ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 2, Issue 4, April 2017

Pharmaco-Informatics: Harnessing The Power Of Bioinformatics In Cancer Research & Management: A Review

[1, 2]Preenon Bagchi, [1]Tara Devi Chetry, [1]Anuradha M
[1]Padmashree Institute of Management & Sciences, Bangalore, Karnataka, India.
[2]Sarvasumana Association, Deoghar Br., Jharkhand, India.

Abstract: **- One of the greatest challenges faced by the cancer researchers is that the disease varies so much from individual to individual. Even the same type of cancer – blood, brain, kidney, pancreas, and so on – can be different subtly. This concludes that a therapy working excellent in one patient may have absolutely no effect in another. Cancer Research worldwide has set up several centers and started collecting 9,000 tumor cells' samples from a wide range of cancer patients and created a DNA database of cancerous cells. Researchers extract DNA from these tumors and scan them for a series of key genes involved in tumor development and compares & cross-checked against a range of cancer treatments, cancerous genes, to create a map of which treatments in particular works best for cancers associated with which particular genes. This is based on the concept of pharmacogenomics & pharmacoinformatics: certain genes predispose individual to respond to certain molecules in certain ways. Doctors can already test a cancer patient for a single known gene, knowing how tumors with that gene respond to a particular molecule. However currently they don't have a way of testing with a broad panel or set of genes. And to compensate the problem, they don't have a way of quicker and more accurate way of sharing information in-between research labs in the same city, across the country or internationally. With the proposed cancer DNA database, a doctor might analyze a patient's cancerous tumor sample and prescribe a detailed tailored treatment plan within a very short period of time. Bioinformatics research is increasing steadily at an exponential rate. DNA sequences are available to researchers with just an Internet connection – along with free bioinformatics tools to explore any sequence data, predict the presence of genes/mutated genes, and compare features shared between various organisms.**

Index Terms**: Cancer, bioinformatics, metabolomics, epigenomics, genome, bio-software**

I. INTRODUCTION

The greatest challenges faced by the cancer researchers is that the disease varies so much from individual to individual that even the same type of cancer – blood, brain, kidney, pancreas, and so on – can be different subtly [1]. This concludes that a therapy working excellent in one patient may have absolutely no effect in another [2]. Cancer Research worldwide has set up several centers and started collecting 9,000 tumor cells' samples from a wide range of cancer patients and created a DNA database of cancerous cells [3]. Researchers extract DNA from these tumors and scan them for a series of key genes involved in tumor development and compares & cross-checked against a range of cancer treatments, cancerous genes, to create a map of which treatments in particular works best for cancers associated with which particular genes [4]. This is based on the concept of pharmacogenomics & pharmacoinformatics: certain genes predispose individual to respond to certain molecules in certain ways [1]. Doctors can already test a cancer patient for a single known gene, knowing how tumors with that gene respond to a particular molecule. However currently they don't have a way of testing with a broad panel or set of genes [2, 3]. And to compensate the problem, they don't have a way of quicker and more accurate way of sharing information in-between research labs in the same city, across the country or internationally [1, 4]. Why Bioinformatics - Enter the healer - bioinformatics. With the proposed cancer DNA database, a doctor might analyze a patient's cancerous tumor sample and prescribe a detailed tailored treatment plan within a very short period of time [2, 3, 4]. As Professor Matthew Seymour, director of the National Cancer Research Network (NCRN) in the UK, recently stated, "We have to get clever about how to target drugs. Medications for cancer have to be personalized because no two cancers are identical." So global researchers brought in a big gun – the bioinformatics.

Bioinformatics research is increasing steadily at an exponential rate. DNA sequences are available to researchers with just an Internet connection – along with free bioinformatics tools to explore any sequence data, predict the presence of genes/mutated genes, and compare features shared between various organisms [5, 6]. Cancer - Cancer is

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 2, Issue 4, April 2017

one of the commonest causes of patient death in the clinic and a complex disease occurring in multiple organs per system, multiple systems per organ, or both, in the body. The poor detects, therapies and prognosis of the disease could be mainly due to the variation of rigorousness, extents, locations, sensitivity and confrontation against medications, cell differentiation and origin, and understanding of pathogenesis. With increasing evidence that the interface and network between genes and proteins play an important role in exploration of cancer molecular mechanisms, it is essential and important to introduce a new perception of Systems Clinical Medicine into cancer research, to integrate systems biology, clinical science, omics-based technology, bioinformatics and computational science to improve diagnosis, therapies and prognosis of diseases [1, 7].

Types of Cancer - There are more than 100 types of cancer. Types of cancer are usually named for the organ or tissue where the cancer form, but they also may be described by the type of cell that formed them [8].

TABLE 1: Different types of cancer [8]

A	Acute granulocytic leukemia
	Acute lymphocytic leukemia
	Acute myelogenous leukemia
	Adenocarcinoma
	Adenosarcoma
	Adrenal cancer
	Adrenocortical carcinoma
	Anal cancer
	Anaplastic astrocytoma
	Angiosarcoma
	Appendix cancer
	Astrocytoma
	Acute granulocytic leukemia
	Acute lymphocytic leukemia (ALL)
	Acute myelogenous leukemia (AML)
	Adenocarcinoma
	Adenosarcoma
	Adrenal cancer
	Adrenocortical carcinoma
	Anal cancer
	Anaplastic astrocytoma
	Angiosarcoma
	Appendix cancer
	Astrocytoma
B	Basal cell carcinoma
	B-Cell lymphoma
	Bile duct cancer
	Bladder cancer
	Bone cancer
	Bone marrow cancer
	Bowel cancer
	Brain cancer
	Brain stem glioma
	Brain tumor
	Breast cancer
	Basal cell carcinoma
	B-Cell lymphoma
	Bile duct cancer
	Bladder cancer
	Bone cancer
	Bone marrow cancer
	Bowel cancer
	Brain cancer
	Brain stem glioma
	Brain tumor
	Breast cancer
C	Carcinoid tumors
	Cervical cancer
	Cholangiocarcinoma
	Chondrosarcoma
	Chronic lymphocytic leukemia (CLL)
	Chronic myelogenous leukemia (CML)
	Colon cancer
	Colorectal cancer
	Craniopharyngioma
	Cutaneous lymphoma
	Cutaneous melanoma
D	Diffuse astrocytoma
	Ductal carcinoma in situ (DCIS)

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 2, Issue 4, April 2017

E	Endometrial cancer
	Ependymoma
	Epithelioid sarcoma
	Esophageal cancer
	Ewing sarcoma
	Extrahepatic bile duct cancer
	Eye cancer
F	Fallopian tube cancer
	Fibrosarcoma
G	Gallbladder cancer
	Gastric cancer
	Gastrointestinal cancer
	Gastrointestinal carcinoid cancer
	Gastrointestinal stromal tumors (GIST)
	General
	Germ cell tumor
	Glioblastoma multiform (GBM)
	Glioma
H	Hairy cell leukemia
	Head and neck cancer
	Hemangioendothelioma
	Hodgkin lymphoma
	Hodgkin's disease
	Hodgkin's lymphoma
	Hypopharyngeal cancer
I	Infiltrating ductal carcinoma (IDC)
	Infiltrating lobular carcinoma (ILC)
	Inflammatory breast cancer (IBC)
	Intestinal Cancer
	Intrahepatic bile duct cancer
	Invasive / infiltrating breast cancer
	Islet cell cancer
J	Jaw cancer

K	Kaposi sarcoma
	Kidney cancer

L	Laryngeal cancer
	Leiomyosarcoma
	Leptomeningeal metastases
	Leukemia
	Lip cancer
	Liposarcoma
	Liver cancer
	Lobular carcinoma in situ
	Low-grade astrocytoma
	Lung cancer
	Lymph node cancer
	Lymphoma
M	Male breast cancer
	Medullary carcinoma
	Medulloblastoma
	Melanoma
	Meningioma
	Merkel cell carcinoma
	Mesenchymal chondrosarcoma
	Mesenchymous
	Mesothelioma
	Metastatic breast cancer
	Metastatic melanoma
	Metastatic squamous neck cancer
	Mixed gliomas
	Mouth cancer
	Mucinous carcinoma
	Mucosal melanoma
	Multiple myeloma
	Mycosis Fungoides
	Myelodysplastic Syndrome
N	Nasal cavity cancer
	Nasopharyngeal cancer
	Neck cancer
	Neuroblastoma

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 2, Issue 4, April 2017

	Neuroendocrine tumors (NETs)
	Non-Hodgkin lymphoma (NHL)
	Non-Hodgkin's lymphoma
	Non-small cell lung cancer
O	Oat cell cancer
	Ocular cancer
	Ocular melanoma
	Oligodendroglioma
	Oral cancer
	Oral cavity cancer
	Oropharyngeal cancer
	Osteogenic sarcoma
	Osteosarcoma
	Ovarian cancer

	Ovarian epithelial cancer
	Ovarian germ cell tumor
	Ovarian primary peritoneal carcinoma
	Ovarian sex cord stromal tumor
P	Paget's disease
	Pancreatic cancer
	Papillary carcinoma
	Paranasal sinus cancer
	Parathyroid cancer
	Pelvic cancer
	Penile cancer
	Peripheral nerve cancer
	Peritoneal cancer
	Pharyngeal cancer
	Pheochromocytoma
	Pilocytic astrocytoma
	Pineal region tumor
	Pineoblastoma
	Pituitary gland cancer
	Primary central nervous system (CNS) lymphoma
	Prostate cancer
R	Rectal cancer

	Renal cell carcinoma
	Renal pelvis cancer
	Rhabdomyosarcoma
S	Salivary gland cancer
	Sarcoma
	Sarcoma, bone
	Sarcoma, soft tissue
	Sarcoma, uterine
	Sinus cancer
	Skin cancer
	Small cell lung cancer (SCLC)
	Small intestine cancer
	Soft tissue sarcoma
	Spinal cancer
	Spinal column cancer
	Spinal cord cancer
	Spinal tumor
	Squamous cell carcinoma
	Stomach cancer
	Synovial sarcoma
T	T-cell lymphoma
	Testicular cancer
	Throat cancer
	Thymoma / thymic carcinoma
	Thyroid cancer
	Tongue cancer

	Tonsil cancer
	Transitional cell cancer
	Transitional cell cancer
	Transitional cell cancer
	Triple-negative breast cancer
	Tubal cancer
	Tubular carcinoma
U	Undiagnosed Cancer
	Ureteral cancer
	Ureteral cancer
	Urethral cancer
	Uterine adenocarcinoma
	Uterine cancer
	Uterine sarcoma

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 2, Issue 4, April 2017

V	Vaginal cancer
	Vulvar cancer

II CANCER BIOINFORMATICS –

Cancer bioinformatics is a important and vital part of the systems clinical medicine in cancer and the core tool and approach to carry out the investigations of cancer in systems clinical medicine and for the development of bioinformatics methods, network biomarkers and precision medicine to explore the potential of clinical applications and improve the outcomes of patients with cancer [1, 7].

Expectations of methodologies - Cancer bioinformatics is one of many ways to focus bioinformatics methods in cancer, according to the specificity of disease metabolisms, signaling, communication, and proliferations [7]. Clinical bioinformatics, an emerging science combining clinical informatics, bioinformatics, medical informatics, information technology, mathematics, and omics science together can be considered to be one of critical elements addressing clinical relevant challenges in early diagnosis, efficient therapies, and predictive prognosis of patients with cancer [7]. There is a necessity to build up cancer bioinformatics-specific methodologies or introduce new and advanced bioinformatics tools to answer the specific question of cancer [9]. Like, the Semantic Web technology was used to recognize high throughput clinical data and develop quantitative semantic models retrieved from Corvus, a data warehouse which provides a uniform interface to various forms of Omics data, based on systematic biological knowledge and by application of SPARQL endpoint [11]. Semantic models, having genomic, transcriptomic and epigenomic data from melanoma samples with Gene Ontology data and regulatory networks constructed from transcription factor binding information, were applied for the interaction between a cell molecular state and its response to anti-cancer therapy [12, 13, 14]. Multivariate assays, a process to illustrate inaccuracy introduced in the assay results from the built-in error in sample preparation and measurement of the contributing factors, were used to help and guide clinicians understanding the application to PAM50 centroid- based genomic predictors for breast cancer management plans and providing the uncertainty information in a usable way. It may be a non-relative query or a prospect expectation how experts in cancer bioinformatics can help clinicians to set up the potential picture of gene or protein interactions and mechanisms correlated with tumor-associated shapes, densities, or locations [15, 16]. A recent article by von der Heyde and Beissbarth in the in BMC Medicine discusses the current insights into methods of cetuximab resistance in head and neck cancers resulting from original analysis of the EGFR pathway [17].

III NEW STRATEGIES OF BIOMARKERS

Cancer bioinformatics is expected to participate in a more significant role in the recognition and validation of biomarkers, specific to clinical phenotypes related to early diagnoses, measurements to scrutinize the progress of the disease and the response to therapy, and predictors for the development of patient's life value [18]. Of gene-, protein-, peptide-, chemical- or physic-based variables in cancer, biomarkers were examined from a single one to many markers, from the expression to functional indication, and from the network to dynamic network. Network biomarkers as a new type of biomarkers with protein-protein interactions were examined with the amalgamation of knowledge on protein annotations, interaction, and signaling pathway [19]. Alterations of network biomarkers can be monitored and evaluated at different stages and time points during the development of diseases, named dynamic network biomarkers, as one of the new strategies. Vibrant network biomarkers were presumed to be associated with clinical informatics, including patient complaints, history, therapies, clinical symptoms and signs, physician's examinations, biochemical analyses, imaging profiles, pathologies and other measurements. Systems clinical medicine is recommended as one of latest approaches for the development of cancer biomarkers [20]. Systems clinical medicine is created as the assimilation of systems biology, clinical phenotypes, high- throughout technologies, bioinformatics and computational science to improve diagnosis, therapies and prognosis of diseases [7, 21]. Cancer biomarkers should have the natures of networks, dynamics, interfaces, and specificities to disease diagnosis, therapy and prognosis. Understanding the interaction between clinical informatics and bioinformatics is the first and critical step to discover and develop the new diagnostics and therapies for diseases [7, 22]. Such approach has been described in other disorders like acute rejection after renal transplantation or lung diseases. In short, human samples from clinical studies under lucid and strict criterion of participating recruitments are collected and collected

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 2, Issue 4, April 2017

with an entire profile of clinical informatics translated from clinical descriptions. Gene and/or protein profiles of defined samples are analyzed and vibrant set-ups and interfaces between genes and/or proteins can be figured out by bioinformatics and systems biology. Selected disease-specific associations and dynamic arrangements of genes and/or proteins in patients are correlated with each of clinical phenotypes by the computational mode, to validate and optimize disorder-special biomarkers [23]. However, a number of challenges in the application of systems clinical medicine are encountered and need to be overcome; e.g. the optimal system to decode the information of clinical descriptions to clinical informatics, bioinformatics analysis oriented with disease severity, extent, location, sensitivity to therapies, and progress, or computational mode to integrate all elements from clinical and high-throughout data for accuracy conclusions [7]s. It is also a challenge to find out the deviation and significance between molecular networks, between networks of molecules and clinical phenotypes, and between gene and/or protein interfaces, in addition to the expression of genes and proteins. It is seen that incorporating protein set-up and molecular interaction data recovers the ability to interpret the specific gene signature in breast cancer patients because R weighted Recursive Feature Elimination and average pathway expression were found to be most effective at generating interpretable signatures [24].

IV THE "OMICS" IN CANCER BIOINFORMATICS –

Compared to past years, currently there are numerous open source projects active in the life science arena, each offering freely available source code that promises to address a specific problem or problem domain of biological sciences in a reusable way [25]. Bioinformatics.org alone hosts over 275 projects, which address a bioinformatics need by definition. In addition, Sourceforge hosts around 750 projects categorized as 'Bioinformatics' including projects such as Generic Model Organism Database (GMOD), Microarray Gene Expression Data society (MGED) and Life Sciences Identifiers (LSID). There are many projects hosted by the developers' home institutions or by other open source-allocated umbrella organizations such as the Open Bioinformatics Foundation. The latter actually hosts some of the toolkits most widely used in the life sciences, such as BioPerl and Biojava. Scenario of each project offering 'stuff' the effectiveness or outcome of which is often not immediately clear, is much reminiscent of a market [25].

Cancer is one of the most complex types of all human disorders. Its complexity lies in:

(1) its rapidly evolving population of cells that flow away from their usual functional states at the molecular, epigenetic and genomic levels,

(2) its enlargement and spreading out to encroach and substitute normal tissue cells; and

(3) its abilities to defend against both endogenous and exogenous measures for preventing or slowing down its growth [26].

Major challenging issues that clinical oncologists deal with are considerable heterogeneity and different genetic & generic backgrounds even within the same type of cancer, but also that most effective medicines tend to lose their efficaciousness within a year, or so. Thus the natural question comes is: what can be the reasons for their losing efficacy? Intuitively this is due to a cancer's capability to evolve speedily, particularly in terms of generating drug-resistant sub- populations, which is facilitated by its abilities to proliferate and to accumulate genomic mutations rapidly [26]. Of the many reasons that our knowledge is so light has been the lack of molecular-level data, fully analysed and mining of which can potentially can reveal the full complexity of an evolving cancer. While large quantities of omic data (in database) such as gen omic , epigen omic, transcript omic , metabol omic and prote omic data have currently been generated by computational biologists for a variety of cancer types, only a few cancer studies have been designed to take full advantage of all the information derivable from the available omic data [27, 28]. Integrative analyses of numerous data types proves to be a boon to gain a full and systems- level understanding about a cancer's evolutionary dynamics, including the elucidation of its true drivers as well as key facilitators at different developmental stages of a cancer. We hypothesize that only when all of the key information hidden in omic data is fully derived and utilized, we can expect a meaningful breakthrough in our understanding of cancer & its diagnosis [27, 28].

The Human Genome Project has been sequenced and the three billion base pairs (bps) of nucleotides comprising a complete human cancer genome are represented in a digital form, directly readable by humans and computers, allowing cancer researchers and clinicians to view and analyze the detailed genetic makeup of a healthy human and a cancer patient [29, 30]. Complementing

and extending the invaluable genome sequence data are the major change the Human Genome Project has brought about and the genetic science is now equipped with two powerful tools: rapid genome-sequence generation and computation-based information discovery from the genomic sequences. With the open accessibility of digitally represented human genomes in hand, scientists have computationally identified the vast majority of the ~20,000 protein encoding genes in our genome, along with large numbers of single-nucleotide polymorphisms (SNPs) and other types of genetic variations across individuals and different ethnic groups as well as various disease groups and targeted sequencing of specific genomic regions deemed to be relevant to certain diseases has led to the identification of numerous genetic markers for various diseases [29, 30]. In addition to the Human Genome Project, a number of closely related genome sequencing projects have been carried to provide a more comprehensive dataset for the human genome(s):

(1) the Human Genome Diversity Project to document genomic differences across different ethnic groups [31];
(2) the Human Variome Project to establish relationships between human genomic variations and diseases [32];
(3) the International HapMap Project to develop a haplotype map of the human genome [33];
(4) the 1000 Genome Project to establish a detailed catalog of all human genetic variations [34]; and
(5) the Personal Genome Project to sequence the complete genomes and establish the matching medical records of 100,000 individuals [35].

All these sequencing projects, along with other related ones, such as the Neanderthal Genome Project [36] and the Chimpanzee Genome Project [37, 38], provided a comprehensive view of the genomes of healthy humans with normal polymorphisms as well as mutations associated with various diseases. The Cancer Genome Atlas (TCGA) represents probably the most motivated cancer-genome sequencing project, which aims to sequence up to 10,000 cancer genomes covering 25 major cancer types by 2014 and make the data publicly available [39]. Such data tends to provide a substantial amount of information about cancer-related genomic mutations and by comparing the genome sequences of a cancer and the matching normal tissue researchers can identify all the genomic changes in the cancer genome, which tends to fall into two categories: simple and complex mutations [40]. Specifically, simple mutations refer to single base-pair mutations and DNA single or double-strand breaks; and complex mutations refer to duplications and deletions (together referred to as copy-number changes), translocations and inversions of genomic segments. Simple mutations can result from exogenous factors such as radiation, air-borne and food-related carcinogens in the environment, as well as from endogenous factors in the microenvironments inside our bodies, including ROS (reactive oxygen species) and other reactive metabolites plus random mutations [41]. Eg., ionizing radiation, including X-rays and gamma rays, can directly cause point mutations and DNA breaks. In addition, a variety of non-radioactive carcinogens have been identified that can damage DNA, including microbes, chemical compounds in the environment and reactive species inside our cells [41]. Free radicals represent a large class of internal, potentially carcinogenic agents that are highly reactive molecules and can partake in undesired reactions, causing damages to cells and particularly to DNA. Infidelity of transcription including repair can also lead to simple mutations. While these carcinogens can produce simple DNA damages, it is the faulty or imprecise DNA replication and repair machineries that lead to the complex mutations, namely undesired duplications, deletions, inversions and translocations of large DNA segments. There are multiple instances that can result in such complex genomic mutations. Eg., under persistent hypoxic conditions, cells tend to use emergency mechanisms to repair simple mutations, but the inaccuracy of such mechanisms can lead to complex mutations [41, 42]. One such mechanism, namely micro homology-mediated end joining (MMEJ) for repairing double-strand DNA breaks, through which unwanted DNA copy number changes, inversions and translocations can result [43]. Usually like the regular repair mechanism for double-strand breaks, MMEJ uses the sister chromosome as the template to replace the region with a break and the difference is that it uses a much shorter homologous region in the sister chromosome, typically 5–25 bps rather than the usual 200 bps required by the normal DNA repair mechanism. Hence the designation is micro homology-mediated. While the advantage in this mechanism is substantially faster than the regular DNA repair machinery, which is needed under certain emergency conditions, it is error prone due to the less stringent requirement for finding the equivalent region in the sister chromosome, thus leading to various complex mutations [44]. This mechanism is used only under highly stressful conditions when the regular DNA repair mechanisms are functionally repressed [45], and hence is often used in

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 2, Issue 4, April 2017

cancer associated environments. Knowledge about how different genomic mutations occur, one could possibly develop computational models to infer the evolutionary history of the mutations observed in a cancer genome from the matching reference genome. Idea is, one can first identify all the genomic differences between a cancer genome and the matching reference genome. For each identified complex mutation, one can apply a mechanistic model to predict how it occurs from the previous generation of the genome, while simple mutations can be assumed to take place randomly according to some stochastic models. It is noteworthy that some of the evolutionary intermediates (mutations) may or may not be present in the cancer genome, due to the possibilities that some portions of the genome might have been deleted during evolution. In addition, it should be emphasized that such an "Omic Data", Information Derivable and Computational Needs approach (even when taking into consideration the other emergency DNA repair mechanisms) always not necessarily yield a unique evolutionary path from the reference to the cancer genome. One possible way to constrain this phylogenetic reconstruction problem to a solution space as small as possible is to find such a path under the parsimony assumption [46], as usually used in phylogenetic reconstruction algorithms. Specifically one can require that the predicted evolutionary path have either the smallest number of generations or the highest consistency with the occurrence probabilities of different types of mutations as documented in the literature. As of now, no one has published such algorithms for making evolutionary path predictions, but the need for such tools is clearly there in order to understand the evolution of a cancer genome. Various types of information may also be derivable from cancer genomes, such as:

(1) oncogenes and tumor suppressor genes that may be specific to a particular cancer type. Eg., gene fusions as in the case of the Philadelphia chromosome for chronic myelogenous leukemia (CML) [47, 48];
(2) potential incorporation of microbial genes into the cancer genomes as in the case of hepatitis B virus genes integrated into the host genome;
(3) biological and metabolic pathways that are enriched with genetic mutations in a particular cancer, leading to the loss of function at the pathway level; and
(4) changes in mutation patterns as the cancer advances.

By systematically identifying the genomes variations of multiple patients of the same cancer type, one can identify biological pathways enriched with such mutations, using analysis tools like DAVID [49] against pathway databases such as KEGG [50, 51, 52], BIOCARTA [53] or cancer- related gene sets [54, 55]. Eg., a study, published in 2007 on genomic mutations observed across 210 cancer types, discovered that the pathway having the highest enrichment with non-synonymous mutations is the FGF (fibroblast growth factor) signaling pathway, revealing one commonality among changes needed by cancer evolution across different cancer types [56]. With such an information, one can further assume which cellular processes need to be terminated or become hyperactive in any specific order as a cancer evolves, hence possibly developing new insights about the evolutionary paths unique to particular cancer types or common among all cancer types. Epigenomic data provide information about all the chemical modifications in the genomic DNA and associated histone proteins in a cell, namely DNA methylation and histone modification, among others. While epigenetic analyses are not new, it is the high- throughput array and sequencing techniques that have made such analyses at a genome scale possible and have clearly advanced our overall capabilities to study cancer. DNA methylation is a process by which a methyl group is added to the carbon 5 position of cytosine residues (C) in CpG dinucleotides and this is accomplished through a group of enzymes known as DNA methyl-transferases, the reactions of which can be reversed by another group of enzymes termed DNA demethylases. While a CpG region is highly methylated, they attract a group of enzymes called histone deacetylases that will initiate chromatin remodeling to change the local structure of the DNA, hence altering its accessibility to large molecular structures such as the transcription machinery, RNA polymerase. Since long CpG islands tend to be associated with the promoters of genes, methylation of such regions represses the expression of the genes [14]. Histones are proteins that bind with DNA to form the basic folding units, denoted as nucleosomes, of chromatin. The packing density of chromatin is closely related to the transcriptional state of a gene, i.e., lower packing the density, higher the transcriptional activity and cells change their chromatin structures through post-translational modifications on the relevant histones, including acetylation, deamination, methylation, phosphorylation, SUMOylation and ubiquitination. The understanding is that interactions between histones and DNA are formed by electrostatic attraction between the positive charges on the histone surface and the negative charges on DNA and

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 2, Issue 4, April 2017

consequently, modifications on histones may change the charges of the surface residues, possibly changing the conformation and the transcriptional accessibility of a folded DNA and ultimately enhancing or repressing expression of the relevant genes [57, 58]. Another mechanism is through recruiting and applying chromatin remodeling ATPases, where histone modifications can lead to disruptions of ATPase attraction to the chromatin, hence altering the DNA's physical accessibility to the RNA polymerase [59]. Various technologies have been developed to reliably capture DNA methylations and histone modifications at a genome level. Among the assays that have been used for detecting methylations is the bisulfite sequencing technique [60]. By converting each methylated C to a T and removing the methylation, the bisulfite method utilizes the current sequencing techniques to produce the modified sequence and then recovers the methylation locations through comparisons between the sequenced Ts and Cs at the same locations in the original DNA and the modified DNA. Histone modification sites can be detected using the ChIP- chip array technique [61], which has been used to identify the binding sites of transcriptional factors. The difference here is to detect the DNA binding sites with histones relevant to the packing of DNA. Comparisons between the identified DNA binding sites under different conditions can lead to the identification of modified chromatin structures. The advancement of sequencing techniques in the past few years has led to the development of the second generation ChIP technique, namely ChIP-seq, which can provide more quantitative and reliable data about histone modification sites as well [62]. From any of the two types of epigenomic data, one needs to infer genes that are primed to be repressed or enhanced transcriptionally at the epigenomic level. These data, in conjunction with other omic-data such as transcriptomic and genomic information, can be used to derive association relationships between epigenomic activities and the cellular as well as micro-environmental states [63]. This leads to identification of possible triggers and regulatory pathways of different epigenomic activities. Information of this type is clearly needed since, although numerous epigenomic effectors such the enzymes for DNA methylation and histone modifications have been identified, very little is known about the regulation of these effectors and under what conditions a specific set of genes will be methylated [64]. The epigenomic level changes can be considered as an intermediate step between functional state changes of effector molecules and the permanent genetic mutations. A number of large- scale epigenomic sequencing projects have been started with similar goals to those of the genome sequencing projects. These projects include: (a) the NIH Roadmap Epigenomics Program, started in 2008 with the aim of producing histone modification data for over 30 types of modifications in a variety of human cell types;
(b) a component of the ENCODE (Encyclopedia of DNA Elements) project launched by the US National Human Genome Research Institute aiming as part of its goal the characterization of the epigenomic profiles of 50 different tissue types [65]; (c) the International Human Epigenome Consortium having its goal to build on and expand the NIH Epigenomics Program to include nonhuman cells and tissues, and to make it a functional international program; and (d) some regional epigenomics projects such as the "Epigenetics, Environment and Health" project in Canada and the Australian Alliance for Epigenetics [64, 65]. Again, one bioinformatics initiative of interest to population scientists is the cancer Biomedical Informatics Grid (caBIG®), a cyberinfrastructure designed to connect all communities in the cancer family—researchers, physicians, and patients—to share cancer genomic data and knowledge [66]. Within caBIG, sharing is predicated on interoperability, the ability of a system to access and use the parts or data of another system and this interoperability requires the development of data and software standards so that systems communicate with one another in a meaningful way to enhance data sharing [66, 67]. Underlying these standards are four principles: "federated," meaning that tools and data are widely distributed and locally controlled; "open development," where tools and infrastructure are built using an open and participatory process; "open access," where tools and resources are freely available; and "open source," where the written code (documentation) is freely available and therefore, the tools and infrastructure of caBIG are open to all in order to foster their reuse well beyond cancer research. For example, epidemiologic risk factor questionnaire and biospecimen data may be collected separately from various clinical databases; caBIG offers the tools and infrastructure to integrate disparate data to test why Indians have a higher incidence of diabetes but lower mortality rate from colon cancer. Several bioinformatics tools & techniques that are available, or in development, through caBIG or elsewhere are highlighted in this paper. Population scientists aim to understand disease patterns and develop approaches for disease prevention, detection, and diagnosis at an early stage to reduce the burden of disease [68]. Research data ranging from the molecular to the

societal level should be aggregated, integrated, and analyzed across all levels and inter-related data are needed across the disease control continuum—from the healthy through the outcome of treatment; for instance, one goal of research in this area is to investigate why African-American women have a lower incidence but higher mortality from breast cancer. But in many such areas, data often exist in a "stove-pipe," but accessible only to the investigators who generates them. Hindrances to data access and use may arise from the need for access to sources that were never intended for health research. A number of human epigenomic databases have been developed as the result of these projects [66, 67, 68].

V CONCLUSION

The bioinformatics initiatives described in this review paper enhances a collaborative and inclusive approach to development of data and tools—many of which have originated from the cancer research community, but with substantial applicability to other areas of health science research [7, 69]. Bioinformatics has changed the research landscape and provided opportunities for scientists to utilize improved methodologies, enhance use of "omic" - data, and rapidly address important research questions. Greater engagement by the population science community in bioinformatics will enhance integration of multidimensional data and new tools will help accelerate research to prevent disease, promote health, detect disease early, and reduce its impact and a substantial amount of information concerning the activities of individual biochemical pathways, their dynamics and the complex relationships among them, and with respect to various micro-environmental factors, is hidden in the very large pool of publicly available cancer omic data, including transcriptomic, genomic, metabolomic and epigenomic data [66, 67, 68]. Powerful statistical analysis techniques can assist immensely in uncovering these information if one poses the right questions. Such focused questions create a framework for hypothesis-guided data mining to check for the validity of the formulated hypothesis, as well as for guiding the formulation of further question fremework, which may ultimately lead to the elucidation of specific pathway database or even possibly causal relationships among the activities of different pathways. More powerful analysis tools for different omic data types are clearly needed in order to address more complex and deeper questions about the available data such as de- convolution of gene-expression information collected on tissue samples consisting of multiple cell types and inference of causal relationships. Integrative analyses of multiple types of omic and computational data proves to be the key to effective data mining and knowledge about genes and proteins, DOI: information discovery [27, 28].

REFERENCES:

1. Adler NE, Page AEK, editors., 2008, Cancer Care Geoffrey J. Faulkner, Subramaniam S. for the Whole Patient: Meeting Psychosocial Health Needs., National Academies Press (US).
2. Katarzyna Tomczak, Patrycja Czerwińska, and Maciej Wiznerowicz, (2015), "The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge", Contemp Oncol (Pozn), 19(1A): A68–A77.

3. Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter; 2002, Molecular Biology of the Cell. 4th edition. Garland Science.

4. Cathie Garnis, Timon PH Buys, and Wan L Lam, 2004, Genetic alteration and gene expression modulation during cancer progression, Mol Cancer. 3: 9.

5. Koonin EV, Galperin MY.Boston, 2003, Sequence - Evolution - Function: Computational Approaches in Comparative Genomics. Kluwer Academic.

6. Wong, KC (2016). Computational Biology and Bioinformatics: Gene Regulation. CRC Press (Taylor & Francis Group).

7. Duojiao Wu, Catherine M Rice, and Xiangdong Wang, 2012, Cancer bioinformatics: A new approach to systems clinical medicine, BMC Bioinformatics. 13: 71.

8. A to Z List of Cancers National Cancer Institute

9. Atul J Butte, 2009, Translational bioinformatics applications in genome medicine, Genome Med.; 1(6): 64.

10. Matthew E Holford, James P McCusker, Kei-Hoi Cheung and Michael Krauthammer, 2012, A semantic web framework to integrate cancer omics data with biological knowledge, BMC Bioinformatics. 13(Suppl 1): S10.

11. Matthew E. Holford, James P. McCusker, Kei-Hoi Cheung, and Michael Krauthammer, 2010, Analysis of Cancer Omics Data In A Semantic Web Framework, Nature Precedings.DOI: 10.1038/npre.2010.5391.1.

12. Muralidhar. S. Talkad, Anil Kumar H.V, Sourav Acharya, Gazala Showkat, Sanober Samreen, Sohini Ghosal, Jeevana Praharsha, Juganta Das, 2013, Therapeutic interventions in Cancer, International Journal of Advancements in Research & Technology, 2(7): 298-316.

13. Andrew D. Rouillard, Gregory W. Gundersen, Nicolas F. Fernandez, Zichen Wang, Caroline D. Monteiro, Michael G. McDermott and Avi Ma'ayan, 2016, The harmonizome: a collection of processed datasets gathered to serve and mine

14. https://doi.org/10.1093/database/baw100 Jian-Liang Li, Joseph Mazar, Cuncong Zhong Govindarajan, Zhan Zhang, Marcel E. Dinger, Gavin Meredith, Christopher Adams, Shaojie Zhang, John S. Mattick, Animesh Ray & Ranjan J. Perera, 2013, Genome-wide methylated CpG island profiles of melanoma cells reveal a melanoma coregulation network, SCIENTIFIC REPORTS | 3 : 2962 | DOI: 10.1038/srep02962.

15. Mark TW Ebbert, Roy RL Bastien, Kenneth M Boucher, Miguel Martín, Eva Carrasco, Rosalía Caballero, Inge J Stijleman, Philip S Bernard and Julio C Facelli, 2011, "Characterization of uncertainty in the classification of multivariate assays: application to PAM50 centroid-based genomic predictors for breast cancer treatment plans", J Clin Bioinforma., 1: 37.

16. Brett Wallden, James Storhoff, Torsten Nielsen, Naeem Dowidar, Carl Schaper, Sean Ferree, Shuzhen Liu, Samuel Leung, Gary Geiss, Jacqueline Snider, Tammi Vickery, Sherri R. Davies, Elaine R. Mardis, Michael Gnant, Ivana Sestak, Matthew J. Ellis, Charles M. Perou, Philip S. Bernard, and Joel S. Parker, 2015, "Development and verification of the PAM50- based Prosigna breast cancer gene signature assay", BMC Med Genomics., 8: 54.

17. Silvia von der Heyde and Tim Beissbarth, 2012, "A new analysis approach of epidermal growth factor receptor pathway activation patterns provides insights into cetuximab resistance mechanisms in head and neck cancer", BMC Med., 10: 43.

18. Andrew Schoenrock, Bahram Samanfar, Sylvain Pitre, Mohsen Hooshyar, Ke Jin, Charles A Phillips, Hui Wang, Sadhna Phanse, Katayoun Omidi, Yuan Gui, Md Alamgir, Alex Wong, Fredrik Barrenäs, Mohan Babu, Mikael Benson, Michael A Langston, James R Green, Frank Dehne, and Ashkan Golshani, 2014, "Efficient prediction of human protein-protein interactions at a global scale", BMC Bioinformatics; 15(1): 383.

19. Vladimir Kuznetsov, Hwee Kuan Lee, Sebastian Maurer-Stroh, Maria Judit Molnár, Sandor Pongor, Birgit Eisenhaber and Frank Eisenhaber, 2012, "How bioinformatics influences health informatics: usage of biomolecular sequences, expression profiles and automated microscopic image analyses for clinical needs and public health", Health Information Science and Systems, 1:2.

20. Wayne Katon, Alfred O. Berg, Anthony J. Robins, And Steven Risse, 1986, "Depression Mecical Utilization and Somatization", Clinical Medicine, 144(5): 564-568.

21. Duncan Ayers and Philip J. Day, 2015, "Systems Medicine: The Application of Systems Biology Approaches for Modern Medical Research and Drug Development", Molecular Biology International, Volume 2015: 1-8

22. Xiangdong Wang, 2011, "Role of clinical bioinformatics in the development of network- based Biomarkers", Journal of Clinical Bioinformatics, 1:28.

23. Brittany D. Conroy, Tyler A. Herek, Timothy D. Shew, Matthew Latner, Joshua J. Larson, Laura Allen, Paul H. Davis, Tomáš Helikar and Christine E. Cutucache, 2014, "Design, assessment, and in vivo evaluation of a computational model illustrating the role of CAV1 in CD4+ T-lymphocytes", Front. Immunol., 5:1-9.

24. Cun Y, Fröhlich H, 2012, "Prognostic Gene Signatures for Patient Stratification in Breast Cancer - Accuracy, Stability and Interpretability of Gene Selection Approaches Using Prior Knowledge on Protein-Protein Interactions. BMC Bioinformatics. 13: 69.

25. Jason E. Stajich and Hilmar Lapp, 2006, "Open source tools and toolkits for bioinformatics: significance, and where are we?", Briefings In Bioinformatics, 7(3):287-296.

26. Douglas Hanahan, Robert A. Weinberg, 2011, "Hallmarks of Cancer: The Next Generation" Cell, 144(5):646–674.

27. Katarzyna Tomczak, Patrycja Czerwińska and Maciej Wiznerowicz, 2015, "The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge", Contemp Oncol (Pozn).; 19(1A): A68–A77.

28. Katherine A. Hoadley, Christina Yau, Denise M. Wolf, Andrew D. Cherniack, David Tamborero, Sam Ng, Max D.M. Leiserson, Beifang Niu, Michael D. McLellan, Vladislav Uzunangelov, Jiashan Zhang, Cyriac Kandoth, Rehan Akbani, Hui Shen, Larsson Omberg, Andy Chu, Adam A. Margolin, Laura J. van't Veer, Nuria Lopez-Bigas, Peter W. Laird, Benjamin J. Raphael, Li Ding, A. Gordon Robertson, Lauren A. Byers, Gordon B. Mills, John N. Weinstein, Carter Van Waes, Zhong Chen, Eric A. Collisson, The Cancer Genome Atlas Research Network, Christopher Benz,corresponding author Charles M. Perou, and Joshua M. Stuart, 2014, Multi-platform analysis of 12 cancer types reveals molecular classification within and across tissues-of-origin, Cell. 158(4): 929–944.

29. Winston Timp, Utkur M. Mirsaidov, Deqiang Wang, Jeff Comer, Aleksei Aksimentiev, and Gregory Timp, 2010, "Nanopore Sequencing: Electrical Measurements of the Code of Life", IEEE Trans Nanotechnol. 9(3): 281–294.

30. Human Genome Project

31. Cavalli-Sforza LL, 2005, "The Human Genome Diversity Project: past, present and future", Nat. Rev. Genet. 6 (4):333–340.

32. Cotton RG, Auerbach AD, Beckmann JS, Blumenfeld OO, Brookes AJ, Brown AF, Carrera P, Cox DW, Gottlieb B, Greenblatt MS, Hilbert P, Lehvaslaiho H, Liang P, Marsh S, Nebert DW, Povey S, Rossetti S, Scriver CR, Summar M, Tolan DR, Verma IC, Vihinen M, den Dunnen JT, 2008, "Recommendations for locus-specific databases and their curation", Hum Mutat 29:2–5.

33. The International HapMap Consortium. The International HapMap Project. Nature, 426: 789- 796. 2003.

34. Service RF, (2006), Gene sequenccing. the race for the $1000 genome. Science 311: 1544-1546.

35. Church, G. M. (2005). "The Personal Genome Project". Molecular Systems Biology. 1 (1): E1– E3.

36. Richard E. Green, Johannes Krause, Adrian W. Briggs, Tomislav Maricic, Udo Stenzel, Martin Kircher, Nick Patterson, Heng Li, Weiwei Zhai, Markus Hsi-Yang Fritz, Nancy F. Hansen, Eric Y. Durand, Anna-Sapfo Malaspinas, Jeffrey D. Jensen, Tomas Marques-Bonet, Can Alkan, Kay Prüfer, Matthias Meyer, Hernán A. Burbano, Jeffrey M. Good, Rigo Schultz, Ayinuer Aximu-Petri, Anne Butthof, Barbara Höber, Barbara Höffner, Madlen Siegemund, Antje Weihmann, Chad Nusbaum, Eric S. Lander, Carsten Russ, Nathaniel Novod, Jason Affourtit, Michael Egholm, Christine Verna, Pavao Rudan, Dejana Brajkovic, Željko Kucan, Ivan Gušic, Vladimir B. Doronichev, Liubov V. Golovanova, Carles Lalueza-Fox, Marco de la Rasilla, Javier Fortea, Antonio Rosas, Ralf W. Schmitz, Philip L. F. Johnson, Evan E. Eichler, Daniel Falush, Ewan Birney, James C. Mullikin, Montgomery Slatkin, Rasmus Nielsen, Janet Kelso, Michael Lachmann, David Reich, and Svante Pääbo, 2010, "A draft sequence of the Neandertal genome. Science", 328(5979): 710–722.

37. Cheng Z, Ventura M, She X, Khaitovich P, Graves T, Osoegawa K, Church D, DeJong P, Wilson RK, Pääbo S, Rocchi M & Eichler EE, 2005, "A genome-wide comparison of recent chimpanzee and human segmental duplications", Nature 437: 88-93.

38. Angata T, Margulies EH, Green ED, and Varki A, 2004, "Large-scale sequencing of the CD33- related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms", Proc. Natl. Acad. Sci. 101: 13251-13256.

39. Cancer-Genome-Atlas-Research 2013.

40. Lynda Chin, William C. Hahn, Gad Getz and Matthew Meyerson, 2011, "Making sense of cancer genomic data", Genes Dev. 25(6): 534–555.

41. Ying Xu, Juan Cui, David Puett, 2014, Cancer Bioinformatics, Springer.

42. Scanlon S, Glazer P (2013) Genetic Instability Induced by Hypoxic Stress. In: Mittelman D (ed) Stress-Induced Mutagenesis. Springer New York, pp 151–181.

43. Lan N. Truong, Yongjiang Li, Linda Z. Shi, Patty Yi-Hwa Hwang, Jing He, Hailong Wang, Niema Razavian, Michael W. Berns, and Xiaohua Wua, 2013, "Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells", Proc Natl Acad Sci U S A. 110(19): 7720–7725.

44. Bentley J, Diggle CP, Harnden P, Knowles MA, Kiltie AE, (2004) DNA double strand break repair in human bladder cancer is error prone and involves micro-homology-associated end-joining. Nucleic acids research 32: 5249–5259.

45. Bindra RS, Crosby ME, Glazer PM (2007) Regulation of DNA repair in hypoxic cancer cells. Cancer Metastasis Rev 26: 249–260.

46. Steel MA, Penny D. 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. Mol. Biol. Evol. 17:839–50.

47. Nowell P., Hungerford D., 1960, A minute chromosome in human chronic granulocytic leukemia. Science. 132:1497.

48. Nowell P.C., Hungerford D.A., 1960, Chromosome studies on normal and leukemic human leukocytes. J. Natl. Cancer Inst. 25:85–109.

49. Huang DW, Lempicki RA, Sherman BT. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44–57.

50. Yuki Moriya, Daichi Shigemizu, Masahiro Hattori, Toshiaki Tokimatsu, Masaaki Kotera, Susumu Goto, Minoru Kanehisa, (2010), "PathPred: an enzyme-catalyzed metabolic pathway prediction server, Nucleic Acids Res, 38 (suppl_2): W138- W143.

51. Minoru Kanehisa, Susumu Goto, Yoko Sato, Miho Furumichi, Mao Tanabe, 2010, "KEGG for integration and interpretation of large-scale molecular data sets", Nucleic Acids Res 40 (D1): D109-D114.

52. Minoru Kanehisa, Susumu Goto, Yoko Sato, Masayuki Kawashima, Miho Furumichi, Mao Tanabe, 2014, "Data, information, knowledge and principle: back to metabolism in KEGG", Nucleic Acids Res 42 (D1): D199-D205.

53. Darryl Nishimura, 2004, "BioCarta", Biotech Software & Internet Report. 2(3): 117-120.

54. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, Stratton MR, Futreal PA, 2011, "COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer:, Nucleic Acids Research., 39(Suppl 1):D945–D950.

55. Zhou C, Licciulli S, Avila JL, Cho M, Troutman S, Jiang P, Kossenkov AV, Showe LC, Liu Q, Vachani A, Albelda SM, Kissil JL, 2013, "The Rac1 splice form Rac1b promotes K-ras-induced lung tumorigenesis", Oncogene 32: 903–909.

56. Greenman C., Stephens P., Smith R., Dalgliesh G.L., Hunter C., Bignell G., Davies H., Teague J., Butler A., Stevens C., Edkins S., O'Meara S., Vastrik I., Schmidt E.E., Avis T., Barthorpe S., Bhamra G., Buck G., Choudhury B., Clements J., Cole J., Dicks E., Forbes S., Gray K., Halliday K., Harrison R., Hills K., Hinton J., Jenkinson A., Jones D., Menzies A., Mironenko T., Perry J., Raine K., Richardson D., Shepherd R., Small A., Tofts C., Varian J., Webb T., West S., Widaa S., Yates A., Cahill D.P., Louis D.N., Goldstraw P., Nicholson A.G., Brasseur F., Looijenga L., Weber B.L., Chiew Y.-E., DeFazio A., Greaves M.F., Green A.R., Campbell P., Birney E., Easton D.F., Chenevix-Trench G., Tan M.-H., Khoo S.K., Teh B.T., Yuen S.T., Leung S.Y., Wooster R., Futreal P.A., Stratton M.R. Patterns of somatic mutation in human cancer genomes. Nature. 2007; 446:153– 158.

57. Strahl BD, Allis CD 2000. The language of covalent histone modifications. Nature 403: 41– 45.

58. Kamakaka RT, Biggins S. Histone variants: Deviants? Genes Develop. 2005;19:295–310.

59. Vignali M., Steger, D.J., Neely, K.E. and Workman, J.L. (2000) Distribution of acetylated histones resulting from Gal4-VP16 recruitment of SAGA and NuA4 complexes. EMBO J., 19, 2629– 2640.

60. Allen S. Yang, Marcos R. H. EsteÂcio, Ketan Doshi, Yutaka Kondo, Eloiza H. Tajara and Jean- Pierre J. Issa, 2004, A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements, Nucleic Acids Research, 32(3)e38: 1-6.

61. Huebert, D. J., Kamal, M., O'Donovan, A., and Bernstein, B. E. (2006). Genome-wide analysis of histone modifications by ChIP-on-chip. Methods 40, 365-369.

62. Benjamin L. Kidder, Gangqing Hu, and Keji Zhao, 2011,ChIP-Seq: Technical Considerations for Obtaining High Quality Data, Nat Immunol. 12(10): 918–922.

63. Michiel E. Adriaens, Peggy Prickaerts, Michelle Chan-Seng-Yue, Twan van den Beucken, Vivian E. H. Dahlmans, Lars M. Eijssen, Timothy Beck, Bradly G. Wouters, Jan Willem Voncken and Chris T. A. Evelo, 2016, "Quantitative analysis of ChIP-seq data uncovers dynamic and sustained H3K4me3 and H3K27me3 modulation in cancer cells under hypoxia", Epigenetics & Chromatin, 9:48.

64. Diane E. Handy, Rita Castro, and Joseph Loscalzo, 2011, "Epigenetic Modifications: Basic Mechanisms and Role in Cardiovascular Disease", Circulation., 123(19): 2145–2156.

65. Joost H.A. Martens, Hendrik G. Stunnenberg, and Colin Logie, 2011, "The Decade of the Epigenomes?", Genes Cancer. 2(6): 680–687.

66. J. Robert Beck, 2010, The Cancer Biomedical Informatics Grid (caBIG,): An Evolving Community for Cancer Research, Biomedical Informatics for Cancer Research, pp 177-200.

67. Jiajia Chen, Fuliang Qian, Wenying Yan, and Bairong Shen,2013, "Translational Biomedical Informatics in the Cloud: Present and Future", Biomed Res Int. 2013: 658925.

68. Michele R. Forman, Sarah M. Greene, Nancy E. Avis, Stephen H. Taplin, Paul Courtney, Peter A. Schad, Bradford W. Hesse, Deborah M. Winn, 2010, "Bioinformatics: Tools to Accelerate Population Science and Disease Control Research", Am J Prev Med, 38(6):646–651.

69. Rebecca S. Jacobson, Michael J. Becich, Roni J. Bollag, Girish Chavan, Julia Corrigan, Rajiv Dhir, Michael D. Feldman, Carmelo Gaudioso, Elizabeth Legowski, Nita J. Maihle, Kevin Mitchell, Monica Murphy, Mayurapriyan Sakthivel, Eugene Tseytlin and JoEllen Weaver, 2015, "A Federated Network for Translational Cancer Research Using Clinical Data and Biospecimens", Cancer Research, 75(24):5194–201.