



Issue 4 : April, 2011

OCG PERSPECTIVE

Is Genomics Paving "the Path to Progress" in Cancer Research?

It's April, and that can only mean one thing at the NCI. No, it's not the DC Cherry Blossom Festival, but the Annual Meeting of the American Association for Cancer Research (AACR).

FEATURED RESEARCHERS



Andrea Califano, Ph.D.

Professor of Systems Biology Chief, Division of Biomedical Informatics Director, Columbia Initiative in Systems Biology Director, Center for the Multiscale Analysis of Genetic Networks (MAGNet) Associate Director for Bioinformatics, Irving Cancer Research Center (ICRC)

CGCI PROGRAM HIGHLIGHTS

New Program Highlight: Burkitt Lymphoma Genome Sequencing Project

A new addition to the Office of Cancer Genomics' (OCG) portfolio of initiatives aimed at improving our understanding of the molecular mechanisms of cancer is the Burkitt Lymphoma Genome Sequencing Project (BLGSP).

CTD² PROGRAM HIGHLIGHTS

Translating Genomic Data into Viable Cancer Therapeutics

(adaptation from NCI Bypass BudgetOpens in a New Tab)

CONNECTING THE DATA

Transitioning from Sanger to the Next Generation of



Sequencing

Next-generation sequencing (NGS), high-throughput sequencing methods that parallelize the sequencing process thereby producing thousands to millions of sequences in a short time span, has recently gained tremendous momentum in cancer genomic research.

OCG PERSPECTIVE

Is Genomics Paving "the Path to Progress" in Cancer Research?

Jaime Guidry Auvil, Ph.D.

It's April, and that can only mean one thing at the NCI. No, it's not the DC Cherry Blossom Festival, but the Annual Meeting of the American Association for Cancer Research (AACR). This year's gathering of oncology-laden minds was inundated with a plethora of multiple symposia, educational and scientific sessions, workshops, talks and poster presentations that revolved around the theme of cancer genomics. The opening plenary session featured the NCI Director, Dr. Harold Varmus emphasizing the importance of defining malignant genomes to provide a foundation to base future cancer research. Appropriately, the NCI [Office of Cancer Genomics \(OCG\)](#) [2] was well-represented throughout the conference, accentuating not only the research programs supported through the office but also the investigators that make OCG initiatives successful. In his talk, Dr. Varmus further alluded to the critical aspect of integrating genomic data discovered for as many cancers as possible, and in particular a select few cancers, which were highlighted in the [2012 NCI Professional Budget Judgment Opens in a New Tab](#) [3] (also known as the Bypass Budget) "Cancer: Changing the Conversation" as more of an immediate priority for the NCI. While most of the adult cancers listed in the Bypass Budget are currently under investigation by a well-known large-scale genomics initiative piloted in [OCG](#) [2], [The Cancer Genome Atlas \(TCGA\) Opens in a New Tab](#) [4], two childhood cancers were also included, neuroblastoma and acute myeloid leukemia (AML). Both of these pediatric cancers are currently being studied as part of the current OCG initiative, [Therapeutically Applicable Research to Generate Effective Treatments \(TARGET\) Opens in a New Tab](#) [5].

TARGETing Cancer Health Disparities

The AACR organized several sessions highlighting [TARGET Opens in a New Tab](#) [5] pediatric projects in both neuroblastoma and childhood leukemia. In addition to an [NCI Opens in a New Tab](#) [6]/[NIH Opens in a New Tab](#) [7] sponsored educational session, which generally informed on the [TARGET Opens in a New Tab](#) [5] initiative as a whole, exome and whole genome sequencing efforts of the neuroblastoma project were featured in a major symposium and two minisymposia. Additionally, data from the pilot project for acute lymphoblastic leukemia (ALL) was at the forefront of a pediatric session on new concepts in organ site research, which focused on potential new therapeutic targets specific for the disease. The [TARGET Opens in a New Tab](#) [5] ALL

project team has already published several interesting papers describing potential drug targets discovered through genomic characterization. Additionally, some [TARGET](#) ^[5] ALL project team members recently reported in [Nature Genetics](#) ^{Opens in a New Tab} ^[8] that pharmacogenomics, the study of genetics as related to drug response, revealed that ancestry alters a child's risk of relapse and survival in certain types of ALL. Specifically, African American and Hispanic ethnicities are associated with poor survival when compared with Caucasian and Asian Americans, and Native American ancestry was found to significantly increase risk of relapse in pediatric ALL. Of great interest is that a single extra phase of chemotherapy reduced the risk of relapse equivalent to that seen among other ethnicities. This finding provides the first validated genomic evidence of a heritable genetic basis for ethnic disparities in cancer survival, and outlines how genomics studies can lead to modifications in therapy that will result in better outcomes for cancer patients.

One area of focus for AACR, [NCI](#) ^{Opens in a New Tab} ^[9] and cancer research at large is the issue of health disparities and the need for proper cancer education of underserved populations. The annual AACR meeting is unique in the cancer field, as it draws not only clinicians and scientists that perform cancer research, but also students and trainees at all levels, science and biotechnology industry and pharmaceutical representatives, patient advocates and even cancer patients. Cancer directly affects more than 1 in 3 Americans personally every year, which makes tumor biology education critical among all ages, ethnicities and cultures in the US. For example, the remarkable results obtained in the [TARGET](#) ^{Opens in a New Tab} ^[5] ALL pharmacogenomics study previously mentioned are of little benefit if the Native American culture is unable to understand the implications, not only of the outcome but of the necessity of participating in such genomics studies that will lead to more effective treatments for all populations. "Culturally targeted and sensitive cancer outreach and education program activities need to be included in all research focusing on genomics and genetically-targeted medicine and treatments, which include Native Americans and any underserved population. And outreach and education about underserved populations is required of and includes researchers and their staff," notes Phyllis Pettit Nassi, MSW, Manager of Special Populations at Native American Outreach. She and countless other advocates attend AACR to learn about what is being done in cancer research as well as to educate the research and medical community about issues underlying their science. [OCG](#) ^[2] understands the need for and further supports the expansion of cancer genomics education to include people of all ages, ethnicities, cultures and education levels. As cancer research enters an age of the cancer genome and epigenome, it will be important to create a robust infrastructure of information that allows each patient, student, researcher and clinician alike to fully grasp the concepts being studied, the resulting data obtained and how to properly interpret findings so that medicine can be improved.

Translating the Data

A key component in using genomics to better understand and treat cancer is the ability to integrate and interpret the massive amounts of data being generated by large-scale genomics initiatives such as [TARGET](#) ^{Opens in a New Tab} ^[5], [TCGA](#) ^{Opens in a New Tab} ^[4], and the [Cancer Genome Characterization Initiative \(CGCI\)](#) ^[10]. CGCI currently has data available through OCG Data Portals for medulloblastoma and diffuse large B-cell lymphoma (DLBCL), and data from HIV+ cancers and Burkitt's

lymphoma projects will be coming in the future. [CGCI](#) [10] investigators recently published their findings in [ScienceOpens in a New Tab](#) [11], noting the overall lack of genetic mutations in medulloblastoma versus adult cancers. The investigators also noted that the mutations present largely appear in genes involved in normal developmental processes. Additionally, a promising therapeutic target for DLBCL was presented by a [CGCI](#) [10] investigator during a minisymposium at AACR.

As with [TARGETOpens in a New Tab](#) [5] and [CGCI](#) [10], another [OCG](#) [2] initiative known as the [Cancer Target Discovery and Development \(CTD²\) Network](#) [12] had investigators prominently featured at this year's AACR as well. A major plenary session featured Dr. Andrea Califano discussing the use of systems biology to study integrative cancer genomics, and 3 other major symposia included talks by network colleagues surrounding the need for proper translation of genomics data to be useful in cancer treatment. The [CTD² Network](#) [12] focuses on the development of novel scientific approaches to accelerate the translation of genomic discoveries into new cancer treatments. The network emphasizes interaction of laboratories with complementary and unique expertise, including bioinformatics, genome-wide loss of function screening and targeted gain-of-function candidate gene validations, judicious use of mouse-based screens and small molecule high-throughput screens. The NCI announced its [intent to publish a request for applications for CTD²Opens in a New Tab](#) [13] in coming months, which will present a timely opportunity to explore new methods for translation of large-scale genomics; a topic heavily emphasized at AACR and with NCI.

The 102nd Annual AACR Meeting was highly successful in spotlighting all that has been accomplished in the 40 years since the National Cancer Act was signed into law. The title on the AACR meeting program, "Innovation and Collaboration: the Path to Progress", clearly defines the goal for the event in which the theme of cancer genomics was so greatly interwoven. The time for defining the cancer genome is here, and [OCG](#) [2] is leading many of the projects taking on that challenge. The [OCG](#) [2] pathway of progress will serve to lay the foundation for the next 40 years of cancer innovation by enhancing the understanding of the molecular mechanisms of cancer, advancing and accelerating genomics science and technology development, and efficiently translating the genomics data to improve cancer prevention, early detection, diagnosis and treatment.

[Back to Top](#)

FEATURED RESEARCHERS

Andrea Califano, Ph.D.

Andrea Califano, Ph.D.



*Professor of Systems Biology
Chief, Division of Biomedical Informatics
Director, Columbia Initiative in Systems Biology
Director, Center for the Multiscale Analysis of Genetic Networks (MAGNet)
Associate Director for Bioinformatics, Irving Cancer Research Center (ICRC)*

Dr. Andrea Califano serves as Principal Investigator of the Cancer Target Discovery and Development (CTD²) Center at Columbia University. Below he describes how his lab works to identify models of cancer regulation through the application of systems biology, a field of interdisciplinary study that examines complex interactions between components of biological systems. Researching cancer from this vantage point affords Dr. Califano and colleagues the opportunity to study physiological processes in a comprehensive and highly integrated manner that incorporates *in vivo (in living organisms)*, *in vitro (in the test tube)*, and *in silico (in the computer) methods*.

Throughout my career, my research interests have been focused on understanding dynamic systems. At the start of my career, I was a physicist working on deterministic chaos. More recently, over the last 20 years, I have been a systems biologist using a variety of physics and knowledge-based methodologies to elucidate molecular mechanisms that are dysregulated in disease.

My lab at Columbia University combines advanced *in silico* reverse engineering methods and high-throughput experimental biology to reconstruct and interrogate cell-context specific gene regulatory networks underlying pathophysiological processes. On the computational side, we have developed a repertoire of algorithms for the dissection of transcriptional, post-transcriptional, and post-translational regulatory interaction networks (or interactomes) in mammalian cells. We have also developed tools for interrogating interactomes to identify genes that are master regulators of specific cancer subtypes or to elucidate the mechanisms of action of small-molecules. We have shown that master regulator genes can constitute both oncogene and non-oncogene addiction points of the cancer cell and that compounds targeting these addictions can be effectively identified using computational approaches.

Our key strength, however, is in the ability to follow-up these computational inferences with rigorous experimental validation using biochemical or functional assays. For example, in collaboration with Dr. Antonio Iavarone, also at Columbia, we computationally predicted and experimentally validated that synergistic activation of two transcription factors, C/EBP and Stat3, is necessary and sufficient to reprogram neural stem cells along an aberrant mesenchymal lineage. The synergistic activation contributes to the worst prognosis of Glioblastoma (GBM) and establishes a synergistic addiction point in these tumors. In fact, their silencing led to collapse of the mesenchymal signature and reduction of tumor aggressiveness *in vivo*. Recently, we have been able to use the high-grade glioma regulatory network to further identify and experimentally validate both genetic alterations that contribute to the activation of C/EBP and Stat3 in more than 70% of mesenchymal GBMs. We have also been able to identify druggable targets and associated drugs (either FDA approved or in clinical studies) that can inhibit the activity of these transcription factors *in vitro*.

Thanks to the [CTD² Network](#) ^[12] initiative, we set up a Translational Cancer Systems Biology Center that combines the strength of several Columbia investigators with expertise in cancer biology, reverse engineering, pooled siRNA screening, and high-throughput screening. These investigators systematically identify and prioritize biomarkers, therapeutic targets, and small compounds for several tumor subtypes, including the mesenchymal phenotype of GBM, glucocorticoid resistance in T-cell Acute Lymphoblastic Leukemia, and the non-oncogene addiction of the ABC subtype of Diffuse Large B-Cell Lymphoma to NF-κB. Our pipeline can be applied to a variety of cancer subtypes, as long as appropriate molecular profile data and cell lines are available. The CTD² Network initiative has allowed us to leverage the highly complementary expertise of the other Network Centers, thus dramatically accelerating progress towards our goals.

The CTD² Network initiative comes at a propitious moment in the progression of cancer research. The availability of large and comprehensive public datasets and functional models is allowing researchers the unprecedented opportunity to create and interrogate highly accurate models of cancer regulation, both *in silico* and *in vivo*. The CTD² Network provides the framework to optimally exploit these models to discover safe and effective compounds, in a patient-centered fashion, and to facilitate their clinical development by providing associated targets, biomarkers, and mechanisms of action.

[Back to Top](#)

CGCI PROGRAM HIGHLIGHTS

New Program Highlight: Burkitt Lymphoma Genome Sequencing Project

A new addition to the Office of Cancer Genomics' (OCG) portfolio of initiatives aimed at improving our understanding of the molecular mechanisms of cancer is the Burkitt Lymphoma Genome Sequencing Project (BLGSP). The BLGSP has been established to develop a databank of the many alterations found in Burkitt lymphoma (BL), an uncommon type of Non-Hodgkin lymphoma that occurs most often in children and young adults. A highly aggressive type of B-cell lymphoma, BL is associated with a chromosomal translocation of the MYC gene. The three main types of BL are sporadic, endemic, and immunodeficiency-associated. Sporadic BL occurs throughout the world, while endemic BL occurs in East Africa. Immunodeficiency-related Burkitt lymphoma is most often seen in AIDS patients. Current chemotherapy regimens for BL are effective in approximately 40-90% of patients, depending on age, stage of the disease, treatment regimen, and site of the treatment facility. Hence, new treatments are needed to improve the efficacy of current regimens and to potentially substitute less toxic agents for the high intensity chemotherapeutic drugs currently given.

A National Cancer Institute project, the BLGSP is supported, in part, with funds donated to the Foundation for the National Institutes of Health by the Foundation for Burkitt Lymphoma Research. The goal is to explore potential genetic changes in

patients with BL that could lead to better prevention, detection and treatment of the cancer. OCG will provide the administrative and analytical infrastructure needed to carry out this project based on lessons learned in managing other major cancer genomics projects. The BLGSP will include adult and pediatric patients with sporadic BL, as well as those with endemic and HIV-positive sporadic BL.

The first major objective of the BLGSP is to accrue tumor and patient-matched control tissues obtained by precise protocols from a relatively large number of clinically well-annotated cases of BL. It is anticipated that this phase of the project may take up to two years. The second objective will be to characterize the alterations of the tumors' genomes (with matched normal as control) and transcriptomes by sequencing the DNA and RNA of each case. Using the data generated, the ultimate goals of the project are to discover the molecular changes that are present in BL patients and then determine how those changes correlate with treatment regimen and outcome. Similar to those developed for other large-scale cancer sequencing projects such as [Therapeutically Applicable Research to Generate Effective Treatments](#)^[5] and the [Cancer Genome Characterization Initiative](#)^[14], data generated from the BLGSP will be submitted to publicly accessible databases for use by the cancer research community.

The BLGSP holds much promise for uncovering new insights into the mechanisms of BL that may lead to prevention strategies and more effective treatments and relief for those living with this disease.

[Back to Top](#)

CTD² PROGRAM HIGHLIGHTS

Translating Genomic Data into Viable Cancer Therapeutics

Daniela Gerhard, Ph.D.

(adaptation from [NCI Bypass Budget](#)^[15])

Comprehensive molecular characterization of cancers is revealing somatic changes that may be part of the root cause of malignant disease. Tumor cells acquire mutations in their genes that, in turn, lead to production of abnormal proteins. This allows the cancer to interrupt a normal physiological process and further manipulate it to promote proliferation and survival of the tumor. Large-scale genomics initiatives, such as those investigated in programs managed by the [NCI Office of Cancer Genomics \(OCG\)](#)^[2], are identifying and characterizing these mutations, thereby uncovering novel mechanisms for cancer initiation and progression, and new ways to treat cancer in the future. One recent finding has been the discovery of a mutation in a gene coding for an enzyme called isocitrate dehydrogenase (IDH). IDH plays an essential role in converting simple carbohydrates into a key molecule involved in producing energy in all normal cells.

There are two IDH genes, IDH1 which works in the cytoplasm and IDH2 that functions in mitochondria. A mutation in IDH1 was initially discovered in gliomas (a tumor

affecting the brain or spinal cord), and subsequent research over the past two years has shown that both IDH1 and IDH 2 enzymes have heterozygous mutations (only 1 of the 2 alleles is affected) in >70% of grade II-III gliomas and secondary glioblastomas (highly lethal brain tumors which arise from low grade gliomas). In addition, it is now known that IDH1/2 are mutated in up to 15% of acute myeloid leukemias (AML) and possibly other cancers at low frequency. These mutations appear to reduce the normal activity of each enzyme (namely the conversion of isocitrate to alpha-ketoglutarate (a-KG)), and give rise to a new function (the conversion of a-KG to (R)-2-hydroxyglutarate (2-HG)). Indeed 2-HG levels are elevated >50-fold in tumor cells from patients with IDH1/2 mutations, motivating further study of 2-HG as a disease biomarker, as well as deeper investigation of the molecular mechanism by which this putative 'oncometabolite' may contribute to disease. Additionally, initial studies suggest that the mutant form of the 2-HG enzyme changes the epigenetic state (modification of gene expression by a mechanism independent of the underlying DNA sequence) of the cancer cells. One compelling implication of these novel findings is that mutations of IDH genes in multiple cancers could preliminarily provide some supportive evidence of Warburg's hypothesis that malignant growth is caused by tumor cells that mainly generate energy by non-oxidative breakdown of glucose (rather than normal oxidative breakdown of pyruvate), and cancer is actually a result of mitochondrial dysfunction.

While identification of altered enzyme activity was pivotal in understanding the cellular impact of IDH1/2 mutations, the dependency of IDH1/2-mutant cancers on the newly-acquired function for initiation of malignant transformation remains mostly uncharacterized. This is largely due to the lack of high-quality small-molecule probes for mutant IDH1/2 activity available to the cancer biology community. As part of the NCI's Cancer Target Discovery and Development (CTD²) [12] Network, the Broad Institute is developing small molecules that inhibit the 2-HG-generating activity of IDH1/2 mutants in cells, providing much needed tools to study the role of IDH mutations in cancer cell biology and to validate mutant IDH1/2 as a possible target for drug development. This research dovetails well with efforts underway within the pharmaceutical industry, as well as within the NCI's Experimental Therapeutics (NExT) program [16]. In less than one year, the Broad-CTD² Center identified highly novel small molecules (derived from innovative synthetic organic chemistry methods) that inhibit the novel activity of one of the IDH1 mutant forms which has 50% inhibition concentration (IC₅₀ of 2uM) and that has ~10-fold selectivity over the normal activity of the wild-type enzyme. The molecules screened have properties that facilitate modifications, so as to become either more specific for a given target (in this case the mutant form of IDH1/2) or confer activity at lower concentrations, which are two features not present in typical compound collections. The Broad-CTD² Center will continue to characterize and optimize the cell-based activity and selectivity of these preliminary leads to develop a high-quality probe that will be made available to other investigators and potentially may serve as a starting point for clinical studies. These compounds will enable thorough investigation by the research community into the role of mutant IDH1/2 in cancer biology, as well as further validate it as potential target for therapy.

[Back to Top](#)

CONNECTING THE DATA

Transitioning from Sanger to the Next Generation of Sequencing

Jinghui Zhang, Ph.D.



Next-generation sequencing (NGS), high-throughput sequencing methods that parallelize the sequencing process thereby producing thousands to millions of sequences in a short time span, has recently gained tremendous momentum in cancer genomic research. NGS is increasingly becoming the technology of choice as it has the capacity to unveil a large number of somatic alteration events in tumors. However, at the dawn of this technology, there was skepticism as to how to make use of NGS data, which can come in a massive bolus of billions of short sequence reads from a single experiment.

Despite its monstrous volume, NGS has a lot in common from the data analysis perspective with traditional Sanger sequencing, a chain termination DNA sequencing method commonly used in automated 1st generation sequencing. The name of the "game" for identifying somatic sequence mutations using either NGS or Sanger technology is to find true variants amid the artifacts. With respect to Sanger sequencing methods, the resulting artifacts could be compared to the size of a splash pool, whereas NGS produces artifacts more akin to the size of an ocean. Many of the "tricks" used to cleanse the artifacts found in Sanger sequencing data sets can be re-applied and recalibrated for NGS analysis, including:

- evaluation of the quality scores
- filtering alignment artifacts caused by poor quality or short read length
- filtering of non-specific mapping caused by paralogous duplications in the human genome
- evaluation of the data consistency (e.g. consistent genotype call in forward and reverse orientations)

However, there is a catch: in order to clean up the "ocean-size" mess, everything must be run on a computing cluster of high computing speed and capacity, thus parallelizing the analysis pipeline is a must.

Neither Sanger sequencing nor NGS is perfect; both are challenged to overcome the common problem that certain regions of the genome are not "covered". In Sanger sequencing, the lack of sequence coverage is usually caused by PCR failure or poor sequencing data in target regions; in NGS, GC bias is a major contributor in uneven coverage of the sequencing data. Therefore, a negative result needs to be evaluated in the context of coverage of a target region. For Sanger sequencing data, coverage is computed by evaluating the high-quality bases collected from both the forward and the reverse reads of the same sample, whereas in NGS the high-quality bases covering at the same site for each sample are calibrated.

In both Sanger and NGS data sets, signals that indicate an important biological event can be initially interpreted as an "artifact". In Sanger sequencing data, a read with a heterozygous insertion/deletion (indel) has the appearance of a poor quality read. In

NGS, a read which spans an indel may become unmappable. Furthermore, a structural variation may be represented as a read pair with discordant mapping, which can also represent an artifact in library preparation or a mapping artifact caused by repetitive regions. In both cases, "abnormal" data may represent a bona fide artifact or be a potentially interesting biological signal, and therefore need to be fully analyzed to avoid "throwing the baby out with the bathwater". One approach for using sequence data that does not correspond to the normal genome to identify interesting biological findings from NGS is the use of the soft-clipped reads which mark poorly-aligned subsequences. My group has developed a novel algorithm we named **CREST** (Clipping REveals STructure) to use the marked soft-clipped reads for directly mapping structural variations using next-generation sequencing data (Wang et al, submitted). This method shows much improved sensitivity and accuracy to find structural variations compared with the traditional methods. NGS's abundance of data has been challenging to interpret and requires the development of novel bioinformatic methods. Nonetheless, the data, once obtained, allows cancer researchers an opportunity to better understand the changes which occur in tumor genomes. These changes will need to be followed up with biological experiments to explain which genetic variations drive tumor development and metastasis. Using the oceans of information generated by NGS to understand the relevant changes in a tumor genome will lead to a better "pool" of treatment options in the future for patients with cancer.

[Back to Top](#)

*/

Source URL: <https://ocg.cancer.gov/news-publications/e-newsletter-issue/issue-4>

Links

[1] <https://ocg.cancer.gov/printpdf/taxonomy/term/61> [2] <https://ocg.cancer.gov> [3] <https://www.cancer.gov/research/annual-plan> [4] <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga> [5] <https://ocg.cancer.gov/programs/target> [6] <https://www.cancer.gov> [7] <https://www.nih.gov> [8] <https://pubmed.ncbi.nlm.nih.gov/21297632/> [9] <https://www.cancer.gov/> [10] <https://ocg.cancer.gov/programs/cgci> [11] <https://pubmed.ncbi.nlm.nih.gov/21163964/> [12] <https://ocg.cancer.gov/programs/ctd2> [13] <https://grants.nih.gov/grants/guide/notice-files/NOT-CA-11-010.html> [14] <https://cgap.nci.nih.gov/cgci.html> [15] <https://www.cancer.gov/about-nci/budget/about-annual-plan?redirect=true> [16] <https://next.cancer.gov/>