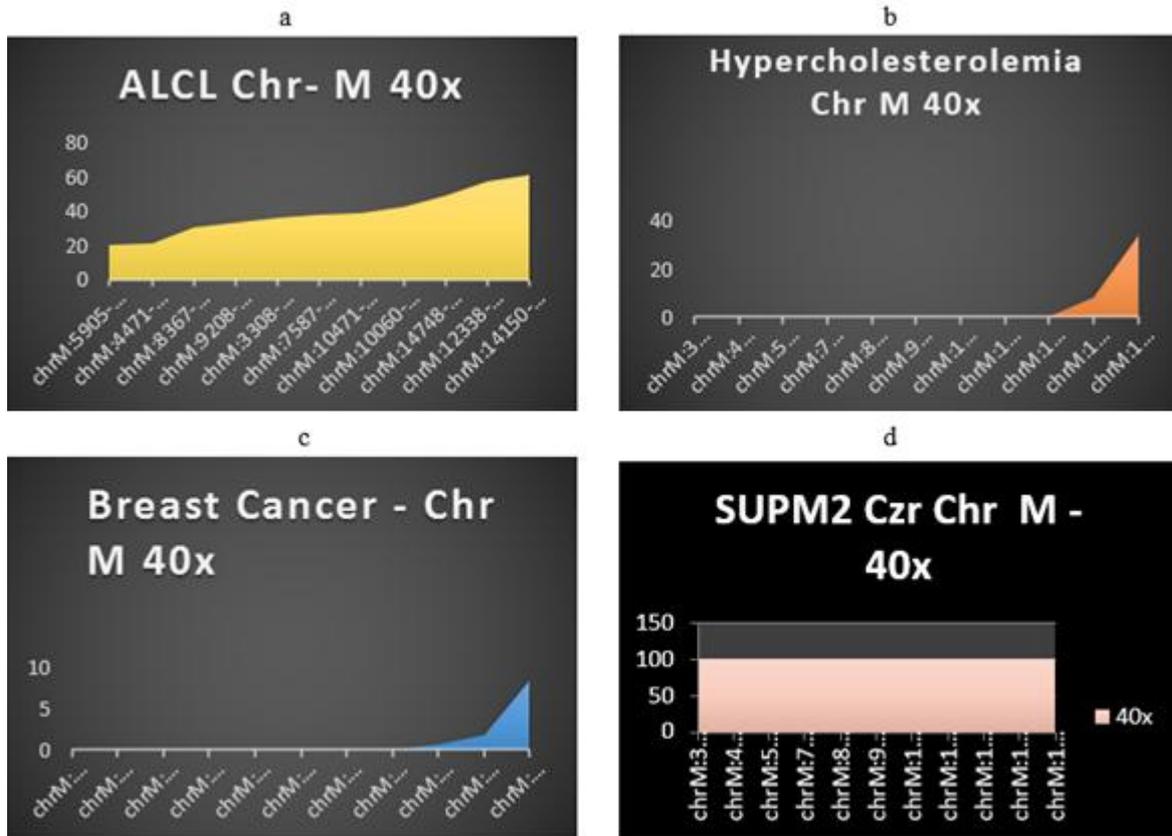
predicted by the ability of the probes on X-axis to bind to its site in the fragments of the library. The relative integrity of the sites will show binding, whereas the degradation of the sites is characterized by absence of binding (b and c).



**Figure7:** Chromosome 13 binding curve for ALCL, Hypercholesterolemia Breast cancer and SUPM2 Czir. The X axis represent the probes coordinates and the Y axis represent the % of binding to a specific locus. The magnitude of CIN is predicted by the ability of the probes on X-axis to bind to its site in the fragments of the library. The relative integrity of the sites will show binding, whereas the degradation of the sites is characterized by absence of binding (c).



**Figure 8:** Chromosome 19 binding curve for ALCL, Hypercholesterolemia Breast cancer and SUPM2 Czir. The X axis represent the probes coordinates and the Y axis represent the % of binding to a specific locus. The magnitude of CIN is predicted by the ability of the probes on X-axis to bind to its site in the fragments of the library. The relative integrity of the sites will show binding, whereas the degradation of the sites is characterized by absence of binding (b and c).



**Figure9:** Chromosome M Saturation curve for ALCL, Hypercholesterolemia Breast cancer and SUPM2 C2r. The X axis represent the probes coordinates and the Y axis represent the % of binding to a specific locus. The magnitude of CIN is predicted by the ability of the probes on X-axis to bind to its site in the fragments of the library. The relative integrity of the sites will show binding, whereas the degradation of the sites is characterized by absence of binding (b and c).

