Using nanopores to sequence DNA

Aleksei Aksimentiev



A few examples





Double stranded DNA (persist. length ~50nm)

The sequence has direction: 5'-AAGCTGGTTCAG-3'

Single stranded DNA (persist. length ~1.5nm)



Double stranded DNA (persist. length ~50nm)

The sequence has direction: 5'-AAGCTGGTTCAG-3'

Single stranded DNA (persist. length ~1.5nm)



- Each human cell has two homologous copies of the genome (from father and mother), each containing $\sim 3,100,000,000$ ($3.1x10^9$) base pairs. Fits on a CD.

- All cell types in a human body have the same genome, but appear different because different part of the genome are "ON"

- Genes (protein coding part) occupy only 1.5% of the sequence. ~23,000 genes, twice as many as in a fruit fly. 50% of the genes have unknown function.

- More than 50 percent are simple repeat sequences CGTCGTCGTCGT..., the "dark matter" of the genome. Difficult to characterize.

- Among individuals, 99.9% of the sequence is similar. > 1,000,000 differences.

The Human Genome Project

Duration: October 1990 - 2003

Discovered <u>ALL</u> 20,000-25,000 human genes



Determined complete sequence of the 3 billion DNA bases

GTGGGTGCATAGCTGTGCTGTAAGTGAAGTGAGGCGGCAGGTGTTGAAAG TCGATGTAGTTCGTAGGTCAGTTGATGTCGATGTGAAATGCTGATGCTAGT GGACAGGGTGACTAGTGAATCGATGCTAGCCTAGCTAGTCAGTGGTGCTA GCTACGATCGATTTCAGGCTGCTGTGGGTGCATAGCTGTGCTGTAAGTGAA GTGAGGCGGCAGGTGTTGAAAGTCGATGTAGTTCGTAGGTCAGTTGATGTC GATGTGAAATGCTGATGCTAGTGGACAGGGTGACTAGTGAATCGATGCTAG CCTAGCTAGTCAGTGGTGCTAGCTACGATCGATTTCAGGCTGCTGTGGGTG CATAGCTGTGCTGTAAGTGAAGTGAGGCGGCAGGTGTTGAAAGTCGATGTA TCGTAGGTCAGTTGATGTCGATGTGAAATGCTGATGCTAGTGGACAGGG GATTTCAGGCTGCTGTGGGTGCATAGCTGTGCTGTAAGTGAAGTGAGGCGG CAGGTGTTGAAAGTCGATGTAGTTCGTAGGTCAGTTGATGTCGATGTGAAAT CAGTGGTGCTAGCTACGATCGATTTCAGGCTGCT CCTAGCTAGTCAGTGGT GTTCGTAGGTCAGTTGATGTCGATGTGAAATGCTGATGCTAGTGGACAGGG GATTTCAGGCTGCTGTGGGTGCATAGCTGTGCTGTAAGTGAAGTGAGGCGG CAGGTGTTGAAAGTCGATGTAGTTCGTAGGTCAGTTGATGTCGATGTGA GATGTAGTTCGTAGGTCAGTTGATGTCGATGTGAAATGCTGATGCTAGTGGA CAGGGTGACTAGTGAATCGATGCTAGCCTAGCTAGTCAGTGGTGCTATTGT GCTACGATCGATTTCAGGCTGCTGTGGGTGCATAGCTGTGCTGTAAGTGAA GTGAGGCGGCAGGTGTTGAAAGTCGATGTAGTTCGTAGGTCAGTTGATGTC GTGAAATGCTGATGCTAGTGGACAGGGTGACTAGTGAATCGATGCTAG

GATGTGAAATGCTGATGCTAGTGGACAGGGTGACTAGTGAATCGATGCTAG CCTAGCTAGTCAGTGGTGCTAGCTACGATCGATTTCAGGCTGCTGTGGGTG CATAGCTGTGCTGTAAGTGAAGTGAGGCGGCAGGTGTTGAAAGTCGATGTA GTTCGTAGGTCAGTTGATGTCGATGTGAAATGCTGATGCTAGTGGACAGGG GATTTCAGGCTGCTGTGGGTGCATAGCTGTGCTGTAAGTGAAGTGAGGCGG CAGGTGTTGAAAGTCGATGTAGTTCGTAGGTCAGTTGATGTCGATGTGAAAT CAGTGGTGCTAGCTACGATCGATTTCAGGCTGCT CCTAGCTAGTCAGTGGT GTTCGTAGGTCAGTTGATGTCGATGTGAAATGCTGATGCTAGTGGACAGGG GATTTCAGGCTGCTGTGGGTGCATAGCTGTGCTGTAAGTGAAGTGAGGCGG CAGGTGTTGAAAGTCGATGTAGTTCGTAGGTCAGTTGATGTCGATGTGA GATGTAGTTCGTAGGTCAGTTGATGTCGATGTGAAATGCTGATGCTAGTGGA CAGGGTGACTAGTGAATCGATGCTAGCCTAGCTAGTCAGTGGTGCTATTGT GCTACGATCGATTTCAGGCTGCTGTGGGTGCATAGCTGTGCTGTAAGTGAA GTGAGGCGGCAGGTGTTGAAAGTCGATGTAGTTCGTAGGTCAGTTGATGTC GATGTGAAATGCTGATGCTAGTGGACAGGGTGACTAGTGAATCGATGCTAG CCTAGCTAGTCAGTGGTGCTAGCTACGATCGATTTCAGGCTGCTGTGGGTG AAACGATGTGAAATGCTGATGCTAGTGGACAGGGTGACTAGTGAATCGATG CTAGCCTAGCTAGTCAGTGGTGCTAGCTACGATCGATTTCAGGCTGCTGTG GGTG GCTGATGCTAGTGGACAGGGTGACTAGTGAATCGATGCTAGCCTAGC CAGTGGTGCTAGCTACGATCGATTTCAGGCTGCT CCTAGCTAGTCAGTGGT

CAGGTGTTGAAAGTCGATGTAGTTCGTAGGTCAGTTGATGTCGATGTGATC GATGTAGTTCGTAGGTCAGTTGATGTCGATGTGAAATGCTGATGCTAGTGGA CAGGGTGACTAGTGAATCGATGCTAGCCTAGCTAGTCAGTGGTGCTATTGT GCTACGATCGATTTCAGGCTGCTGTGGGTGCATAGCTGTGCTGTAAGTGAA GTGAGGCGGCAGGTGTTGAAAGTCGATGTAGTTCGTAGGTCAGTTGATGTC GATGTGAAATGCTGATGCTAGTGGACAGGGTGACTAGTGAATCGATGCTAG CCTAGCTAGTCAGTGGTGCTAGCTACGATCGATTTCAGGCTGCTGTGGGTG AAACGATGTGAAATGCTGATGCTAGTGGACAGGGTGACTAGTGAATCGATG CTAGCCTAGCTAGTCAGTGGTGCTAGCTACGATCGATTTCAGGCTGCTGTG GGTG GCTGATGCTAGTGGACAGGGTGACTAGTGAATCGATGCTAGCCTAGC CAGTGGTGCTAGCTACGATCGATTTCAGGCTGCT CCTAGCTAGTCAGTGGT GATGTGAAATGCTGATGCTAGTGGACAGGGTGACTAGTGAATCGATGCTAG CCTAGCTAGTCAGTGGTGCTAGCTACGATCGATTTCAGGCTGCTGTGGGTG CATAGCTGTGCTGTAAGTGAAGTGAGGCGGCAGGTGTTGAAAGTCGATGTA GTTCGTAGGTCAGTTGATGTCGATGTGAAATGCTGATGCTAGTGGACAGGG GATTTCAGGCTGCTGTGGGTGCATAGCTGTGCTGTAAGTGAAGTGAGGCGG CAGGTGTTGAAAGTCGATGTAGTTCGTAGGTCAGTTGATGTCGATGTGAAAT CAGTGGTGCTAGCTACGATCGATTTCAGGCTGCT CCTAGCTAGTCAGTGGT GTTCGTAGGTCAGTTGATGTCGATGTGAAATGCTGATGCTAGTGGACAGGG GATTTCAGGCTGCTGTGGGTGCATAGCTGTGCTGTAAGTGAAGTGAGGCGG

... and ~ 3,000,000 more pages!

(one month to show 24/7)

Just four letter:

~715 Mb

DNA code is billion times more efficient



A





Conventional DNA sequencing

Nobel Prize in Chemistry 1980

As the DNA is synthesized, nucleotides are added on to the growing chain by the DNA polymerase.

The reactions start from the same nucleotide and end with a specific base







Fluorescence-based sequence gel

http://bbrp.llnl.gov

Electric readout of the DNA sequencing



Coulter counter



patented in 1953 by W.H. Coulter



1/current (i.e. resistance) measurements when a particle enters the microscopic pore.

folk.uio.no/anderne/research.html

The nanopore technology



A voltage bias is imposed across the membrane

The nanopore technology



- Compatible with several detection schemes
- No limit on the read length

The nanopore technology



Sequencing DNA by measuring ionic current





Nature Reviews Drug Discovery 1, 77-84 (January 2002)

The ionic current blockade reveals the sequence of the confined nucleotides

Sequencing DNA using MspA



How it is actually done



Nature Biotechnology (2012) doi:10.1038/nbt.2171

Problems:

Insertion and deletion errors associated with skip and backsteps at 10-24%

Enzyme work best at 0.1.-0.2M salt, ionic current detection works best at 1M

Oxford Nanopore Technologies



MiniION: ~800 parallel detection wells

Read length: up to 100,000 nucleotides (2 strands of lambda phage genome)

Unknown pore (hemolysin, MspA, other?)

Unknown enzyme (better polymerase? Helicase?)

Accuracy: 96%

Homopolymer blockades in MspA



Liz Manrao ... J Gundlach, U Washington

Plos One 2011, 6

MD simulation neutravidin-anchored ssDNA in MspA

All-Atom Molecular Dynamics Simulations of Nanopore Systems

Massive parallel computer Blue Waters (UIUC): ~200,000 CPUs







Time scale: $\sim 0.1-100 \ \mu s$ Length scale:10K - 100M atoms or (< 50 nm)³Time resolution:2 fsSpacial resolution:0.1 A

Interaction between atoms is defined by molecular force field

Nanoscale 2:468 (2010)

Setting up an simulation is like cooking





Computing conductance of α -hemolysin with molecular dynamics



Protein + lipid bilayer membrane + 1M water solution of KCl = ~300,000 atoms Average electrostatic potential map

Current-voltage curve of α -hemolysin

Biophys. J. 88:3745 (2005)



$$I(t) = \frac{1}{\Delta t L_z} \sum_{i=1}^{N} q_i (z_i (t + \Delta t) - z_i (t))$$

Instantaneous current

Current-voltage curve of α -hemolysin

Biophys. J. 88:3745 (2005)





MD simulations of current blockades in MspA





Molecular origin of the current blockade



Unstructured (bulk-like) water: more than 2.5A away from protein or DNA

Correlation between Current and Water: Pearson Coefficient Poly(dT): 0.86 Poly(dC): 0.90 Poly(dA): 0.85



Electric readout of the DNA sequencing



Building Amorphous SiO₂



Water-silica interaction



Wednesday, November 27, 13



Biophysical Journal 87:2086 (2004)





TclBC and *TclForces*: basic knowledge of *Tcl*

The best of the two worlds



Lipid-silica interactions



Venkatesan et al., Biomedical Microdevices 13: 671 (2011)

Using nanopores to trap DNA



Jeff Comer

Using nanopores to trap DNA E_{z} (V/nm) 0.25 8 6 0.2 4 2 0.15 2 (uu) 2 0 2 2 2 2.2 nm 0.1 -4 0.05 -6 -8 -2 2 x (nm)3.0 V 5.0 V E-Field is high only near the Threshold for translocation constriction Nanopore Trap Nano Letters 5, 1883 (2005)

Stepping and sequencing one bp at a time

J. Comer and A. Aksimentiev, Journal of Physical Chemistry C 116: 3370 (2012)

W. Timp, J. Comer and A. Aksimentiev, Biophysical Journal Letters 102, L37-L39 (2012)



A voltage pulse steps dsDNA by a single basepair

The thinner, the better!



Graphene nanopores





Nano Letters 12:4117 (2012)



Ionic current blockades can reveal the DNA sequence



Atomic-Resolution Brownian Dynamics simulations of ionic current blockades in graphene nanopores



Wells, Belkin, Comer, Aksimentiev, Nano Letters 12:4117 (2012)

Sequencing by transverse current







Scientific Reports 1:46 (Kawai Lab)



Local heating promotes single-file translocation

M Belkin et al. ACS Nano 7:6816 (2013) ≥¹⁰⁰ 80 Single-file conformation 60 П $T_{n.v.} = 368 \text{ K}$ 40 O room temperature 20 single-file 0.05 0.15 0.25 0.3 0.1 0.2 External bias, U₀ [V] <u>not</u> single-file

Local heating:



increase DNA mobility 20 fold

enables controlled displacement at 10mV biases

DNA origami



Rothemund, Paul W. K. Nature (2006)

ACS Nano, Article ASAP **DOI:** 10.1021/nn303767b

Molecular sensing with origami nanopores



Plasmonics tweezers for nanopore sensing

Experiment: Cees Dekker, Magnus Jonsson (TU Delft)











Jeff Comer Maxim Belkin Rogan Carr David Wells Manish Shankla Swati Bhattacharya Jejoong Yoo Chen-Yu Li Chris Maffeo



Swati Bhattacharya



Jejoong Yoo

David Wells



Maxim Belkin



Chris Maffeo





VMD and NAMD





US Army Corps of Engineers. Construction Engineering Research Education



ACS.