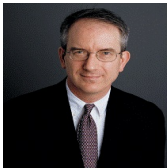# Issue 10 : October, 2013

NCI GENOMIC PROGRAM HIGHLIGHTS

A Roadmap for the Center for Cancer Genomics: An Interview with Dr. Louis Staudt

Dr. Louis Staudt, a member of the National Academy of Sciences, is a leading expert in lymphoma research within NCI's intramural research program. He was recently named the Director of the Center for Cancer Genomics (CCG), the organization that encompasses the Office of Cancer Genomics. In this short interview, Dr.

FEATURED RESEARCHERS

Dr. Marco Marra: Pioneer and Visionary in Cancer Genomics Research

Dr. Marco Marra is a highly distinguished genomics and bioinformatics researcher. He is the Director of Canada's Michael Smith Genome Sciences Centre at the BC Cancer Agency and holds a faculty position at the University of British Columbia.

CTD² PROGRAM HIGHLIGHTS

Decoding the Functional Dimension of the Cancer Genome: Protein-Protein Interaction Networks

Comprehensive molecular characterization of human cancers driven by large-scale genomics initiatives, such as Therapeutically Applicable Research to Generate Effective Treatment (TARGET) and Cancer Genome Characterization Initi
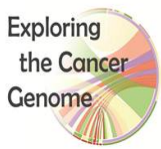
TARGET PROGRAM HIGHLIGHTS

Childhood and Adult Acute Myeloid Leukemia: Genetically Distinct Diseases

Childhood cancers constitute a diverse group of malignancies that are diagnosed in patients ranging in age from newborns to young adults.

Although they are quite rare compared to adult malignancies, pediatric cancers, as a whole, are the leading cause of death from disease in children.

NCI GENOMIC PROGRAM HIGHLIGHTS

A Roadmap for the Center for Cancer Genomics: An Interview with Dr. Louis Staudt

Jessica Mazerik, Ph.D.

Dr. Louis Staudt, a member of the National Academy of Sciences, is a leading expert in lymphoma research within NCI's intramural research program. He was recently named the Director of the Center for Cancer Genomics (CCG), the organization that encompasses the Office of Cancer Genomics. In this short interview, Dr. Staudt discusses the objectives, challenges, and future directions of the Center.

What are the lessons learned from The Cancer Genome Atlas (TCGA) [2] and Therapeutically Applicable Research to Generate Effective Treatments (TARGET) [3], two large-scale genome characterization projects that are near completion? How will CCG build on these insights for future studies?

We have to embrace the complexity of cancer. In order to make therapeutic progress, we need to fully understand the different molecular subtypes and the pathways that are activated in each one. Using a multi –ome approach to view mutations, copy number alterations, RNA and miRNA expression differences, and methylation pattern changes will help identify a constellation of molecular abnormalities that may reveal which biological pathways are affected.

We also must understand that the number of tumors analyzed is important. Statistical analysis shows that if we sequence 500 cases of a particular cancer type, as we have done at TCGA, we observe molecular events that occur in as low as 5-10% of patients. However, there are recurring mutations that affect only 1% of patients with a particular cancer type. Identifying and understanding these mutations and abnormalities may also lead to possible therapeutic interventions. Until we are able to describe rare lesions in common cancers, our analysis is not finished. To

analyze enough tumor samples to accomplish this, we need to take advantage of more affordable advanced technologies and build on what we have learned already from TCGA and TARGET.

What are major challenges facing CCG and the cancer genomics community?

The genetic changes we have identified need to be integrated with functional insight. We know there are many recurrent mutations within a cancer subtype, but we do not understand the biological context of many of those mutations. Do they contribute to tumorigenesis? At what stage of tumor development do they occur? Do these alterations lead to aberrant signaling of particular pathways? One goal within CCG that will help address these questions is to develop new models for the functional study of human cancers. Cell lines, which are limited in number and do not always genetically reflect the primary tumor, are the workhorses for cancer biologists. In the spirit of the CTD^2 initiative, we want to develop more functionally relevant cell lines to study known genetic lesions. Newly available technologies, like organoid cultures, will allow for cell lines to be derived from stem-like progenitors within primary patient material. Because organoids are heterogeneous cell populations with both malignant and supporting stromal cells, they may recapitulate tumor biology more accurately. CCG would like to scale up this technology to provide cancer researchers with similar models of predefined and characterized genetic abnormalities. These tools will allow for functional testing using experimental techniques such as RNA interference.

How will CCG safeguard patient privacy, while simultaneously providing researchers access to clinical data that is critical to making scientific discoveries?

We will perpetuate all of the privacy regulations and permissions currently required to use TARGET and TCGA data. However, we need to improve the model of informed consent, because it creates barriers to research. The current procedure for consent limits the ability of cancer researchers to compare data across different studies. This is especially problematic in pediatric cancer research, where use of the data is restricted to the study of pediatric diseases and cannot include comparative analyses between pediatric and adult samples. To solve this problem, CCG and the cancer community should employ a new "library card" model for informed consent. This model would allow qualified and responsible cancer researchers broad access to genomic samples for many types of studies. Cancer patients typically respond positively when asked about their willingness to participate in this potential "library card" procedure. TCGA is a step in the right direction because a researcher is granted access to all TCGA samples at once without having to separately ask for access to unique tumor types.

What are your visions and future directions for CCG?

The plan going forward is to marry the different NCI divisions and centers that are doing genomics research into one functional unit. This genomic unit will pair closely with clinical trials in order to more rapidly and efficiently develop precision medicine. As part of this partnership, CCG will participate in two NCI clinical trial initiatives, the Exceptional Responders Initiative and the Adjunct Lung Cancer Enrichment Marker

Identification and Sequencing Trial (ALChEMIST). The Exceptional Responders project aims to understand the genetic basis for why some patients have dramatic positive responses to therapies, while others do not. The ALChEMIST aims to identify patients with *EGFR* or *ALK* alterations and treat them with targeted inhibitors.

Additionally, the NCI genomics unit/clinical trials partnership will increase the scope of ongoing analyses at CCG by providing additional biopsies for study. It will build on the successes from TCGA and TARGET by providing opportunities to test new and investigational drugs on molecularly profiled cancers.

Can you talk briefly about how the human genome sequence has impacted your studies on diffuse large B-cell lymphoma (DLBCL)?

It has been a pretty exciting ride. When we started my lab, scientists were getting the first glimpses of the human genome. There was a big project at NCBI called UniGene that clustered human sequence data into individual gene models . We used those primitive sequences to make specialized microarrays that interrogated 5,000 – 6,000 genes involved in lymphocyte function, because we figured these genes would be significantly expressed in lymphoma. This "Lymphochip" was a great tool for identifying subtypes of lymphoma. More recent genomic analysis has determined these subtypes have unique genetic abnormalities and clinical outcomes. Identifying these subtypes would have been easier if we had waited 12 years for next generation sequencing technology, but we were too impatient.

Because we now understand the signaling pathways that are important for lymphomagenesis, we can test drugs that target those pathways. Ibrutinib is a drug that targets a component of the B-cell receptor (BCR) signaling pathway. We predicted the ABC subtype of DLBCL would respond to ibrutinib, because its growth depends on BCR signaling. This prediction was validated by recent clinical trials where 41% of ABC DLBCL cases responded to the drug. These results are promising and suggest we are on the right track in finding viable treatments for lymphoma.

FEATURED RESEARCHERS

Dr. Marco Marra: Pioneer and Visionary in Cancer Genomics Research

Shannon Behrman, Ph.D.

Dr. Marco Marra is a highly distinguished genomics and bioinformatics researcher. He is the Director of Canada's Michael Smith Genome Sciences Centre [4] at the BC Cancer Agency and holds a faculty position at the University of British Columbia. The Centre is a state-of-the-art sequencing facility in Vancouver, Canada, with a major focus on the study of cancers. Many of their research projects are undertaken in collaborations with other Canadian and international

institutions.

Dr. Marra and his laboratory participate in OCG's <u>Cancer Genome Characterization Initiative (CGCI)</u> [5] and <u>Therapeutically Applicable Research to Generate Effective Treatments (TARGET)</u> [3]. For example, they have done extensive genomic analysis on non-Hodgkin lymphoma in CGCI, revealing a complex combination of recurrent mutations, including mutations in genes encoding epigenetic modifiers. The most recent findings from this study were published May 2013 in *Blood*[1]. For this issue of the OCG e-News, we spotlight Dr. Marra in an interview where he discusses his background in genome research, the key to the Centre's success, and the challenges in relating genomic studies to treatment outcome.

What role did you play in the Human Genome Project?

The Genome Sequencing Center at Washington University in St. Louis, now known as The Genome Institute, was a major contributor to the sequencing of the human genome (Human Genome Project). I started my postdoctoral fellowship there in 1994 to begin sequencing the *C. briggsae* genome. Sequencing and assembling smaller genomes with better characterized gene functions allowed us to determine the right systems and tools needed to assemble the sequence of the human genome. Shortly after I started, I was asked by Dr. Waterston, my post-doctoral advisor, to develop and implement technologies to map the human genome and to sequence expressed sequence tags (ESTs). I, along with a team of scientists, used bacterial artificial chromosome (BAC) clones as the sequencing substrate for the human genome itself. The overlapping sequences from the BAC clones, together with the information from the ESTs, were used for the reconstruction and assembly of the whole genome.

Following your work on the human genome, you moved to Vancouver and eventually became Director of the Genome Sciences Centre. What makes the Centre successful in its mission to improve our understanding of the changes that occur in cancer and other disease genomes?

Our founding and continued philosophy is to go beyond creating an inventory of disease-specific alterations to understanding how those alterations affect underlying disease biology. To that end, we integrate informatics and biology in our laboratories. This integration allows us to complement efforts from other groups doing genomics research. Additionally, we are forward-thinking with our research objectives. We try to address the most pressing problems in cancer, such as "How do cancers become resistant to treatment?" and "Why are some cancers sensitive to treatment?" We'd like to understand how alterations in cancer genomes can provide answers. Much work has been done cataloging alterations in malignancies, but more work is needed to relate such catalogs of genetic alterations directly to treatment outcome. That is a hard problem that demands more than just the technology; it demands a strong interface with the clinic.

Does the Centre have infrastructure that connects laboratory and clinical research?

Yes. The centralized infrastructure of the British Columbia Cancer Agency (BCCA) facilitates the integration of cancer genomics with clinical research. The BCCA is a government entity that delivers cancer treatment and cancer control strategies to

the population of British Columbia and captures information on all cancer diagnoses in the province (about 25,000 new cases of cancer annually). Approximately 19,000 of those cases are treated at a BCCA facility, enabling the collection of complete clinical information on each patient. This framework provides important research opportunities that connect genomic information to treatment and treatment outcomes.

How does the Centre keep up with the technological and informatics demands of cancer genomics research?

Well, it's a bit of a struggle. Funds are scarce and technology advances rapidly. Depending on the cost, level of effort, and resources required, each new technology can either be additive or disruptive to our current technological infrastructure. As a result, we take a conservative, cost-effective approach to technology investment. We assess new instruments, and combine our assessments with those from our colleagues, to inform whether or not we systematically adopt a certain technology. Our aim isn't to be at the absolute leading-edge of technology; it's to use the best quality technology to do leading-edge clinically relevant genomics. We are able to accomplish this through our ability to successfully compete for financial support from organizations such as the National Cancer Institute the Canada Foundation for Innovation, and Genome Canada & Genome British Columbia, among other funders.

Recently your group integrated whole genome sequencing data with transcriptome data in non-Hodgkin lymphoma. What are the advantages of taking this type of multi–ome approach to cancer genome discovery?

A multi-ome approach provides an opportunity for a more detailed understanding of malignancies. As we've learned from our research in NHL, having multiple data types offers important clues to disease etiology that individually may be cryptic. Sequencing the whole genomes of NHL allowed us to detect novel relevant mutations and altered pathways (e.g. B-cell homing, which is linked to B-cell maturation) that were not identified by transcriptome or gene expression analyses alone. It also enabled a more accurate calculation of mutation prevalence in lymphomas. By combining transcriptome and whole genome data, we identified one of the most frequently mutated genes in follicular lymphoma. *MLL2,* a gene involved in histone modification, is inactivated in about 90% of follicular lymphomas (FL) and 30-60 % of diffuse large B-cell lymphomas (DLBCL). Frequent mutations in another histone modifying gene, *EZH2*, were also detected in both FL and DLBCL in our multi –ome analyses. These results suggest that *MLL2* and *EZH2* are fundamentally important in lymphoma biology, and they both play roles in regulating gene expression.

Taken together, sequencing transcriptomes and genomes led directly to the discovery that the epigenome may play a major role in lymphomas. Now we must interrogate the epigenome if we want a more complete understanding of how mutations like *MLL2* and *EZH2* are influencing the disease.

What are the underlying challenges of integrative analysis that your group has encountered in their study of NHL?

We need to understand how the NHL cancer genome evolves to help refine treatment

strategies in the future. Presumably, the tumor's environment influences the evolution of the disease, but we don't know how. We also need to recognize in our sequencing analyses that there is no such thing as THE genome, even within a tumor sample. Tumors are a complex community of cells, which include individuals that have different genotypes and different biological properties. Thus, multiple samples from each tumor are likely to be more informative of biology than a single sample. Analysis of longitudinally collected samples, linked to treatment data, will inform on mechanisms and patterns of tumor evolution.

Understanding both tumor evolution and intra-tumoral genetic heterogeneity will help unravel some of the mysteries of treatment-resistant cancer. Treatment resistance can occur when a drug selects for resistant genotypes in a sub-population of cells in a tumor (it can also occur when a tumor does not respond to a particular treatment). What is the selective pressure applied by the treatment? What does the selective pressure act on? How can we avoid creating a treatment-resistant super-cancer? These are all pressing questions that we as researchers must answer in order to make a greater impact on patient outcomes.

To gain a deeper understanding of selective pressure exerted by treatment, we need better models of cancer progression under treatment and better strategies for sequencing analysis of tumors. Our current approach to sequencing analysis is to take a consensus view of genomic alterations at one point in time. This approach may not be all that relevant to treatment outcome, because it is not reflective of the genomic composition of the malignancy under selective pressure of treatment. A tumor's genomic composition changes over time due to selection, high mutation rate, and other factors. This challenge must be addressed in future genomics studies.

References

1. Morin RD, Mungall K, Pleasance E, Mungall AJ, Goya R, Huff R, Scott DW, Ding J, Roth A, Chiu R, Corbett RD, Chan FC, Mendez-Lago M, Trinh DL, et al.(2013) Mutational and structural analysis of diffuse large B-cell lymphoma using whole genome sequencing. *Blood* 122(7):1256-65 (PMID: 23699601)

Back to Top

CTD² PROGRAM HIGHLIGHTS

Decoding the Functional Dimension of the Cancer Genome: Protein-Protein Interaction Networks

Haian Fu, Ph.D.

Comprehensive molecular characterization of human cancers driven by large-scale genomics initiatives, such as <u>Therapeutically Applicable Research to Generate Effective Treatment (TARGET)</u> [3] and <u>Cancer Genome Characterization Initiative (CGCI)</u> [5], has led to the generation of vast amounts of tumor-derived sequencing data. Integrated analysis of such data has revealed driver genes that are essential for the initiation and progression of cancer, and has also re-defined the molecular subtypes of cancers. However, it remains a daunting challenge to determine how these genomic changes can be exploited to rapidly develop new therapies.

The NCI <u>Cancer Target Discovery and Development (CTD^2)</u> [6] Network strives to accelerate the translation of these enormous volumes of information into efficacious genomics-based treatments for cancer patients. In doing so, the CTD^2 Network utilizes growing cancer genomics data to identify and exploit tumor dependencies with high-throughput functional genomics and systems biology approaches. The Emory Center, which is part of the Network, focuses specifically on delineating the functional effects of genetic alterations by interrogating the protein-protein interaction (PPI) networks of tumors for therapeutic target discovery.

Translating Genomics Data into PPI Networks for Therapeutic Manipulation

Cellular signaling pathways and networks are made up of many interacting proteins which respond to external and internal signals/changes. Often the proteins interact together (via PPIs) to carry out their function. PPIs act as "communication centers" that integrate and propagate biological signals through these networks to exert their impact on cell fate. Therefore, the effect of a misregulated or mutated protein can spread along a signaling network to alter the activity of normal downstream proteins, inducing a pathological phenotype. For example, activating mutations in epidermal growth factor receptors often lead to persistent stimulation of crucial regulatory proteins, such as AKT, ERK and STAT, which drive uncontrolled cell survival, proliferation, and tumor growth. Accordingly, disrupting PPIs at the critical junctions of growth signaling networks may attenuate tumorigenesis.

By defining PPI networks in tumors, researchers may identify novel therapeutic strategies. Many of the cancer driver mutations revealed by cancer genome projects are found in gene products that are traditionally challenging to target therapeutically, such as tumor suppressors and genes that encode proteins with no enzymatic activity. Targeting the nodes and hubs downstream of inactivated tumor suppressors, for example, could restore tumor suppressive function. Thus, understanding how cancer driver mutations are integrated within growth control signaling networks may present new opportunities for pathway perturbation and novel therapeutic discovery in tumors, even with "undruggable" targets.

Discovering Cancer Targets in Re-Wired PPI Networks using a High-Throughput Approach

To leverage the tremendous advances in genomics for accelerated therapeutic discovery, the Emory Center maps and interrogates signaling networks of molecularly characterized cancers to identify important PPIs for potential therapeutic targeting

(Figure 1). We examine the interconnectivity of candidate gene products identified from genomic initiatives through high-throughput PPI screening assays. PPIs identified based on statistical parameters are used to build maps of PPI networks. We have confirmed known PPIs and also identified new PPIs among "cancer driver" gene products and their associated proteins. Using our PPI network maps, we systematically scan and compare the interaction profiles of wild-type driver genes with their mutant counterparts. Our goal is to discover mutation-dependent re-wired PPI networks that may be specifically targeted in tumors.

Figure 1. *The Emory Center's approach to accelerating genomics-based cancer target and therapeutic discovery*
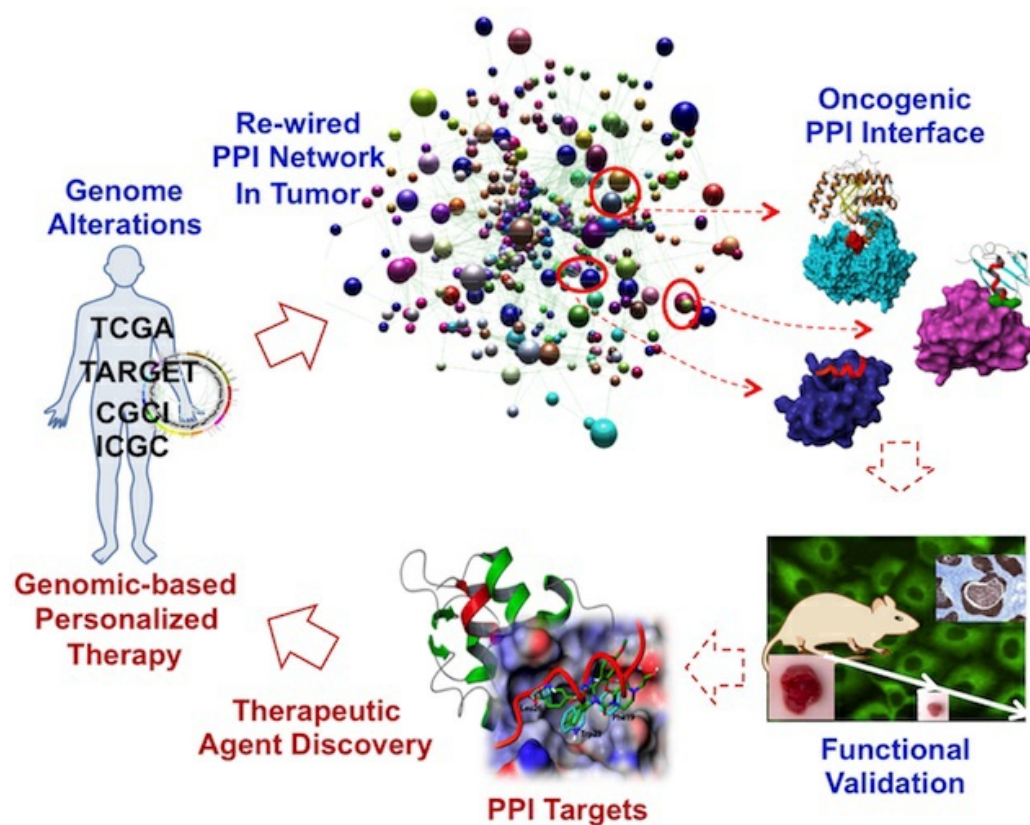
To uncover the potential therapeutic significance of the identified oncogenic PPIs, computational and experimental approaches are combined to rapidly define minimal PPI interfaces (PPI targets) and discover protein fragments that disrupt the intended PPIs. These fragments, or antagonist peptides, are used to determine if selected PPIs are required for tumorigenesis and progression in cancer cell lines. Innovative tumor models developed in the CTD^2 network will be employed for oncogenic PPI validation. Supporting evidence from studies that employ these models along with patient-derived information will be used to advance the definition of a particular PPI as a potential drug target and/or biomarker.

Functional peptide antagonists that are identified through this screening process may be further developed as therapeutic agents through innovative technologies, such as peptide stapling (chemically stabilized peptides that are cell permeable) and nanoparticles. In addition, the molecular features of antagonist peptides that allow for interaction with PPI interfaces can also be used for *in silico* screening to design

new therapeutic agents. The defined peptides and PPI interfaces can provide the basis for novel high-throughput screening for the discovery of small molecule modulators. This targeted approach is expected to help develop therapeutic agents with much improved tumor selectivity, and thus, fewer side effects.

Integrating PPI Discoveries with the CTD^2 Network

Emory's genomics-based PPI network interrogation approach is an integral component of the CTD^2 network, which allows timely data sharing and close collaboration. For example, powerful informatics tools developed in CTD^2 centers integrate PPI results with cancer cell response profiling data from genome-wide RNA interference (RNAi) and pharmacological agent screens to identify molecular signatures for tumor dependency. Together, integrated analysis of data sets from multiple platforms, including RNAi, chemical agents, and PPI screens, will accelerate the discovery and translation of novel cancer targets to patient care.

Placing cancer driver genes in the context of signaling PPI networks offers unique opportunities to design the most promising therapeutic strategies for protein targets, with or without enzymatic activity. Such opportunities rely heavily on vigorous functional validation in relevant cancer models, and linking the PPI node and hub functions to patient outcomes in a team science setting. Together, we can leverage our expertise and technology platforms to interrogate the once "undruggable" space of PPI interfaces and develop the next generation of pathway-perturbation agents to selectively disrupt re-wired oncogenic signaling networks for genomic-based precision medicine.

TARGET PROGRAM HIGHLIGHTS

## Childhood and Adult Acute Myeloid Leukemia: Genetically Distinct Diseases

Soheil Meshinchi, M.D., Ph.D.

Childhood cancers constitute a diverse group of malignancies that are diagnosed in patients ranging in age from newborns to young adults. Although they are quite rare compared to adult malignancies, pediatric cancers, as a whole, are the leading cause of death from disease in children. Coordinated efforts from large cooperative groups, such as the Children's Oncology Group (COG), have helped reduce mortality more than 50% between 1975 and 2002[1]. Despite this considerable improvement in overall outcome, mortality rates have plateaued in recent years. Furthermore, conventional treatment regimens are particularly harsh on developing children, because the protocols are largely modified adult protocols. In children, they often lead to devastating effects, such as developmental delays and infertility. Therefore, a novel approach to treatment is needed. Because targeted therapies can be less toxic and more effective than current cancer treatments, agents directed at underlying genomic abnormalities have the potential to improve both survival and quality of life of pediatric patients.

Many childhood cancers appear as genetically distinct diseases from their adult counterparts and will benefit from independent genomic studies and novel therapeutic strategies. Acute myeloid leukemia (AML), a heterogeneous group of myeloid cancers, is a prime example of a cancer type that demonstrates different genomic signatures between pediatric and adult patients despite phenotypic similarities (Figure 1). For instance, most AML patients less than 2 years of age have MLL translocations (11q23) (~60% of all AML < 1 year) and it declines with increasing age (<5% in adults). *MLL* is a methyltransferase gene involved in the regulation of hematopoietic differentiation and proliferation. Conversely, the normal karyotype (NL) AML is much less common in young children, but increases in prevalence with age (<10% in children under 2 years; 40% in adults). Patients with NL lack identifiable cytogenetic alterations, and are enriched in *FLT3* and *NPM* somatic mutations. *FLT3* regulates stem cell differentiation and proliferation, and *NPM* participates in many basic cellular processes, including biosynthesis of ribosomes and regulation of centrosome duplication. Similar differences in age-related prevalence of Core Binding Factor (*CBF*) alterations (*inv(16)* or *t(8;21)* translocations) have also been observed. Finally, recent studies identified novel clinically relevant genomic alterations (*IDH1*, *DNMT3A*) in adult AML that were absent in childhood AML (data not shown in the figure), providing more support that these cancers are distinct diseases with distinct underlying etiologies.

Figure 1. *Prevalence of specific karyotypic and genomic alterations in different age groups of AML patients*[2]
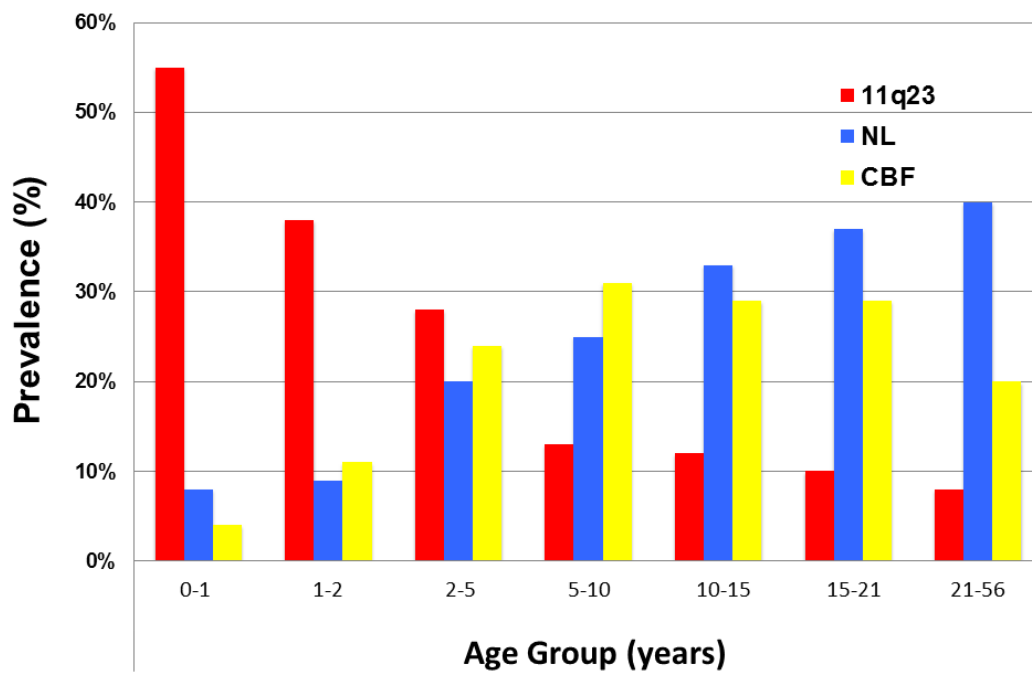
Age-associated variations in AML incidence are also detected (Figure 2). A decline in AML incidence is observed over the first decade of life from its highest rate (infant) to its lowest rate (9 years of age). This trend starts to reverse by age 10, where there is a slow increase in incidence up to age 40. This is followed by two successive bursts in AML incidences in older populations up to the highest recorded levels at age 84.

Figure 2. *SEER data on AML incidence in different age groups from newborn to 84 years of age*[3]
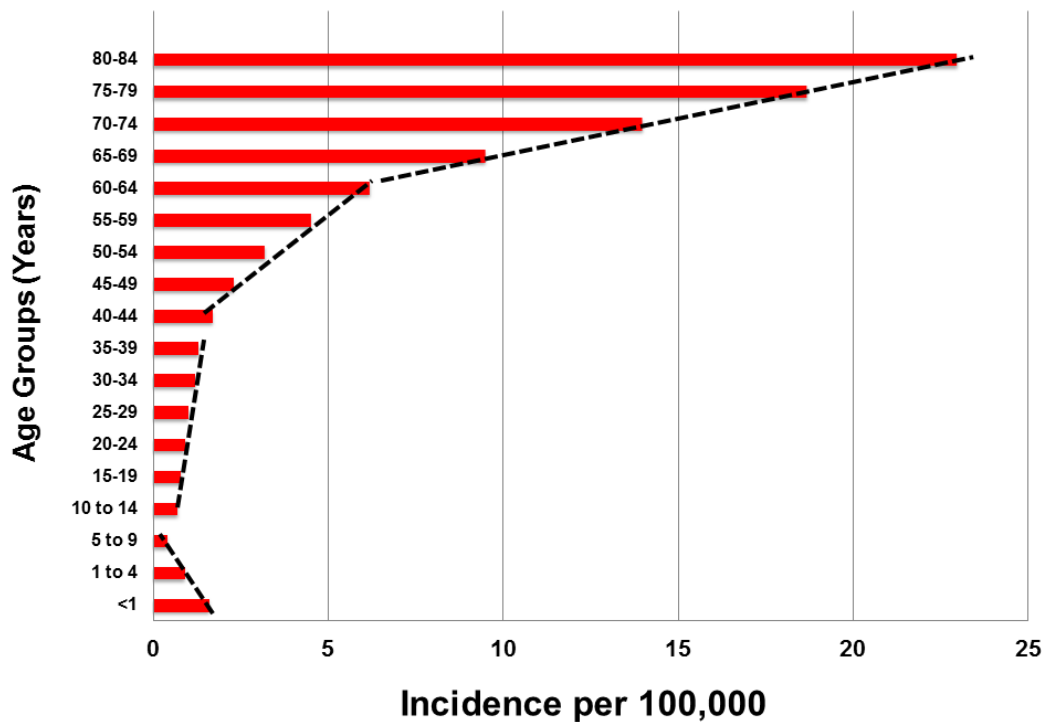
The differences in both the genomic makeup and frequency of AML between children and adults suggest that different molecular mechanisms contribute to AML tumorigenesis through various stages of life. This age-associated molecular variation underscores a need to develop therapies that target the underlying molecular aberrations identified in both children and adult AML patients.

In the last decade, there have been a number of large-scale genomic studies interrogating cancer genomes for clinically actionable genomic alterations. Some of them exclusively study childhood cancers (e.g. Therapeutically Applicable Research to Generate Effective Treatments [3]; TARGET) or adult cancers (e.g. The Cancer Genome Atlas [2]; TCGA), while others study both (e.g. Cancer Genome Characterization Initiative [5]; CGCI). Through extensive sequencing and genome-wide analysis, the pediatric initiative TARGET will help identify novel molecules for clinical intervention in five pediatric cancer types, including AML, neuroblastoma, kidney tumors, osteosarcoma, and acute lymphoblastic leukemia. TARGET will soon provide comprehensive data and analyses on the whole genomes, transcriptomes (RNA-seq, miRNA-seq) and epigenomes of these diseases. This multi –omic information will be integrated to determine potential therapeutic targets as well as biomarkers for risk identification in pediatric patients to improve outcomes.

As more genomic data on various malignancies are being generated, it is important for investigators to consider the age of patients in their interpretation of the data. As we are learning with AML, the stage of development can be a factor in the underlying disease processes of tumorigenesis. Moreover, frequency of mutation varies between age groups, highlighting the need for targeted therapies based on these differences. Molecular differences identified between certain age cohorts should be accounted for in future clinical studies to ensure that appropriate treatment is available for all patients.

References:

1. https://www.cancer.gov/types/leukemia/hp/child-aml-treatment-pdq?redirec... [7]
2. Horan JT, Hasle H, Meshinchi S. Acute Myeloid Leukemia. In: Smith FO, GH R, JM aR, editors. Hematopoietic Cell Transplantation in Children with Cancer: Springer; 2013.
3. Altekruse S, Kosary C, Krapcho M, Neyman N, Aminou R, Waldron W, et al. SEER Cancer Statistics Review, 1975-2007, National Cancer Institute. Bethesda, MD.

Back to Top

EXPLORING CANCER GENOMES

cBioPortal: A Web Platform of Gene-Based Data Exploration

Gene Gillespie, Ph.D.
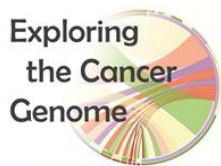
The OCG e-News "Exploring Cancer Genomes" series highlights visualization and analysis tools used by the cancer research community to extract actionable hypotheses from large-scale genomic datasets. In the latest edition, we spotlight the cBioPortal for Cancer Genomics, a tool developed at Memorial Sloan Kettering Cancer Center's Computational Biology Center (cBio).

Large-scale genomic studies, such as Therapeutically Applicable Research to Generate Effective Treatments (TARGET) [3] and The Cancer Genome Atlas (TCGA) [2], are generating an unprecedented amount of data on a significant number of tumors. Much of this raw data is made available to the research community, but the sizes of these datasets pose an enormous barrier to analysis. To address this shortcoming, Memorial Sloan Kettering Cancer Center's Computational Biology Center (cBio) has developed the cBioPortal for Cancer Genomics [8]. The cBioPortal is a web application platform that facilitates data exploration and analysis through the use of a variety of visualization and analytical tools. Some of the tools in cBioPortal, such as the network mapper Cytoscape [9], were developed by external sources, and some were developed by the cBio engineers themselves. The tools are threaded together seamlessly through cBioPortal's user-friendly interface that was designed to be accessible to all researchers, regardless of their level of informatics expertise.

The cBioPortal allows users to interrogate datasets across genes, samples, and data types, giving them the opportunity to examine a number of different biologically and/or clinically relevant hypotheses. The cBioPortal currently hosts more than 40 datasets from TCGA and other large-scale genomic studies, and makes them available for bulk download. **Data from OCG's TARGET Initiative will be added to its database in the next year**. The data types from the 13,000+ tumor samples include mutations, copy number alterations, mRNA expression changes, and DNA methylation values, as well as clinical parameters, such as disease-free survival.

cBioPortal provides "gene-based" visualizations and analyses, so that users can find

altered genes and/or networks within a study of interest or across all studies. Initiating a search is simple with the cBioPortal's four-step "Query" web interface (<u>Figure 1</u>). Users select the desired: 1) cancer study, 2) genomic profile(s), such as mutations and copy number alterations, 3) patient case set, and 4) gene sets of interest. Gene sets can be entered manually or selected from pre-loaded cancer pathways derived from <u>Pathway Commons</u> [10]. Following the submission of a query, the cBioPortal generates condensed, easy-to-navigate data summary readouts that are organized in a series of clickable tabs (<u>Figure 2</u>). A few of these readouts are highlighted below. Visit the <u>cBioPortal website</u> [8] to explore all of them.

Figure 1: The cBioPortal query Web interface

OncoPrint and Mutual Exclusivity

The OncoPrint is a graphical representation of alterations in genes across multiple tumor samples (<u>Figure 2</u>). Individual patient cases are designated by columns, and individual genes are designated by rows. To simplify analysis, genes are discretely classified as either "altered" or "unaltered" in each patient case. Different colors, shapes, and symbols (e.g. arrowheads) distinguish the type of alteration(s), such as homozygous deletions, amplifications, differential expression, and mutations.

**Figure 2**: OncoPrint of genomic alterations in samples from a TCGA lung squamous cell carcinoma study
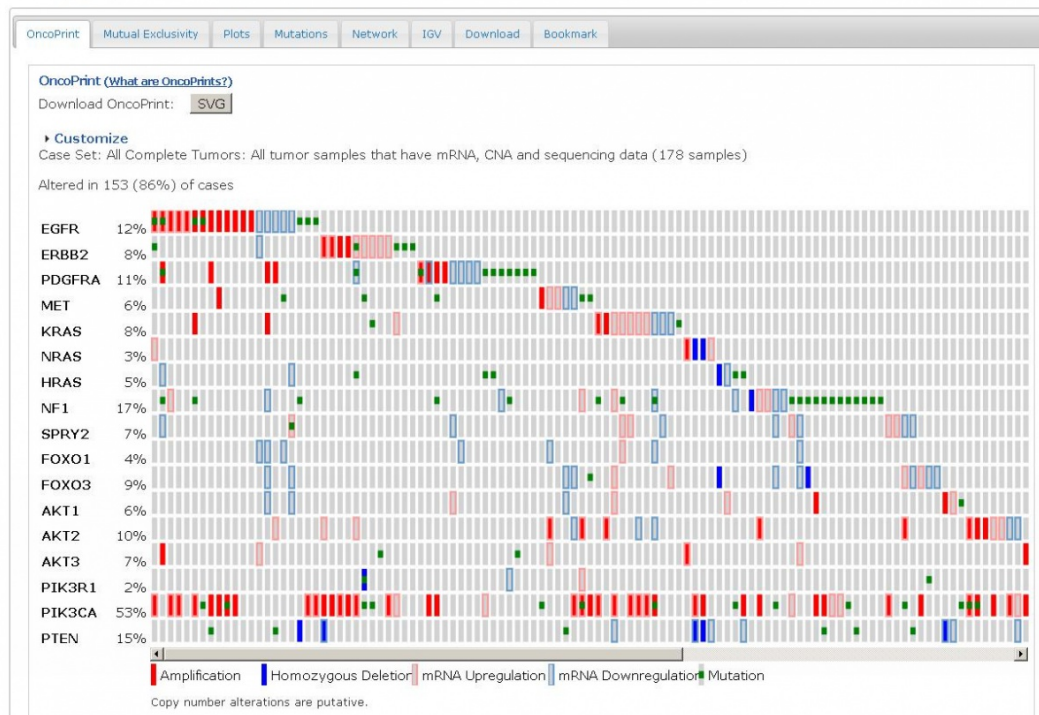
The OncoPrint is especially useful for visualizing alterations across a set of cases and identifying trends, such as co-occurrence and mutual exclusivity. Genes that do not co-occur in the same patients may ultimately be in the same pathway. Identifying these genes exposes opportunities for synthetic lethal therapies. The significance level of co-occurrence and mutual exclusivity, calculated by the statistical method Mutual Exclusivity Modules (MEMo), is found under the Mutual Exclusivity tab.

When gene queries are performed across all studies, the Portal returns a graphical summary of alterations, as well as OncoPrints for every individual study. The summary is a histogram that shows the percentage or total number of samples per study in which at least one of the query genes is altered (Figure 3).

Figure 3: *Histogram summarizing percent of tumor samples with at least one gene altered in selected gene network for all cancer studies in the cBioPortal*
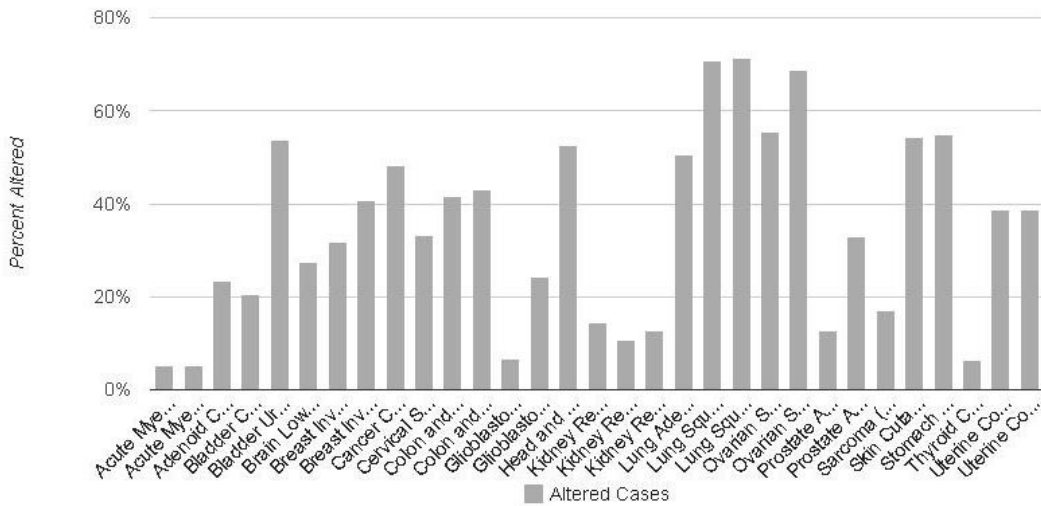
Plots

The OncoPrint displays discrete gene values (i.e., whether a gene is altered or not based on a defined threshold) for all data types, even for data with continuous values such as mRNA expression. The Plots tab offers a way to visualize continuous and/or discrete data for one or two genes across samples in pairwise comparison plots. For instance, mRNA expression of one gene (e.g. TP53) can be plotted against copy number alterations from the same locus (Figure 4). Alternatively, mRNA expression of one gene can be plotted against the mRNA expression of another.

Figure 4: A plot comparing TP53 mRNA levels to putative copy alterations from GISTIC
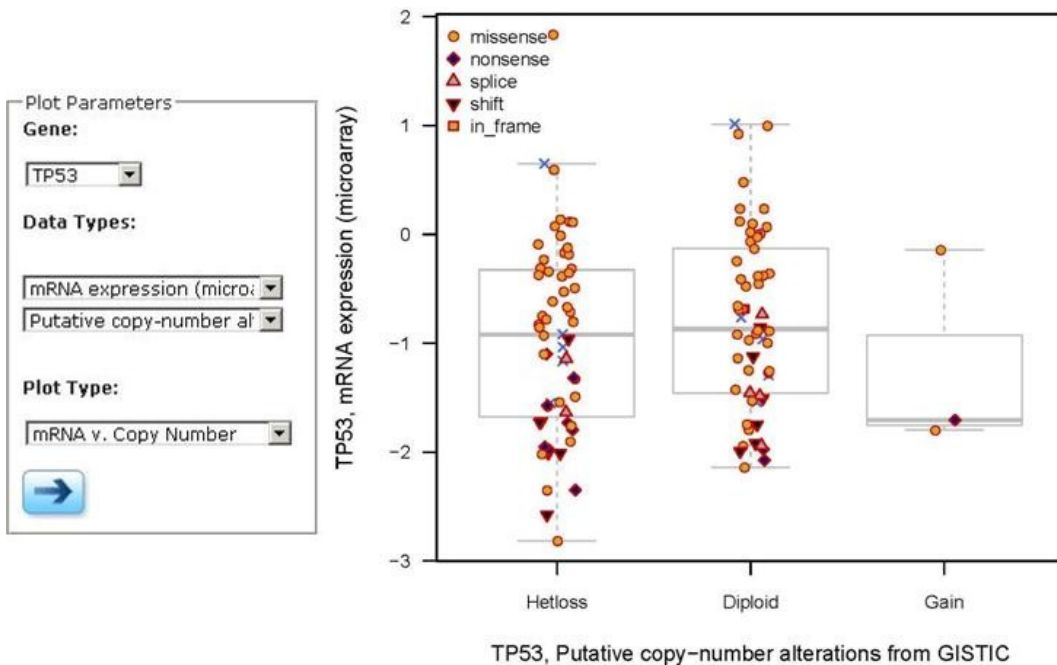
Network

Under the Network tab, the Portal constructs a map consisting of the query genes and up to 50 of the most highly altered neighboring genes in the selected cancer study (Figure 5). This map uses network and interaction data derived from the Pathway Commons Project [11]. The 50 gene limit helps minimize the complexity of each map, but the full network can be viewed using the independent application Cytoscape [9].

Figure 5: An interactive network map of query genes (p53 signaling pathway) and highly altered neighboring genes from a TCGA lung squamous cell carcinoma study
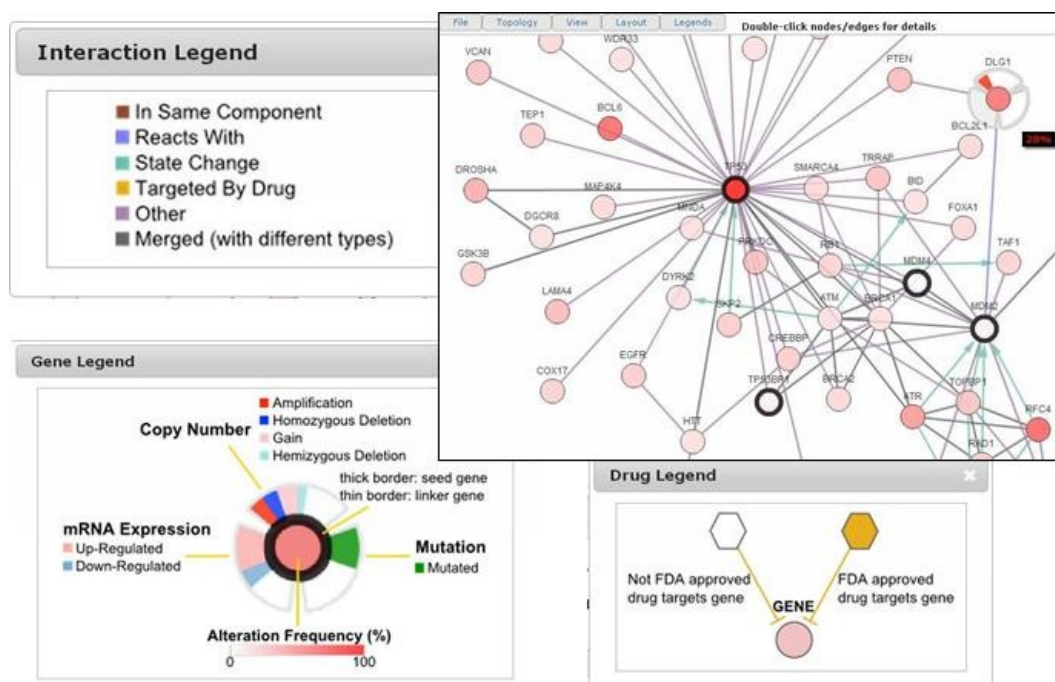

The cBioPortal network map contains multiple layers of genomic information that can be filtered using its numerous features. Each node represents a single gene. The color-coded discs that appear when hovering over a node display the various genomic data, including copy number alterations, mutations, and mRNA expression differences. The edges connecting network nodes are color-coded based on the type of interaction they represent (e.g., a brown edge indicates that Genes A and B are in the same complex). From the Genes & Drugs menu, the user may also choose to display drug-target interactions in the network. This information is curated from other sources, including DrugBank, KEGG Drug, NCI Cancer Drugs, and Rask-Andersen et al. 2011.

Mutations

Not all mutations have functional consequences at the protein level. The Mutations tab helps researchers predict how specific mutations may affect the function of individual genes before initiating functional studies. It contains a summary graphic and table of all nonsynonymous mutations [12] identified in each gene. The graphic display (Figure 6) shows the position and frequency of the mutations overlaid on protein domains from Pfam (Protein family database from Wellcome Trust Sanger

Institute) [13], so that users can visually inspect the potential impact a mutation has on a functional domain. The table (not shown) provides a host of information about each mutation, such as amino acid change, type of mutation, predicted functional impact as calculated by the Mutation Assessor application, and a hyperlink to the 3D protein structure with mutation highlighted (in Mutation Assessor).

Figure 6: *Graphical summary of TP53 nonsynonymous mutations from TCGA lung squamous cell carcinoma study mapped across the gene*
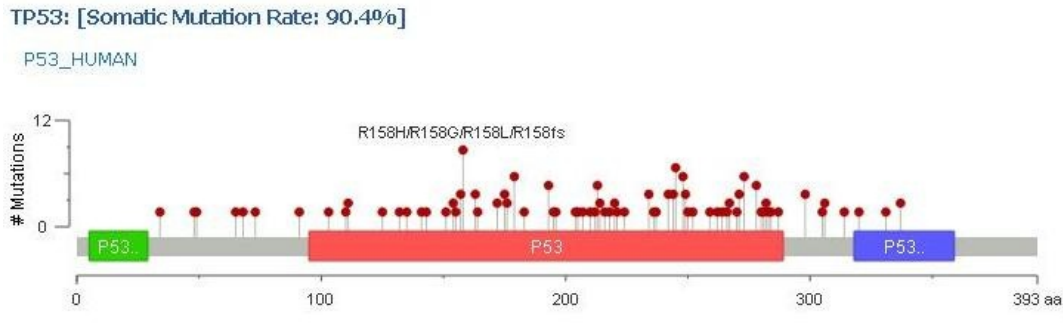
Patient View

 The Patient View page condenses all relevant molecular and clinical data available from one patient's tumor into multiple visualizations. This tool provides users the unique opportunity to examine patient-specific questions, such as which altered genes might influence response to treatment. Access this page by clicking on the case ID hyperlinks found on the OncoPrint or the Mutations table. The links on the OncoPrint appear when hovering over a patient-specific genomic alteration.

 The cBioPortal has many options for saving, downloading, and sharing results from a query. Each of the readouts, a few of which are highlighted above, can be saved at unique URLs and/or e-mailed for convenient sharing. In addition to the visualization and analysis readouts, the Portal also generates simple tab-delimited files of each query for use in other applications. Furthermore, cBioPortal facilitates data sharing and transfer of Portal data into other programs, such as the Integrative Genomics Viewer [14], which may be more suitable for certain types of analyses.

 cBioPortal offers a variety of easy-to-understand readouts for researchers that may not necessarily have a background in bioinformatics and computational science. It facilitates data integration and interpretation, which is a critical in developing novel treatments that will improve the lives of cancer patients.

 Researchers may download and install cBioPortal locally. Please contact cbioportal@googlegroups.com [15] for information. To learn more about this and many of the other features of this tool, visit the cBioPortal website (http://www.cbioportal.org [8]).

 References:
 1)   https://cbio.mskcc.org/tools/cancer-genomics/index.html [16]
 2)   Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The

cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery* May;2(5):401-4 (2012)

3)   Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling* Apr 2;6(269):pl1 (2013)

4)   Schroeder MP, Gonzalez-Perez A, Lopez-Bigas N. Visualizing multidimensional cancer genomics data. *Genome Medicine* Jan 31;5(1):9, (2013)

5)   Storrs, Carina. Combing the Cancer Genome. *The Scientist Magazine* (2012)

Back to Top

*/