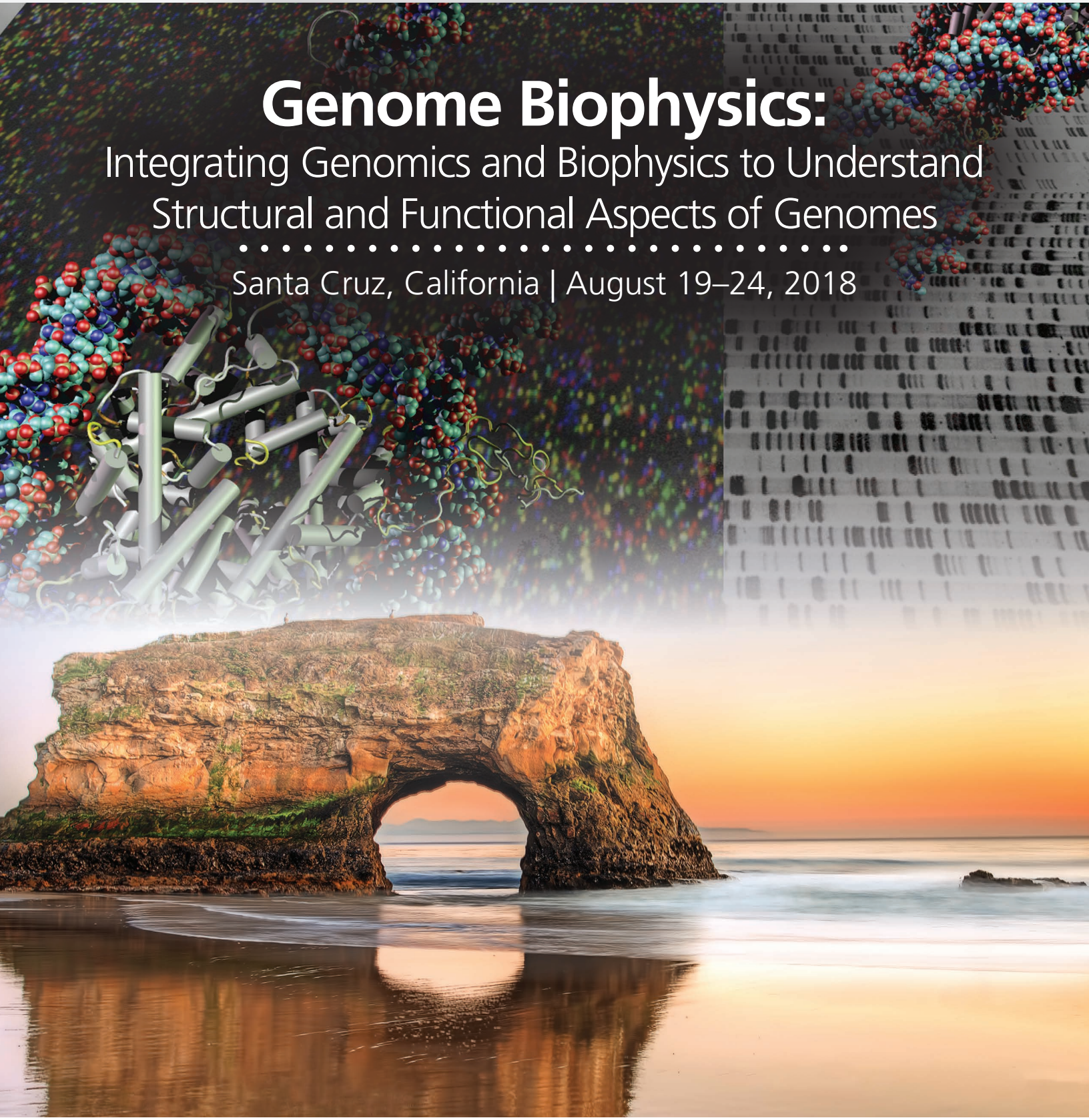
Genome Biophysics:

Integrating Genomics and Biophysics to Understand Structural and Functional Aspects of Genomes

Santa Cruz, California | August 19–24, 2018



BURROUGHS WELLCOME FUND

illumina®

Oxford NANOPORE Technologies™

Biophysical Society

Organizing Committee

Sarah Harris, University of Leeds

Stephen Leven, University of Texas at Dallas

Julia Salzman, Stanford University

Massa Shoura, Stanford University

Sponsorship Provided In Part By:

August 2018

Dear Colleagues,

We would like to welcome you to the Biophysical Society Thematic Meeting on ***Genome Biophysics: Integrating Genomics and Biophysics to Understand Structural and Functional Aspects of Genomes***. We have assembled an exciting program, which aims to bring together scientists across disciplines to explore the long-overdue application of biophysical methods in genomics, emphasizing structural and functional aspects of genome and transcriptome dynamics. The program will cover a wide range of topics including: extremophile genomes, highly compact genomes, circular and micro RNAs, DNA viruses and viroids, just to name a few.

The program features 26 invited speakers, 10 short talks selected from contributed posters, and 22 contributed posters. Over 60 participants from around the world will be in attendance to share and discuss their ideas. We hope that the meeting will not only provide a venue for exchanging recent exciting progress, but also promote fruitful discussions and foster future collaborations.

Situated on the northwestern edge of scenic Santa Cruz, Chaminade Resort & Spa offers breathtaking views of the Monterey Bay and Santa Cruz Mountains amid the resort's historic Mission style design. The property was originally opened as the Chaminade Boys High School by the Society of Mary in 1930. The high school was closed in 1940, and the property continued to be used as an educational facility until 1979, when it was purchased by the Hildreths, Taylors, and Swansons. In 1985, the Chaminade Resort & Spa opened its doors, and has been a relaxing retreat for over 30 years.

Thank you all for joining this meeting, and we look forward to enjoying this event with you!

Sincerely,

Sarah Harris, Stephen Levene, Julia Salzman, Massa Shoura
The Organizing Committee

Table of Contents

GENERAL INFORMATION

Registration/Information Location and Hours

Registration will be located outside the Santa Cruz Room on the second floor. Registration hours are as follows:

Sunday, August 19	15:00 – 18:00
Monday, August 20	8:30 – 17:00
Tuesday, August 21	8:30 – 17:00
Wednesday, August 22	8:30 – 17:00
Thursday, August 23	8:30 – 17:00
Friday, August 24	8:30 – 12:00

Instructions for Presentations

(1) Presentation Facilities:

A data projector will be available in the Santa Cruz. The data projector is connectable to VGA and HDMI hookups. Speakers are required to bring their own laptops and adaptors. Speakers are advised to preview their final presentations before the start of each session. It is recommended to have a backup of the presentation on a USB drive, in case of any unforeseen circumstances.

(2) Poster Sessions:

1) The poster session will be held in the New Brighton Room.

2) Posters will be mounted on available wall space. ***Without exception, all posters must be vertically oriented with the dimensions of 3' wide by 4'6" high (91.5 cm x 138 cm) to allow adequate space between posters.*** Poster boards will follow the same numbering scheme as listed in the E-book.

3) Poster boards require pushpins or thumbtacks for mounting. Authors are expected to bring their own mounting materials.

4) There will be formal poster presentations on Wednesday. Odd-numbered posters will be displayed from 16:20 – 17:05, and even-numbered posters will be displayed from 17:05 – 18:00.

5) During the assigned poster presentation sessions, presenters are requested to remain in front of their poster boards to meet with attendees.

6) All presenters must remove their poster by 18:00 on the day of their scheduled presentation. Posters left uncollected at the end of the evening will be disposed.

Meals and Coffee Breaks

There will be a one-hour Welcome Reception on Sunday, August 19 from 6:00 – 7:00 PM. This reception will be held in the Courtyard Terrace.

Coffee breaks will be served in the Seacliff Lounge & Terrace. Beverages and snacks will be available daily from 8:30 AM – 4:30 PM.

Meals are included starting with dinner on the day of arrival, August 19, through lunch the day of departure, August 24. Breakfast, lunch, and dinner will be served in the Sunset Restaurant, which is located on the first floor.

Internet

Wi-Fi will be available throughout all areas of Chaminade Resort & Spa. Access is complimentary and no password is required.

Smoking

Please be advised that smoking is not permitted inside the Chaminade Resort & Spa or the meeting facilities. Smoking is permitted in designated outside areas.

Name Badges

Name badges are required to enter all scientific sessions, poster sessions, and social functions. Please wear your badge throughout the conference.

Contact Information

If you have any further requirements during the meeting, please contact the meeting staff at the registration desk from August 19-24 during registration hours.

In case of emergency, you may contact the following:

Dorothy Chaconas, BPS Director of Meetings
dchaconas@biophysics.org

Front Desk, Chaminade Resort & Spa
1-831-475-5600

Genome Biophysics: Integrating Genomics and Biophysics to Understand Structural and Functional Aspects of Genomes

Santa Cruz, California
August 19-24, 2018

Sunday, August 19, 2018

15:00 – 18:00	**Registration/Information**	**Santa Cruz Room**
18:00 – 19:00	**Welcome Reception**	**Courtyard Terrace**
19:00 – 20:00	**Dinner**	**Sunset Restaurant**

Monday, August 20, 2018

8:30 – 17:00	**Registration/Information**	**Santa Cruz Room**
Session I	**Junk or Not Junk: Structure and Sequence of Coding and Noncoding DNA** Sarah Harris, University of Leeds, United Kingdom, Chair	
9:00 – 9:45	Sergei Mirkin, Tufts University, USA ***RNA-DNA Hybrids Promote the Expansion of Friedreich's Ataxia (GAA)n Repeats via Break-induced Replication***	
9:45 – 10:30	David Levens, NIH, USA ***The Regulatory Roles of DNA Topology and Conformation in Mammalian Gene Expression***	
10:30 – 10:50	**Coffee Break**	**Seacliff Lounge & Terrace**
10:50 – 11:10	Anton Goloborodko, MIT, USA* ***A Pathway for Mitotic Chromosome Formation***	
11:10 – 11:55	Charles Dorman, Trinity College, Ireland ***Bacterial Decision-making Operating Through Tuneable Binary Genetic Switches***	
12:00 – 14:00	**Lunch**	**Sunset Restaurant**
Session II	**Biophysical Approaches to Understanding Chromatin Structure** Stephen Levene, University of Texas at Dallas, USA, Chair	
14:00 – 14:45	Wilma Olson, Rutgers University, USA ***Contributions of DNA Sequence in 3D Genomic Architectures***	
14:45 – 15:30	Tamar Schlick, New York University, USA ***Modeling Gene Elements at Nucleosome Resolution***	
15:30 – 15:50	**Coffee Break**	**Seacliff Lounge & Terrace**

15:50 – 16:35	Andrzej Stasiak, University of Lausanne, Switzerland ***Transcription-induced Supercoiling and TADs Formation***	
16:35 – 17:20	Karsten Rippe, Heidelberg University, Germany ***Establishing Chromatin Subcompartments That Are Both Stable and Plastic***	
17:20 – 18:00	**Nature Hike**	**Chaminade Red & Blue Trails**
18:00 – 20:00	**Dinner**	**Sunset Restaurant**

Tuesday, August 21, 2018

8:30 – 17:00	**Registration/Information**	**Santa Cruz Room**
Session III	**Exploring the Physical Genome I** Wilma Olson, Rutgers University, USA, Chair	
9:00 – 9:45	Javier Arsuaga, University of California, Davis, USA ***Biophysical Models of DNA Organization inside Viral Capsids***	
9:45 – 10:30	Shinichi Morishita, University of Tokyo, Japan ***Understanding Tandem Repeats and Methylation with Long-read Sequencing***	
10:30 – 10:50	**Coffee Break**	**Seacliff Lounge & Terrace**
10:50 – 11:10	Thomas Bishop, Louisiana Tech University, USA* ***G-Dash: A Genomics Dashboard that Unites Physics and Informatics Studies of Chromatin***	
11:10 – 11:55	Alexandra Zidovska, New York University, USA ***The "Self-stirred" Genome: Bulk and Surface Dynamics of the Chromatin Globule***	
12:00 – 14:00	**Lunch**	**Sunset Restaurant**
Session IV	**Exploring the Physical Genome II** Stephen Levene, University of Texas at Dallas, USA, Chair	
14:00 – 14:45	Xaiver Darzacq, University of California, Berkeley ***Nuclear Organization and Transcription Regulation Mechanisms Studied by Live Cell Imaging***	
14:45 – 15:30	Martin Depken, Delft University of Technology, The Netherlands ***Bottom-up Physical Modelling for CRISPR//Cas Target Prediction***	
15:30 – 15:50	**Coffee Break**	**Seacliff Lounge & Terrace**
15:50 – 16:35	Laura Landweber, Columbia University, USA ***RNA-programmed Genome Rearrangement in the Ciliate Oxytricha***	
16:35 – 16:55	Katerina Kraft, Stanford University, USA* ***Genomic Rearrangement Induced Gene Activation by Architectural Stripes***	

17:00 – 18:00	David Schwartz, University of Wisconsin, USA ***From Big DNA Molecules to Big Data*** **Keynote Speaker**	
18:00 – 20:00	**Dinner**	**Sunset Restaurant**

Wednesday, August 22, 2018

8:30 – 17:00	**Registration/Information**	**Santa Cruz Room**
Session V	**Fundamental Limits of Sequencing Accuracy, Sensitivity, and Uniqueness: What Do, and What Don't, We Know?** Marc Salit, NIST, USA, Chair	
9:00 – 9:45	Marc Salit, NIST, USA ***Metrology of Genome-scale Measurements: Standards and Systematics to Get Comparability and Confidence***	
9:45 – 10:05	Miten Jain, University of California, Santa Cruz, USA* ***A Reference Human Transcriptome Based on Native RNA Sequencing***	
10:05 – 10:25	Idan Gabdank, Stanford University, USA* ***Portable and Reproducible Computational Analyses***	
10:25 – 10:50	**Coffee Break**	**Seacliff Lounge & Terrace**
10:50 – 11:10	Stephen Lincoln, Invitae, USA* ***Complex Genetic Variants: Implications for Clinical Sequencing Methods and Validation Approaches***	
11:10 – 12:00	**Formal Discussions**	
12:00 – 14:00	**Lunch**	**Sunset Restaurant**
Session VI	**Single-Cell Genomics and Single-Molecule Sequencing** Tim J. Stevens, MRC Laboratory of Molecular Biology, United Kingdom, Chair	
14:00 – 14:45	Tim J. Stevens, MRC Laboratory of Molecular Biology, United Kingdom ***Capturing the 3D Folds of Whole Mammalian Genomes in Single Cells***	
14:45 – 15:30	Bo Wang, Stanford University, USA ***Self-assembling Manifolds in Single-Cell RNA Sequencing Data***	
15:30 – 16:15	Cristian Micheletti, SISSA Trieste, Italy ***Nanopore Translocation of Knotted DNA***	
16:15 – 18:00	**Poster Session**	**New Brighton**
18:00 – 20:00	**Dinner**	**Sunset Restaurant**
20:00	**Pool Night**	**Resort Pool**

Thursday, August 23, 2018

8:30 – 17:00	**Registration/Information**	**Santa Cruz Room**
Session VI I	**The RNA World: Structure, Dynamics, and Interaction** Andrew Fire, Stanford University, USA, Chair	
9:00 – 9:45	Andrew Fire, Stanford University, USA ***Long-term RNA-based Transmission of Biological States***	
9:45 – 10:30	Alan Lambowitz, University of Texas at Austin, USA ***Thermostable Group II Intron Reverse Transcriptases (TGIRTs) and Their Use in RNA-seq***	
10:30 – 10:50	**Coffee Break**	**Seacliff Lounge & Terrace**
10:50 – 11:10	Priscilla L. Boon, National University of Singapore* **How Dengue Capsid Protein Assists in Organizing Dengue Virus Genomic RNA**	
11:10 – 11:55	Sarah Woodson, Johns Hopkins University, USA ***Sequential Folding of RNA***	
12:00 – 14:00	**Lunch**	**Sunset Restaurant**
Session VIII	**Genomics of Gene Regulation** Massa Shoura, Stanford University, USA, Chair	
14:00 – 14:45	Nadav Ahituv, University of California, San Francisco, USA ***Functional Characterization of Gene Regulatory Elements***	
14:45 – 15:30	Polly Fordyce, Stanford University, USA ***Quantitative Mapping of Transcription Factor Binding Energy Landscapes***	
15:30 – 15:50	**Coffee Break**	**Seacliff Lounge & Terrace**
15:50 – 16:35	Zeba Wunderlich, University of California, Irvine, USA ***Noise in the Shadows***	
16:35 – 16:55	Ariel Afek, Duke University, USA* ***Mismatched Base-pairs Locally Distort DNA Structure and Can Induce Increased DNA-binding by Transcription Factor Proteins***	
16:55 – 17:15	Alexander Wood, Newcastle University, United Kingdom* ***What Gene Expression Noise Tells about the Spatiotemporal Organization of Gene Regulatory Networks***	
17:15 – 18:00	**Illumina Workshop**	
18:00 – 20:00	**Dinner**	**Sunset Restaurant**

Friday, August 24, 2018

8:30 – 12:00	**Registration/Information**	**Santa Cruz Room**
Session IX	**Extreme Genomes** Massa Shoura, Stanford University, USA, Chair	
9:00 – 9:45	Ami Bhatt, Stanford University, USA ***Culture-free Microbial Genome Assembly and Tracking in Hospitalized Patients***	
9:45 – 10:05	Jason DeRouchey, University of Kentucky, USA* ***DNA in Tight Spaces: Linking Structure, Stability, and Protection in Sperm Chromatin***	
10:05 – 10:50	Joanna Kelley, Washington State University, USA ***Eukaryotic Genome Evolution in Extreme Environments***	
10:50 – 11:20	**Closing Remarks** Co-Organizers: Sarah Harris, Stephen Levene, Massa Shoura	
12:00 – 14:00	**Lunch**	**Sunset Restaurant**

** Contributed talks selected from among submitted abstracts*

SPEAKER ABSTRACTS

RNA-DNA Hybrids Promote the Expansion of Friedreich's Ataxia (GAA)n Repeats via Break-induced Replication

Sergei Mirkin
Tufts University, USA

No Abstract

The Regulatory Roles of DNA Topology and Conformation in Mammalian Gene Expression

David L Levens[1]
[1]National Cancer Institute, Laboratory of Pathology, CCR, Bethesda, Maryland, United States

As a undimensional matrix, DNA encodes RNA and the cis-elements that bind gene regulatory proteins to direct transcription and replication. Via these bound factors, DNA sequence also instructs the 3-d folding of the genome. Beyond this static coding by DNA sequence, the double helix undergoes dynamic changes in structure and topology in response to applied mechanical forces. The torque that untwists DNA enabling the bases to serve as a template during transcription and replication, is dynamically propagated through the DNA fiber. Eventually these dynamic supercoils must be accommodated by stress absorbing alternative conformations of DNA and chromatin, dissipated by transmission to remote regions or off DNA ends (telomeres or breaks), or removed by the action of topoisomerases. Melting of DNA at susceptible sequences licenses the formation of non-B DNA structures and absorbs torsional stress. These alternative structures in turn, can modify gene activity by binding structure and/or sequence-selective single-stranded DNA binding proteins, such as the Far Upstream Element, that interact with the transcription machinery, or by controlling chromatin structure via positioned nucleosomes or modifying the elastic moduli of chromatin. Thus, torsional stress is not merely a by-product of genetic processes, but has the capacity via mechanical feedback to regulate those same processes. In turn the transcription machinery and transcription factors directly modify topoisomerase activity to tune torsional stress that on the one hand may impedes or even arrests transcriptional elongation, but that also may encourage DNA melting. For example, both RNA polymerase and the general amplifier of transcription MYC, also stimulate topoisomerase 1 to diminish the dynamic supercoils that otherwise would oppose transcription elongation.

A Pathway for Mitotic Chromosome Formation

Johan H. Gibcus[1], Kumiko Samejima[2], **Anton Goloborodko**[3], Itaru Samejima[2], Natalia Naumova[1], Johannes Nuebler[3], Masato T. Kanemaki[4], Linfeng Xie[5], James R. Paulson[5], William C. Earnshaw[2], Leonid A. Mirny[3], Job Dekker[1,6].
[1]University of Massachusetts Medical School, Worcester, MA, USA, [2]University of Edinburgh, Edinburgh, United Kingdom, [3]Massachusetts Institute of Technology, Cambridge, MA, USA, [4]National Institute of Genetics, Shizuoka, Japan, [5]University of Wisconsin-Oshkosh, Oshkosh, WI, USA, [6]Howard Hughes Medical Institute, Chevy Chase, MD, USA.
During mitosis, cells compact their chromosomes into dense rod-shaped structures to ensure their reliable transmission to daughter cells. Our work explores how cells achieve this compaction. We integrate genetic, genomic, and computational approaches to characterize the key steps in mitotic chromosome formation from the G2 nucleus to metaphase, and we identify roles of specific molecular machines, condensin I and II, in these major conformational transitions.

In this study, we perform time-resolved analyses of mitotic chromosome structure in engineered chicken DT-40 cells which express an analog-sensitive CDK1 and thus enable their synchronous release into mitosis. We probe the chromosome organization by microscopy and Hi-C. We elucidate the role of condensin I and II complexes in chromosome organization using engineered cell lines that enable a rapid depletion of the subunits of these complexes prior to mitotic entry. Finally, we use obtained data to develop polymer models of chromosomes that we examine analytically and by computer simulations.

As a result, we delineate a detailed pathway of mitotic chromosome folding that unifies many previous observations. In prophase, condensins mediate the loss of interphase organization and the formation of arrays of consecutive loops. In prometaphase, chromosomes adopt a spiral staircase–like structure with a helically arranged axial scaffold of condensin II at the bases of chromatin loops. The condensin II loops of ~400kb are further compacted by condensin I into clusters of smaller nested loops, ~80kb each, that are additionally collapsed by chromatin-to-chromatin attraction. The combination of nested loops distributed around a helically twisted axis plus dense chromatin packing achieves the 10,000-fold compaction of chromatin into linearly organized dense mitotic chromosomes.

Bacterial Decision-making Operating Through Tunable Binary Genetic Switches

Charles J Dorman[1]
[1]Trinity College Dublin, Microbiology, Dublin, Dublin, Ireland

Populations of genetically identical bacterial cells manage to generate cell-to-cell physiological diversity through stochastic processes operating at the level of gene expression. This seems to be an important strategy when confronted with an unpredictable and potentially hostile external environment: if things change suddenly, at least a few members of the population may be prepared and will survive the new challenge to transmit their genes vertically to future generations. We have studied a genetic switch in the model bacterium *Escherichia coli* that seems to operate in a random way, switching on or off a set of genes that encode a surface structure that attaches *E. coli* to solid surfaces. Swimming (planktonic) *E. coli* can switch to surface attachment if they no longer have sufficient energy to keep moving. The attached lifestyle also lends itself to physical and chemical stress survival. On closer inspection, we have found that the genetic switch can become biased toward its 'on' state in response to a deteriorating environment, overriding random switching. This biasing involves adjustments to the topology of the DNA in the bacterium and contributions by DNA structuring proteins that drive more and more of the switches across the population into the 'on' state. In addition to describing the molecular mechanisms at work in this specific switch, evidence will be presented that variable DNA topology is used quite generally to bias switching outcomes in bacteria, including pathogenic bacteria.

Contributions of DNA Sequence in 3D Genomic Architectures

Wilma K. Olson
[1]Rutgers, the State University of New Jersey, Chemistry and Chemical Biology, Piscataway, New Jersey, United States

One of the critical unanswered questions in genome biophysics is how the primary sequence of DNA bases influences the global properties of very long chain molecules. The local sequence-dependent features of DNA found in high-resolution structures introduce irregularities in the disposition of adjacent residues that facilitate the specific binding of proteins and modulate the global folding and interactions of the double helix. The contributions of DNA sequence reveal themselves in reduced molecular representations whereby the orientation and displacement of paired bases or successive base pairs are described in terms of a set of rigid-body parameters. The importance of this treatment lies in its utility in linking atomic observables with macromolecular properties, and in bridging the gap between the size of systems that can be studied at full atomic detail with events, such as protein-mediated looping, that take place at the mesoscale level. The local sequence also contributes to the positions of nucleosomes on DNA and the properties of the interspersed DNA linkers. Like the patterns of base-pair association within DNA, the arrangements of nucleosomes in chromatin modulate the properties of longer polymers. The spatial arrangements of interacting nucleosomes along short, well-defined arrays provide a basis for linking the mesoscale features of chromatin to higher-order structures. A nucleosome-level depiction of chromatin reduces the complexity of the system along the same lines as a base-pair level depiction of DNA and makes it possible to bridge the gap between chromatin â€˜secondary structuresâ€™ and even longer polymers.

Modeling Gene Elements at Nucleosome Resolution

Tamar Schlick
New York University, USA

No Abstract

Transcription-induced Supercoiling and TADs Formation

Andrzej Stasiak
University of Lausanne, Switzerland

No Abstract

Establishing Chromatin Subcompartments That Are Both Stable and Plastic

Karsten Rippe
Heidelberg University, Germany

No Abstract

Biophysical Models of DNA Organization Inside Viral Capsids

Javier Arsuaga[1]
[1]University of California, Davis, California, USA

The three dimensional organization of genomes is a key player in multiple biological processes including the genome packaging and release in viruses. The genome of some viruses, such as bacteriophages or human herpes, is a double stranded DNA (dsDNA) molecule that is stored inside a viral protein capsid at a concentration of 200 mg/ml-800mg/ml and an osmotic pressure of 70 atmospheres. The organization of the viral genome under these extreme physical conditions is believed to be liquid crystalline but remains to be properly understood. A general picture of this organization has been recently given by cryoelectron microscopy (cryoEM) studies that show a series of concentric layers near the surface of the viral capsid followed by a disordered arrangement of DNA fibers near the center of the capsid.

In this talk I will present results from three different approaches to study the problem of dsDNA packing in bacteriophages. The first approach complements the cryoEM observations and uses the formation of knots inside viral capsids as a probe for DNA packing. These results suggest that DNA knots are highly likely upon confinement and that the DNA molecule is chirally organized inside the viral capsid. The second approach aims at identifying the possible sources of the chiral organization of the genome and employs methods from random knotting and brownian dynamics and suggest that the DNA packing motor can account for the suggested chirality of the genome. The third approach uses continuum mechanics models to rigorously describe cryoEM observations as the minima of a liquid crystalline phase. The emergent picture of these approaches suggest that DNA is in a chirally organized liquid crystalline phase in which knots may be the product of liquid crystal defects.

Understanding Tandem Repeats and Methylation with Long-read Sequencing

Shinichi Morishita
[1]University of Tokyo, Kashiwanoha 5-1-5, Kashiwa, Chiba, Japan

We address the problem of understanding previously uncharacterized genomic regions such as centromeres, long gaps, short tandem repeat expansions, diploid methylomes/transcriptomes, and plasmids/phages in metagenomes. With our collaborators, integrating the merits of long read sequencing technologies (PacBio Sequel, ONT nanopore, 10X Chromium, and Hi-C), we found:

- We sequenced VC2010, a single uniform strain and a non-mutagenized clonal derivative of N2 populations. We used four state-of-the-art genome assemblers to generate independent assemblies. We found that these assemblers were complementary each other to fill gaps in others, producing a nearly complete genome with two gaps. Most of filled gaps were tandem repeat expansions of length 10K-100Kb that could be closed by nanopore reads.
- We compared genome assemblies (~800Mb in size) of three medaka (Japanese killifish) inbred strains that diverged ~18 million years ago. In medaka, centromeric monomers in non-acrocentric chromosomes evolved significantly faster than those in acrocentric chromosomes.
- Abnormal expansions of TTTCA and TTTTA repeats in intron 4 of SAMD12 were associated with benign adult familial myoclonic epilepsy (BAFME). Transcriptional abortion was observed at the repeat expansions.
- Gene expression is regulated by DNA methylation in two homologous chromosomes separately in personal diploid human genomes. Despite its importance, however, observing DNA methylation in individual homologous chromosomes independently (diploid methylomes) has been technically challenging because homologous chromosomes are extremely highly similar (identity of ~99.9%). We developed a statistical method for uncovering complex diploid methylomes by integrating PacBio and 10X methods.
- We processed fecal DNA samples from 12 individuals using PacBioâ€™s single-molecule real-time (SMRT) sequencing, and identified 71 plasmids and 11 phages including crAssphases, half of which were unknown but actually prevalent in several different countries. With SMRT sequencing, we also observed DNA methylation motifs shared between plasmids/phages and their hosts, allowing us to assign plasmids/phages to their hosts.

G-Dash: A Genomics Dashboard That Unites Physics and Informatics Studies of Chromatin

Thomas C. Bishop[1]; Zilong Li[1]; Ran Sun[1];
[1]Louisiana Tech University, Physics, Ruston, Louisiana, United States

Historically, bioinformatics and computational biology are recognized as distinct endeavors. The underlying theories, experiments, software and computing resources differ significantly. We demonstrate that these differences can be overcome by exploiting existing data standards, algorithms, and web based tools. We define a genomics dashboard as a console that can track, analyze or display genomics information and that provides controllers or other means for manipulating chromatin structure. We present G-Dash as a prototype of the genomics dashboard concept. G-Dash unites our Interactive Chromatin Modeling(ICM) webserver's capabilities with the Dalliance Genome Browser(DGB), utilizes a simple coarse grained model of chromatin implemented in LAMMPS for structure relaxation and displays 3D models with JSmol. All atom models of individual nucleosomes can also be generated. Thus, data obtained from public databases through the genome browser (e.g. experimentally or theoretically determined nucleosome positions) can be used to manipulate nucleosomes (add, delete and move), assign unique conformational states (e.g. tetrasome, octasome, chromatosome), display 3D structural data as tracks in the genome browser (e.g. Roll, Slide or Twist) or map informatics data onto a 3D structure (e.g. nucleosome positions, DNA damage sites, functional annotations). We demonstrate how experimentally determined maps of nucleosome positions for *Saccharomyces cerevisiae* can be used to assemble a computational karyotype. Models of the MMTV, CHA1, HIS3 and PHO5 promoters highlight important observations: experimentally determined nucleosome positions are insufficient to achieve tight packing of chromatin and sequence specific material properties of DNA (conformation and flexibility) can affect chromatin bending and looping. As a tool G-Dash supports cross-validation of physical modeling and informatics approaches and provides a means of investigating structure-function relationships for a genome. See the "Genome Dashboard" tab at http://dna.engr.latech.edu.

The "Self-Stirred" Genome: Bulk and Surface Dynamics of the Chromatin Globule

Alexandra Zidovska[1]
[1]New York University, Department of Physics, New York City, New York, United States

Chromatin structure and dynamics control all aspects of DNA biology yet are poorly understood. In interphase, time between two cell divisions, chromatin fills the cell nucleus in its minimally condensed polymeric state. Chromatin serves as substrate to a number of biological processes, e.g. gene expression and DNA replication, which require it to become locally restructured. These are energy-consuming processes giving rise to non-equilibrium dynamics. Chromatin dynamics has been traditionally studied by imaging of fluorescently labeled nuclear proteins and single DNA-sites, thus focusing only on a small number of tracer particles. Recently, we developed an approach, displacement correlation spectroscopy (DCS) based on time-resolved image correlation analysis, to map chromatin dynamics simultaneously across the whole nucleus in cultured human cells [1]. DCS revealed that chromatin movement was coherent across large regions (4-5μm) for several seconds. Regions of coherent motion extended beyond the boundaries of single-chromosome territories, suggesting elastic coupling of motion over length scales much larger than those of genes [1]. These large-scale, coupled motions were ATP-dependent and unidirectional for several seconds. Following these observations, we developed a hydrodynamic theory [2] as well as a microscopic model [3] of active chromatin dynamics. In this work we investigate the chromatin interactions with the nuclear envelope and compare the surface dynamics of the chromatin globule with its bulk dynamics [4].

Nuclear Organization and Transcription Regulation Mechanisms Studied by Live Cell Imaging

Xavier Darzacq[1]
[1]Division of Genetics, Genomics & Development, Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA

While the vast majority of the molecules involved in transcription regulation are known and the reaction can be reconstituted in vitro from purified components, the basic rules governing transcription regulation remain poorly understood. Transcription initiation is very inefficient in vitro, and our goal is to understand how cells exploit the spatial organization of the nucleus to increase the rate of this reaction. Studying the heterogeneity in nuclear distribution of RNA polymerase and other essential transcription factors we seek to understand how nuclear organization can act on transcription regulation. Local nucleoplasmic protein accumulations, which we call hubs, are enriched in various transcription factors by interactions mediated by intrinsically disordered protein domains such at the C-terminal domain of the RNA polymerase II catalytic subunit. I will discuss how local polymerase hubs are regulated and how they affect the dynamics and kinetics of their constituent proteins. Also, I will discuss our attempts to reconstitute artificial activation hubs and visualize their effect on single-molecule transcription in vitro.

Bottom-up Physical Modelling for CRISPR//Cas Target Prediction

Martin Depken
Delft University of Technology, The Netherlands

No Abstract

RNA-programmed Genome Rearrangement in the Ciliate Oxytricha

Laura Landweber
Columbia University, USA

No Abstract

Genomic Rearrangement Induced Gene Activation by Architectural Stripes

Katerina Kraft[1,2]; Andreas Magg[2]; Verena Heinrich[2]; Christina Riemenschneider[2]; Robert Schoepflin[2]; Stefan Mundlos[2];
[1]Stanford, , Stanford, California, United States
[2]MPIMG, , Berlin, Berlin, Germany

Precise spatiotemporal gene expression is essential for normal organismal development and homeostasis and requires regulation at many levels. At one level, cis-regulatory elements, enhancers, drive the cell-type- and time-specific expression of developmental genes. These long-range enhancers confer their activity on genes via chromatin looping and this 3d chromatin structure constrains enhancer activity to the target genes. In recent years many architectural units have been defined based on hic, a number of which regulate enhancer-promoter communication. This includes architectural stripes, which were characterized in a recent in vitro study. Here we induce serial genomic invertions in vivo in embryonic limb buds, and observe the formation of tissue specific architectural stripes by capture hic. We find the enhancer at the inversion point is able to communicate aberrantly with several promoters inside an architectural stripe leading to congenital limb malformation. Deletion of the stripe anchor point results in stripe collapse, leading to significant expression changes and rescue of the skeletal phenotype. This study sheds light on the mechanism by which structural variations inducing architectural stripes control several important developmental genes within the stripe. Moreover, it suggests a general mechanism explaining connection between chromatin architecture and gene expression. This work opens up the discussion of enhancer-promoter specificity, a new uncharacterized field.

From Big DNA Molecules to Big Data

David C Schwartz;
[1]University of Wisconsin-Madison, Chemistry; Genetics, Madison, Wisconsin, United States

Contemporary genome analysis uses single nucleic acid molecules, either directly measured, or measured after amplification, for gaining new biological insights into single cells and human populations in ways that are now functionalizing the non-genic portion of the human genome. The challenge of achieving comprehensive analysis that structurally elucidates the entire human genome requires addressing the massively pervasive repetitive elements of human genomes using single molecules analytes within integrated high-throughput systems. This challenge led to our development of Optical Mapping systems. As part of this vision, we have advanced a nascent biophysical approach by taking steps to migrate molecular discoveries into systems capable of grappling with the complex regions harbored by human and cancer genomes. Within this context, I will provide an introduction to human and cancer genomes and describe the history of Optical Mapping with emphasis given to detailing the many genomic challenges requiring synergistic development efforts across many fields to highlight an example where molecular discoveries were advanced via system design and, in turn, where systems were advanced by new molecular insights. Given this contextualized background, I will then describe research vignettes from our group and offer some thoughts about what the future may hold for new breakthroughs in genomics via biophysical approaches.

Metrology of Genome-Scale Measurements: Standards and Systematics to Get Comparability and Confidence

Marc Salit
NIST, USA

No Abstract

A Reference Human Transcriptome Based on Native RNA Sequencing

Miten Jain[1]; Hugh E Olsen[1]; Benedict Paten[1]; Angela Brooks[1]; Mark Akeson[1];
[1]UC Santa Cruz, BME, Santa Cruz, California, United States

The Nanopore RNA consortium is an international consortium of Oxford Nanopore MinION and GridION users. In 2017, the consortium generated a dataset consisting of 13 million native RNA and 24 million cDNA strand reads based on poly-A RNA isolated from the human reference cell line GM12878. This dataset is publicly available on GitHub: https://github.com/nanopore-wgs-consortium/NA12878/blob/master/RNA.md. The median read identity for RNA strand reads was around 86%, and we observed aligned read lengths of up to 22 kb (116 exons). We also observed a strong correlation (R=0.875) between native RNA and cDNA datasets, and that 73% of annotated human reference transcripts were captured by the native RNA data. We anticipate this dataset will serve as a resource to the community for native RNA sequencing. We will present updates from the consortium work on analysis of these data that will include characterization of poly-A tail lengths using nanopore ionic current dwell time, assessment of full-length transcripts, detection of novel isoforms, and detection of base modifications using signal-level analysis.

Portable and Reproducible Computational Analyses

Idan Gabdank[1]; Esther Chan[1]; Jason Hilton[1]; Weiwei Zhong[1]; Seth J Strattan[1]; Yunhai Luo[1]; Ulugbek K Baymuradov[1]; Timothy R Dreszer[1]; Otto A Jolanki[1]; Keenan Graham[1]; Kathrina C Onate[1]; Nicholas Luther[1]; Zachary A Myers[1]; Stuart R Miyasato[1]; Forrest Tanaka[1]; Philip Adenekan[1]; Karthik Kalyanaraman[1]; Benjamin C Hitz[1]; Michael J Cherry[1];
[1]Stanford, Genetics, Palo Alto, California, United States

The Encyclopedia of DNA Elements (ENCODE) project is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The main goal of the project is to identify all of the functional elements in the human genome. The ENCODE Data Coordination Center (DCC) collects, curates, and disseminates the results, methods, and raw data for a variety of complex assays and analyses that have been performed to identify and validate these elements.
In order to achieve transparent, reproducible, and comparable analysis results the ENCODE DCC has established a framework for the development and implementation of computational processing pipelines. Docker, for software containerization, and WDL, for workflow description, are used to develop modular and scalable pipelines that run identically on multiple compute platforms. Continuous integration methodologies are used to automate pipeline testing and deployment. Standardization of the computational methodologies for analysis and quality control leads to comparable results from different ENCODE collections - a prerequisite for successful integrative analyses.
ENCODE uniform processing pipelines are available or in development for the analysis of ChIP-seq, RNA-seq, DNase-seq, ATAC-seq, HiC, ChIA-PET, and WGBS assays. The pipelines are open-source and have unified documentation.
The ENCODE DCC codebase is at https://github.com/ENCODE-DCC

Complex Genetic Variants: Implications for Clinical Sequencing Methods and Validation Approaches

Stephen Lincoln[1]; Andrew Fellowes[2]; Shazia Mahamdallie[3]; Shimul Chowdhury[4]; Eric Klee[5]; Justin Zook[6]; Rebecca Truty[1]; Russell Garlick[7]; Marc Salit[8]; Nazneen Rahman[3]; Stephen Kingsmore[4]; Robert Nussbaum[1]; Matthew Ferber[5]; Brian Shirts[9]

[1]Invitae, San Francisco, California, United States; [2]Peter MacCallum Cancer Center, Melbourne, Victoria, Australia; [3]Institute of Cancer Research, London, United Kingdom; [4]Rady Children's Hospital, San Diego, California, United States; [5]Mayo Clinic, Rochester, Minnesota, United States; [6]NIST, Gaithersburg, Maryland, United States; [7]Seracare, Gaithersburg, Maryland, United States; [8]Stanford University, Palo Alto, California, United States; [9]University of Washington, Seattle, Washington, United States

Objective: We evaluated the clinical prevalence of complex genetic variants and developed novel resources to help improve methods for detecting such variants. Background: DNA sequencing methods are often validated or benchmarked for their ability to detect simple single nucleotide variants (SNVs), small insertions and deletions (indels) and large copy number changes. Such variants are prevalent in every person, but most are not medically important. It is presently difficult to evaluate the performance of sequencing methodologies for more complex variant types. Such data are critical to determine which approaches may be most appropriate for specific medical tests. Methods: We analyzed over 80,000 patients undergoing clinical genetic testing using high-sensitivity methods. Guided by these data, a pilot specimen containing a diverse set of 22 challenging variants was engineered, validated, and provided to collaborating laboratories who sequenced it using 10 different NGS based workflows. Results: In our patient cohort, between 9 and 19% of the medically important (i.e. pathogenic) variants were of types that are technically challenging. These variants included small copy number changes, structural variants, large or complex indels, and repeat associated variants. Such variants were prevalent in genes critical to cancer risk assessment, cardiology, neurology, and pediatrics. In the interlaboratory study, most of the “easy” SNVs and indels in the pilot specimen were uniformly detected. However, only 10 of the 22 challenging variants were detected by all tests, and just 3 tests detected all 22. Many, but not all of these limitations were bioinformatic in nature and most were previously known but not well documented. Conclusions: The high prevalence of complex medically important variants is an under-recognized problem, incompletely addressed by current off-the-shelf DNA sequencing methods and control reagents. Approaches such as ours may help improve the standardization, quality, and transparency of clinical genetic tests.

Capturing the 3D Folds of Whole Mammalian Genomes in Single Cells

Tim J Stevens[1]; David Lando[2]; Xiaoyan Ma[2]; Srinjan Basu[2]; Ernest D Laue[2];
[1]MRC Laboratory of Molecular Biology, Cell Biology, Cambridge, Cambridgeshire, United Kingdom
[2]University of Cambridge, Biochemistry, Cambridge, Cambridgeshire, United Kingdom

We have calculated 3D structures of entire mammalian genomes using single-cell Hi-C data. Using up to 200,000 DNA ligation events per haploid genome, a particle-on-a-string representation of chromosomes, with segment sizes down to 25 kb, is folded to generate a precise, packed, whole-genome structure. This allow us to know where, within certain limits, all the different DNA sequences of an entire genome reside in the 3D volume of the nucleus. By studying genome structures from several cells it is clear that different G1-phase genomes have completely different chromosomal arrangements and that smaller-scale features like TADs and loops are observable but somewhat dynamic. Nonetheless, all structures show consistent organisational principles: an overall 3D genome organisation is observed that segregates inactive and gene sparse regions from the transcriptionally active regions, thus helping to confirm the origin of the A/B compartment pattern observed in population Hi-C. ChIP-seq and RNA-seq data from cell populations suggest that factors and markers associated with gene activity have greater 3D co-localisation where there is increased transcription and at the interior inter-chromosomal interfaces. Also, detailed analysis of transcription factor (TF) binding sites reveals two co-localising groups of TFs with distinct cellular roles. These and similar observations suggest that whole genome structures can become an important resource for investigating genome function and testing molecular hypotheses.

Self-assembling Manifolds in Single-cell RNA Sequencing Data

Alexander J Tarashansky[1]; Yuan Xue[1]; Pengyang Li[1]; Stephen R Quake[2,4]; **Bo Wang**[1,3];
[1]Stanford University, Bioengineering, Stanford, California, United States
[2]Stanford University, Applied Physics, Stanford, California, United States
[3]Stanford University School of Medicine, Developmental Biology, Stanford, California, United States
[4]Chan Zuckerberg Biohub, , San Francisco, California, United States

Analysis of single-cell transcriptomes remains an open challenge in that existing algorithms all have limitations in their ability to select features that can resolve subtle differences in cell types. Here we present the self-assembling manifolds (SAM) algorithm, a fully unsupervised method for dimensionality reduction and marker gene identification. SAM employs a novel feature selection strategy in which it iteratively rescales gene expression, weighting genes according to their ability to separate distinct groups of cells or cell states. Benchmarking on 48 published datasets against other state-of-the-art methods reveals that SAM consistently improves manifold reconstruction, cell clustering and marker gene identification, especially in datasets that contain cells in dynamic transitions or cell groups that are only distinguishable through subtle differences. We use SAM to analyze the stem cells from the parasitic flatworm, *Schistosoma mansoni*, which infects more than 250 million people worldwide. SAM is able to identify new stem cell subpopulations in juvenile parasites and their respective associated marker genes which we validate using fluorescent in-situ hybridization. In comparison, other existing methods fail to capture any of these populations. Taken together, we show that SAM is particularly useful for unsupervised, parameter-free analysis of scRNA-seq data from tissues and organisms with little to no *a priori* knowledge to gain novel biological insights.

Nanopore Translocation of Knotted DNA

Cristian Micheletti
SISSA Trieste, Italy

No Abstract

Long-term RNA-based Transmission of Biological States

Andrew Fire
Stanford University, Palo Alto, California, USA

This talk will explore the diversity and complexity of rules by which RNAs (i) can serve as a long-term carrier for cellular and organismal state information (biochemistry), (ii) could serve as a long-term carrier for cellular and organismal state information (bio-engineering), and perhaps (iii) do serve as a long-term carrier for cellular and organismal state information (biology).

Thermostable Group II Intron Reverse Transcriptases (TGIRTs) and Their Use in RNA-seq

Alan M. Lambowitz[1]
[1]University of Texas at Austin, , Austin, Texas, United States

Group II intron reverse transcriptases (RTs) are ancient enzymes that are evolutionary ancestors of retroviral RTs and have novel biochemical properties advantageous for RNA-seq and other biotechnological applications. These properties include higher fidelity and processivity than retroviral RTs, and a novel template-switching activity that enables seamless attachment of RNA-seq adapters to nucleic acid templates without RNA tailing or ligation, enabling RNA-seq library construction from small amounts of starting material. Thermostable group II intron RTs (TGIRTs) from bacterial thermophiles combine these properties with the ability to reverse transcribe at high temperatures, which help melt out impeding RNA secondary structures. We have used TGIRTs to develop new RNA-seq methods (TGIRT-seq), which enable: (i) quantitative profiling of protein-coding and long ncRNAs in the same RNA-seq as tRNAs and other small ncRNAs; (ii) give full-length, end-to-end reads of tRNAs and other structured small ncRNAs; and (iii) enable high-throughput mapping of post-transcriptional modifications by distinctive patterns of misincorporation. Validation of TGIRT-seq using fragmented human reference RNAs with ERCC spike-ins demonstrated advantages compared to the widely used TruSeq v3 method, including better quantitation, higher strand-specificity, less bias, more uniform 5â€™- to 3â€™-gene coverage, and detection of more splice junctions, particularly near the 5â€™ ends of genes, even from fragmented RNAs. The ability of TGIRT-seq to construct comprehensive RNA-seq libraries from small amounts of RNA is useful for analyzing extracellular RNAs in human plasma and exosomes, with potential liquid biopsy applications. A crystal structure of a full-length TGIRT bound to template-primer substrate and incoming dNTP identified novel structural features that contribute to its beneficial properties, revealed a close evolutionary relationship between group II intron RTs and RNA-dependent RNA polymerases, and indicated that retroviral RTs are degenerate enzymes that have lost regions and active-site features that contribute to the high processivity and fidelity of group II intron RTs.

How Dengue Capsid Protein Assists in Organizing Dengue Virus Genomic RNA

Priscilla L. Boon[1,2,3], Roland G. Huber[2], Ana S. Martins[4], Ivo C. Martins[4], Yue Wan[5].Peter J. Bond[2,3].
[1]NUS Graduate School of Integrated Science and Engineering, National University of Singapore, Singapore, Singapore, [2]Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore, [3]Department of Biological Sciences, National University of Singapore, Singapore, Singapore, [4]Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal, [5]Genome Institute of Singapore, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore.

Dengue virus is an enveloped positive-sense single-stranded RNA virus. Its ~11 kb genome consists of one open reading frame encoding three structural proteins, including envelope (E), membrane (M), and capsid (C) proteins, along with seven non-structural proteins. In its infectious mature state, the genome is packaged within the viral envelope, composed of E and M proteins embedded within a lipid bilayer vesicle. The RNA genome and multiple copies of the C protein form the nucleocapsid. The C protein is a positively charged, partially disordered alpha-helical protein with a unique fold. It exists as a homodimer in solution and has been shown to associate with RNA and lipid-droplets in-vivo. The nucleocapsid is conventionally thought to comprise of a randomly coiled genome encapsulated by an amorphous outer shell of C proteins. However, we show here that the RNA genome of the dengue virus is decidedly more organized than expected. Using a combination of chemical modification, next-generation sequencing, structural biology, and multiscale molecular dynamics (MD) simulations, we investigate the inherent conformational dynamics of the C protein, and its preference for secondary structural elements and specific binding sites on the viral genome. Studies of the isolated C protein dimer in vitro and in silico reveal that it exhibits clamp-like motions that occlude a hydrophobic patch that may mediate its interaction with lipid bilayers. We also model the interactions of the capsid protein with different RNA secondary structural elements and explore its effect upon genome packaging. Using this collection of information, we proceed to build a 3D model of the nucleocapsid and, in turn, a complete dengue virion. This "virtual virus" represents a platform for studying various aspects of the viral life cycle in the future, including fusion, genome uncoating, and viral assembly.

Sequential Folding of RNA

Sarah Woodson
Johns Hopkins University, USA

No Abstract

Functional Characterization of Gene Regulatory Elements

Nadav Ahituv
University of California, San Francisco, USA

No Abstract

Quantitative Mapping of Transcription Factor Binding Energy Landscapes

Polly Fordyce
Stanford University, USA

No Abstract

Noise in the Shadows

Zeba Wunderlich
University of California, Irvine, USA

No Abstract

Mismatched Base-pairs Locally Distort DNA Structure and Can Induce Increased DNA-binding by Transcription Factor Proteins

Ariel Afek, Honglue Shi, Atul Rangadurai, Hashim Al Hashimi, Raluca Gordan.
Duke, Durham, NC, USA.

The local structure of genomic DNA can vary drastically from the ideal B-form double helix, and one cause for structural deformations is the pairing of non-complementary bases (i.e. mis-paired bases, or mismatches). DNA mismatches are frequently formed by spontaneous base deamination, replication errors, and genetic recombination. Mismatches alter DNA structure and the functional groups available in the major/minor grooves, which can affect interactions with regulatory transcription factors (TFs). Currently, very little is known about the effects of mismatches and other DNA structural changes on TF binding.

We present Saturation Mismatch Binding Assay (SaMBA), the first assay to characterize the effects of mismatches on TF-DNA binding in high-throughput. For genomic sequences of interest, SaMBA generates DNA duplexes containing all possible single-base mismatches, and quantitatively assesses the effects of the mismatches on TF-DNA interactions.

We applied SaMBA to measure binding of 20 TFs (covering 14 structural families) to thousands of mismatched sequences, and mapped the impact of mismatches on these TFs. Interestingly, for all tested factors we found that DNA mismatches within TF binding sites can significantly increase TF binding levels compared to the wild-type sequences. Furthermore, for several TFs we have identified non-specific genomic regions that become strongly bound after certain mismatches are introduced. Structural analyses of mismatches that increase TF binding revealed that these mismatched are oftentimes distorting the naked DNA such that its structure becomes similar to that of bound DNA sites, thus explaining the increased binding measured in our assay.

In addition to providing new insights into the role of DNA structure in protein-DNA recognition, our finding of increased TF binding to mismatched DNA is important for understanding DNA repair and the formation of mutations in the genome, directions that we currently investigating.

What Gene Expression Noise Tells About the Spatiotemporal Organization of Gene Regulatory Networks

Alexander Wood[1]; Ruud Stoof[1]; Angel Goni-Moreno[1];
[1]Newcastle University, School of Computing, Newcastle upon Tyne, Newcastle upon Tyne, United Kingdom

Gene expression noise is not only the consequence of mere random fluctuations, but also a signal that reflects physical dynamics of upstream regulatory machinery.Many bacteria exploit such noisy signals to deploy metabolic bet-hedging and/or division of biochemical labor, permitting clonal cells to exhibit different phenotypes and thus making populations more resistant to environmental shocks. This suggests that each gene and each protein of a given genetic network might need a specific location within the volume of a cell for optimal performance â€“ an issue that has received relatively little attention. We studied promoter activity in regard to its distance to its regulating transcription factor source. Experiments with engineered bacteria in which that distance is minimized or enlarged highlighted the implication of spatial effects in gene expression patterns. In particular, it was found that gene expression noise was greater when the controlling transcription factor expression site was moved a greater distance from the promoter.Computational analysis suggested that the information about the geometrical architecture of genetic networks is finely encoded in gene expression signals. This approach allowed deconvolution of experimental data into spatial mechanistic information of gene regulation and provided a basis for selecting programmable noise levels in synthetic regulatory circuits.

Culture-free Microbial Genome Assembly and Tracking in Hospitalized Patients

Ami Bhatt
Stanford University, USA

No Abstract

DNA in Tight Spaces: Linking Structure, Stability and Protection in Sperm Chromatin

Jason DeRouchey.
University of Kentucky, Lexington, KY, USA.

Packaged DNA is ubiquitous in nature and the laboratory with examples ranging from chromatin, viruses, sperm cells, bacterial nucleoids, artificial viruses and gene therapy constructs. Sperm nuclei are one of the best examples of in vivo maximum DNA compaction and therefore an ideal model system to study biophysically. Despite intense research, the physical mechanisms underlying tight packaging of DNA remain poorly understood especially at the molecular level. Spermiogenesis is a unique multi-step process resulting ultimately in the replacement of histones by protamines in sperm nuclei to a final volume roughly 1/20th that of a somatic nucleus. The near crystalline organization of DNA in mature sperm is thought crucial for both DNA delivery and the protection of genetic information due to the absence of DNA repair. Using osmotic stress coupled to small-angle X-ray scattering(SAXS) to directly measure intermolecular forces, we have done extensive work to understand how cations modulate DNA-DNA forces in the condensed phase and the interrelationships between cation chemistry, packaging densities and compaction. Recent experiments aimed at understanding the various biological implications for both protamine-DNA packaging and correlations to infertility and oxidative stress in sperm chromatin will also be discussed.

Eukaryotic Genome Evolution in Extreme Environments

Joanna Kelley
Washington State University, USA

No Abstract

POSTER ABSTRACTS

POSTER SESSION

Wednesday, August 22, 2018
16:15 – 18:00
New Brighton Room

Posters are available for viewing only during their scheduled date of presentation. Below are the formal presentation times. The presenters listed below are required to remain in front of their poster boards to meet with attendees during the designated times.

Odd-Numbered Boards 16:20 – 17:05 | Even-Numbered Boards 17:05 – 18:00

Backman, Vadim	**1-POS**	**Board 1**
Bai, Lu	**2-POS**	**Board 2**
Barbensi, Agnese	**3-POS**	**Board 3**
Chuang, Frank	**4-POS**	**Board 4**
Huber, Roland	**5-POS**	**Board 5**
Huertas, Jan	**6-POS**	**Board 6**
Johnson, Stephanie	**7-POS**	**Board 7**
Motai, Yuichi	**8-POS**	**Board 8**
Munoz, Victor	**9-POS**	**Board 9**
Nordenskiöld, Lars	**10-POS**	**Board 10**
Norouzi, Seyed Davood	**11-POS**	**Board 11**
Porter, Lauren	**12-POS**	**Board 12**
Qui, Xiangyun	**13-POS**	**Board 13**
Reymer, Anna	**14-POS**	**Board 14**
Spakowitz, Andrew	**15-POS**	**Board 15**
Sumikama, Takashi	**16-POS**	**Board 16**
Wan, Biao	**17-POS**	**Board 17**
Wedemann, Gero	**18-POS**	**Board 18**
Yoo, Jejoong	**19-POS**	**Board 19**
Young, Robert	**20-POS**	**Board 20**
Zhan, Y. Ada	**21-POS**	**Board 21**
Ziraldo, Riccardo	**22-POS**	**Board 22**

Posters should be set up on the morning of Monday, August 20 and removed by 18:00 on Wednesday, August 22. All uncollected posters will be discarded.

1-POS Board 1

Bridging Chromatin Nanoimaging and Molecular Modeling – What Can We Learn About Chromatin Packing as a Regulator of Transcription?

Vadim Backman[1].
[1]Robert H. Lurie Comprehensive Cancer Center, Evanston, IL, USA, [2]Northwestern University, Evanston, IL, USA.

The behavior of organisms is determined by both their genetic code and their capacity to explore a transcriptional landscape of thousands of genes to create new functional states. The talk will discuss the development of two synergistic platforms to study 3D chromatin structure and its regulation of transcription: The chromatin nanoimaging platform co-registers 3D electron tomography for chromatin imaging at the scale of individual DNA strands, spectroscopic photo-localization optical nanoscopy for imaging of targeted genes and transcriptional regulators, and partial wave spectroscopic microscopy for the quantification of the statistics of chromatin packing and contact probability at genomic length-scales from ~kbp to ~10Mbp in live cells, real-time, label-free in hundreds of cells simultaneously with single cell resolution. The computational platform enables predictive modeling of gene transcription in the context of a realistic chromatin nanoscale structure from atomistic details of DNA to whole chromatin and is based on a multi-scale analysis, which includes molecular dynamics, Brownian dynamics, and molecular theory. By combining the imaging and computational platforms, we show how differential chromatin packing works across length-scales from individual DNA strands to topologically associated domains as a key regulator of intercellular transcriptional heterogeneity, transcriptional diversity, and gene-network heterogeneity, which in turn affect cells' transcriptional access to their genomic space and have implications on cellular processes that depend on cells' ability to adapt or respond to external stresses. Carcinogenesis provides a particularly significant testbed. We show that the regulation of the physical structure of chromatin packing allows for the predictable modulation of transcriptional heterogeneity in cancer cells during chemotherapeutic response to achieve near-complete cancer cell killing *in vitro*.

2-POS Board 2

Systematic Study of Nucleosome-Depleting Factors in Budding Yeast

Lu Bai.
Pennsylvania State University, University Park, USA.

Nucleosomes present a barrier for the binding of most transcription factors (TFs). However, special TFs known as nucleosome-depleting factors (NDFs) can access embedded sites and cause nucleosome depletion. Here, we developed a novel high-throughput method in yeast to identify NDFs among 104 TFs and systematically characterized the impact of orientation, affinity, location, and copy number of their binding motifs on nucleosome depletion level. Using this assay, we identified ~20 new NDF motifs and divided the nuclear TFs into three groups with strong, weak, and no nucleosome-depleting activities. Further studies revealed that tight DNA binding is the key property that underlies strong NDF activity, and this activity has cell-cycle dependence. These observations lead to a model where NDFs function by binding to transiently exposed motifs during DNA replication and preventing nucleosome formation. Overall, our study presents a new framework to functionally characterize NDFs and elucidate the mechanism of nucleosome invasion.

3-POS Board 3

Modelling the Unknotting Functionof Topoisomerases and Knot Adjacency

Agnese Barbensi[1]; Dorothy Buck[2]; Heather A Harrington[1]; Daniele Celoria[1]
[1]University of Oxford, Oxford, United Kingdom
[2]University of Bath, Bath, United Kingdom

Action of type II topoisomerases on covalently closed DNA molecules can change the topology, resulting in a range of different knot types. Here we model the configuration space of a knotted DNA molecule as a graph. The vertices are planar projections (i.e. shadows) of the molecule, which can be thought as a closed polymer in the space, and these are connected by edges representing local deformations and inter-segmental passages. This diagram focused approach is applied to investigate knot adjacency and the unknotting function of topoisomerases. We complement and synthesise earlier work by Stasiak and collaborators by looking at neighboring subspaces in the graph of the configurations, modeled as a network of grid diagrams with increasing complexity. We suggest a grid-based calculation as a new and computationally convenient method for investigating unbiased knotting probability biopolymers. Furthermore, in this setting we are able to emulate simulations by Zechiedrich and collaborators in the lattice model to investigate the role of local juxtaposition geometry for Topoisomerases action.

4-POS Board 4

KSHV as Functional Probes for Image-Based Biophysical Studies of 4D Nucleome

Frank Chuang, Yoshihiro Izumiya.
University of California, Davis, Sacramento, CA, USA.

In the course of studying the epigenetic regulation of Kaposi's sarcoma-associated herpesvirus (KSHV) activation in latent infected cells, we have developed a model system that we believe is useful for studying the spatiotemporal (4D) organization of the host cell nucleome. The system is comprised of human B-cell lymphoma lines that are silently infected with KSHV to mimic the latent state. Viral replication is then reactivated by the expression of a single protein, K-Rta (KSHV Replication and Transcription Activation). With this engineered cell line, we have examined several aspects of KSHV gene regulation. Through chromosome conformation capture analyses (Capture Hi-C) we found that the structure of KSHV genomes shift to increase looping at K-Rta binding sites with a notable enhancement at PAN RNA (long non-coding) promoter regions. Rapid immunoprecipitation mass spectrometry of endogenous protein (RIME) was used to generate an extensive list of proteins that interact with K-Rta and cellular RNA Pol II, which could help us to identify proteins that may be responsible for the reorganization of KSHV episomes during reactivation. Most recently, we have used 4D fluorescence imaging to observe that reactivation stimulation causes a significant fraction of KSHV episomes (which are initially distributed widely throughout the nucleus) to aggregate with cellular RNA pol II and form large viral transcription/replication complexes. The mechanism that facilitates this process is not well understood – nor is the basis for heterogeneity in the response of KSHV episomes to reactivation stimulation. Since the viral episomes are genetically identical, we postulate that their individual responses must be modulated by physical or biochemical variations in the local environment. In this way, KSHV episomes could serve as nanoscale probes to map functional regions of the host cell nucleome in a new image-based and biophysical approach.

5-POS Board 5

Identifying Viral Genome Structures Throughout the Viral Life Cycle by Chemical Modification, Sequencing and Simulation

Roland G. Huber[1], Adelene Sim[1], Yue Wan[2], Peter J. Bond[1,3].
[1]n/a, Singapore, Singapore, [2]Bioinformatics Institute (BII), Singapore, Singapore, [3]Genome Institute Singapore (GIS), Singapore, Singapore, [4]National University Singapore (NUS), Singapore, Singapore.

Viral genomes are necessarily highly dynamic, as they undergo many structural changes throughout a complex life cycle consisting of translation, replication, packaging and uncoating phases. To understand the structure and dynamics of these genomes, we apply a broad range of biophysical techniques to study local and genome-scale features of Dengue and Zika viral RNA. Recently, the envelope structures of several viruses have been determined to near-atomistic resolution using Cryo-EM. However, a detailed understanding of the viral genome arrangements still eludes us, as the genome is not as rigidly structured as the viral envelope.
While it is well known that these viruses contain non-coding regulatory structural elements within their genomic RNA especially in the 5' and 3' regions, we show that the coding regions contain additional structures that are both highly conserved within and across serotypes, and these structures are functionally important for the viral life cycle. To this end, we employ chemical structure probing techniques in conjunction with next-generation sequencing, such as Shape-MaP and SPLASH, to derive information on the fold of the viral genomes at different phases in their life cycle, allowing us to accurately identify which structural elements are present at what stage. Using this information as constraints, we then proceed to use coarse-grained molecular modeling and simulation techniques to create detailed three-dimensional models that in turn allow us to understand the spatial arrangement of structural elements within the viral RNA. We proceed to integrate these results with our prior work on the structure and dynamics of the viral envelope to derive a complete model of a fully assembled, infectious virion in its mature state.

6-POS Board 6

Sequence-dependent Nucleosome Dynamics: Implications for Transcription Factor Binding

Jan Huertas[1], Caitlin M. MacCarthy[1], Hans R. Schöler[1,2], Vlad Cojocaru[1,3].
[1]Max Planck Institute for Molecular Biomedicine, Münster, Germany, [2]University of Münster, Münster, Germany, [3]University of Münster, Münster, Germany.

Transcription factors are proteins that bind to DNA to regulate gene expression. In most cases, accessibility to DNA is a prerequisite for their function. However, in the nucleus the DNA is packed into chromatin, which is often inaccessible. The fundamental unit of chromatin is the nucleosome, in which 147 DNA basepairs are wrapped around a core of eight histone proteins.

Interestingly, a series of transcription factors, known as pioneers, are able to bind to closed chromatin states, recognizing their binding sites even in the presence of nucleosomes. They can help open chromatin, increase DNA accessibility, and support binding of other transcription factors.

For example, Oct4, a master regulator of stem cell pluripotency, is able to bind native nucleosomes in a sequence specific manner. We confirmed this experimentally and showed that Oct4 does not bind to nucleosomes with DNA sequences engineered to optimize the positioning of the histone core. Binding to such sequences was not observed even when canonical Oct4 binding sites were introduced using a systematic structural modeling approach.

To understand the nucleosome properties that facilitate Oct4 binding, we performed 3µs of all-atom simulations of three nucleosomes with different DNA sequences, each with a characteristic Oct4 binding profile. Remarkably, we found that the nucleosome mobility correlates with the number of Oct4 factor binding sites available. We identified differences in dynamics and structural properties of the three nucleosomes, most of which are located in the regions known to be important for nucleosome unwrapping. Interestingly, the regions containing the Oct4 binding sites were characterized by higher mobility. We validated our findings by probing the stability of the nucleosomes in thermal unwrapping experiments. These findings suggest that nucleosome dynamics are determinant for the ability of pioneer transcription factors to recognize their binding sites in closed chromatin.

7-POS Board 7

Regulation of Rapid Nucleosome Sliding by Substrate Cues and Accessory Subunits in the Yeast INO80 Chromatin Remodeling Complex

Stephanie L. Johnson[1], Coral Y. Zhou[1], Laura J. Lee[1], Matthew J. Johnson[2], Adam D. Longhurst[1], Geeta J. Narlikar[1].
[1]University of California, San Francisco, San Francisco, CA, USA, [2]Google Brain, San Francisco, CA, USA.

Chromatin remodeling complexes are molecular motors that catalyze diverse structural rearrangements to nucleosomes. As such, they are involved in many fundamental gene regulatory processes through their manipulation of genome structure. The INO80 family is the most recently discovered of the four main remodeler families. Despite INO80's roles in essential processes such as DNA damage repair and the establishment of promoter architectures, the biophysical mechanisms of INO80-family remodelers remain some of the most elusive. Here we describe a mechanistic dissection of the large, multi-subunit yeast INO80 complex, using a combined ensemble biochemistry/single-molecule biophysics approach. We show that (i) INO80 exhibits a switch-like response to flanking DNA length, which is "tuned" by an auto-inhibitory module in the complex; (ii) once sliding is initiated, INO80 moves the nucleosome rapidly at least 20 bp without pausing to re-assess flanking DNA length; and (iii) INO80 can change the direction of nucleosome sliding without dissociation. We present a model for INO80 as a highly processive remodeling motor that is tightly regulated by both substrate cues and non-catalytic subunits of the complex, which may specialize INO80 for its particular in vivo roles.

8-POS Board 8

Establishment of the Three-dimensional Chromatin Architectures in Medaka Embryogenesis

Ryohei Nakamura[1]; **Yuichi Motai**[2]; Masahiko Kumagai[4]; Haruyo Nishiyama[1]; Neva C Durand[3]; Erez L Aiden[3]; Shinichi Morishita[2]; Hiroyuki Takeda[1]
[1]The University of Tokyo, Department of Biological Sciences, Bunkyo-ku, Tokyo, Japan
[2]The University of Tokyo, Department of Computational Biology and Medical Sciences, Kashiwa, Chiba, Japan
[3]Baylor College of Medicine, Department of Molecular and Human Genetics, Houston, Texas, United States
[4]National Agriculture and Food Research Organization, , Tsukuba, Ibaraki, Japan

Mammalian genomes have two distinct three-dimensional (3D) architectures in nuclei; compartments and loop domains, while fruit fly genome exhibits compartments but no loop domains. This difference sheds light on the conservation of the 3D features of chromatin across species, motivating us to understand the 3D structural differences in vertebrates. We generated the loop-resolution Hi-C contact maps in medaka (Japanese killfish) fibroblast cells, and observed both of the two structures in medaka like mammals, indicating that the principal mechanisms underlying chromatin architecture in mammals have been conserved since the emergence of bony fish.
Similar to other epigenetic features, the 3D genome architectures are also reprogrammed in early embryogenesis. According to the study of mouse embryo, both compartments and loop domains emerge during zygotic genome activation (ZGA). ZGA occurs at the one- to two-cell stage in mouse; however, in fishes and amphibians, ZGA takes place at the mid-blastula stage (256 to 1024-cell stage). Due to such extensive variability in the timing and duration of ZGA, the order of ZGA and the reprogramming of 3D chromatin structures remains uncertain. Here, using medaka, we generated loop-resolution Hi-C maps at 12 time points spanning embryogenesis. As the mammalian studies, we find the small contact domains and compartmentalization appears during ZGA simultaneously with the emergence of chromatin accessibility variation for each locus.

9-POS Board 9

Eukaryotic Transcription Factors Can Track and Control Their Target Genes Using DNA Antennas

Victor Munoz[1,2]; Milagros Castellanos[2];
[1]University of California Merced, Bioengineering, Merced, California, United States
[2]IMDEA Nanoscience, Biophysics, Madrid, Madrid, Spain

Eukaryotic transcription factors (TF) function by binding to short 6-10 bp DNA recognition sites located near each of their many target genes, which are scattered through vast genomes. Such process surmounts enormous specificity, efficiency and celerity challenges using a molecular mechanism that remains unknown. Combining biophysical experiments, theory and bioinformatics we dissected the interplay between the DNA-binding domain of Engrailed, a Drosophila TF, and the regulatory regions of its target genes. Remarkably, Engrailed binding affinity is greatly amplified by the DNA regions flanking the recognition site, which contain long tracts of degenerate binding-site repeats. Such DNA organization operates as an antenna that attracts TF molecules in a promiscuous exchange between myriads of weaker binding sites. The antenna ensures a local TF supply, enables gene tracking and fine control of basal site occupancy. This mechanism illuminates puzzling gene expression data, and suggests simple engineering strategies to control gene expression.

10-POS Board 10

Hierarchical Multiscale Simulation of DNA Condensation at Mesoscale Level

Lars Nordenskiöld[1]; Tiedong Sun[1]; Alexander Mirzoev[1]; Nikolay Korolev[1]; Alexander Lyubartsev[2];
[1]Nanyang Technological University, School of Biological Sciences, SINGAPORE, South-West, Singapore
[2]Stockholm University, Materials and Environmental Chemistry, STOCKHOLM, Stockholms län, Sweden

DNA condensation at mesoscale level, induced by multivalent ions is of substantial importance for packing of DNA *in vivo* with many applications in biology, biotechnology and polymer physics. Rigorous modeling of this process with all-atom molecular dynamics (MD) simulations is presently impossible due to size and time scale limitations. Here, we present a hierarchical approach for systematic multiscale coarse-grained (CG) simulations of DNA condensation induced by the three-valent cobalt(III)-hexammine ($CoHex^{3+}$). On the basis of all-atom MD simulations, we extract solvent-mediated effective potentials for a CG model of DNA and simulate DNA aggregation in the presence of $CoHex^{3+}$. Further coarse-graining to a "super-CG" DNA model enables simulations of DNA condensation at mesoscale level. Modeling a 10 kbp-long DNA molecule results in formation of a toroid with distinct hexagonal packing in agreement with Cryo-EM observations. The approach uses no adjustable parameters and is applied on DNA up to megabase dimensions. It may be generalized to modeling chromatin up to chromosome size.

11-POS Board 11

Topological Variability and Conformational Dynamics of Chromatin Fibers

Seyed Davood Norouzi, Victor B. Zhurkin.
NIH, Bethesda, USA.

The DNA double helix is repeatedly supercoiled in nucleosomes. The core of a typical nucleosome contains 145-147 bp of DNA making ~1.7 left superhelical turns around a histone octamer. The nucleosomes cores are connected by DNA linkers forming nucleosome arrays, which in turn, form higher order structures that determine packaging, accessibility, and DNA proximities in chromatin loops. Unlike nucleosome with available X-ray structure, the organization of DNA in the so-called 30-nm fiber is still under debate. By performing an exhaustive computational analysis of stereo-chemically feasible chromatin fibers with inter-nucleosome linkers, "L", we found two families of topoisomers characterized by different linking number, Lk, values. Depending on the linker, the most favorable topoisomer belongs either to the 'known' family with Lk ≈ –1.5 and L=10n, or the 'novel' family with Lk ≈ –1 and L=10n+5. The fiber configurations observed by X-ray or cryo-EM belong to the 'known' family with high superhelical density (and L=20, 30, 40 bp). Based on these in silico findings we made several testable predictions, among them different degrees of DNA supercoiling in fibers with variable linker length (verified by topological gel assay), different flexibility of the two types of fibers (consistent with sedimentation velocity data), and a correlation between the local inter-nucleosome spacing and the level of transcription in different parts of the yeast genome (confirmed by comparing nucleosome positioning maps and transcription data). In terms of conformational dynamics and accessibility, we found in silico evidence of regulation of chromatin compaction by histone epigenetic modifications. At the local level, we found evidence of breathing of nucleosome ends by ~10 bp which increases the accessibility of DNA to transcription factors including tumor suppressor protein p53. Overall, combining experimental data and computational models helps us unveil spatio-temporal dynamics of chromatin.

12-POS Board 12

Lauren Porter, Loren L. Looger.
HHMI, Sterling, USA.

As the number of sequenced genomes increases, the need for computational techniques that identify features of those genomes increases also. Currently, we are developing a computational technique for identifying proteins that switch folds, i.e. remodel their secondary structures in response to cellular changes. Previously, we identified 96 instances of biologically relevant fold switching. These instances occur in response to physiological changes, increase the functionality of the proteome, and occur in all kingdoms of life. Additional evidence suggests that fold-switching proteins are likely to occur more frequently in nature than indicated by the current population of structures in the Protein Data Bank (PDB). Together, these results indicate that protein fold switching is likely more common than currently believed, suggesting that many important functions of proteins remain unknown. Thus we are developing a computational technique that identifies more fold-switching proteins. To do this, we identified two of their characteristic features: incorrect secondary structure predictions and likely independent folding cooperativity. Using these features, our technique correctly identified the fold-switching regions of 14/16 proteins with experimental evidence for fold switching ($p < 7.5*10\text{-}6$, Fisher's exact test). Additionally, it identified 64/93 proteins ($p < 4.7*10\text{-}13$, Fisher's exact test) that were likely to switch folds based on homology. Out of the 29 false negatives, 22 result from secondary structure predictions that seem to be biased by overpopulation of common structure in the PDB. We are currently working to decrease these false negatives and winnow down false positives. These improvements—combined with experimental validation of our technique—could lead to the rapid identification of proteins with heterogeneous structural and functional ensembles, which has implications for genomics and human health.

13-POS Board 13

Elucidate Chromatin Folding with Solution X-ray/Neutron Scattering

Xiangyun Qiu.
George Washington University, Washington, DC, USA.

Knowledge of the three-dimensional structure of chromatin, an active participant of all gene-directed processes, is required to decode its (epi)genetics-structure-function relationships. Albeit often simplified as “beads-on-a-string”, chromatin possesses daunting complexity in its intricate intra- and inter-nucleosome interactions, as well as the myriad types of molecules acting on it. On the other hand, the folding of chromatin from an extended chain of nucleosomes is highly constrained, e.g., by rather bulky nucleosomes and semi-rigid linker dsDNAs. Further given the well-defined nucleosome and dsDNA structures at the nanometer scale, this creates an opportunity for low-resolution structural methods such as small angle scattering to obtain mesoscale structures of chromatin, which can be further refined computationally to yield atomistic structures of chromatin. Here we present results from our recent studies of recombinant nucleosome arrays with solution small angle x-ray and neutron scattering (SAXS/SANS) and ensemble structure modeling.

14-POS Board 14

Regulatory Role of DNA Sequence in Mechanics of DNA Supercoiling Transitions

Anna Reymer[1]
[1]University of Gothenburg, Chemistry and Molecular Biology, Gothenburg, Västra Götalands län, Sweden

Ability of DNA to dynamically change its superhelical state is central to many biological functions, including regulation of gene expression, repair, and packaging in the cell. To address conformational mechanics of DNA during supercoiling transitions we designed a new molecular modelling tool, which can be used in complement with standard all-atom molecular dynamics software. The tool controls the torsional state of DNA between any two base pairs, without restricting any other DNA helical parameter. The tool can be applied to DNA molecules of any length and curvature, alone or in complex with other molecules. This allows for the first time to study DNA in conditions resembling its in vivo state, where DNA's topology is substantially restricted. We applied the tool to a number of linear DNA molecules, including several with methylated cytosine bases, changing their superhelical density from -0.2 to +0.2, which corresponds to under- or overwinding by 6 degrees per base pair step. DNA's response to the torsional stress appears discontinuous - certain dinucleotides are capable of absorbing most of the torsional stress. These "twist-capacitor" dinucleotides, as we call them, modify their twist through coupled conformational changes in DNA backbone. This allows the rest of DNA to remain close to canonical B-form, despite the overall torsional stress. We observe similar trends also in methylated DNA sequences. Overall cytosine methylation stiffens the DNA molecules. But the effect is strongly modulated by the surrounding sequence, where the epigenetic mark could hinder either under- or overwinding molecular transitions. These findings constitute a new aspect of how DNA sequence contributes to biological regulation.

15-POS Board 15

Modeling the Physical Processes Underlying Epigenetic Regulation

Andrew J. Spakowitz[1]
[1]Stanford University, Chemical Engineering, Stanford, California, United States

Historically, the central dogma of genetics asserted that DNA sequence holds all of the information that orchestrates cellular function. However, mounting experimental evidence shows the pivotal role that protein and DNA modifications play in the expression of the genome. In other words, two organisms with identical genetic information may have vastly different behavior due to chemical modifications in their genome packaging. This notion of epigenetic regulation represents a paradigm change in how we think about genetic traits. Aberrations in epigenetic markers lead directly to a range of diseases, including various cancers, developmental disorders, obesity, and diabetes. Research in our lab focuses on biological processes involving DNA to establish a predictive theoretical model that offers new and critical insight into the role of physical forces involved in epigenetic regulation. In this talk, we present a multi-scale approach to modeling the segregation of chromosomal DNA into condensed regions called heterochromatin. This effort leverages our field-theoretic model for predicting copolymer morphology, resulting in a framework that can translate epigenetic modifications at a single nucleosome to genome-scale segregation.

16-POS Board 16

Computational Modeling of Chromosomes and Carbon Nanotubes as a Probe of Atomic Force Microscopy

Takashi Sumikama[1]; Keisuke Miyazawa[2]; Taku Higasayama[2]; Kyosuke Nakayama[2]; Masayuki Harada[2]; Takeshi Fukuma[1,2];
[1]Kanazawa University, Nano Life Science Institute (WPI-NanoLSI), Kanazawa, Ishikawa, Japan
[2]Kanazawa University, Division of Electrical Engineering and Computer Science, Kanazawa, Ishikawa, Japan

Chromosomes are central DNA molecules coding genetic information of life. Recently developed methods such as Hi-C clarified that chromosomes in the interphase take fractal globular structures in each chromosome territory; however, their fine structure at the molecular or nanometer scale is still unclear. The atomic force microscope (AFM) is capable of observing molecules at the nanometer scale even in liquid environment, thus is one of the promising tools to see the chromosomes in nucleus. Accordingly, we are planning to measure the chromosome structures by using the AFM, but the AFM measurements of soft matters such as DNA fluctuating in solution are still challenging. In prior to the AFM measurement, we are going to theoretically calculate the AFM image using computational simulations. To do so, we have to develop computational models of chromosomes and carbon nanotubes (CNT) used as the AFM probe. There already exist some literatures describing chromosome as a polymer simulation, while some other groups developed polymer simulations for CNT. Here, we combine these two simulations to simulate chromosomes and CNT simultaneously. In this presentation, a movie of CNT going in and out of chromosomes, mimicking the measurement of chromosomes by the AFM, will be shown. The AFM images are going to be calculated by our simulation. Also, we will analyze what kind of motions of chromosomes are induced by such observation to reveal their effects on the AFM images.

17-POS Board 17

Modeling Dynamics of Supercoiling and Applications to RNA Polymerase

Biao Wan, Xinliang Xu, Jin Yu.
Beijing Computational Science Research Center, Beijing, China.

The quantitative description of DNA structures and its dynamics is of importance in understanding many processes in cells. Double helical structure endows DNA with the twisting rigidity and elasticity that allow supercoiling an important means to mechanically communicate in essential genetic processes, such as transcription and replication. For the in-extensible DNA chain model, one of coarse-grained representations of DNA, namely the 'discrete worm-like chain' (DWLC) has offered numerous predictions about properties of supercoiled DNA. In the model, DNA is regarded as a series of elastic segments with potentials of stretching bonds, bending and twisting angles, accompanied by the electrostatic interactions. Its dynamics is realized through Brownian dynamics (BD).

During transcription, RNA polymerases (RNAPs) generate supercoiling along DNA by exerting torques on it. The increasing resisting torque from DNA in turn slows down the RNAP as the positive supercoiling is built up, which can even cease the transcription when the resisting torque reaches a threshold, called stalling torque (‘Ma J and Wang M D. Science, 2013, 340:1580.’). By performing the BD simulations of the DWLC, we studied the supercoiling relaxation under certain torques and tensions. We found that the relaxation depends on the torque and the tension imposed on the DNA, and also depends on the length of the DNA. Based on these results, we could investigate whether DNA supercoiling during transcription is built up in a quasi–static process. For equilibrium cases, detailed results including phase diagrams were shown. For the cases out of equilibrium, however, the non-equilibrium effects caused by the torque can also be demonstrated to be important in transcription. Since the theoretical description of the supercoiling relaxation and propagation are still lack of, we also tried to bring some insights from simulations of the supercoiling generation in the RNAP transcription.

18-POS Board 18

Modeling Effects of Nucleosome Positioning in Chromatin

Gero Wedemann.
University of Applied Sciences Stralsund, Stralsund, Germany.

In higher organisms, DNA is complexed with proteins into a structure named chromatin. The basic packaging unit of chromatin is the nucleosome in which DNA is wrapped around a histone octamer. For the structure of chromatin many models were proposed. Experiments indicate that chromatin exhibits different kinds of packaging in distinct activation states. Packaging and activation states are closely linked to positions of nucleosomes on the DNA which are actively regulated.
Determination of the positions of nucleosomes from experimental data is not straight forward, because of experimental uncertainty and data averages over a heterogeneous cell population. We developed a method based on Monte Carlo combined with a simulated annealing scheme to determine occupancy maps. Applying parallelization techniques, this method can determine occupancy maps genome wide. This method exhibits several advantages over existing tools for synthetic nucleosome maps and real data.
To improve the understanding of the interplay between nucleosome positions and chromatin structure we applied computer simulations of a coarse-grained chromatin model including fundamental physical properties such as elasticity, electrostatics and nucleosome interactions. We simulated chromatin models based on experimentally derived nucleosome positions from cells at different stages of cell differentiation and after stimulation with a cytokine for different loci. Simulation results revealed a significant influence of nucleosome positions on the three dimensional structure of chromatin.

19-POS Board 19

Epigenetically Controlled Phase Behaviors of DNA Revealed by Multi-scale Molecular Dynamics Simulations and Single-Molecule Experiments

Jejoong Yoo[1], Hajin Kim[3,1], Taekjip Ha[4], Aleksei Aksimentiev[2].
[1]Institute for Basic Science, Pohang, South Korea, [2]University of Illinois at Urbana-Champaign, Urbana, IL, USA, [3]Ulsan National Institute of Science and Technology, Ulsan, South Korea, [4]Johns Hopkins University, Baltimore, MD, USA.

Historically, it was believed that the phase behavior would not play a role in biology as important as in physics. However, recent findings that biological protein or RNA systems can experience the phase separation are dramatically changing our general views on the phase behavior in biology. Especially, those findings raise a possibility that DNA molecules might also exhibit phase behaviors, which can directly affect the organization of human chromosomes that consist of compacted heterochromatin and loosely folded euchromatin. Here, we demonstrate that DNA in a polyelectrolyte solution of polyamine or lysine peptides can indeed form separate phases depending on sequences and epigenetic modifications by combining multi-scale molecular dynamics (MD) simulations and single-molecule experiments. First, we show that the condensation force of hundreds base pair-long DNA molecules monotonically increases as the AT content of DNA increases. Interestingly, methylations of cytosine and/or lysine peptides could enhance the formation of condensed phase. Further, we also show that the compactness of kilo-base pairs-long bacterial genes is controlled by the AT contents of the genes: the higher AT contents the smaller radius of gyration. The fact that the compactness of synonymous genes can be simply controlled by adjusting the AT content suggests a new evolutionary role of wobble bases in gene activities. Overall, our findings suggest a novel epigenetic mechanism that can contribute to the mechanical regulation of nucleus-scale organization of chromosomes.

20-POS Board 20

Sequence Dependence Studies of a 336-bp DNA Circle of Various Topology Through Energy Minimization Calculations

Robert T Young[1]; Pamela J Perez[1]; Wilma K Olson[1];
[1]Rutgers University, Chemistry and Chemical Biology, New Brunswick, New Jersey, United States

Computational modeling of DNA has aided in understanding how the double helical structure can deform with and without the assistance of proteins. Such deformations allow for genetic and protein regulation as well as higher-order organization within the cell. A contributing factor to such deformability lies in the primary nucleotide sequence of DNA as some nucleotides have different intrinsic characteristics and helical configurations based on the local sequence context. For example, a pyrimidine-purine (YR) dinucleotide step has a greater degree of deformation when compared to pyrimidine-pyrimidine (YY) or purine-purine (RR) steps. Structural deformations, such as domains that are bent or kinked, give rise to supercoiled structures and even disruptions to the hydrogen-bonding network between bases. We explore such sequence dependence through minimum-energy optimization calculations of various DNA conformational rest states and initial conditions, some of which include bent domains within closed circular structures. Calculations were performed on a 336-base pair sequence found in literature on supercoiled diversity among topological variations, with the minimization rest state values based on dimer-specific roll and twist parameters. Initial conditions that were optimized were either of a smooth planar circle, structures generated from molecular dynamic simulations, or figure-8-like circles constructed from protein-mediated DNA loops. Initial optimization calculations were conducted without constraining the initial conditions, the presence of splayed base pairs found in the molecular dynamics-based structures required further optimization calculations where bent and under/over wound domains were constrained in sequence lengths of integral helical turns. These studies provide a step forward in understanding the effect of sequence on naturally-occurring circular DNA structures of various base pair composition and topology as well as the distinction between physical and biological optimal configurations.

21-POS Board 21

CTCF-mediated Genome Organization Distinguishes Primary Human Erythroid Cells

Y. Ada Zhan[1], Liuyang Cai[1], Vincent Schulz[2], Patrick Gallagher[2], Yijun Ruan[1], Jeff Chuang[1].
[1]The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA, [2]Yale University, New Haven, CT, USA.

Although they share the same genome, the hundreds of cell types in the human body have distinct gene expression profiles. Genome organization plays an important role in bringing distal regulatory elements, such as enhancers, and promoters into proximity in support of cell type specific expression. CTCF has been found essential in mediating chromatin interactions and higher order spatial chromosome folding, and CTCF binding sites usually mark the boundaries of topologically associated domains (TADs). Moreover mounting evidence has revealed the indispensable roles of CTCF at sub-TAD level in segregating enhancer usage or bringing enhancers to their targets in a cell type specific manner. In this study, we investigated how CTCF regulates erythroid gene expression in comparison with B cells and T cells. We found that CTCF binding and CTCF-mediated interactions are elevated at promoter regions (TSS ±5kb) of up-regulated genes in erythroid cells. Although the binding and interactions of CTCF are weaker at cell type-specific than constitutive loci, the erythroid-specific interactions have significant associations with erythroid GWAS SNPs and with erythroid-specific genes. For example, SNP rs2703485 was found in an anchor of an erythroid CTCF loop linking enhancers to the KIT promoter, and it is also bridged to rs218264 and rs218265 via another erythroid CTCF loop. The C and T allele were found solely at rs2703485 and rs218265 sites respectively in our erythroid CTCF ChIA-PET experiment, suggesting this allele is necessary for loop formation. Interestingly, at rs2703485 the C allele is only dominant in the 1000 genomes AFR population. Previous GWAS studies have concluded that the three SNPs have association with red cell phenotype like mean corpuscular volume. Herein we showed they are topologically linked together via CTCF loops and regulate KIT expression.

22-POS Board 22

Quantitative Gel Electrophoresis for Fragment Library Preparation

Riccardo Ziraldo[1], Stephen D. Levene[1], Massa J. Shoura[2], Andrew Z. Fire[2].
[1]The University of Texas at Dallas, Richardson, TX, USA, [2]Stanford University, Stanford, CA, USA.

Gel electrophoresis has been a convenient and cost-effective tool for the characterization of biopolymers. However, many details of the underlying physical mechanisms that govern electrophoretic mobility as a function of molecular size, shape, and mechanical flexibility remain poorly understood. In the case of species that migrate as distinct, well-resolved bands, the band's integrated intensity is both a qualitative and quantitative measure of the mass of a particular molecular species in the sample. For gel-electrophoretic patterns that approach that of a continuum, however, extracting useful information about size distributions requires careful assumptions regarding the underlying sample composition. We have developed a versatile and tunable software tool to fit univariate functions to gel-mobility profiles using minimal information from duplex-DNA standards having both simple and complex size distributions. We apply the tool to quantitative characterization of fragment libraries for next-generation DNA sequencing (NGS), optimization of which depends on knowledge of the size distribution of randomly fragmented DNA samples. We show that the tool provides reasonable average values for fragment size when selecting a portion of the continuous profile for library preparation. In addition to characterizing fragment distributions for library preparation based on a line-profile continuum, the tool is applicable to quantifying complex topologies of circular DNA.