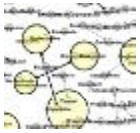
# Bioinformatics and Medicine

Jeremy Yang

Associate Instructor
IU School of Informatics and Computing

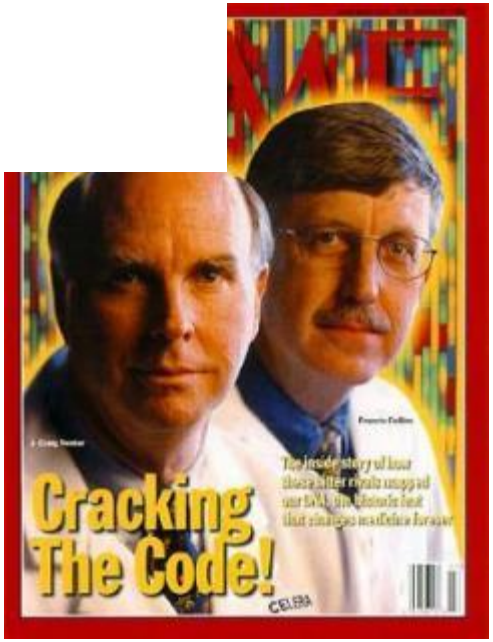
*Instructor: Prof. Joanne Luciano*

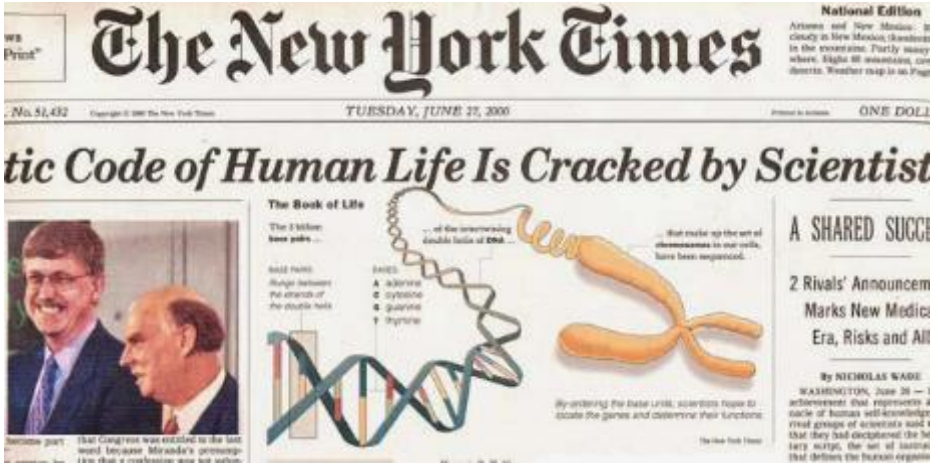


DATA SCIENCE FOR DRUG DISCOVERY, HEALTH AND TRANSLATIONAL MEDICINE (DSDHT) INFO I-590

# Bioinformatics and Medicine: Hype vs Reality

*"What more powerful form of study of mankind could there be than to read our own instruction book"* – Francis S. Collins, Director, NIH.


HGP: Craig Venter and Francis Collins, 2000

# Bioinformatics and Medicine: Gleevec, a motivating example

## Gleevec: the Breakthrough in Cancer Treatment

By: Leslie A. Pray, Ph.D. © 2008 Nature Education

Citation: Pray, L. (2008) Gleevec: the Breakthrough in Cancer Treatment. *Nature Education* 1(1):37

"Some say it's a miracle drug. Others call it a silver bullet. Gleevec, also marketed internationally as Glivec and sometimes referred to by its chemical name imatinib, entered the medical world with a bang. This medication was initially approved for use by the U.S. Food and Drug Administration (FDA) in 2001 for the treatment of chronic myelogenous leukemia (CML), a rare form of cancer that affects certain types of white blood cells. Since its initial approval, Gleevec has also been approved for use in patients with several types of gastrointestinal tumors. Currently, scientists continue to study the drug's effectiveness not only in various cancers, but also in other diseases, such as stroke (Su *et al*., 2008)."

From Scitable, by Nature Education, NPG,
http://www.nature.com/scitable/topicpage/gleevec-the-breakthrough-in-cancer-treatment-565

# Bioinformatics and Medicine: Gleevec

**Gleevec: the Breakthrough in Cancer Treatment**

By: Leslie A. Pray, Ph.D. © 2008 Nature Education

Citation: Pray, L. (2008) Gleevec: the Breakthrough in Cancer Treatment. *Nature Education* 1(1):37

- Gleevec inhibits a tyrosine kinase, a protein involved in cell division.
- Acute myelogenous leukemia (AML) associated with chromosomal abnormality in leukocytes (1950s, cytogenetics), a somatic mutation involving the BCR and ABL genes (1980s, sequencing).
- The mutated ABL is an oncogene, BCR+ABL expressed in the abnormal tyrosine kinase which causes AML.

From Scitable, by Nature Education, NPG,
http://www.nature.com/scitable/topicpage/gleevec-the-breakthrough-in-cancer-treatment-565

# Bioinformatics and Medicine: Gleevec

## Gleevec: the Breakthrough in Cancer Treatment

By: Leslie A. Pray, Ph.D. © 2008 Nature Education

Citation: Pray, L. (2008) Gleevec: the Breakthrough in Cancer Treatment. *Nature Education* 1(1):37

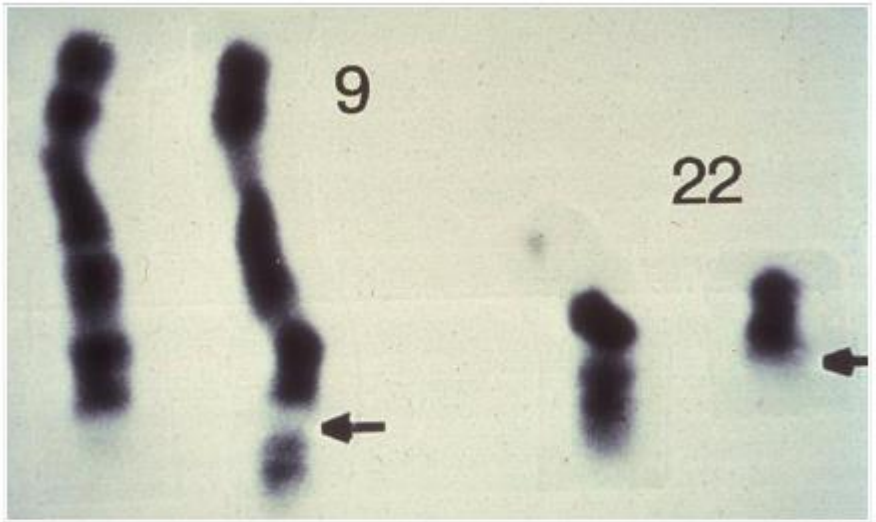
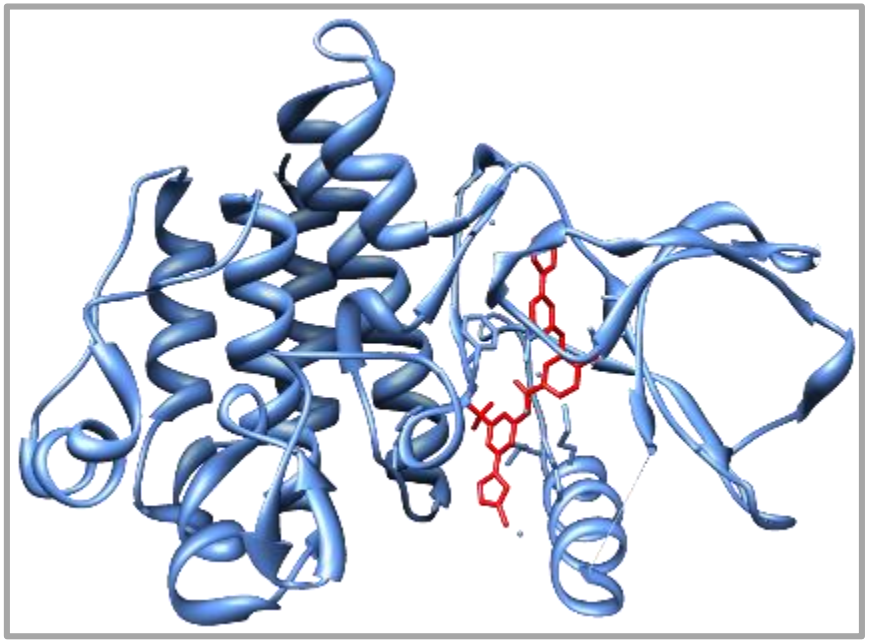



**Figure 1 : The Philadelphia chromosome**

Discovered in 1960, the diagnostic karyotypic abnormality for chronic myelogenous leukemia is shown in this picture of the banded chromosomes 9 and 22. Shown is the result of the reciprocal translocation of 22q to the lower arm of 9 and 9q (c-abl to a specific breakpoint cluster region [bcr] of chromosome 22 indicated by the arrows).



The Philadelphia chromosome

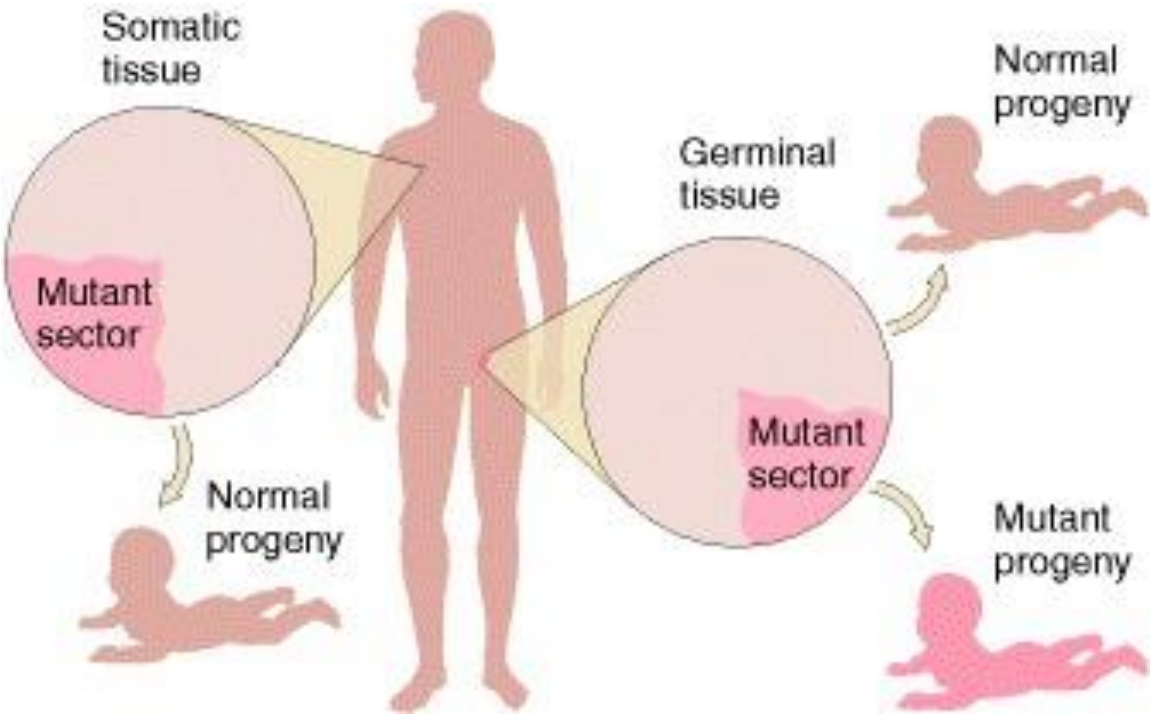


ABL1 with Gleevec in ATP binding site

From Scitable, by Nature Education, NPG,
http://www.nature.com/scitable/topicpage/gleevec-the-breakthrough-in-cancer-treatment-565

# Bioinformatics and Medicine: Germline vs. Somatic Mutations

"Genes and chromosomes can mutate in either somatic or germinal tissue, and these changes are called somatic mutations and germinal mutations, respectively."




From: An Introduction to Genetic Analysis. 7th ed., Griffiths, et al., W. H. Freeman; 2000, https://www.ncbi.nlm.nih.gov/books/NBK21894/.

# Bioinformatics and Medicine: Germline vs. Somatic Mutations

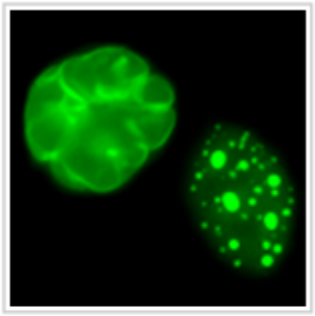
Vol 463|14 January 2010|doi:10.1038/nature08658 nature

## A comprehensive catalogue of somatic mutations from a human cancer genome

Erin D. Pleasance[1]*, R. Keira Cheetham[2]*, Philip J. Stephens[1], David J. McBride[1], Sean J. Humphray[2], Chris D. Greenman[1], Ignacio Varela[1], Meng-Lay Lin[1], Gonzalo R. Ordóñez[1], Graham R. Bignell[1], Kai Ye[3], Julie Alipaz[4],

**All cancers carry somatic mutations. A subset of these somatic alterations, termed driver mutations, confer selective growth advantage and are implicated in cancer development, whereas the remainder are passengers. Here we have sequenced the genomes of a malignant melanoma and a lymphoblastoid cell line from the same person, providing the first comprehensive catalogue of somatic mutations from an individual cancer. The catalogue provides remarkable insights into the forces that have shaped this cancer genome. The dominant mutational signature reflects DNA damage due to ultraviolet light exposure, a known risk factor for malignant melanoma, whereas the uneven distribution of mutations across the genome, with a lower prevalence in gene footprints, indicates that DNA repair has been preferentially deployed towards transcribed regions. The results illustrate the power of a cancer genome sequence to reveal traces of the DNA damage, repair, mutation and selection processes that were operative years before the cancer became symptomatic.**

# Bioinformatics and Medicine



## Genes and Disease

Editor(s): Alexandre Vieira | Subscribe

This topic room focuses on mechanisms of disease. In doing so, it explores why some individuals are affected by specific conditions, such as polydactyly, spina bifida, and cancer. In addition, it discusses what scientists have done and what tools they have developed to investigate these conditions in the effort to better treat or prevent them. However, this topic room does not aim to provide information on every human disease. Rather, its goal is to cultivate interest and awareness of the complex relationship between human genetics and various disease states.

Based on their genetic contribution, human diseases can be classified as monogenic, chromosomal, or multifactorial. Monogenic diseases are caused by alterations in a single gene, and they segregate in families according to the traditional Mendelian principles of inheritance. Chromosomal diseases, as their name implies, are caused by alterations in chromosomes. For instance, within an individual's genome, some chromosomes may be missing, extra chromosome copies may be present, or certain portions of chromosomes may be deleted or duplicated. Finally, the vast majority of human diseases can be categorized as multifactorial. These conditions are also referred to as complex diseases, and they are

From Scitable, by Nature Education, NPG,
http://www.nature.com/scitable/topic/genes-and-disease-17

# Bioinformatics and Medicine

## Birth Defects: Causes and Statistics

By: Ingrid Lobo, Ph.D. (*Write Science Right*) & Kira Zhaurova, M.S. (*Nature Education*) 
Citation: Lobo, I. & Zhaurova, K. (2008) Birth defects: causes and statistics. *Nature Education* 1(1):18

Every year, an estimated 7.9 million infants (6% of worldwide births) are born with serious birth defects. Although some congenital defects can be controlled and treated, an estimated 3.2 million of these children are disabled for life. Moreover, birth defects are the leading cause of infant mortality in the United States. But where do these defects come from? Although some birth defects are inherited, others are a product of harmful environmental factors known as teratogens, and still others are multifactorial, resulting from a complex interaction of genetic and environmental influences. However, in approximately half of all birth defect cases, the causes are unknown (Christianson *et al.*, 2006).

Genetic causes of birth defects fall into three general categories: chromosomal abnormalities, single-gene defects, and multifactorial influences. Prenatal environment can play a major role in the development of defects in all three categories, especially those linked to multifactorial causes.

http://www.nature.com/scitable/topicpage/birth-defects-causes-and-statistics-863

# Bioinformatics and Medicine

## Huntington's Disease: The Discovery of the Huntingtin Gene

By: Heidi Chial, Ph.D. (*Write Science Right*) © 2008 Nature Education
Citation: Chial, H. (2008) Huntington's disease: The discovery of the Huntingtin gene. *Nature Education* 1(1):71

Huntingtin was the first disease gene mapped to a specific chromosome. How did scientists do it and what have we learned since then?

Aa Aa Aa

What if a simple blood test could tell you with absolute certainty that you would suffer from a deadly neurodegenerative disease late in life? What if the same test could tell you your chances of passing this disease on to your children? What if this disease had no possible treatment or cure? Such is the current state of Huntington's disease (HD), an adult-onset autosomal dominant disorder. Today, researchers can literally "measure" the HD-associated gene, called huntingtin (*HTT*), by determining the number of repeats of a set of three specific bases within this gene. An individual with 40 or more of these repeats in one copy of the *HTT* gene will certainly suffer from HD.

### HD: Discovery, Inheritance Patterns, and Phenotypes

HD is a rare, adult-onset, autosomal dominant, progressive neurodegenerative disease. George Huntington (Figure 1) was the first person to provide a comprehensive description of adult-onset HD in 1872; he was only 22 years old at the time. Huntington described the autosomal dominant inheritance pattern of this condition, which is


THE
MEDICAL AND SURGICAL REPORTER.
ORIGINAL DEPARTMENT.
Communications

http://www.nature.com/scitable/topicpage/huntington-s-disease-the-discovery-of-the-851

# Bioinformatics and Medicine

## Proto-oncogenes to Oncogenes to Cancer

By: Heidi Chial, Ph.D. (*Write Science Right*) © 2008 Nature Education

Citation: Chial, H. (2008) Proto-oncogenes to oncogenes to cancer. *Nature Education* 1(1):33

What drives cancer cells to grow and divide uncontrollably turning into cancer? Studies of proto-oncogenes reveal some clues about how normal cellular processes mutate and go awry.

Aa Aa Aa

One out of every two men and one out of every three women will develop cancer during their lifetime (American Cancer Society, 2008). The current list of known cancer genes includes 70 genes associated with germline mutations and 342 genes associated with somatic mutations. Generally speaking, however, mutations in two basic classes of genes—proto-oncogenes and tumor suppressor genes—are what lead to cancer. In fact, a recent high-throughput study of proto-oncogene mutations in 1,000 different tumor samples representing 17 different types of cancer showed that mutations in a set of 14 proto-oncogenes are associated with a high propensity for cancer. (See Figure 1 for a depiction of the association between these genes and certain forms of cancer.) Moreover, this study also revealed that the 14 proto-oncogenes in question are associated with diverse cellular functions (Thomas *et al.*, 2007). But what drives cancerous cells to grow and divide uncontrollably and escape cell death, and just how are proto-oncogenes involved in this process?

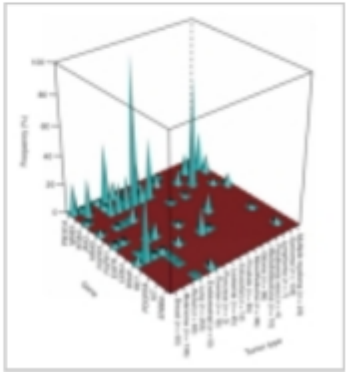
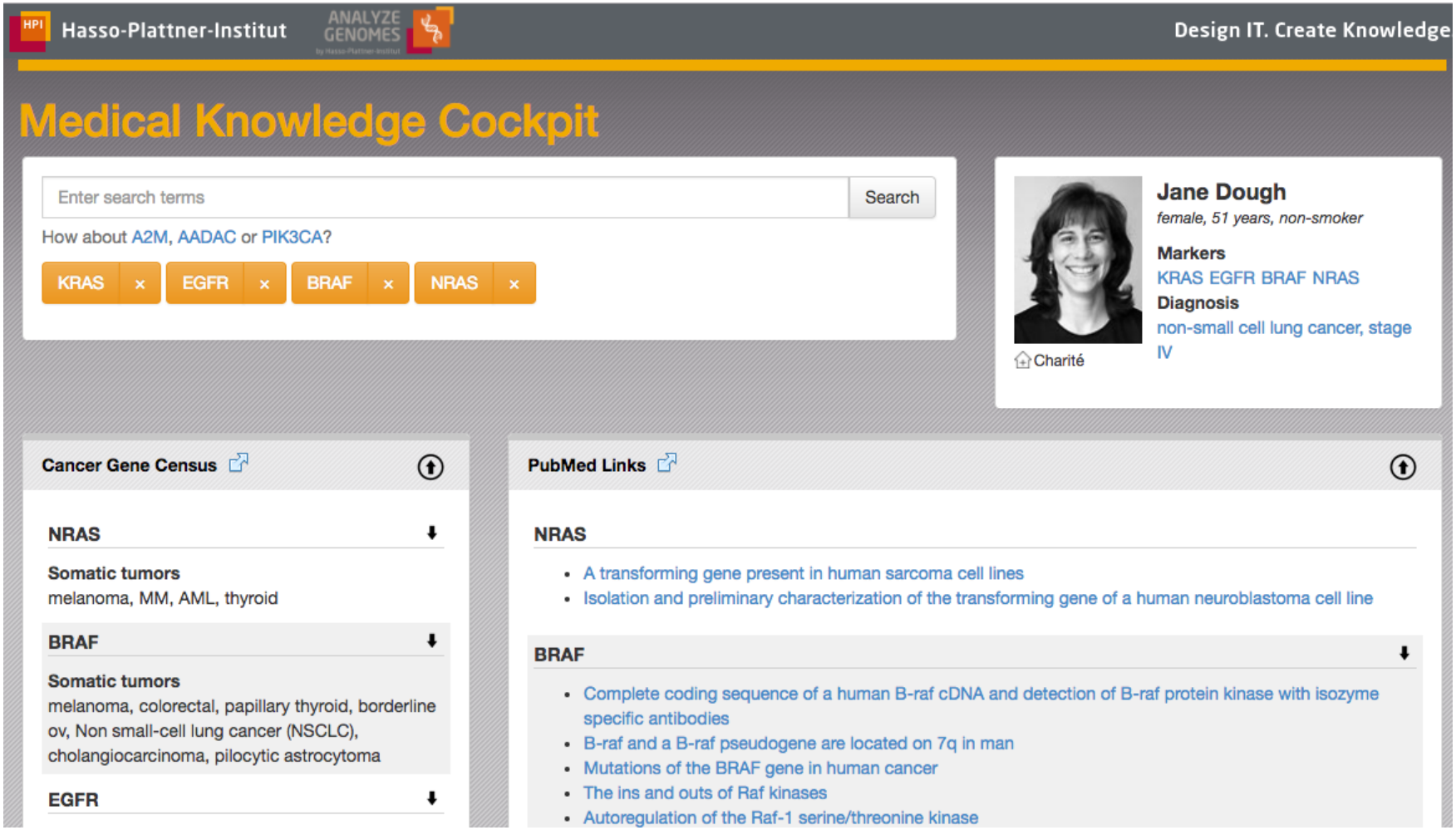


Figure 1

http://www.nature.com/scitable/topicpage/proto-oncogenes-to-oncogenes-to-cancer-883

# Bioinformatics and Medicine




Code of Life, when computer science meets genetics, HPI MOOC, https://open.hpi.de/courses/ehealth2016, https://we.analyzegenomes.com/

# Bioinformatics and Medicine: Data Science Applications

- Bibliology & WWW programming
- GWAS, Genome Wide Association Studies
  - SNPs vs phenotypes (disease)
- Gene Expression data analyses
  - Present vs functional
- Sequence alignment & similarity analyses (e.g. BLAST)
- Cross-species analysis via gene homology/orthology
- Gene-gene/protein-protein interaction network analysis
- Pathways, reactomics, metabolomics

# Bioinformatics and Medicine: Take home messages

- Bioinformatics addresses the molecular basis of health and disease.
- Genomic variations between individuals, tissues and cells can help explain and address disease.
- Bioinformatics and Data Science can be used across the translational spectrum, from basic to clinical science, in medical practice and personal health.