The Detailed 3D Multi-Loop Aggregate/Rosette Chromatin Architecture and Functional Dynamic Organization of the Human and Mouse Genomes

Tobias A. Knoch
Biophysical Genomics & Erasmus Computing Grid
Erasmus Medical Center
TA.Knoch@taknoch.org
The Detailed 3D Multi-Loop Aggregate/Rosette Chromatin Architecture and Functional Dynamic Organization of the Human and Mouse Genomes

Towards a Holistic Understanding of Genomes!
Integral Models of Cell Nuclear Organization I

Already Rabl and Boveri were aware of the obvious fact that the organization of genomes has to be consistent from the sequence level to the morphology of the whole cell nucleus. Although they might be different in detail their common seem is recursive folding and clustering thereof with variation/modification and dynamics accounting for different nuclear states and function.
Dynamic and Hierarchical Genome Organization

The different organization levels of genomes bridge several orders of magnitude concerning space and time. How all of these organization levels connect to processes like gene regulation, replication, embryogenesis, or cancer development is still unclear?
Selective Chromosome Interaction Capture (T2C)

T2C is a novel selective high-resolution high-throughput chromosome interaction capture, in which the relation between region size, resolution, interaction frequency range, and sequencing depth can be designed towards the goal of the experiment. T2C reaches the limit of the “genomic” uncertainty principle and statistical mechanics.
Selective Chromosome Interaction Capture (T2C)

T2C is a novel selective high-resolution high-throughput chromosome interaction capture, in which the relation between region size, resolution, interaction frequency range, and sequencing depth can be designed towards the goal of the experiment. T2C reaches the limit of the “genomic” uncertainty principle and statistical mechanics.

HS IGF locus
~2.1 Mbp
Selective Chromosome Interaction Capture (T2C)

T2C is a novel selective high-resolution high-throughput chromosome interaction capture, in which the relation between region size, resolution, interaction frequency range, and sequencing depth can be designed towards the goal of the experiment. T2C reaches the limit of the "genomic" uncertainty principle and statistical mechanics.

HS IGF locus ~2.1 Mbp
Selective Chromosome Interaction Capture (T2C)

T2C is a novel selective high-resolution high-throughput chromosome interaction capture, in which the relation between region size, resolution, interaction frequency range, and sequencing depth can be designed towards the goal of the experiment. T2C reaches the limit of the "genomic" uncertainty principle and statistical mechanics.

HS IGF locus

~2.1 Mbp
Stable Consensus Architecture of Genomes

Due to the high signal-to-noise ratio of T2C reaching 5-6 orders of magnitude interaction maps reveal definitely an extremely high degree of similarity between different species, cell types, or functional states, thus functional differences are variation of a stable theme persisting through the cell cycle.
Stable Consensus Architecture of Genomes

Due to the high signal-to-noise ratio of T2C reaching 5-6 orders of magnitude interaction maps reveal definitely an extremely high degree of similarity between different species, cell types, or functional states, thus functional differences are variation of a stable theme persisting through the cell cycle.
Fine Structure of Loop Aggregates/Rosettes

Depending on the resolution, the loops within a domain and their arrangement in loop aggregates/rosettes can be shown as well as the details of how the loops are organized at their base as well as their aggregated rosette core: parallel loop fibres exist at the loop base with ~6kbp and these form the core.
Fine Structure of Loop Aggregates/Rosettes

Depending on the resolution, the loops within a domain and their arrangement in loop aggregates/rosettes can be shown as well as the details of how the loops are organized at their base as well as their aggregated rosette core: parallel loop fibres exist at the loop base with ~6kbp and these form the core.

~ 800 kbp
Fine Structure of Loop Aggregates/Rosettes

Depending on the resolution, the loops within a domain and their arrangement in loop aggregates/rosettes can be shown as well as the details of how the loops are organized at their base as well as their aggregated rosette core: parallel loop fibres exist at the loop base with ~6kbp and these form the core.
Fine Structure of Loop Aggregates/Rosettes

Depending on the resolution, the loops within a domain and their arrangement in loop aggregates/rosettes can be shown as well as the details of how the loops are organized at their base as well as their aggregated rosette core: parallel loop fibres exist at the loop base with ~6kbp and these form the core.

~ 800 kbp

~ 380 kbp
Fine Structure of Loop Aggregates/Rosettes

Depending on the resolution, the loops within a domain and their arrangement in loop aggregates/rosettes are shown as well as the details of how the loops are organized at their base as well as their aggregated core: parallel loop fibres exist at the loop base with ~6kbp and these form the core.

Proof of Principle! The 3D Architecture of Chromosomes can be Directly Derived Visually!!!
Simulation of Single Chromosomes

The 30 nm chromatin fiber is modeled as a polymer chain with stretching, bending, and excluded volume interactions. Monte Carlo and Brownian Dynamic methods lead to thermodynamical equilibrium configurations.

All models form chromosome territories with big voids and different chromatin morphologies. Experimental territory and subcompartment diameters agree best with an MLS model with 80 to 120 kbp loops and linkers.

RW/GL model, loop size 5 Mbp, after ~80.000 MC and 1000 relaxing BD steps. Large loops intermingle freely and reach out of the chromosome territory, thus forming no distinct features like in MLS model.

MLS model, loop size 126kbp, linker size 126 kbp, after ~50.000 MC and 1000 relaxing BD steps. Here rosettes form subcompartments as separated organizational and dynamic entities.

RW/GL model, loop size 126 kbp, after ~80.000 MC and 1000 relaxing BD steps. Large loops intermingle freely thus forming no distinct features like in MLS model.
Simulation of Single Chromosomes

The 30 nm chromatin fiber is modeled as a polymer chain with stretching, bending, and excluded volume interactions. Monte Carlo and Brownian Dynamic methods lead to thermodynamical equilibrium configurations. All models form chromosome territories with big voids and different chromatin morphologies. Experimental territory and subcompartment diameters agree best with an MLS model with 80 to 120 kbp loops and linkers.

RW/GL model, loop size 5 Mbp, after ~80,000 MC and 1000 relaxing BD steps. Large loops intermingle freely and reach out of the chromosome territory, thus forming no distinct features like in MLS model.

MLS model, loop size 126 kbp, linker size 126 kbp, after ~50,000 MC and 1000 relaxing BD steps. Here rosettes form subcompartments as separated organizational and dynamic entities.

RW/GL model, loop size 126 kbp, after ~80,000 MC and 1000 relaxing BD steps. Large loops intermingle freely thus forming no distinct features like in MLS model.
Simulation of Whole Nuclei with all 46 Chromosomes

Starting with some metaphase arrangement of cylindrical chromosomes, interphase nuclei with a 30 nm fiber resolution and at thermodynamical equilibrium are created in 4 steps using simulated annealing and Brownian Dynamics methods with stretching, bending, excluded volume and a spherical boundary interactions.

The chromosome territory position depends on their metaphase position and is reasonably stable.
Simulation of Whole Nuclei with all 46 Chromosomes

Starting with some metaphase arrangement of cylindrical chromosomes, interphase nuclei with a 30 nm fiber resolution and at thermodynamical equilibrium are created in 4 steps using simulated annealing and Brownian Dynamics methods with stretching, bending, excluded volume and a spherical boundary interactions.

The chromosome territory position depends on their metaphase position and is reasonably stable.
Starting with some metaphase arrangement of cylindrical chromosomes, interphase nuclei with a 30 nm fiber resolution and at thermodynamical equilibrium are created in 4 steps using simulated annealing and Brownian Dynamics methods with stretching, bending, excluded volume and spherical boundary interactions.

The chromosome territory position depends on their metaphase position and is reasonably stable.
Simulated Interaction Maps

Simulated spatial distance maps as well as simulated interaction maps result in the representation of every parameter variation, and also exhibit the fine-structure describing the loop base as well as rosette core. Thus from the quasi-fibre to the entire chromosome the architecture can be understood in detail.
Simulated Interaction Maps

Simulated spatial distance maps as well as simulated interaction maps result in the representation of every parameter variation, and also exhibit the fine-structure describing the loop base as well as rosette core. Thus from the quasi-fibre to the entire chromosome the architecture can be understood in detail.
Variation of a Consensus Architecture Scheme

The difference between different cell types, functional states or even species is minor despite depending on the region. From this, the chromatin fibre conformation, loop position, and their association into loop aggregates/rosettes can be derived, simulated by polymer models and finally visualized.
Variation of a Consensus Architecture Scheme

The difference between different cell types, functional states or even species is minor despite depending on the region. From this, the chromatin fibre conformation, loop position, and their association into loop aggregates/rosettes can be derived, simulated by polymer models and finally visualized.
DNA Sequence Organization

Determination of the concentration fluctuation function $C(l)$ and its local slope the correlation coefficient $\delta(l)$ are an indication for the i) degree of long-rang scaling behaviour, ii) general multi-scaling, and iii) fine-structure features, which all are connected to all levels of genome organization and especially also the three-dimensional genome architecture.

\[
C(l) = \sqrt{\langle (c_l - \bar{c}_L)^2 \rangle_s}
\]

\[
C(l) = \frac{1}{L-l+1} \sum_{s=1}^{L-l} \left( \frac{1}{l} \sum_{k=1}^{l} n - \frac{1}{L} \sum_{k=1}^{L} N \right)^2
\]

\[
C(l) = \frac{1}{Ll} \sqrt{\frac{1}{L-l} \sum_{s=1}^{L-l} \left[ \left( \sum_{k=1}^{l} Ln \right) - \left( \sum_{k=1}^{L} lN \right) \right]^2}
\]
DNA Sequence Organization

Determination of the concentration fluctuation function $C(l)$ and its local slope the correlation coefficient $\delta(l)$ are an indication for the i) degree of long-rang scaling behaviour, ii) general multi-scaling, and iii) fine-structure features, which all are connected to all levels of genome organization and especially also the three-dimensional genome architecture.
DNA Sequence Organization

Determination of the concentration fluctuation function $C(l)$ and its local slope the correlation coefficient $\delta(l)$ are an indication for the i) degree of long-rang scaling behaviour, ii) general multi-scaling, and iii) fine-structure features, which all are connected to all levels of genome organization and especially also the three-dimensional genome architecture.
DNA Sequence Organization

Determination of the concentration fluctuation function $C(l)$ and its local slope, the correlation coefficient $\delta(l)$, are an indication for the degree of long-range scaling behavior. (i) General multi-scaling, and (ii) fine-structure features, which all are connected to all levels of genome organization and especially also the three-dimensional genome architecture.

Figure: Nucleosomes from the Sequence 'Ground States' and Show the Holistic View.
Scaling analysis show again the entire bandwidth of architectural effects in an aggregated manner. Beyond, they show the scale bridging of the structures and the evolutionary holistic entanglement between the 3D architecture and the DNA sequence organization itself.
Scaling Analysis

Scaling analysis show again the entire bandwidth of architectural effects in an aggregated manner. Beyond, they show the scale bridging of the structures and the evolutionary holistic entanglement between the 3D architecture and the DNA sequence organization itself.
Scaling Analysis

Scaling analysis show again the entire bandwidth of architectural effects in an aggregated manner. Beyond, they show the scale bridging of the structures and the evolutionary holistic entanglement between the 3D architecture and the DNA sequence organization itself.
Scaling Analysis

Scaling analysis show again the entire bandwidth of architectural effects in an aggregated manner. Beyond, they show the scale bridging of the structures and the evolutionary holistic entanglement between the 3D architecture and the DNA sequence organization itself.
Scaling analysis show again the entire bandwidth of architectural effects in an aggregated manner. Beyond, they show the scale bridging of the structures and the evolutionary holistic entanglement between the 3D architecture and the DNA sequence organization itself.

The 3D Architecture and the DNA Sequence Entangled!
From Fiber Topology to Nuclear Morphology

Chromosome territories form in the RW/GL and the MLS model. However, only the MLS model leads distinct subcompartments and low chromosome and subcompartment overlap. Best agreement is reached for an MLS model with 80 to 120 kbp loops and linkers in nuclei with 8 to 10 µm diameter.

The simulated nuclear morphology reflects the chromosome fiber topology of different models in detail.

A: MLS in 6 µm nucleus
  I: 63 kbp loops, 63 kbp linkers
  II: 63 kbp loops, 252 kbp linkers
  III: 126 kbp loops, 252 kbp linkers

B: MLS in 8 µm nucleus
  I: 126 kbp loops, 126 kbp linkers
  II: 84 kbp loops, 126 kbp linkers

C: MLS in 10 µm nucleus
  126 kbp loops, 126 kbp linker, not totally relaxed

D: RW/GL in 12 µm nucleus
  5 Mbp loops not totally relaxed
In vivo Morphology & Chromatin Distribution

The stable expression of fusions between histones and autofluorescent proteins and the integration into nucleosomes allows the minimal invasive investigation of the structure and dynamics of chromatin.

The clustered morphology in detail favour an MLS like chromatin topology.
Counting Nucleosomes *In Vivo*

Counting nucleosomes in living cells with a combination of fluorescence correlation spectroscopy (FCS) and confocal laser scanning microscopy (CLSM) reveals not only the free unbound histone component but also the concentration in absolute numbers of bound histones. Thus, the absolute concentration distribution of histones can be determined and reveals again the typical expected distribution of aggregated chromatin loops.
Counting Nucleosomes In Vivo

Counting nucleosomes in living cells with a combination of fluorescence correlation spectroscopy (FCS) and confocal laser scanning microscopy (CLSM) reveals not only the free unbound histone component but also the concentration in absolute numbers of bound histones. Thus, the absolute concentration distribution of histones can be determined and reveals again the typical expected distribution of aggregated chromatin loops.

Also The Dynamics Can Only Be Based On A Quasi-Fibre Forming Rosettes!
The compacted chromatin quasi-fibre, folds into loop-aggregates connected by a linker!

Every structural level of nuclear organization including its dynamics is connected and represented in all the other levels in a holistic systems genomics manner.

- The 3D genomes architecture consists of chromatin quasi-fibres (5±1 nuc. / 11 nm, \( L_p \) of 80-120 nm), forming stable loop aggregates/rosettes (~40-100 kbp loops, ~60 kbp linkers).

- The dynamics of genomes follows the 3D genome architecture in detail and determines in an inseparable entanglement with the architecture genome function.

- From the single base pair to the entire cell nucleus, all genomic levels represent all other levels and by modification a code is present and used to store genetic information.

- Genomes have a consensus organization with only small variation from the basic theme on each compaction level of the genome and these small variations determine genome function.

- Genome organization and function cannot be determined or understood from a single organizational level but only in a holistic systems genomics manner integrating all parts of the system.

- The genome behaves on the basis of a genomic statistical mechanics with a genomic uncertainty principle attached!
Acknowledgements

Thanks go to all the lab local lab members, those people who supported this work in the last decades, the institutions providing their infrastructure, and the national and international computing infrastructures.

Special thanks go to the reviewers, the EraSysBio+ initiative and the national and EU funding bodies.
Acknowledgements

Thanks go also to all those people who supported this work in the last decades, the institutions providing their infrastructure, and the national and international computing infrastructures.
Acknowledgements

Thanks go also to all those people who supported this work in the last decades, the institutions providing their infrastructure, and the national and international computing infrastructures.
Acknowledgements

Thanks go also to all those people who supported this work in the last decades, the institutions providing their infrastructure, and the national and international computing infrastructures.
The Detailed Three-Dimensional Organization
of the
Human and Mouse Genome

Knoch, T. A.,

12th Dutch Chromatin Meeting, University Library De Uithof Utrecht, Utrecht, The Netherlands, 30th October, 2014.

Abstract

The dynamic three-dimensional chromatin architecture of genomes and the obvious co-evolutionary connection to its function – the storage and expression of genetic information – is still one of the central issues of our time. With a systems genomics combination of a novel superior selective high-throughput high-resolution chromosomal interaction capture (T2C), a novel FCS, polymer models, architectural and DNA sequence scaling analysis, we determined the 3D architecture and dynamics with molecular resolution from some to the mega-base pair level spanning 6 orders of magnitude (!): for several genetic loci, different species, cell type and states we find a chromatin quasi-fibre, folding into loops forming aggregates/rosettes, connected by chromatin linkers. Whereas T2C measures architectural parameters, the FCS approach allows to measure in vivo the dynamics of the architecture. Beyond, we find the same fine-structured multi-scaling behaviour in the architecture and the DNA sequence, thus both are tightly evolutionarily entangled. Hence, we determined the three-dimensional organization and dynamics for the first time in a consistent system genomics manner from several angles which are all in agreement as well as additionally also with the heuristics of the research of the last 170 years. This is of fundamental importance for genome understanding, diagnosis, and treatment.

Corresponding author email contact: TA.Knoch@taknoch.org

Keywords:
Genome, genomics, genome organization, genome architecture, structural sequencing, architectural sequencing, systems genomics, co-evolution, holistic genetics, genome mechanics, genome statistical mechanics, genomic uncertainty principle, multilism genotype-phenotype, genome function, genetics, gene regulation, replication, transcription, repair, homologous recombination, simultaneous co-transfection, cell division, mitosis, metaphase, interphase, cell nucleus, nuclear structure, nuclear organization, chromatin density distribution, nuclear morphology, chromosome territories, subchromosomal domains, chromatin loop aggregates, chromatin rosettes, chromatin loops, chromatin quasi fibre, chromatin density, persistence length, spatial distance measurement, histones, H1.0, H2A, H2B, H3, H4, mH2A1.2, DNA sequence, complete sequenced genomes, molecular transport, obstructed diffusion, anomalous diffusion, percolation, long-range correlations, fractal analysis, scaling analysis, exact yard-stick dimension, box-counting dimension, lacunarity dimension, local nuclear
dimension, nuclear diffuseness, parallel super computing, grid computing, volunteer computing, polymer model, analytic mathematical model, Brownian Dynamics, Monte Carlo, fluorescence in situ hybridization (FISH), targeted chromatin capture (T2C) confocal laser scanning microscopy, fluorescence correlation spectroscopy, spatial precision distance microscopy, super-resolution microscopy, two dimensional fluorescence correlations spectroscopy (2D-FCS) auto-fluorescent proteins, CFP, GFP, YFP, DsRed, fusion protein, in vivo labelling, information browser, visual data base access, holistic viewing system, integrative data management, extreme visualization, three-dimensional virtual environment, virtual paper tool, human ecology, e-human grid ecology, society, social systems, e-social challenge, inverse tragedy of the commons, grid phenomenon, micro-sociality, macro-sociality, autopoietic tragedy of social sub-systems, micro subsystems, macro subsystems, micro operationality, macro operationality, grid psychology micro riskmanagement, macro riskmanagement.

**Literature References**


