# Cancer Development And Complexity 2017

*Lake Como School of Advanced Studies, May 23 - 26, 2017*



## Home

**Villa del Grumello, Como, Italy, May 23-26, 2017**

Cancer is a complex disease involving several intertwined phenomena and events, which collude to unleash the tumor cells inherent programs to proliferate, live, and move; thus, it is the malfunction of the biomolecular machinery responsible for the "checks and balances", normally governed by various complex feedback loops among a population of various cell types. Breakdown of this machinery leads to uncontrolled growth of a cell population being selected by evolutionary pressure that ultimately costs the very survival of the host.

Understanding the many intricacies of all these interactions at the sub cellular, cellular and tissue levels has greatly benefitted from the ever-improving applications of algorithmic, statistical and mathematical modeling tools. Moreover, during the past 15 years, new measurement technology for gene expression and, more recently, "deep" genome sequence data, have produced vast amount of data, waiting to be analyzed to deliver new interpretations. The design of novel "wet" experiments and appropriately matched algorithmic, statistical and mathematical modeling tools are expected to become the key to successful oncological science and practice.

The Workshop and School on Cancer Development and Complexity seeks to convene researchers from various related disciplines to explore multiple facets of the challenges posed by cancer a "disease of the systems." The workshop will provide opportunities for the researchers to exchange new ideas and viewpoints, forge new collaborations and train the next generation of young scientists.

Participants are encouraged to present their work in two sessions and poster presentations that will be held during the workshop.

The program will provide an introduction to both cancer biology and mathematical and statistical methods used in analyzing the datasets currently being produced by several laboratories around the world. Next the program will provide an opportunity to interact with world renowned cancer and bioinformatics researchers and the chance for attendees to present their current work. Finally, all the attendees will receive a certificate of completion of the School.

### School Directors

- Marco Antoniotti, BIMIB, Dipartimento di Informatica, Sistemistica e Comunicazione, Universita` degli Studi di Milano Bicocca, Milan, Italy.
- Bud Mishra, Courant Institute of Mathematical Sciences, and Tandon School of Engineering, NYU, New York, NY, USA.

### Keynote speakers

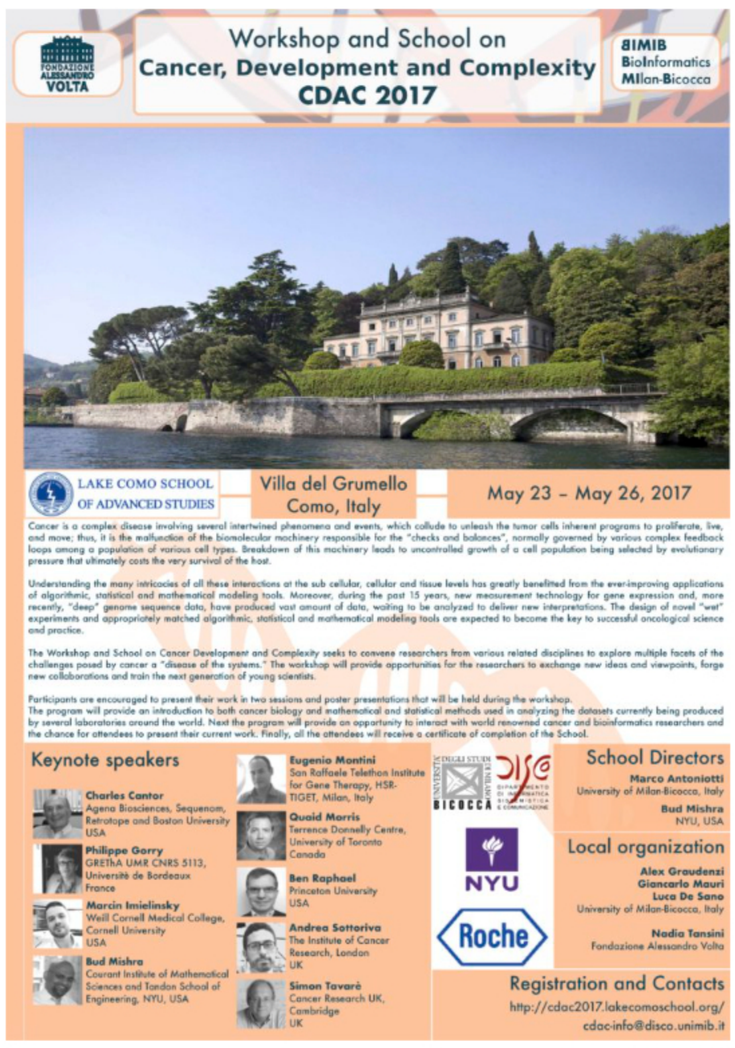
- Charles Cantor, Agena Biosciences, Sequenom, Retrotope and Boston University, USA
- Philippe Gorry, GREThA UMR CNRS 5113, Universitè de Bordeaux, France
- Marcin Imielinsky, Weill Cornell Medical College, Cornell University, New York, USA.
- Bud Mishra, Courant Institute of Mathematical Sciences and Tandon School of Engineering, NYU, New York, USA.
- Eugenio Montini, San Raffaele Telethon Institute for Gene Therapy, HSR-TIGET, Milan, Italy
- Quaid Morris, Banting and Best Department of Medical Research Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Canada.
- Ben Raphael, Princeton University, Princeton, USA.
- Andrea Sottoriva, The Institute of Cancer Research, London, UK.
- Simon Tavarè, Cancer Research UK, Cambridge, UK.

### Local Organization

- Alex Graudenzi (BIMIB)
- Giancarlo Mauri (BIMIB)
- Luca De Sano (BIMIB)
- Nadia Tansini (Fondazione Alessandro Volta)

### Registration

Please visit the Registration page to sign up.




Sponsored by

Roche

## Keynote Speakers and Lectures

**Charles Cantor** (Agena Biosciences, Sequenom, Retrotope and Boston University, USA)

***Tutorial: Is precision medicine possible?***

***Research: Sensitive detection of low levels of cancer-specific DNA sequence difference***

---

**Philippe Gorry** (GREThA UMR CNRS 5113, Université de Bordeaux, France)

***Tutorial: Pharma Competitive Intelligence for Dummies***

*Pharma-biotech industry is facing a changing world with many market challenges. In that context, technology scouting is evolving toward pharma competitive intelligence and would take advantage of the information flow during the drug R&D cycle times. Use of open sources databases for drug R&D milestone and versatile data visualization tool can help to visualize trends across market, bringing real-time clinical data, across geographies and demographics. It could help biotech startups as well academic laboratories to manages flow of competitive intelligence information, learn about new technologies in the market, that could to benefit research laboratories by improving discovery platforms or by identifying public-private partnership opportunities.*

***Research Topic: "Economic Dimensions of Precision Medicine in Oncology"***

*Precision medicine is a new paradigm for disease treatment and prevention, taking in account genes, environment, and lifestyle for each patient. In oncology, it is the use of genomic information about person's tumor in order to individualize treatments. It is rapidly gaining prominence in OECD countries, and it is supported by the pharma industry, moving from blockbuster to niche buster drug model, with the support of the Orphan Drug legislation. Today, the increased availability of targeted therapies with their high cost, and their use for chronic disease, raises a debate about accessibility, cost-effectiveness, and reimbursement. So, if precision medicine approaches become part of routine healthcare, paying for precision drugs would be a future challenge by checking traditional models of health protection systems. We will take advantage of on-going research on the Orphan Drug market to explore the economics dimensions of niche buster drug model on innovation dynamics, intellectual property protections, R&D return on investment and welfare impact.*

---

**Marcin Imielinsky** (Weill Cornell Medical College, Cornell University, New York, USA)

***Tutorial: Deciphering junctions, copy number, and walks in cancer whole genomes with linked-reads***

*The structural structure of cancer DNA can be represented as a genome graph of nodes representing reference genomic intervals (DNA alleles) and edges representing junctions (3-5 phosphodiester bonds). A subset of edges in the genome graph are variant– representing signed pairs of genomic locations that were previously distant (e.g. in the reference genome) and are made adjacent (e.g. in a cancer) through DNA breakage and fusion. The copy number and rearrangement states of cancer genomes are constrained by junction balance, a constraint that arises from from the physical intuition that every copy of every (non-telomeric) interval in the genome must have a left and a right neighbor. In practical terms, junction balance requires the copy number of nodes in the genome graph to be consistent with the copy number of incoming and outgoing edges. I demonstrate how this inference can be performed on standard cancer whole genome sequencing (WGS) data through the analysis of read depth subject to junction balance constraints. We implement this robustly as a mixed-integer quadratic program optimization (JaBbA, https://github.com/mskilab/JaBbA) which also allows for loose ends that represent missing junctions (e.g. rearrangements occurring in unamappable genomic regions) and result in copy number changes unaccounted for by graph topology. I also present an R toolkit for analyzing, exploring, and visualizing annotated cancer genome graphs (gGnome, https://github.com/mskilab/JaBbA). One application of gGnome is decomposition of genome graphs into walks representing possible phases of rearranged alleles. Using standard WGS, the phase of these alleles cannot be often unambiguously reconstructed at long distances (e.g. 100Kbp-1Mbp). I demonstrate how the analysis of walks in the context of long-range sequencing data, such as linked reads arising from the barcoding and sequencing of short reads arising from high molecular weight (10-500 Kbp) DNA fragments via the 10X Chromium platform, can be leveraged to resolve long distance phases and loose ends. These phased reconstructions of can be used to infer the consequences of complex structural variations and understand the evolution of complex rearranged loci across multiple tumor samples from individual patients.*

***Research: Highly prevalent noncoding somatic indel hotspots target lineage-defining genes in human cancer***

*Whole genome sequencing analysis of lung adenocarcinomas revealed indel hotspots in surfactant protein genes (SFTPB1, SFTPB, and SFTPC). Extrapolation to other solid cancers demonstrated highly recurrent and tumor-type-specific indel hotspots targeting the noncoding regions of highly expressed genes defining certain secretory cellular lineages: albumin (ALB) in liver carcinoma, gastric lipase (LIPF) in stomach carcinoma, and thyroglobulin (TG) in thyroid carcinoma. The sequence contexts of indels targeting lineage-defining genes were significantly enriched in the AATAATD DNA motif and specific chromatin contexts, including H3K27ac and H3K36me3. Our findings illuminate a prevalent and hitherto unrecognized mutational process linking cellular lineage and cancer. The findings may also represent a novel noncoding driver phenomenon that is under selection in multiple cancer types.*

---

**Bud Mishra** (Courant Institute of Mathematical Sciences and Tandon School of Engineering, NYU, New York, USA)

***Tutorial: "Learning Cancer Clocks"***

***Research: "Genomics Technologies: Try Again! Fail Again!! Fail Better!!!"***

*As the genomic analysis of tumors and their usage in causality/progression analysis have gained momentum, we are frustrated by the problem of not being able to create data that reflect tumor's heterogeneity (single cell/single molecule), genomic instability (resulting in aneuploidy, structural changes such as fusion), mobility (circulating tumor cells and cell free DNA), epigenetic, etc. We will describe a solution based on a new approach: SubOptical Mapping, which is ultra-long, fast (low latency/high throughput), accurate and potentially "highly disruptive." The talk will focus on various questions related to data analysis and their applications: Image Analysis, Algorithms for Variant Detection, Haplotype Phasing, Whole Genome Off-target Activities (e.g., with CRISPR assays), Transcriptomics (isoforms, small copy number), Sequence Assembly and Clinical Genomics. Additional topics may be discussed upon request.*

---

**Eugenio Montini** (San Raffaele Telethon Institute for Gene Therapy, HSR-TIGET, Milan, Italy)

***Tutorial: "Vector marking as tool to study cancer clonality and anticancer drug resistance"***

***Topic: "Lentiviral vector mediated insertional mutagenesis for the discovery of novel culprits of anticancer drug resistance"***

---

**Quaid Morris** (Banting and Best Department of Medical Research Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Canada)

***Tutorial: "A user's guide to subclonal reconstruction"***

***Research topic: "Reconstructing the evolutionary history of cancers using DNA-seq data"***

---

**Ben Raphael** (Princeton University, Princeton, USA)

***Tutorial: Identification of Cancer Genes, Pathways, and Networks***

*A key challenge in interpretation of cancer genome sequencing data is to distinguish driver mutations responsible for cancer development from random passenger mutations. We will survey a number of computational approaches to predict driver mutations. These include: (1) methods to identify recurrently mutated genes across cohorts of cancer genomes/exomes; (2) approaches that examine clustering of mutations in pathways and interaction networks; (3) algorithms that consider correlations between mutations including mutual exclusivity and co-occurrence. We will demonstrate how these approaches have been used in a number of large-scale cancer sequencing projects.*

***Research: Algorithms for Inferring Evolution and Migration of Tumors***

*Cancer is an evolutionary process driven by somatic mutations that accumulate in a population of cells that form a primary tumor. In later stages of cancer progression, cells migrate from a primary tumor and seed metastases at distant anatomical sites. I will describe algorithms to reconstruct this evolutionary process from DNA sequencing data of tumors. These algorithms address challenges that distinguish the tumor phylogeny problem from classical phylogenetic tree reconstruction, including challenges due to mixed samples and complex migration patterns.*

---

**Andrea Sottoriva** (The Institute of Cancer Research, London, UK)

***Research: "Quantifying clonal selection in human cancer using next-generation sequencing data"***

*Recent studies have identified prevalent subclonal architectures within many cancer types. However, the temporal evolutionary dynamics that produce these subclonal architectures remain largely unknown. Although we can detect subclones in cancer, measuring what these subclones are doing remains challenging. We developed a computational model of tumour evolution based on branching processes that allows measuring the dynamics of clonal selection using high throughput sequencing data from bulk cancer cell populations. Application of our method to high-depth sequencing data from gastric and lung cancers revealed that detectable subclones consistently emerged early during tumour growth and had considerably large fitness advantages (>20% growth advantage). Our quantitative framework provides new insight into the evolutionary history of cancers by facilitating the measurement of fundamental evolutionary parameters in individual patients.*

***Tutorial: "Evolutionary modelling and Bayesian inference on cancer genomic data"***

*Genomics has revolutionized cancer research and has led to the generation of astonishing amounts of data. Such large-scale multidimensional datasets contain an overwhelming amount of information that is often hard to make sense of. In particular, there is the need to integrate genomic information into a solid mechanistic framework based on cancer evolution that allows understanding what has happened in individual patients over time (the big hidden variable in cancer). In evolutionary biology, mathematical modeling and in particular population genetics have helped making sense of genetic data for decades, however such wealth of literature is often underexploited in cancer. In this tutorial we present the use of established statistical inference frameworks, in particular Approximate Bayesian Computation, that allow integrating complex mathematical and computational models of tumour evolution with commonly available genomic data, offering unprecedented insight into the dynamics of cancer evolution.*

---

**Simon Tavaré** (Cancer Research UK, Cambridge, UK)

***Tutorial: "An introduction to Approximate Bayesian Computation"***

***Research: "How to study tumours in 3.5D"***

# Cancer Development And Complexity 2017

*Lake Como School of Advanced Studies, May 23 - 26, 2017*



## Program

The schedule may change at the last minute. Please refer to the Keynote speakers and Lectures page for the title and the abstracts of each tutorial and lecture.

| CDAC 2017 | Tue May 23 | Wed May 24 | Thu May 25 | Fri May 26 | |
|---|---|---|---|---|---|
| 09:00 | | Bud Mishra - Tutorial | Simon Tavaré - Lecture | Philippe Gorry - Lecture | 09:00 |
| 10:00 | | Charles Cantor - Lecture | Marcin Imielinsky - Lecture | Ben Raphael - Lecture | 10:00 |
| 11:00 | | Coffee break | Coffee break | Coffee break | 11:00 |
| 11:30 | | Simon Tavarè - Tutorial | Eugenio Montini - Lecture | Andrea Sottoriva - Lecture | 11:30 |
| 12:30 | Registration | Andrea Sottoriva - Tutorial | Quaid Morris - Lecture | Bud Mishra - Lecture | 12:30 |
| 13:30 | Lunch break | Lunch break | Lunch break | Lunch break | 13:30 |
| 14:30 | Welcome by Bud Mishra and Marco Antoniotti | Marcin Imielinsky - Tutorial | Participants presentations | Round table | 14:30 |
| 15:30 | Ben Raphael - Tutorial | Participants presentations | Participants presentations | | 15:30 |
| 16:30 | Coffee break | Coffee break | Coffee break | | 16:30 |
| 16:45 | Quaid Morris - Tutorial | Eugenio Montini - Tutorial | Philippe Gorry - Tutorial | | 16:45 |
| 17:45 | | | | | 17:45 |
| | | | Social Dinner | | 20.00 |

# Cancer Development And Complexity 2017

*Lake Como School of Advanced Studies, May 23 - 26, 2017*



## Registration

Registration fees include participation to the school, social events, coffee breaks and school material.

- € 350,00 early registration, until April 15th, 2017.
- € 320,00 early registration, until April 15th, 2017, for participants affiliated with the four Universities supporting the Lake Como Schools: Milano Statale, Pavia, Milano Bicocca and Insubria.
- € 400,00 full registration, after April 15th.

All fees include VAT.

TO REGISTER, PLEASE FILL IN THE FORM BELOW:

Registrations are closed.

# Cancer Development And Complexity 2017

*Lake Como School of Advanced Studies, May 23 - 26, 2017*



## Accommodation

The first registrants may ask to be hosted directly at the Guesthouse of Villa Del Grumello (in a shared accommodation) for an additional cost of 33.00 EUR V.A.T. included, per night up to capacity (10 places).

The accommodation can be paid on site at the check in – accommodation is available from May 22rd (arrival day) to May 26th (departure date), 2017.

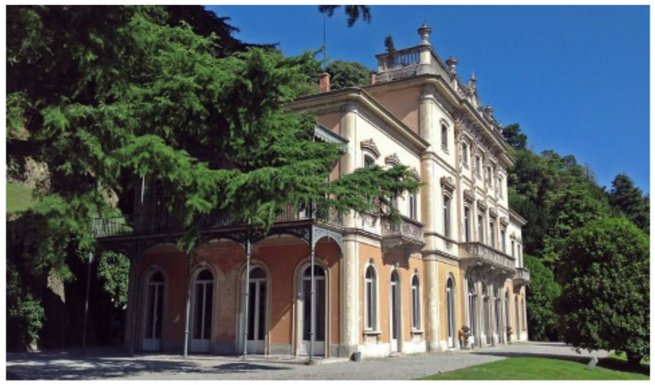
For any help or additional information:
please contact the Organizing Secretariat email: nadia.tansini@fondazionealessandrovolta.it

# Cancer Development And Complexity 2017

*Lake Como School of Advanced Studies, May 23 - 26, 2017*

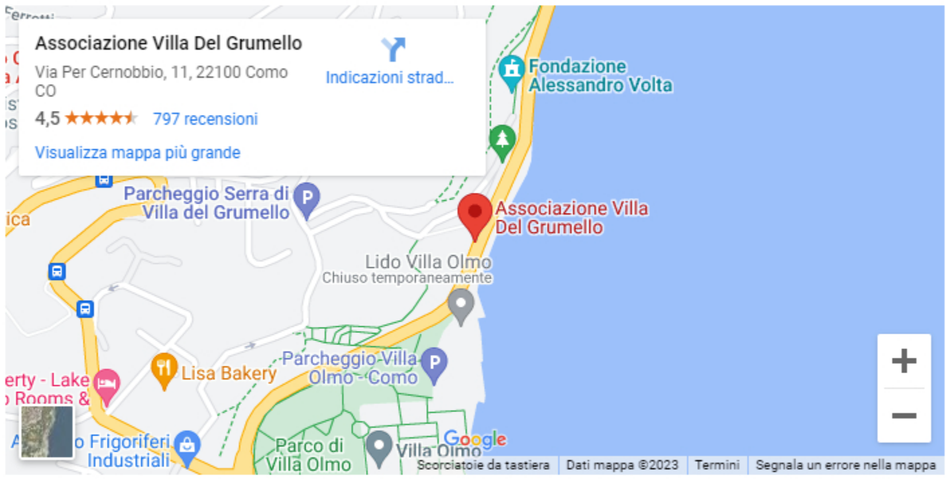


## Venue



The school will be held in the facilities offered by the Fondazione Alessandro Volta, Como (Italy), and specifically at Villa del Grumello, a nice villa close to Como: the center of the town can be reached in a ten minutes walk along the border of the lake.

HOW TO GET THERE: http://lakecomoschool.org/contact/travel-info/




View larger map