

Genomic Era & the Zika Virus

Kevin Dick, PhD Biomedical Engineering
Carleton University

Tuesday 6th November, 2018

Outline

Introduction

Genomics

Proteomics

Multi-Omics

Unfathomable Complexity

Sequencing Genomes

Genomic Space Race

How “Big” is a Genome?

Technology: High-Throughput Sequencing

Post-Genomic Era

Technology: CRISPR

Studying the Zika Virus

Technology: PIPE Predictor

GO Term Analysis

Predicted Interactions

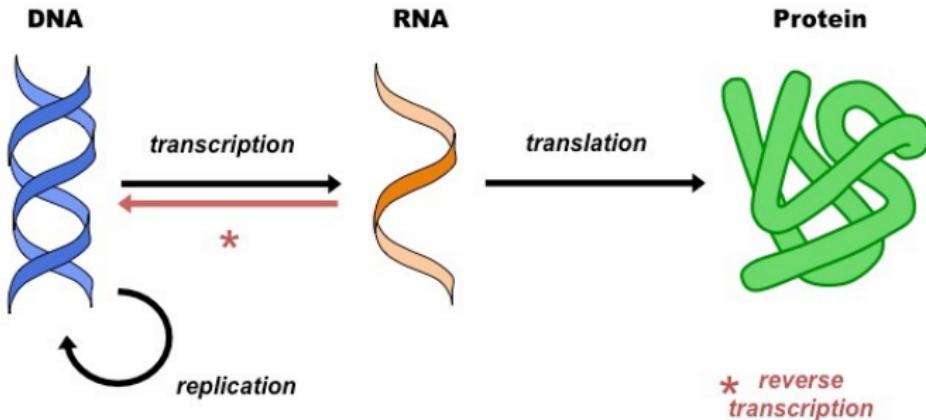
Designed Peptides

Proposed Competitive Pathways

Conclusions



Introduction: Central Dogma



Introduction: Genomics

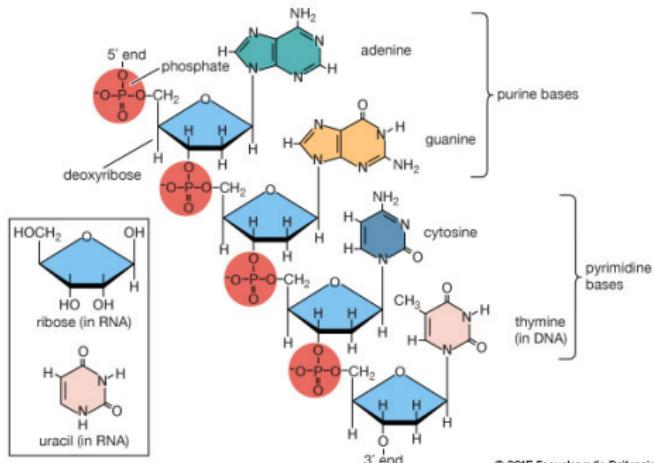
Genomics: The science of studying genes.

- **Before:** Dedicated to the **determination of DNA sequences**
- **Then:** Expanded to more functional studies;
e.g. studying expression profiles of genes and proteins
- **Today:** Genomic data is **growing exponentially**

Introduction: Genomics

Genomics: The science of studying genes.

- **Before:** Dedicated to the **determination of DNA sequences**
- **Then:** Expanded to more functional studies;
e.g. studying expression profiles of genes and proteins
- **Today:** Genomic data is **growing exponentially**



© 2015 Encyclopædia Britannica, Inc.

Introduction: Proteomics

Proteomics: The large scale study of proteins.

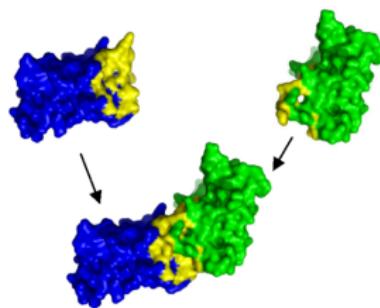
- Proteins perform the “**work**” in a cell
- When working correctly, cells function “**as designed**”
- Diseases arise when they **don't work properly**
- Proteins interact together (PPIs) to produce **complex dynamics**

Introduction: Proteomics

Proteomics: The large scale study of proteins.

- Proteins perform the “**work**” in a cell
- When working correctly, cells function “**as designed**”
- Diseases arise when they **don't work properly**
- Proteins interact together (PPIs) to produce **complex dynamics**

Understanding PPIs enables us to understand cellular functions and design therapeutics to treat diseases.



Introduction: Multi-Omics; Nature vs. Nurture

Multi-Omics Approaches to Disease,[3]

- **Transcriptomics:** Examines RNA levels genome-wide (which ones?, how much?)

Introduction: Multi-Omics; Nature vs. Nurture

Multi-Omics Approaches to Disease,[3]

- **Transcriptomics:** Examines RNA levels genome-wide (which ones?, how much?)
- **Metabolomics:** Quantifies multiple small molecule types (e.g. amino acids, fatty acids, carbohydrates, etc.)

Introduction: Multi-Omics; Nature vs. Nurture

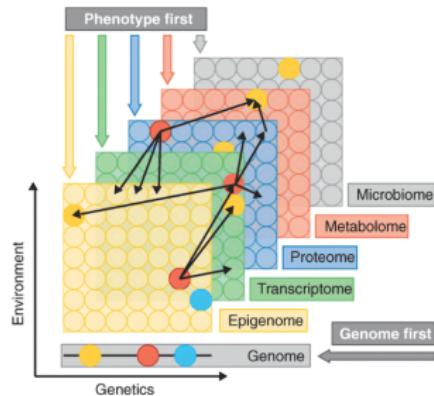
Multi-Omics Approaches to Disease,[3]

- **Transcriptomics:** Examines RNA levels genome-wide (which ones?, how much?)
- **Metabolomics:** Quantifies multiple small molecule types (e.g. amino acids, fatty acids, carbohydrates, etc.)
- **Microbiomics:** Study all the microorganisms of a given community together.

Introduction: Multi-Omics; Nature vs. Nurture

Multi-Omics Approaches to Disease,[3]

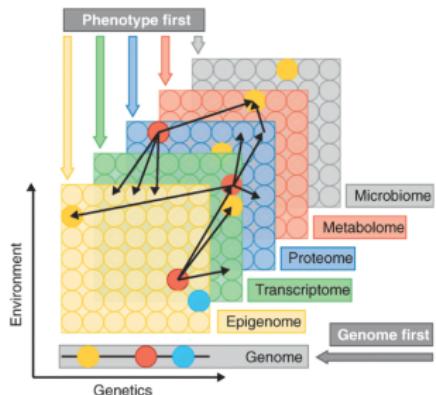
- **Transcriptomics:** Examines RNA levels genome-wide (which ones?, how much?)
- **Metabolomics:** Quantifies multiple small molecule types (e.g. amino acids, fatty acids, carbohydrates, etc.)
- **Microbiomics:** Study all the microorganisms of a given community together.



Introduction: Multi-Omics; Nature vs. Nurture

Multi-Omics Approaches to Disease,[3]

- **Transcriptomics:** Examines RNA levels genome-wide (which ones?, how much?)
- **Metabolomics:** Quantifies multiple small molecule types (e.g. amino acids, fatty acids, carbohydrates, etc.)
- **Microbiomics:** Study all the microorganisms of a given community together.



Holy Grail of Biology: Tame the complexity of these multi-omics to detect, measure, and, ultimately, cure diseases.

Life is Unfathomably Complex

*How many **base pairs** in the Human Genome?*

Life is Unfathomably Complex

How many base pairs in the Human Genome?

3,200,000,000 Base Pairs

...ACGTTAGCTAGCTACGAACGCTCGACTAGCTACTAGCTACGATGCTCTGCATAC...

Life is Unfathomably Complex

How many base pairs in the Human Genome?

3,200,000,000 Base Pairs

...ACGTTAGCTAGCTACGAACGCTCGACTAGCTACTAGCTACGATGCTCTGCATAC...

How many proteins in Humans?

Life is Unfathomably Complex

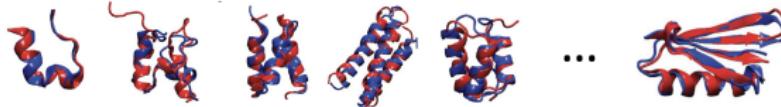
How many base pairs in the Human Genome?

3,200,000,000 Base Pairs

...ACGTTAGCTAGCTACGAACGCTCGACTAGCTACTAGCTACGATGCTCTGCATAAC...

How many proteins in Humans?

23,000 Proteins



Life is Unfathomably Complex

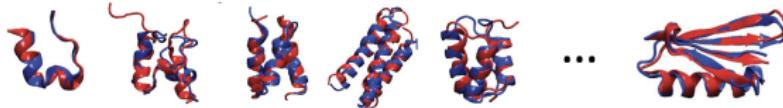
How many base pairs in the Human Genome?

3,200,000,000 Base Pairs

...ACGTTAGCTAGCTACGAACGCTCGACTAGCTACTAGCTACGATGCTCTGCATAAC...

How many proteins in Humans?

23,000 Proteins



How do we start studying it?

Life is Unfathomably Complex

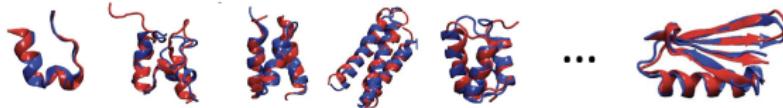
How many base pairs in the Human Genome?

3,200,000,000 Base Pairs

...ACGTTAGCTAGCTACGAACGCTCGACTAGCTACTAGCTACGATGCTCTGCATAAC...

How many proteins in Humans?

23,000 Proteins



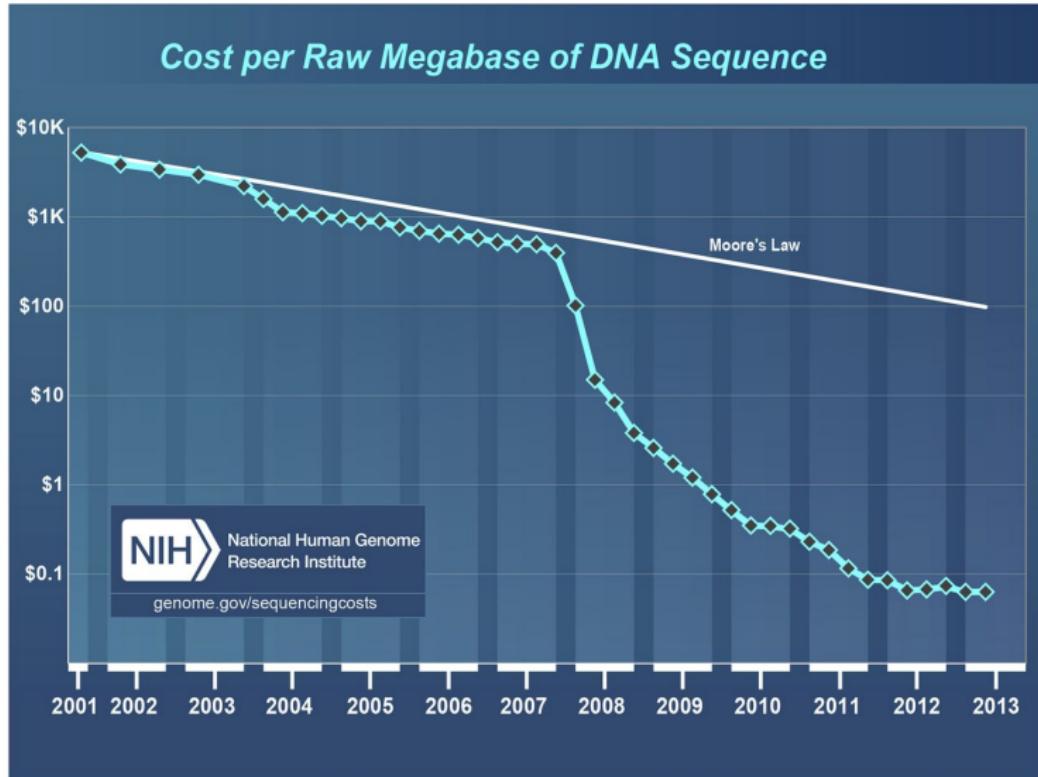
How do we start studying it?

Sequencing!

Sequencing Genomes

Has anyone heard of the Human Genome Project?

Declining Costs for Genome Sequencing



Sequencing Genomes: The “Genomic Space Race”



Sequencing Genomes: The “Genomic Space Race”



Significance: Speed up by 5 years! At $\frac{1}{10}$ the cost!

Sequencing Genomes: The “Genomic Space Race”



Significance: Speed up by 5 years! At $\frac{1}{10}$ the cost!

How? A tremendously risk and unproven approach to piecing together the proteome.

Sequencing Genomes: The “Genomic Space Race”

The Human Genome Project Approach

Hierarchal Shotgun Sequencing

The Celera Approach

Whole Genome Shotgun Sequencing

Sequencing Genomes: The “Genomic Space Race”

The Human Genome Project Approach

Hierarchal Shotgun Sequencing

- Cut up DNA into 150K overlapping seqs.

The Celera Approach

Whole Genome Shotgun Sequencing

Sequencing Genomes: The “Genomic Space Race”

The Human Genome Project Approach

Hierarchal Shotgun Sequencing

- Cut up DNA into 150K overlapping seqs.
- Grown in bacteria and expressed, informing researchers about *where* this sequence was in the genome.

The Celera Approach

Whole Genome Shotgun Sequencing

Sequencing Genomes: The “Genomic Space Race”

The Human Genome Project Approach

Hierarchal Shotgun Sequencing

- Cut up DNA into 150K overlapping seqs.
- Grown in bacteria and expressed, informing researchers about *where* this sequence was in the genome.
- Send fragment to lab to piece genome together

The Celera Approach

Whole Genome Shotgun Sequencing

Sequencing Genomes: The “Genomic Space Race”

The Human Genome Project Approach

Hierarchal Shotgun Sequencing

- Cut up DNA into 150K overlapping seqs.
- Grown in bacteria and expressed, informing researchers about *where* this sequence was in the genome.
- Send fragment to lab to piece genome together
- Very methodological

The Celera Approach

Whole Genome Shotgun Sequencing

Sequencing Genomes: The “Genomic Space Race”

The Human Genome Project Approach

Hierarchal Shotgun Sequencing

- Cut up DNA into 150K overlapping seqs.
- Grown in bacteria and expressed, informing researchers about *where* this sequence was in the genome.
- Send fragment to lab to piece genome together
- Very methodological
- Make data public as it is discovered!

The Celera Approach

Whole Genome Shotgun Sequencing

Sequencing Genomes: The “Genomic Space Race”

The Human Genome Project Approach

Hierarchal Shotgun Sequencing

- Cut up DNA into 150K overlapping seqs.
- Grown in bacteria and expressed, informing researchers about *where* this sequence was in the genome.
- Send fragment to lab to piece genome together
- Very methodological
- Make data public as it is discovered!

TOTAL TIME: 6 Years

The Celera Approach

Whole Genome Shotgun Sequencing

Sequencing Genomes: The “Genomic Space Race”

The Human Genome Project Approach

Hierarchal Shotgun Sequencing

- Cut up DNA into 150K overlapping seqs.
- Grown in bacteria and expressed, informing researchers about *where* this sequence was in the genome.
- Send fragment to lab to piece genome together
- Very methodological
- Make data public as it is discovered!

TOTAL TIME: 6 Years

The Celera Approach

Whole Genome Shotgun Sequencing

- Skip mapping phase altogether

Sequencing Genomes: The “Genomic Space Race”

The Human Genome Project Approach

Hierarchal Shotgun Sequencing

- Cut up DNA into 150K overlapping seqs.
- Grown in bacteria and expressed, informing researchers about *where* this sequence was in the genome.
- Send fragment to lab to piece genome together
- Very methodological
- Make data public as it is discovered!

TOTAL TIME: 6 Years

The Celera Approach

Whole Genome Shotgun Sequencing

- Skip mapping phase altogether
- Cut up genome into a soup of overlapping bits

Sequencing Genomes: The “Genomic Space Race”

The Human Genome Project Approach

Hierarchal Shotgun Sequencing

- Cut up DNA into 150K overlapping seqs.
- Grown in bacteria and expressed, informing researchers about *where* this sequence was in the genome.
- Send fragment to lab to piece genome together
- Very methodological
- Make data public as it is discovered!

TOTAL TIME: 6 Years

The Celera Approach

Whole Genome Shotgun Sequencing

- Skip mapping phase altogether
- Cut up genome into a soup of overlapping bits
- Sequence each fragment

Sequencing Genomes: The “Genomic Space Race”

The Human Genome Project Approach

Hierarchal Shotgun Sequencing

- Cut up DNA into 150K overlapping seqs.
- Grown in bacteria and expressed, informing researchers about *where* this sequence was in the genome.
- Send fragment to lab to piece genome together
- Very methodological
- Make data public as it is discovered!

TOTAL TIME: 6 Years

The Celera Approach

Whole Genome Shotgun Sequencing

- Skip mapping phase altogether
- Cut up genome into a soup of overlapping bits
- Sequence each fragment
- **CRAZY PART:** Piece back together using only the overlaps.

Sequencing Genomes: The “Genomic Space Race”

The Human Genome Project Approach

Hierarchal Shotgun Sequencing

- Cut up DNA into 150K overlapping seqs.
- Grown in bacteria and expressed, informing researchers about *where* this sequence was in the genome.
- Send fragment to lab to piece genome together
- Very methodological
- Make data public as it is discovered!

TOTAL TIME: 6 Years

The Celera Approach

Whole Genome Shotgun Sequencing

- Skip mapping phase altogether
- Cut up genome into a soup of overlapping bits
- Sequence each fragment
- **CRAZY PART:** Piece back together using only the overlaps.
- Very tough without a place to start

Sequencing Genomes: The “Genomic Space Race”

The Human Genome Project Approach

Hierarchal Shotgun Sequencing

- Cut up DNA into 150K overlapping seqs.
- Grown in bacteria and expressed, informing researchers about *where* this sequence was in the genome.
- Send fragment to lab to piece genome together
- Very methodological
- Make data public as it is discovered!

TOTAL TIME: 6 Years

The Celera Approach

Whole Genome Shotgun Sequencing

- Skip mapping phase altogether
- Cut up genome into a soup of overlapping bits
- Sequence each fragment
- **CRAZY PART:** Piece back together using only the overlaps.
- Very tough without a place to start
- **BUT** can use the HGP public data!

Sequencing Genomes: The “Genomic Space Race”

The Human Genome Project Approach

Hierarchal Shotgun Sequencing

- Cut up DNA into 150K overlapping seqs.
- Grown in bacteria and expressed, informing researchers about *where* this sequence was in the genome.
- Send fragment to lab to piece genome together
- Very methodological
- Make data public as it is discovered!

TOTAL TIME: 6 Years

The Celera Approach

Whole Genome Shotgun Sequencing

- Skip mapping phase altogether
- Cut up genome into a soup of overlapping bits
- Sequence each fragment
- **CRAZY PART:** Piece back together using only the overlaps.
- Very tough without a place to start
- **BUT** can use the HGP public data!

TOTAL TIME: 3 Years

Sequencing Genomes: The “Genomic Space Race”

The Human Genome Project Approach *Hierarchal Shotgun Sequencing*

- Cut up DNA into 150K overlapping seqs.
- Grown in bacteria and expressed, informing researchers about *where* this sequence was in the genome.
- Send fragment to lab to piece genome together
- Very methodological
- Make data public as it is discovered!

TOTAL TIME: 6 Years

The Celera Approach *Whole Genome Shotgun Sequencing*

- Skip mapping phase altogether
- Cut up genome into a soup of overlapping bits
- Sequence each fragment
- **CRAZY PART:** Piece back together using only the overlaps.
- Very tough without a place to start
- **BUT** can use the HGP public data!

TOTAL TIME: 3 Years



Sequencing Genomes: How “Big” is a Genome?



Sequencing Genomes: How “Big” is a Genome?



Physical Representation:

- 130 Volumes; double-sided
- 4pt font; 43,000 char/page
- **X:** 7 volumes; **Y:** only 1

Sequencing Genomes: How “Big” is a Genome?



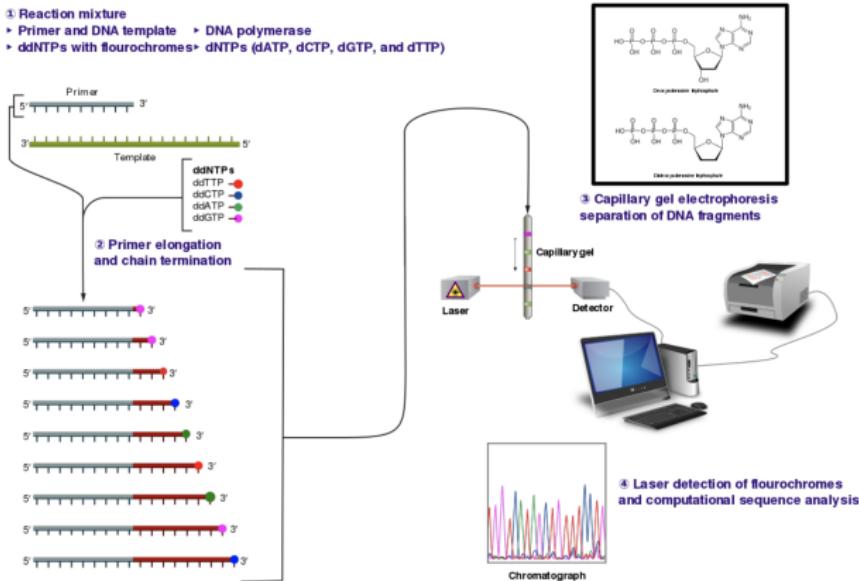
Physical Representation:

- 130 Volumes; double-sided
- 4pt font; 43,000 char/page
- **X:** 7 volumes; **Y:** only 1



Technology: Next-Generation Sequencing (NGS)

Many *high-throughput sequencing* technologies exist today.
From years, to months, to weeks, to days; increasing **read depth**!



Illumina Sequencing Video



The Post-Genomic Era

Personalized Medicine

Drug Design



Gene Therapy/Genome Editing

“Designer Babies”



Ethical ramifications for child-bearing mothers

Disease Eradication

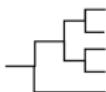
Sequencing Organisms

Understanding Viruses/Bacteria



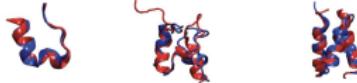
Studying Evolution

1000 Genome Project
Sequencing Life



Studying Complexity of Life: Proteins!

Complexity Arises from Protein Diversity
~15,000 more proteins between human and yeast



The Post-Genomic Era

Personalized Medicine

Drug Design



Gene Therapy/Genome Editing

“Designer Babies”



Ethical ramifications for child-bearing mothers

Disease Eradication

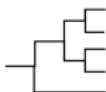
Sequencing Organisms

Understanding Viruses/Bacteria



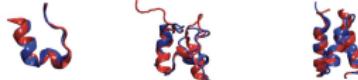
Studying Evolution

1000 Genome Project
Sequencing Life



Studying Complexity of Life: Proteins!

Complexity Arises from Protein Diversity
~15,000 more proteins between human and yeast



The Legal Framework has lagged tremendously
Protocols & Policies to protect genomic information are *far from resolved.*
Substantial health care benefit is possible.
Required to advance research more rapidly!

Example Technology: CRISPR [1]

Video Overview

Example Technology: CRISPR [1]

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

The CRISPR-Cas9 system has been adapted to edit genomes with **high precision** and **low off-target effect**.

Example Technology: CRISPR [1]

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

The CRISPR-Cas9 system has been adapted to edit genomes with **high precision** and **low off-target effect**.

Somatic Cells

The cells in your body that do not get inherited by your offspring.

Example Technology: CRISPR [1]

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

The CRISPR-Cas9 system has been adapted to edit genomes with **high precision** and **low off-target effect**.

Somatic Cells

The cells in your body that do not get inherited by your offspring.

- Skin, Bones, Cartilage, Muscles
- Lungs, Heart, Liver, etc...

Example Technology: CRISPR [1]

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

The CRISPR-Cas9 system has been adapted to edit genomes with **high precision** and **low off-target effect**.

Somatic Cells

The cells in your body that do not get inherited by your offspring.

- Skin, Bones, Cartilage, Muscles
- Lungs, Heart, Liver, etc...

Germline Cells

The cells in your body that do get inherited by your offspring.?

Example Technology: CRISPR [1]

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

The CRISPR-Cas9 system has been adapted to edit genomes with **high precision** and **low off-target effect**.

Somatic Cells

The cells in your body that do not get inherited by your offspring.

- Skin, Bones, Cartilage, Muscles
- Lungs, Heart, Liver, etc...

Germline Cells

The cells in your body that do get inherited by your offspring.?

- **Females:** Ovum/Eggs
- **Males:** Spermatozoa/Sperm

Example Technology: CRISPR [1]

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

The CRISPR-Cas9 system has been adapted to edit genomes with **high precision** and **low off-target effect**.

Somatic Cells

The cells in your body that do not get inherited by your offspring.

- Skin, Bones, Cartilage, Muscles
- Lungs, Heart, Liver, etc...

Germline Cells

The cells in your body that do get inherited by your offspring.?

- **Females:** Ovum/Eggs
- **Males:** Spermatozoa/Sperm

Tremendous Controversy: CRISPR could be used to **eradicate diseases** (e.g. Zika, Malaria), **genetically engineer babies**, and drastically **increase crop yields** modifying **future ecosystems**.

However, we know very little about any downstream consequences; there is no going back.

Example Technology: CRISPR [1]

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

The CRISPR-Cas9 system has been adapted to edit genomes with **high precision** and **low off-target effect**.

Somatic Cells

The cells in your body that do not get inherited by your offspring.

- Skin, Bones, Cartilage, Muscles
- Lungs, Heart, Liver, etc...

Germline Cells

The cells in your body that do get inherited by your offspring.?

- **Females:** Ovum/Eggs
- **Males:** Spermatozoa/Sperm

Tremendous Controversy: CRISPR could be used to **eradicate diseases** (e.g. Zika, Malaria), **genetically engineer babies**, and drastically **increase crop yields** modifying **future ecosystems**.

However, we know very little about any downstream consequences; there is no going back.

Should we use it?

CRISPR Overview

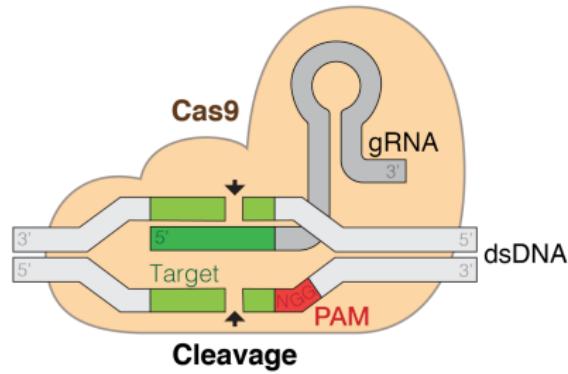


Figure 1: High-Level Depiction of the CRISPR-Cas9 System; [1].

Studying the Zika Virus

Does anyone know how a drug works in the body?

Studying Human-Zika Virus Interactions

**Designing anti-Zika virus peptides derived from predicted
human-Zika virus protein-protein interactions, [4]**

Studying Human-Zika Virus Interactions

Designing anti-Zika virus peptides derived from predicted human-Zika virus protein-protein interactions, [4]

Motivation

Predicting putative PPIs to design a competitive peptide that would mitigate symptoms.

Studying Human-Zika Virus Interactions

Designing anti-Zika virus peptides derived from predicted human-Zika virus protein-protein interactions, [4]

Motivation

Predicting putative PPIs to design a competitive peptide that would mitigate symptoms.

- Zika Virus is an understudied organism

Studying Human-Zika Virus Interactions

Designing anti-Zika virus peptides derived from predicted human-Zika virus protein-protein interactions, [4]

Motivation

Predicting putative PPIs to design a competitive peptide that would mitigate symptoms.

- Zika Virus is an understudied organism
- Causes neonatal microcephaly (newborns with small, underdeveloped heads/brains)

Studying Human-Zika Virus Interactions

Designing anti-Zika virus peptides derived from predicted human-Zika virus protein-protein interactions, [4]

Motivation

Predicting putative PPIs to design a competitive peptide that would mitigate symptoms.

- Zika Virus is an understudied organism
- Causes neonatal microcephaly (newborns with small, underdeveloped heads/brains)
- Part of a family of well-studied organisms (Dengue virus, West Nile virus, Japanese Encephalitis virus, etc.)

Studying Human-Zika Virus Interactions

Designing anti-Zika virus peptides derived from predicted human-Zika virus protein-protein interactions, [4]

Motivation

Predicting putative PPIs to design a competitive peptide that would mitigate symptoms.

Results Overview

Four candidate Zika-Human PPIs potentially linked to neonatal microcephaly.

- Zika Virus is an understudied organism
- Causes neonatal microcephaly (newborns with small, underdeveloped heads/brains)
- Part of a family of well-studied organisms (Dengue virus, West Nile virus, Japanese Encephalitis virus, etc.)

Studying Human-Zika Virus Interactions

Designing anti-Zika virus peptides derived from predicted human-Zika virus protein-protein interactions, [4]

Motivation

Predicting putative PPIs to design a competitive peptide that would mitigate symptoms.

- Zika Virus is an understudied organism
- Causes neonatal microcephaly (newborns with small, underdeveloped heads/brains)
- Part of a family of well-studied organisms (Dengue virus, West Nile virus, Japanese Encephalitis virus, etc.)

Results Overview

Four candidate Zika-Human PPIs potentially linked to neonatal microcephaly.

- **RNASET2-NS3:** Cleaved peptide leads to white brain matter mutation (in zebrafish)

Studying Human-Zika Virus Interactions

Designing anti-Zika virus peptides derived from predicted human-Zika virus protein-protein interactions, [4]

Motivation

Predicting putative PPIs to design a competitive peptide that would mitigate symptoms.

- Zika Virus is an understudied organism
- Causes neonatal microcephaly (newborns with small, underdeveloped heads/brains)
- Part of a family of well-studied organisms (Dengue virus, West Nile virus, Japanese Encephalitis virus, etc.)

Results Overview

Four candidate Zika-Human PPIs potentially linked to neonatal microcephaly.

- **RNASET2-NS3:** Cleaved peptide leads to white brain matter mutation (in zebrafish)
- **ENO2-NS5:** Neurodevelopment - cell maturation

Studying Human-Zika Virus Interactions

Designing anti-Zika virus peptides derived from predicted human-Zika virus protein-protein interactions, [4]

Motivation

Predicting putative PPIs to design a competitive peptide that would mitigate symptoms.

- Zika Virus is an understudied organism
- Causes neonatal microcephaly (newborns with small, underdeveloped heads/brains)
- Part of a family of well-studied organisms (Dengue virus, West Nile virus, Japanese Encephalitis virus, etc.)

Results Overview

Four candidate Zika-Human PPIs potentially linked to neonatal microcephaly.

- **RNASET2-NS3:** Cleaved peptide leads to white brain matter mutation (in zebrafish)
- **ENO2-NS5:** Neurodevelopment - cell maturation
- **TRAF4-M:** Neurodevelopment - membrane receptor component

Studying Human-Zika Virus Interactions

Designing anti-Zika virus peptides derived from predicted human-Zika virus protein-protein interactions, [4]

Motivation

Predicting putative PPIs to design a competitive peptide that would mitigate symptoms.

- Zika Virus is an understudied organism
- Causes neonatal microcephaly (newborns with small, underdeveloped heads/brains)
- Part of a family of well-studied organisms (Dengue virus, West Nile virus, Japanese Encephalitis virus, etc.)

Results Overview

Four candidate Zika-Human PPIs potentially linked to neonatal microcephaly.

- **RNASET2-NS3:** Cleaved peptide leads to white brain matter mutation (in zebrafish)
- **ENO2-NS5:** Neurodevelopment - cell maturation
- **TRAF4-M:** Neurodevelopment - membrane receptor component
- **CEP63-NS3:** Loss of CEP63 in mice results in microcephaly

Sequence-Based Protein-Protein Interaction Prediction

Machine Learning: A suite of method used to devise complex models and algorithms that lend themselves to prediction.

Sequence-Based Protein-Protein Interaction Prediction

Machine Learning: A suite of method used to devise complex models and algorithms that lend themselves to prediction.

W.R.T. this study:

Leverage the set of previously validated PPIs between humans and the *Flaviridae* family of viruses to predict **new** interactions with the Zika Virus.

Sequence-Based Protein-Protein Interaction Prediction

Machine Learning: A suite of method used to devise complex models and algorithms that lend themselves to prediction.

W.R.T. this study:

Leverage the set of previously validated PPIs between humans and the *Flaviridae* family of viruses to predict **new** interactions with the Zika Virus.

Data: Human and Zika protein sequences from the reference proteomes (Uniprot), and Human-Virus data (VirusMentha).

Sequence-Based Protein-Protein Interaction Prediction

Machine Learning: A suite of method used to devise complex models and algorithms that lend themselves to prediction.

W.R.T. this study:

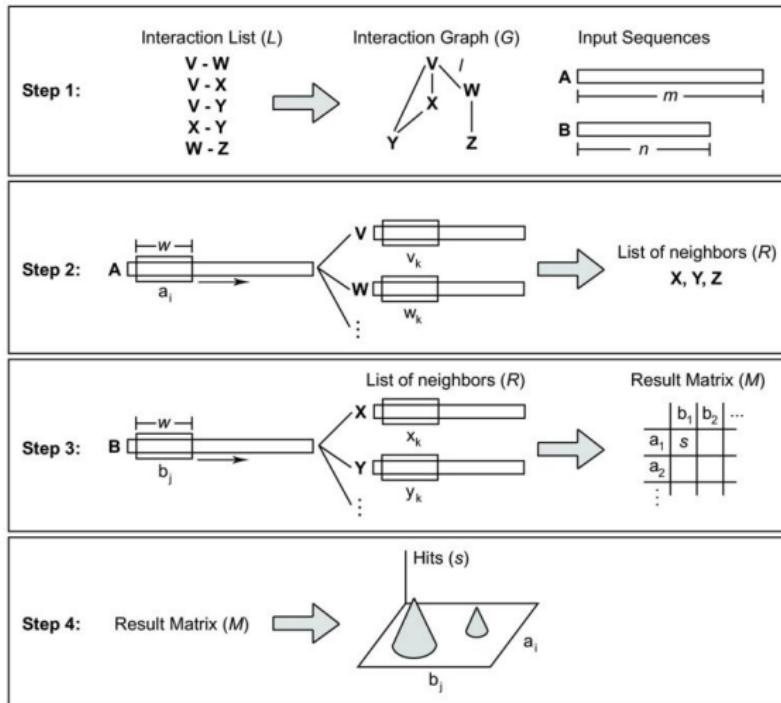
Leverage the set of previously validated PPIs between humans and the *Flaviridae* family of viruses to predict **new** interactions with the Zika Virus.

Data: Human and Zika protein sequences from the reference proteomes (Uniprot), and Human-Virus data (VirusMentha).

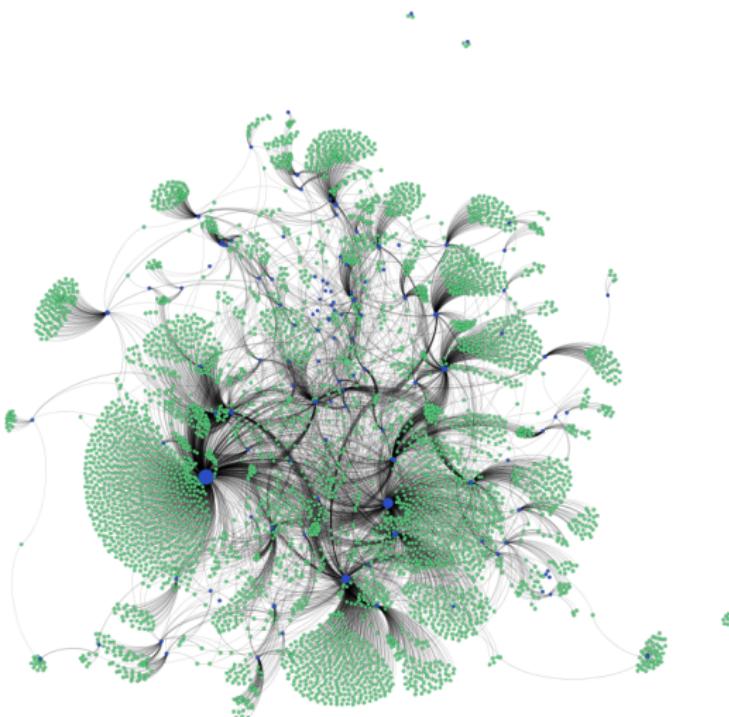
**PIPE: Protein-Protein
Interaction Prediction Engine, [5]**
*Carleton University Bioinformatics
Research Group*

**DeNovo Host-Virus PPI
Predictor, [2]**
*Virginia Technical University
Bioinformatics Research Group*

Technology: Protein-Protein Interaction Engine



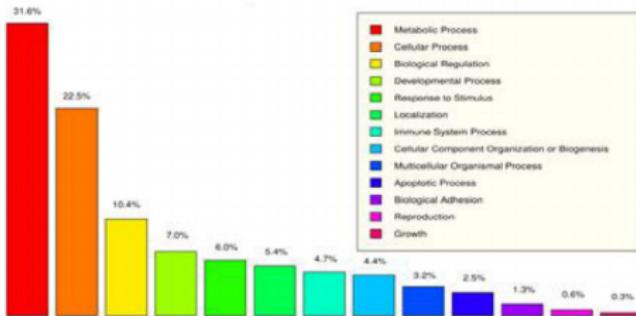
Example Interactome



GO Term Analysis

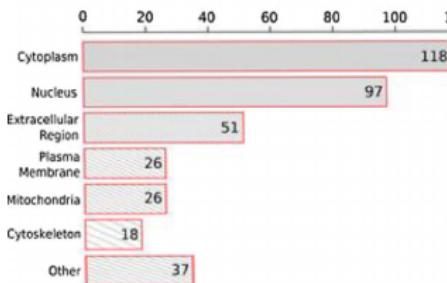
A

Biological Process Classification



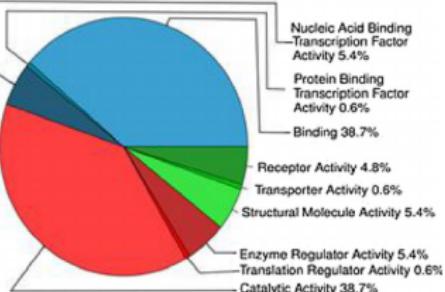
B

Cellular Component Classification



C

Molecular Function Classification



Predicted Interactions

Table 1

The 25 human protein candidates determined by PIPE (P) or DeNovo (N), respective ZIKV interactors, function (F), associated disease phenotype (D), and supporting literature.

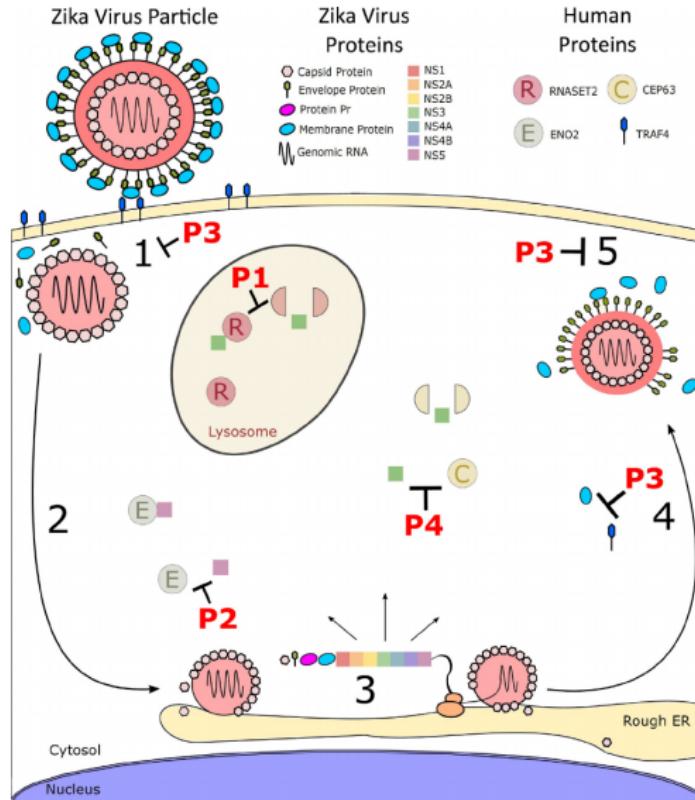
Human Proteins	ZIKV Proteins	Function or Disease Phenotype	References
ENO1 (P)	NS5, NS3, NS1 ^a , M ^a	Affects cell proliferation, differentiation (F)	Schmeichel et al. (1980), Eriksson et al., (1998a) and Gould et al. (1999)
ENO2 (P)	NS5, NS3, NS1 ^a , M ^a	Affects cell proliferation, differentiation (F)	Schmeichel et al. (1980), Eriksson et al., (1998a) and Gould et al. (1999)
ENO3 (P)	NS3, NS1 ^a , E ^a	β-enolase deficiency (D)	Comi et al. (2001)
RNASET2 (P)	NS3 ^a	Microcephaly (D)	Henneke et al. (2009), Staretz-Chacham et al. (2009), Tonduti et al. (2016) and Haud et al., 2010a
CALR3 (N)	NS4B, M	Male fertility, spermatogenesis, gamete fusion (F)	Tokuhiro et al., 2015, Irie et al. (2011) and Dun et al. (2012)
NME1 (P)	NS3	Induction of fibronectin (F)	Novak et al. (2015)
NME3 (P)	NS5, NS3, M	Cell motility in somatic cells, spermatozoa (D)	Bauer et al. (2012)
RNF151 (N)	Pr, M	Spermatogenesis (F)	Nian et al. (2007)
RNF125 (P)	NS3	Regulation of HIV-1 replication (F)	Shoji-Kawata et al. (2007)
FUND1C1 (P)	NS3	Activates hypoxia-induced mitophagy (F)	Liu et al. (2012)
BCLG (P)	NS5, NS3, M	Apoptosis factor (F)	Hu and Kavanagh (2003)
TRAF4 (P)	NS5, NS3, M ^a	Impaired neural crest development, folding (D)	Kalkan et al. (2009)
MLPH (P)	NS3	Griselli syndrome (D)	Van Gele et al. (2009)
PIAS3 (P)	NS3	SUMOylation of photoreceptors (F)	Onishi et al. (2010)
CAMTA2 (P)	NS5, NS3, M ^a , C ^a	Decreased cardiac growth (D)	Song et al. (2006)
AZI2 (P)	NS5, NS3, M ^a	Neurodevelopment (D)	Fukasaka et al. (2013)
MATR3 (P)	NS3	Cardiac development (F)	Quintero-Rivera et al. (2015)
CEP63 (P)	NS5, NS3	p53-dependent microcephaly (D)	Marjanovic et al. (2015)
RIAM (P)	NS5, NS3	Leukocyte adhesion deficiency (D)	Calderwood (2015)
DYX1C1 (N)	NS4B, M	Dyslexia (D)	Dahdouh et al. (2007)
SNAP25 (N)	NS4B, M	Huntington's Disease (D)	Smith et al. (2007)
YWHAE (N)	NS4B	Neurocognitive, Cerebrospinal fluid marker (D)	Morales et al. (2013)
COX17 (N)	NS4B, NS2A	Cardiomyopathy, hepatic failure (D)	Hamza and Gitlin (2002)
SPZ1 (N)	NS4B, M	Spermatogenesis (F)	Hsu et al. (2004)
DNAJA1 (N)	Pr, M	Testis development, spermatogenesis (F)	Hu et al. (2004)

Designed Peptides

<i>Human Protein</i>	<i>ZIKV Protein</i>	<i>PPI-Site; Human</i>	<i>PPI-Site; ZIKV</i>
RNASET2 MRPAALRGALLGCLCALLLGLGAKDRLRNDHEWKKLMLVQ HWPTPECEKINDCRDPDFWYIWTWLWPKDRCNRSPWPFNL EIEKDLLPEMRAVYDWPDIHFSRNSRFWKEWEKHGTCAAGV DALNSOKKYFGSRSLLEYLRELDLNSVLKLGLGKPSINYYQVADF KDALARVYGVIPKIQCLPPSQDEEVOTIGIELCLTRQDQQLQNC TEPGEOPSPKQEWLNGANGAESLRVCEDPGVYPFPKPKTKH	NS3 SGALWDVPAPKVEVKKGGETIDGVYRVMTTRRLGGSTQ VGVGVMQEGVFTHTMWHTVKGAALRSGEGRLLPVW GDVKQDLSVSYCGPWKLDAAWDGLSEVOLLAVPPIGE BARNQJTLPGIFKTKDQGIIQAVALDYPAGTSQSPILDK CGRVIGLYGNGVVVIKNGSYSSAATDQKREEEPTVECF EPSSMLKKKQTLVLDLPGAGKTRPLPEIvreAIIKKR LRTVILAPTRVVAEAEIIEALRGLPVRYMMTTAVNVTH SGTEIVMDLCHATTFSRLLQPIFVPPVNLNIMDEAHFT DPSSIAARGYISTREVMEGAAAIIFTATPPGTRDAPFD SNSPIMDTEVEYPERAASSGSFDWVTDHSGKTWVWFPS VRNGNEIAACLTAKAGRKRVQLSRSKTFTEPQTKNQE WDFVTTIDSEMIGANFKADRVIDSRRCLKPVILDGERV ILAGPMVPTHASAOQRGRGRGRGRNPNGDQEYMGGGC AETDEGHAFWLEARMILLDNITYQDGLIASLYRPE VAAIEGEPKLKTEQRVTFELMVKRDGPVWLAYQVAS AGITYTDHRWCDFDTNTNTIMEEDSPVFAEVWTKYGKE VLPKRWMDARVCSDHIAALKSFKEPAAGKR	P1 (188-215) SQDDEEVQTIGIEL CLTKQDQQQLQN CTEP	(90-178) DGLSEVQLLA VPPGERARNI QTLPGIFKTK DGDIGAVALD YPATGSGSPI LDKCGRVIGL YGNGVVKNG SYVSAITQGK REEETPVEC
ENO2 MSIEKIWAREELDSRGNPTEVVDLYTAGLFRAAVPSGAST GIYEALELIRDGDQKRLGKGVLKAVDHINSTIAP ALISSGLS VVEQEQLDNLMEILEDGTENKS KFGANAILGVSLAVKAGA AERELPLYRHIAQJLQNSDILIPVPAFNVNGSHAGNKL MQEFMLPVGAEFSRDAMLRGAEEVYHGLVVKDQYKGDA TNVGDEGGFAPNILENSEAELVKEAIDKAGTYKEVIVGMDV AASEFYRDGKYLDDLFKSPIDPSRYYTGDOLGALYFDRVDY PVVSEIDPFDQDOWAAWSKFTANVQIQVGDLLTVTPNPKRE RAVEEKACNCLLKVNOIQSVEIAQACKLAQENGWGMV SHIRSGEDEDITADLVVGLCTQIKTGPACRCSRERLAKNYQLM RIEFEARFAGHNFRNPSV	NS5 GGGTGETLGEWKAIRLNQMSALEFYSYKKSQITEVC REEARRALKVGATGGHVSRSGSAKIRWLEGRGKQD PYGVKVVDLGCGRGGWSYVAAITRKVQEVGRVYTKGG FGHEEPMVQFQVYKQVLLRKGQVDFHMMVMEAPCDT LCDQEGESSSVEETRTLRLVLSMVGDWLEKRPGAPCI KVLCPVYDQKQVLLRKGQVDFHMMVMEAPCDT EMYVWSGAKSNIIKSVSITTSQLLGRMDGFRPRVKYE EDVNLSLGRTRAVASCAEPNPKRGRERIIRNEAHT WFLENHNPYRTRWVAGHSYEAPTOGSASSVNLVWV RLSKPWDVUTGVTGTMIAIMDTPTVQGQVRWGRD PDQFEGTRQVNINIVSVSWKWEILGRKRPRVCTKEEF NKVRNSNAALGAJFEEKEWIKTAEVAENDPFWALWD REREHILRGECHSCVYNNMIGIREKQGEFGKAKGS RAJWYNWL GARFLPEALGFLNEDWVNGMRENSGGG VEGLGLQRLGYILEEMNNAPGKMYMADDTAGWDT ISKFDLNEALINTNQMEEGHRLALAVIKYTQNQKV KVLRAEAGGGKTVMDISRODQRGSQGVVTYALNTFT LVLVQLRNMEAEYELMQLWLLRKEPVTRWLQSN GWDRRLKRMVASGSDDCVKVPIDFRAHALRFLNDMG KVRKDQEWPKGSTGWSNWEVEPCSHHIFNKLYLKDGF RSIVVPCRHODELIGRARVSPGAGWSIRETACLAKSYA OMWOLYFHRDLRLMANAICSAVPPWDVPTPGRRTW SHBGKGEWATTTEDMLMVWNRVRWEIENDHMEDKPTV TKWTDPYLGKREDL WCGSLIGRPRTTWAENIKDTV NMVRIIGDEEKKYMDYLSLTVQVRYLGEEGGSTPQVL	P2 (76-105) ALISSGLS V EQEKLQDNLML ELDGTEKS K	(0-248) GGGTGETLGEKWKA RLNQMSALEFYSYKK SGITEVCREEARRAL KDGVATGGHAVSRGS AKIRWLEERGYLQPYG KVVDLGCGRGGWSY AATIRKVQEVGRYTKG GPGHEEPMVLVQSYGWN IVRLKSGVDVFHMAAE PCDTLLCDIGESSSSPEV EETRTLRLVLSMVGDWLE KRPGAFCKVLCPVYST MMETMERLQRRHGGGL VRVPLCRNSTHEMYVV SGAKSNIIKSVSITTSQ LGRMDGPR
TRAF4 MIPGDFKFLPKRKKRNLCPCLGKPMRPEPVOSTCGHRCDF TCLQEQFLSIEGVFKPCPDEQJ AKIYDPELEVQVQLGLP CIISEEGCRWNSQPLRHOGLHNTCSVNPICPNCRCPMKLRS	Membrane AVTLPHSRTRKLQTRQSQTWLSREYTKHLIKVENWIFR NPGFALAVAVIAWLGLSSTSQKVIYLVMILLI APAYS	P3 (453-482) MCA LVSQRQ	(11-57) RKLQTRSQ WLESREYTK



Proposed Competitive Pathways



Conclusions & Future Directions

1. Generated the **first-ever** comprehensive **interactome** of **Zika-Human** predictions; useful for future studies.
2. Subsequent **wet-lab validation** of these predicted interactions.
3. Development of the **proposed peptides** and testing **binding affinity**.

References

-  Le Cong et al. "Multiplex genome engineering using CRISPR/Cas systems". In: *Science* (2013), p. 1231143.
-  Fatma-Elzahraa Eid, Mahmoud ElHefnawi, and Lenwood S. Heath. "DeNovo: virus-host sequence-based protein–protein interaction prediction". In: *Bioinformatics* 32.8 (2016), pp. 1144–1150.
-  Yehudit Hasin, Marcus Seldin, and Aldons Lusis. "Multi-omics approaches to disease". In: *Genome biology* 18.1 (2017), p. 83.
-  Tom Kazmirschuk et al. "Designing anti-Zika virus peptides derived from predicted human-Zika virus protein-protein interactions". In: *Computational biology and chemistry* 71 (2017), pp. 180–187.
-  Sylvain Pitre et al. "PIPE: a protein-protein interaction prediction engine based on the re-occurring short polypeptide sequences between known interacting protein pairs". In: *BMC bioinformatics* 7.1 (2006), p. 365.

Genomic Era & the Zika Virus

Kevin Dick, PhD Biomedical Engineering
Carleton University

Tuesday 6th November, 2018